SCUOLA DI DOTTORATO
UNIVERSITÀ DEGLI STUDI DI MILANO-BICOCCA

Department of Psychology

PhD program in Psychology, Linguistic, and Cognitive Neuroscience
Cycle XXXV

Curriculum in Mind, Brain, and Behaviour

Processing faces and voices in early infancy: new insights into the typical and atypical development of brain networks comprising the social brain.

PhD candidate: Polver Silvia

Registration number: 754217

Tutor: Prof. Hermann Bulf

Coordinator: Prof. Simona Sacchi

ACADEMIC YEAR 2021/2022

Table of contents

Abstract....4
Abstract italiano....5
General Introduction....6
Chapter 1. Developmental trajectories of low-level auditory processing in the infant brain. 11
1. Introduction....11
2. Methods....14
2.1. Participants....14
2.2. Stimuli and procedure....14
2.3. EEG data acquisition and preprocessing....15
2.4. ERP analysis....16
2.5. Source localization....18
3. Results....19
3.1. Early negative difference....20
3.2. Late negative difference....23
4. Discussion....26
Chapter 2. How does it sound? Social saliency affects voice cortical tracking in low risk and HL-ASD infants. 30
1. Introduction....30
2. Methods....34
2.1. Participants....34
2.2. Stimuli....35
2.3. EEG data collection and preprocessing....35
2.4. Voice-brain coherence computation and statistical analyses....36
3. Results....37
4. Discussion....43
Chapter 3. Emotional faces guide sensory and higher-order processing during development. 47
1. Introduction....47
2. Methods....49
2.1. Participants....49
2.2. Stimuli....50
2.3. Procedure....51
2.4. EEG recordings....51
2.5. Preprocessing and graph metrics computations....52
2.6. PCA and PCR....53
3. Results....55
4. Discussion....57
Chapter 4. Multimodal temporal binding of emotional vocalizations: typical and atypical brain development. 60

1. Introduction 60
2. Methods 65
2.1. Participants 65
2.2. Stimuli 66
2.3. EEG data collection and preprocessing 67
2.4. Source reconstruction 68
2.5. Time-varying connectivity analysis 69
3. Results 70
4. Discussion 75
General Discussion 80
References 85

Abstract

The first months of life are characterized by striking developmental changes. As the infant grows, through processes of interactive specialization between the environment and the brain, unspecific and suboptimal neural responses specialize and lay the basis for the development of the brain networks observed in adults. Within this context, particular sets of brain areas specialize progressively for the processing of social stimuli, becoming what is known as the social brain. This feature highlights the central role that social processing has across the first year of life, when social cues represent the primary mean through which preverbal infants interact with and absorb information from their surroundings. Hence, the emergence of the social brain may represent a turning point separating typical from atypical development, especially in disorders whose core features include disruptions in social processing, such as autism spectrum disorder (ASD; Grossman & Johnson, 2007).

The aim of this thesis is to define brain networks involved in social processing across developmental stages and to investigate the departures from typical development seen in infants at higher likelihood of developing ASD (HL-ASD). To this end, in Chapter 1 we first isolated the developmental axis of brain networks involved in low-level auditory processing through source reconstruction techniques. In Chapter 2 we manipulated social saliency in voice processing in both low risk (LR) and HL-ASD infants, investigating the encoding of voices in the brain through cortical tracking. Following the same line of thought, in Chapter 3 we focused on the early processing of emotional faces in the brain via decoding techniques. As faces and voice mostly co-occur in everyday life, through source reconstruction and time-varying connectivity measures, we then investigated how the integration of multimodal emotional vocalization happens in the brain of adults and of HL-ASD infants in Chapter 4. While some of the data we present must be considered preliminary due to the Covid-19 pandemic, our results uncover the presence of early intercommunications between networks subserving sensory processing and higher-order cortical areas in response to social saliency. Specifically, we propose that the brain is hard-wired to social cues, and social cues scaffold networks interactions through reciprocal feedbacks between sensory and associative regions belonging to the social brain. This tuning is seemingly absent in HL-ASD infants across studies, as we found a lack of synchronization between networks involved in social processing and perceptual and cognitive systems.

Overall, our result support the progressive specialization of brain networks. Specifically, while brain activations may be undifferentiated in the beginning, with social experience patterns of activations specialize to specific sets of social cues, such as faces and voices and their integration. We thus hypothesise that the loops connecting sensory and higher-order areas influence the networks involved in social and perceptual processing and their deviations from typical development.

Abstract italiano

I primi mesi di vita del bambino sono caratterizzati da radicali cambiamenti evolutivi che pongono le basi per lo sviluppo successivo. A livello cerebrale, nel corso dei primi mesi di vita risposte non specifiche e subottimali si specializzano e, man mano che il bambino cresce, questo processo di specializzazione getta le basi per lo sviluppo dei network cerebrali osservati negli adulti. In tale contesto gruppi di aree cerebrali si specializzano progressivamente per l'elaborazione di stimoli sociali, andando a formare quel che viene definito con il termine social brain. L'emergere del cervello sociale rappresenta un importante spartiacque tra le traiettorie di sviluppo tipiche e quelle atipiche, specialmente in disturbi fortemente caratterizzati da deficit di natura sociale, quali il disturbo dello spettro autistico (ASD; Grossman & Johnson, 2007).

Lo scopo di questa tesi è quindi di caratterizzare i network cerebrali che supportano l'elaborazione di stimoli sociali nella prima infanzia e di indagare tali network all'interno di traiettorie di sviluppo atipiche, quali sono quelle osservate in bambini con maggiore probabilità di sviluppare ASD (HL-ASD). Nel Capitolo1 abbiamo indagato lo sviluppo dei network cerebrali coinvolti nell'elaborazione di semplici stimoli uditivi attraverso tecniche di ricostruzione di sorgente. Nel Capitolo 2 ci siamo focalizzati sull'elaborazione della salienza sociale delle voci in bambini a basso rischio (LR) e in bambini HL-ASD tramite tracking corticale. Nel Capitolo 3 ci siamo focalizzati sull'elaborazione di volti emozionali attraverso tecniche di decoding. Dato che volti e voci sono spesso associati, nel Capitolo 4, attraverso tecniche di ricostruzione di sorgente e di connettività time-varying, abbiamo indagato come avviene nel cervello l'integrazione di vocalizzazioni emozionali multimodali.

Nonostante i dati presentati siano a tratti preliminari a causa della pandemia di Covid-19, i nostri risultati rivelano la presenza di capillari intercomunicazioni tra network sensoriali e aree associative in risposta a cambiamenti nella salienza sociale degli stimoli. Nello specifico, il cervello sembra essere specificamente orientato all'elaborazione di segnali sociali e, a loro volta, i segnali sociali strutturano le interazioni tra network attraverso feedback reciproci tra le aree appartenenti al social brain. Questa regolazione sembra non essere presente nei bambini HL-ASD negli studi presentati, e tale dato potrebbe contribuire a spiegare i deficit associati all'ASD.

Complessivamente, quindi, i nostri risultati supportano la presenza di meccanismi progressivi di specializzazione, in cui risposte inizialmente non differenziate si specializzano tramite l'esposizione a segnali sociali, quali volti e voci, per mezzo dei loop descritti, caratterizzando sia lo sviluppo tipico che quello atipico.

General Introduction

In the first year of life the infant is subjected to dramatic development. As the infant develops, interactions with the environment become increasingly complex and at the forefront of this complexity is social behaviour (Grossmann & Johnson, 2007.; van der Velde, White, & Kemner, 2021). Indeed, since learning from conspecifics is one of the greatest adaptations of our species, the infant in the first months of life is required to reach the fundamental milestones enabling the development of social and learning abilities (Grossmann & Johnson, 2007). As social interactions become more complex, underlying brain areas and their interactions develop accordingly (van der Velde et al., 2021). Indeed, while the newborn brain shares many features with the adult brain, it is unoptimized due to an overabundance of connections (Huttenlocher & Dabholkar, 1997; van der Velde et al., 2021). Across infancy and childhood brain networks are thus subjected to massive changes, in the form of pruning of unnecessary and the strengthening of long-range crucial connections (Huang et al., 2015; Tóth et al., 2017; van der Velde et al., 2021). Behaviourally, these changes are reflected in the emergence of more complex behaviours (van der Velde et al., 2021). Within this context, the interactive specialization model provides a useful framework to interpret the development of the interrelationships between brain areas (Grossmann & Johnson, 2007; Johnson, 2011; van der Velde et al., 2021). According to the interactive specialization, early in development brain areas are unspecialized and respond to wide arrays of non-specific cues (van der Velde et al., 2021). As the infant grows, repeated interactions between brain areas in response to sets of stimuli lead to a narrowing in the specificity of responses (Jones, Venema, Lowy, Earl, & Webb, 2015; van der Velde et al., 2021). Thus, behavioural development arises not from the physiological maturation of a single area, but through the collective synchronizations of wide-scale networks (Grossmann & Johnson, 2007; Jones et al., 2015; van der Velde et al., 2021). In this context, certain brain regions become gradually specialized for processing particular features of social stimuli and are thus recruited into what is called the social brain (Grossman & Johnson, 2007; Jones et al., 2015).

The social brain is a network of brain areas underlying complex social behaviours (van der Velde et al., 2021; Grossman & Johnson 2007). Main hubs of the social brain are the fusiform gyrus, the superior temporal sulcus (STS), and the orbitofrontal cortex (Elsabbagh & Johnson, 2016; van der Velde et al., 2021). Interestingly, in questioning the early emergence of these brain regions as social brain hubs, it has been proposed that their early development is due to their extensive connections with sensory areas, positioning them as integration hubs devolved to inputs requiring highly temporally precise processing (Elsabbagh & Johnson, 2016). Unfortunately, most research comes from

adults, leaving developmental trajectories mostly obscure (Elsabbagh & Johnson, 2016). However, understanding how the social brain emerges and develops is of fundamental importance for social, educational, and clinical policies (Grossmann, 2015; Grossmann & Johnson, 2007). While adults and infants show similarities in their processing of social information, highlighting the early functioning and thus relevance of brain mechanisms involved in the perception of conspecifics (Grossman & Johnson, 2007), social processing is less specialized in infants and requires the involvement of wider brain networks (Grossman & Johnson, 2007). Then, activity in social brain areas becomes gradually more fine-tuned during development due to the strengthening of activation patterns (Grossman & Johnson, 2007). This view has implications for atypical development, especially in disorders characterized by disruptions in social processing, such as autism spectrum disorder (ASD; Grossman & Johnson, 2007).

Social communication and interaction impairments comprise core features of ASD (Bradshaw et al., 2021). ASD is characterized by deficits in social interactions, communication, and by restricted interests and/or repetitive behaviors (American Psychological Association, 2013). While a stable diagnosis is mostly achieved after three years of life, early signs of the disorder emerge during the first year (Bradshaw et al., 2021; Lloyd-Fox et al., 2013; Righi, Tierney, Tager-Flusberg, & Nelson, 2014; Riva et al., 2018; Shic, Macari, & Chawarska, 2013). However, thus far reliable vulnerabilities in the first year have been difficult to identify (Bradshaw et al., 2021; Lloyd Fox et al., 2013; Shic et al., 2013). In order to detect early abnormalities, the focus is increasingly on infant siblings of children diagnosed with ASD, referred as high-likelihood for ASD (HL-ASD) infants (Riva et al., 2018; Lloyd Fox et al.,2013). Indeed, ASD is a complex heritable condition with estimates of inheritance for siblings of approximately 18.7% (Riva et al., 2018; Ozonoff et al., 2011; Shic et al., 2013). In this sense, studies on siblings provide the unique opportunity to identify biomarkers of the disorder aiding early diagnosis and interventions (Riva et al., 2018, 2022; Shic et al., 2013). Moreover, prospective studies have shown the complex interweaving of neurodevelopmental and behavioural trajectories in the first year of life, which would have remained elusive focusing only one developmental time point (Klin & Jones, 2015; Bradshaw et al., 2021). This suggests a complex interaction between brain structures and connectivity, genetic liability, and environment that characterizes infants that will develop ASD during critical developmental windows (Bradshaw et al., 2021). Considering the centrality of social impairments in ASD, a key question regards when, how, and why atypical orientation toward social stimuli emerges in ASD (Shic et al., 2013). Understanding the ontogeny of both typical and atypical early social trajectories gives us a unique window on the developmental

progression that should lead to communicative and social competence at later stages and what happens in response to deviations (Shic et al., 2013).

From a neural point of view, socio-communication dysfunctions in ASD have been attributed to atypical development in the networks underling the social brain (Elsabbagh & Johnson, 2016; Lloyd Fox et al., 2013). As the brain is hardwired to respond to salient stimuli, especially social ones, early failures in these mechanisms could interfere with the achievement of critical socio-communicative milestones (Elsabbagh & Johnson, 2016; Lloyd Fox et al., 2013). Then, the cascading effects of these failures could preclude more advanced learning (Elsabbagh & Johnson, 2016; Lloyd Fox et al., 2013). Nonetheless, it is important to emphasize that the debate regarding the source of socio-communicative impairments in ASD is still very active. Indeed, deficits could be also explained by abnormalities in basic functions across sensory and cognitive domains (Ronconi, Molteni, & Casartelli, 2016). In this context a challenging question regards whether the integration of complex stimuli derive from perceptual complexity per se, or given that social stimuli are inherently complex, they are the result of a failure in assigning appropriate salience to social stimuli. For example, there are reports of reduced orientation to social stimuli, such as faces and voices, compared to non-social stimuli, such as objects and sounds, in infants later diagnosed with ASD (Blasi et al., 2015; Elsabbagh & Johnson, 2016; Gervais et al., 2004; Maslennikova, Portnova, & Martynova, 2022). However, other reports suggest that impairments in face processing arise only during complex, dynamic, and emotionally laden conditions (Shic et al., 2013; Lloyd Fox et al., 2013). Thus, in order to make sense of seemingly contrasting results it is necessary to consider how information is integrated across multiple brain regions and in response to a wide range of stimuli (Elsabbagh & Johnson, 2016). Despite its crucial importance, there is relatively little research on how the binding of social and/or perceptual contingencies occurs in HL-ASD and its neural underpinnings. This is surprising given that from a neuroconstructivist point of view tiny variations in basic processes at initial developmental stages can lead to marked differences in final ones, producing a wide range of individual differences, including atypical outcomes (Karmiloff-smith, 2009). Specifically, due to the context-dependent nature of development, and because early on brain regions are highly interconnected, disruptions in one part of the system often have cascading effects over time on other parts of the system, affecting interactions with the environment (Karmiloff-smith, Scerif, & Thomas, 2002). Thus, as the majority of research has been conducted on adults, notwithstanding the wealth of information it gave us on the neural mechanisms underlying complex social processing (Grossmann, 2015), the question still stands. How the processes enabling the brain to interact with the environment develop? What are their critical developmental precursors? What happens when something goes wrong?

The overall aim of this thesis is to characterize social processing across development at increasing social complexity and to describe the deviations from typical development represented by HL-ASD infants. In **Chapter 1, Developmental trajectories of low-level auditory processing in the infant brain**, we addressed basic auditory processing by manipulating sounds' duration in 4- and 9-months-old infants. More in depth, in this study we applied age-appropriate anatomical templates (O'Reilly, Larson, Richards, & Elsabbagh, 2021) to spatially localized brain networks involved in low-level auditory processing across development. Ours is one of the few studies attempting this, as fMRI studies are complex to conduct in developmental samples. Nevertheless, improving functional localization of activity in the infant brain is critical for our understanding of neural developmental trajectories and their deviations from the norm (Grossmann & Johnson, 2007; Sperdin & Schaer, 2016). Moreover, to address complexity it is first necessary to start with the basics. In **Chapter 2, How does it sound? Social saliency affects voice cortical tracking in low risk developing and HL-ASD infants**, we increased the level of complexity through the investigation of cortical tracking of the human voice in low risk (LR) and HL-ASD 18-months-old infants. We presented infants with three types of voices positioned along a social salience continuum: the voice of an actress, the mother's voice, and a vocal synthesiser narrating a nursery rhyme. Hence, by manipulating prosodic information and identity we set out to answer whether HL-ASD infants are generally less sensitive to conspecifics communicational cues, especially of close ones such as their own mother, or they are characterized by general failures in the perceptual processing of complex information. Voice processing was investigated through the novel perspective of cortical tracking of external stimuli. By assuming that the brain is able to quickly react to environmental stimuli through the tuning of the oscillations mediating communications between distant brain areas (Jones et al., 2015; van der Velde et al., 2021), we provided a psychophysiological basis of voice processing in both LR and HL-ASD infants. In **Chapter 3, Decoding functional brain network involved in infants' processing of emotional faces**, we investigated face processing in 7-months-olds in response to static and dynamic emotional faces. By assuming that the infant brain is specialized enough to already show the subdivisions of the visual system involved in static and dynamic visual processing (Pitcher, Duchaine, & Walsh, 2014), we decoded the brain networks involved in the processing of emotional faces. Thus, following the same line of thought as in Chapter 2, we investigated whether infants brain activity contains enough information to differentiate the face-sensitive networks observed in adults through decoding techniques. Having thoroughly investigated auditory and visual unimodal perception in the previous chapters, we addressed multimodal binding of socio-emotional stimuli in 9-months-old HL-ASD and LR infants in **Chapter 4, Multimodal temporal binding of emotional vocalizations in**

typical and atypical brain development. In this study we presented participants with synchronous and asynchronous AV recordings of two actresses crying, laughing, and coughing to both LR and HL-ASD infants and to an adult sample, while applying source reconstruction and time-varying connectivity techniques. Indeed, since the exact brain mechanisms involved in intersensory binding across development are still largely unknown, but such mechanisms are fairly well-known in adults, we included an adult sample to which refer results. We expected HL-ASD infants to show delays and impairments in the binding of asynchronous stimuli confirming impairments in temporal synchronization across neural networks in ASD (Myers, Lense, & Gordon, 2019). In Figure 1 are summarized the studies reported in each chapter and the analytical strategies applied.

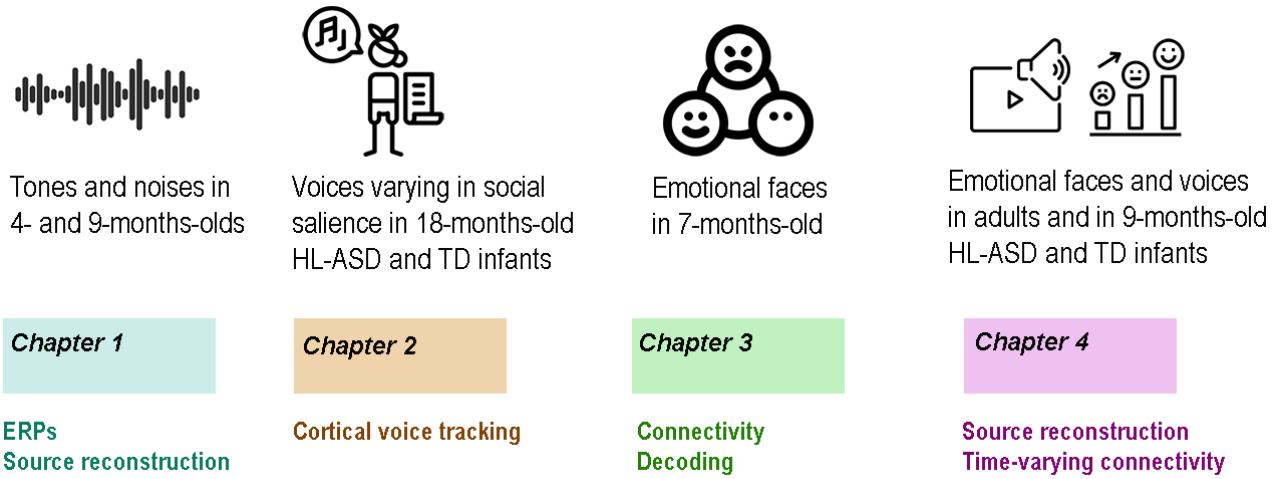


Figure 1. *Visual representation of the studies with related age of samples and analysis performed.*

Chapter 1

Developmental trajectories of low-level auditory processing in the infant brain.[1]

As we outlined in the introduction, seemingly insignificant impairments in early infancy can have great cascading effects on developmental processes. This may result in the considerable impairments characterizing neurodevelopmental disorders such as ASD (Elsabbagh & Johnson, 2016). Indeed, higher-order developmental acquisitions are heavily influenced by competence in underlying sensory systems (Vouloumanos & Curtin, 2014). For this reason, before delving deeper into higher-order social processes and their role in the dysfunctions observed in ASD, it is especially important to address the functioning and timing of their building blocks. We shall do so using state-of-the-art analytical methods, in order to depict neural organizations scaffolding auditory sensory perception. Indeed, to single out cortical contributions we will apply source reconstruction techniques, an area of research remained largely unexplored in developmental literature. We will do so while remaining grounded in the well-known ERPs literature, to aid results interpretation and their relation to previous reports.

1. Introduction

One of the main characteristics of auditory information is that sound unfolds in time. Therefore, the integration of information across time is crucial for perception on multiple timescales (Johnson & Hautus, 2010). Infants are sensitive to a number of temporal sound features that also characterize language at birth, such as gaps, sound onset and offset, and changes in sound duration (Benasich et al., 2006; Kushnerenko, Fellman, et al., 2001; Háden, Honing, Török, & Winkler, 2015). Therefore, these abilities are thought to be innate and to support the foundation of language acquisition (Benasich et al., 2006; van den Heuvel et al., 2015; Kudo, Nonaka, Mizuno, Mizuno, & Okanoya, 2011). Similarly to adults, infants utilize temporal cues, such as phoneme duration, for segmenting sound sequences (Obrig, Rossi, Telkemeyer, & Wartenburger, 2010), including finding words within continuous speech (Kudo et al., 2011). In this sense, the ability to process rapidly changing sounds has been suggested to be a critical skill for language development (Benasich, Thomas, Choudhury, & Leppänen, 2002), while the ability to perform fast and

[1] Parts of this chapter are included in Polver, S., Háden, G.P., Bulf, H., Winkler, I., Tóth, B. Early maturation of sound duration processing in the infant brain. Scientific reports. (under revision)

fine-grained acoustic analyses is critical for speech decoding (Cantiani et al., 2016; Tallal & Gaab, 2006). Moreover, neurophysiological findings suggest that neurons in early auditory regions exhibit complex spectro-temporal receptive fields and participate in high-level encoding of auditory features at multiple timescales (Gervain & Geffen, 2019). Despite their importance, however, the mechanisms underlying the processing of auditory temporal features are still poorly understood.

The present study aimed to investigate the early development of the processing of an auditory temporal feature tightly linked to time: sound duration. To this end, auditory event-related brain potentials (ERP) and their neural generators elicited by tones and noises of different duration were compared between 4 and 9-month-old infants. ERPs have been long used to study the neural substrate of auditory temporal processing and their development in infancy (Benasich et al., 2006; Kushnerenko, Fellman, et al., 2001; Kushnerenko, Van den Bergh, & Winkler, 2013). Specifically, some previous ERP studies have shown that the newborn's brain is sensitive to sound duration as well as unexpected changes in the duration of a repeating sound (Kushnerenko, Fellman, et al., 2001; Kushnerenko, Cheour, et al., 2001; Pihko et al., 1999). The adult ERP components sensitive to sound duration comprise the P1-N1-P2-N2 complex (Kushnerenko, Fellman, et al., 2001). The adult P1–N1–P2–N2 sequence begins at around 40-50 ms after stimulus onset and lasts until ca. 150-250 ms (Ponton, Eggermont, Kwong, & Don, 2000). Similar waveforms have been observed in newborns, infants, and in 5 to 10 years old children (Kushnerenko, Fellman, et al., 2001). However, it is debated whether the underlying functional mechanisms and brain areas involved are the same as those observed in adults (Choudhury & Benasich, 2011). Indeed, several maturational mechanisms are at play across development and can affect ERPs (Korpilahti et al., 2007). Among others we can cite advancing myelination and increasing synaptic connections (Korpilahti et al., 2007).

From what we do know, generally neonatal ERPs exhibit a large positive waveform at midline starting at about 100 ms and ending 450 ms from stimulus onset, followed by a low-amplitude negative deflection at approximately 450-600 ms (N450). These waveforms are thought to reflect precursors of the P2 and N2 (Kushnerenko et al., 2002). At about 3 months of age, the P2 is reportedly split by the appearance of a negative deflection, and between 3 and 6 months, another deflection peaking at about 250-280 ms (N250) appears. The two negative waveforms are assumed to represent precursors of the adult N1 and N2, respectively (Kushnerenko et al., 2002). By 6- to 9 months of age the separation between the N1 and the N2 precursor waveforms becomes well-defined, making the infantile ERPs' morphology resemble the P1-N1-P2-N2 complex (Choudhury & Benasich, 2011; Fellman et al., 2004; Jing & Benasich, 2006). Thus, the developmental trajectories of ERP waveforms point toward the presence of underlying developmental

milestones from 3 to 9 months of age (Choudhury & Benasich, 2011; Fellman et al., 2004; Jing & Benasich, 2006). It is worth noting, however, that scalp recordings do not directly reflect the underlying generators. Sound perception activates multiple pathways in the brain. Hence, scalp-recorded waveforms may represent summed activity from several generators (Bishop, Hardiman, Uwer, & Von Suchodoletz; 2007; Conte & Richards, 2022). Moreover, different ERP generators may have different maturational courses (Kushnerenko et al., 2002). Therefore, discerning and locating these generators and addressing their developmental trajectory provides important information for anchoring functional changes in brain maturation and plasticity. In this context, individuating brain areas involved in low-level sensory processing can provide the basis for discussing higher-order brain processes, as we shall do in the following chapters.

In adults, generators of the P1-N1-P2-N2 complex have been found to reside in the primary auditory cortex (Heschl's gyrus), the superior temporal gyrus (STG), in thalamo-reticular systems, in the planum temporale (encompassing Wernicke's region), and in supratemporal auditory cortices (Bishop et al., 2007; Hautus & Johnson, 2005; Mayhew, Dirckx, Niazy, Iannetti, & Wise, 2010; Snyder, Alain, & Picton, 2006). Interestingly, some of these regions, such as the STS are also involved in sensory integration mechanisms, suggesting a tight link between sensory and associative cortices (Stevenson et al., 2016). Regarding developmental populations, source analysis revealed that presenting syllable to 6-month-old infants activated bilaterally the auditory and frontal cortices, and the anterior cingulate cortex (Ortiz-Mantilla, Hämäläinen, & Benasich, 2012). 6-months-old infants also show activity in supratemporal and frontal areas in response to sound discrimination (Hämäläinen, Ortiz-Mantilla, & Benasich, 2011). However, until recently, the lack of age-appropriate templates has complicated the application of source localization in infants, because the quality of age-appropriate head models determines the accuracy of source localization (Conte & Richards, 2022).

In this study, we investigated duration-sensitive auditory ERP waveforms and their sources to establish their developmental changes between 4 and 9 months of age. We chose these two time points, because they bracket important milestones in the development of ERP morphology (see above). Infants were presented with four sounds differing in duration (200 vs. 300 ms) and sound quality (harmonic tones vs. white noise). We expected to find significant changes in ERP morphology between the two age-groups in accordance with previous descriptions of ERP development. Indeed, in contrast to the simpler ERP component structure at 4 months, by 9 months of age we expected to find precursors of the adult ERP complex in the form of a positive deflection, accompanied by two negative ones coinciding with the N250 and the N450 latencies (Kushnerenko et al., 2002). We then asked whether either or both waveforms show sensitivity to sound duration

and whether sensitivity to sound duration is dependent on the quality of the sound (harmonic tones vs. noise segments).

In this thesis we rely on the recently released infant templates (O'Reilly et al., 2021) derived from the Neurodevelopmental MRI Database (Sanchez, Richards, & Almli, 2012) and made ready for use by their implementation in the Brainstorm software (Tadel, Baillet, Mosher, Pantazis, & Leahy, 2011). We expected to find the main sources of duration-sensitive activity in brain areas belonging to the auditory brain network, including primary auditory cortex (PAC), and activity in network areas previously shown to be sensitive to sound duration, such as the STG and inferior frontal areas (Bueti, Dongen, Van & Walsh, 2008), in line with what reported in the Introduction. Further, in line with our expectations for scalp-recorded ERPs, we hypothesized that 9-months-old infants show stronger activity in a more distributed network comprising fronto-central areas compared to 4-months-old infants in the time window(s) of the effect(s) found over the scalp.

2. Methods

2.1. Participants

Infants were recruited as part of a broader longitudinal study carried out at Institute of Cognitive Neuroscience and Psychology, Research Centre for Natural Sciences, Budapest, Hungary to assess the role of infant-directed speech in speech acquisition. In this chapter, we will refer only to two separate age cohorts: the first one comprised sixty-four (20 males) 4-months-old, the second cohort sixty-three (19 males) 9-months-old healthy full-term infants. All infants were firstborns, none of them twins. The average age at the recording time was 4.22 months (SD = 0.2) for the 4-months-old and 9.52 (SD = 0.36) for the 9-months-old group. Data recorded from thirty-one 4-months-old infants and from eighteen 9-months-old infants was excluded due to the presence of excessive artifacts. Thus, thirty-three 4-months-old and forty-five 9-months-old infants' data were included in the final sample. Due to the longitudinal nature of the study, 9 infants in the final sample were tested at both 4 and 9 months. However, given the small number of infants being tested at both time points, we did not investigate longitudinal effects in this study.

Recordings were carried out at the sound attenuated infant laboratory and informed consent was obtained from one (mother) or both parents. The study was conducted in full accordance with the Declaration of Helsinki and all applicable national and international laws, and it was approved by the United Ethical Review Committee for Research in Psychology (EPKEB), Hungary.

2.2. Stimuli and procedure

All infants were presented with shorter (200 ms) or longer (300 ms) sounds grouped into two separate blocks each consisting of 150 short and 150 long sounds delivered in random order. One of the stimulus blocks presented tones of 500 Hz base frequency with 3 harmonics of 50, 33, and 25 percent amplitude, respectively, summed linearly together ('tone' condition). The other stimulus block comprised frozen white noise segments ('noise' condition). All sounds were presented at approximately 70 dB SPL loudness and the sounds were attenuated by 5 ms long raised-cosine onset and offset ramps. The onset-to-onset interval was 800 ms. The tone condition was always delivered first, and there was a short rest of 30-60 seconds between conditions. Stimuli were created and delivered using Matlab (MathWorks Inc., Natick, MA, USA) and Psychtoolbox (Kleiner, Brainard, & Pelli, 2007) software. The sound signal from the computer was amplified by a Yamaha A-S301 amplifier (Hamamatsu, Japan) and presented through a pair of speakers (Boston Acoustics A25, Woburn, MA, USA). The speakers were positioned ca. 1.75 meters in front of the participant, 70 cm apart from each other.

The infant sat comfortably in her mother's lap while the experimenter employed toys to keep the infant facing toward the loudspeakers and his/her attention away from the electrode net. The mother was listening to music through closed can audiometric headphones to isolate her from the experimental stimulation. If the infant became fussy, the playback was stopped, and the experimenter attempted to pacify him/her. If pacifying was successful the playback restarted from the beginning, otherwise, the experiment was discontinued. Because some infants did not complete both stimulus blocks, thirty-three 4-month-old and forty-five 9-month-old infants' data was analysed for the tone condition, and twenty-three 4-months-old and forty-one 9-months-old infants' data for the noise condition.

The experiment reported here was presented first within a session combining multiple experiments. Data from the other experiments and the associations across results will be published separately after collecting the complete data set. As the current experiment was always the first one presented, the other experiments are not listed.

2.3. EEG data acquisition and preprocessing

EEG was recorded with a 60-channel HydroCel GSN net (64 channel v1.0 layout, geodesic; channels 61-64 were connected to the ground in the small pediatric caps used) and a GES 400 DC amplifier passing the digitized signal to a computer running the NETSTATION v5.4.1.1 software (both Electrical Geodesics, Eugene, OR, USA). Signals were recorded online at a sampling rate of 500 Hz with the Cz reference (DC, no on-line filtering). Electrode impedance during recording was kept below 50 kΩ.

The recorded signals were imported to MATLAB (Mathworks, Natick, MA; ver. 2021a) and processed using EEGLAB (Delorme & Makeig, 2004). First, data were band-pass filtered between 0.1 Hz and 30 Hz using a finite impulse response (FIR) filter (Kaiser

windowed, Kaiser β=5.65, filter length 4530 points). Data were then cleaned using the Artifact Subspace Reconstruction (ASR) algorithms in EEGLAB (Plechawska-Wójcik, Tokovarov, Kaczorowska, & Zapała, 2019): electrodes were removed and stored for subsequent interpolation if either they were not functioning for longer than 5 seconds, or if their correlation with neighbouring electrodes was under 0.8. Correlation was computed for each electrode with the RANdom SAmple Consensus (RANSAC) reconstruction. The reason behind the latter choice was the assumption that nearby channels should be characterized by fairly similar activity. Signals from the remaining electrodes were then entered into a Principal Component Analysis (PCA) with a 5-s sliding window, classifying PCs into high-variance (10 SD from clean data) or normal-variance sets. High-variance parts of the data, which correspond to artifacts, were then reconstructed based on normal-variance subspace components. Infants' data were rejected either if more than 9 channels were removed or if more than 2 removed channels were neighbours. This led to the exclusion of twenty-eight 4-months-old and fourteen 9-months-old infants' data from the analysis. To further remove artifacts, an Independent Component Analysis (ICA) was performed on 45 PCA-reduced components after re-referencing them to the average reference. The topographical distribution and frequency contents of the ICA components were visually inspected, and the components constituting oculomotor, cardiac, muscular, or line artifacts were manually removed. The maximum number of rejected components was set at 11. After these steps, electrodes removed in the first artifact rejection step were spherically interpolated with the help of the complete electrode matrix.

2.4. ERP analysis

From the continuous EEG recordings, epochs were extracted between -100 and 800 ms relative to the sound onset, separately for the two stimulus durations (short- and long-duration sounds), stimulus types (tones and noises), and age groups (4- and 9-month-olds). The average voltage in the -100 and 0 ms window served as the baseline. A threshold of +/- 150 μV was set to reject epochs with abrupt amplitude changes, while thresholds of 100 μV over the whole epoch for slope and 0.5 r^2 were set to remove epochs with step-like artifacts and linear trends. Infants' data was not used if the remaining number of epochs for each stimulus duration was under 75 for either stimulus type. This led to the additional exclusion of three 4-months-old and four 9-months-old infants' data. The final sample thus consisted of thirty-three 4-months-old infants, of which 10 did not complete the noise condition, and forty-five 9-months-old infants, of which 4 did not complete the noise condition. Epochs of each remaining infant were averaged separately for each stimulus duration, stimulus type, and age group. The mean number of epochs are reported in Table 1. No significant differences were found between average epoch numbers of conditions and group.

Table 1*. Mean number of epochs remaining after preprocessing reported separately for stimulus duration, stimulus type, and age group.*

	Short tones	Long tones	Short noises	Long noises
4-months-old	133.53 (SD = 15.6)	133.96 (SD = 15.13)	31.04 (SD = 19.69)	131.78 (SD = 18.16)
9-months-old	134.82 (SD = 21.47)	134.17 (SD = 18.95)	135.04 (SD = 30.24)	136.87 (SD = 30.08)

For ERP analysis, we first calculated difference waveforms by subtracting the average ERP elicited by short stimuli from the average ERP elicited by the corresponding long ones. In accordance with previous findings (Čeponienė et al., 2002, 2004; Kushnerenko et al., 2001, 2002), we observed two successive negative peaks in the difference waveforms (Figure 1 and 2; early and late negativity: N250 and N450, respectively in Kushnerenko et al., 2002). For component latencies, the right frontal (F4) grand average (GA) waveform was visually inspected. F4 was selected as the signal from this electrode typically shows the highest difference in amplitude (Kushnerenko et a., 2002).

Looking at Figure 1A and at Figure 2A the early negative difference peak elicited by 4-month-olds barely passes the baseline. However, judging from the voltages surrounding the peak, the negative waveform may be riding on a positive shift or a wide positive waveform. Therefore, we decided to measure the early component's amplitude by subtracting from the average voltage measured from a 48 ms long window centered on the peak (the most negative point in the GA response at F4) from the average voltage in two windows of 48 ms duration, one on each side of the peak, separately for each stimulus type and age group. The latencies of the two baseline side-windows were determined by starting from adjacent positions on both sides of the peak, then separately sliding the windows in 24 ms steps away from the peak until their average voltage in both side-window was significantly lower (Student's t-test; $p < 0.01$) than that of the peak window. After establishing the windows (separately for the two stimulus-types and two age groups; see Figure 1A and Figure 2A), the early peak's amplitude was measured from the same windows in each infant's individual average.

Amplitude measurement windows of the late negative waveform were established from the GA of the 9-months-old group at F4, separately for tone and noise segments (Figure 1A and Figure 2A). We chose this procedure because the late negative component was only shown clearly in the responses of the 9-months-old infants. The peak was again

set as the most negative point. Amplitudes for the late peak were then measured as the average voltage in 48 ms long windows centered on the GA peak for all infants (see Figure 1A and Figure 2A), separately for responses to tones and noise segments.

Statistical analyses of the difference waveform amplitudes were performed separately for the early and late components and the two stimulus types (tone, noise) by mixed-mode ANOVAs with within-subject factors of anterior-posterior (frontal, central, parietal) and laterality (left, central, right), and the between-subject factor of age group (4-, 9-months-olds). The scalp distribution factors (anterior-posterior and laterality) were based on a matrix of 3 × 3 electrodes: F3, Fz, F4, C3, Cz, C4, P3, Pz, and P4. For post-hoc testing of significant interactions between age group and the anterior-posterior and/or laterality factor(s) ANOVAs were separately performed for each level of the scalp distribution factors. The false discovery rate (FDR) correction was applied to p-values to account for multiple comparisons (Benjamini & Hochberg, 1995). The p values reported reflect the FDR corrected values. The alpha level was set to 0.05. Only significant results are reported. The partial eta squared (ηp^2) measure of effect size is also reported.

2.5. Source localization

The Brainstorm toolbox (Tadel et al., 2011) was used to perform EEG source reconstruction, following the protocol of previous studies (Stropahl, Bauer, Debener, & Bleichner, 2018). Age-appropriate templates for both groups were used to derive default anatomical regions (O'Reilly et al., 2021), which, along with the default electrode locations, were entered into the forward boundary element head model (BEM) provided by the openMEEG algorithm (Gramfort, Papadopoulo, Olivi, & Clerc, 2011). The tissue conductivity values were those used in O'Reilly et al. (2021) in 7-months-old infants (0.33 S/m for the gray matter, 0.0041 S/m for the skull, and 0.33 S/m for the scalp).

For the modelling of time-varying source signals (current density) of all cortical voxels, a minimum norm estimate inverse solution was applied after appropriate noise covariance estimation. Matrices were then scaled through dynamical Statistical Parametric Mapping normalization (Dale et al., 2000). Averaging current density across voxels yielded time series for 62 cortical areas (region of interest ROIs), defined by the standardized parcellation scheme introduced by Desikan and Killiany (2006). Finally, for all ROIs, we extracted the average signal for the corresponding ERP peak windows, separately for each sound duration (short, long), stimulus type (tones, noise), and age group (4, 9 months).

Based on visual inspection of the source activity on the cortical maps, we selected the bilateral ROIs from the Desikan-Killiany parcellation scheme (Desikan et al., 2006) for subsequent statistical analysis. Frontal regions: all three anatomical partitions (opercular, orbital, and triangular) of the Inferior Frontal Gyrus (opIFG, orbIFG, triIFG) and the Orbital

Frontal Gyrus (OFG) in its medial and lateral parts; parietal regions: the Supramarginal Gyrus (SMG) and the Inferior Parietal Gyrus (IPG); temporal regions: the Superior Temporal Gyrus (STG), the Inferior Temporal Gyrus (ITG), the Middle Temporal Gyrus (MTG), the bank of the Superior Temporal Sulcus (STS), and primary auditory cortex (PAC).

For statistical analyses, average source amplitudes in the selected regions were tested separately for the early and late time windows, for tones and noise segments, and ROIs by between-subjects ANOVAs with the age group factor (4-, 9-months-olds). All other aspects of the statistical analysis are identical to that employed for ERPs.

3. Results

Figure 1 shows the grand average ERP difference responses and their neural generators for tones (panels A, B, and C), while Figure 2 shows the same for noise segments (panels A, B, and C) at 4 and 9-month of age.

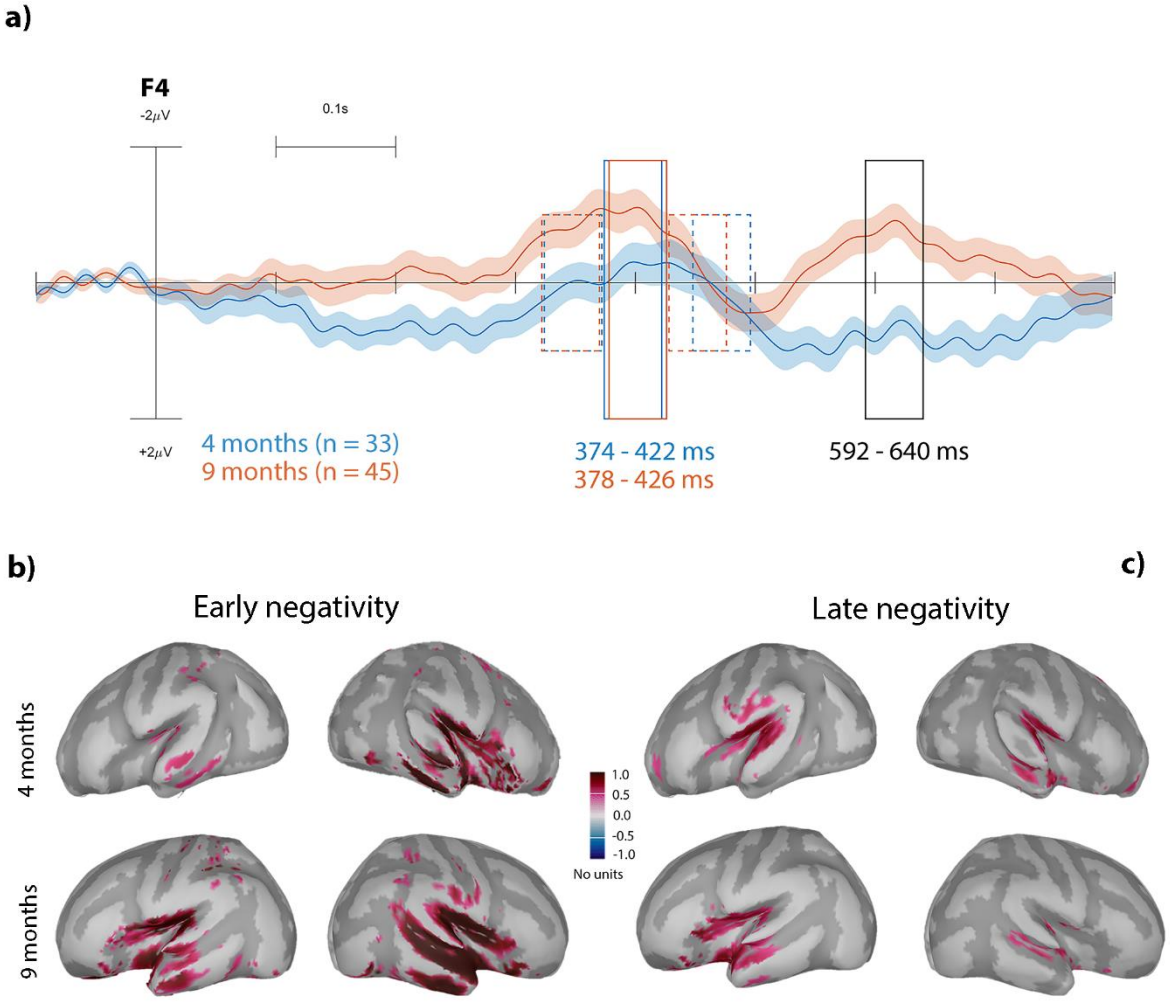


***Figure 1**. Grand average ERP difference waveforms elicited by tones. Panel A) overplots the right frontal (F4) ERP responses of the two age groups (marked by line colours). The early and the late measurement windows are shown with rectangles drawn with solid lines (the exact range shown under the rectangle). The rectangles with dashed lines refer to the reference interval used for calculating the amplitudes for the early waveform. Panels B) and C) show the distribution of the*

average current source density (CSD) over the inflated cortex during the early (panel B) and late (panel C) duration-sensitive ERPs elicited by tones. The upper row presents the 4-months old infants' data, the lower that of the 9-month-olds. Colour scale is at the center of the cortical distribution panels.

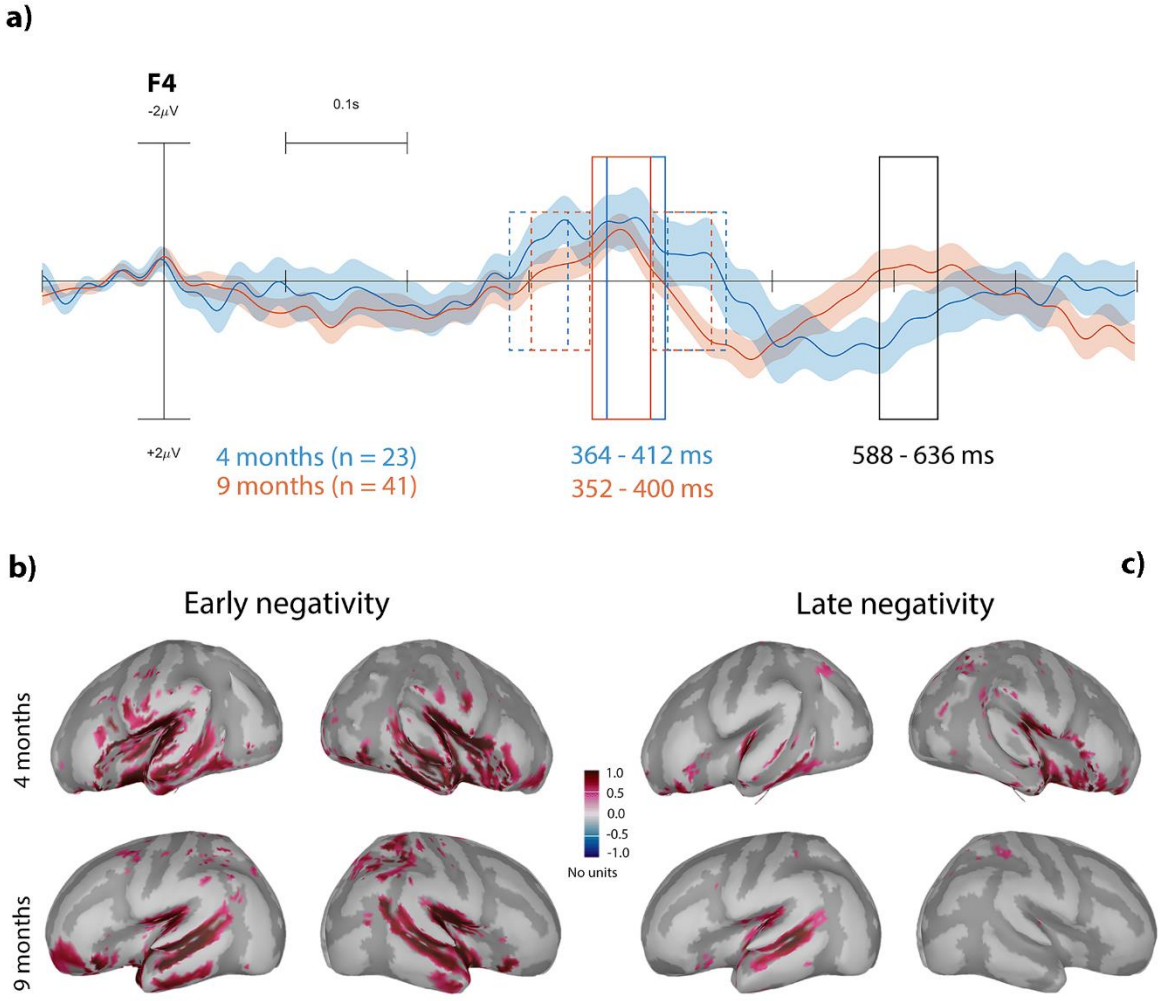


***Figure 2**. Grand average ERP difference waveforms elicited by noise segments. Panel A) overplots the right frontal (F4) ERP responses of the two age groups (marked by line colors). The early and the late measurement windows are shown with rectangles drawn with solid lines (the exact range shown under the rectangle). The rectangles with dashed lines refer to the reference interval used for calculating the amplitudes for the early waveform. Panels B) and C) show the distribution of the average current source density (CSD) over the inflated cortex during the early (panel B) and late (panel C) duration-sensitive ERPs elicited by noise. The upper row presents the 4-months old infants' data, the lower that of the 9-month-olds. Color scale is at the center of the cortical distribution panels.*

3.1. Early negative difference

For tones, the amplitude measurement window for the peak of the early negativity was set from 374 to 422 ms at 4 months and from 378 to 426 ms at 9 months (Figure 1A). Results of the ANOVA showed significant main effects of anterior-posterior ($F(2,684) = 28.89$, $p < 0.0001$, $\eta p^2 = 0.08$) and age group ($F(1,684) = 12.9$, $p = 0.0008$, $\eta p^2 = 0.02$), and a significant interaction between anterior-posterior and age group ($F(2,684) = 8.14$, $p = 0.0008$, $\eta p^2 = 0.02$). The mean amplitude of the 4-months-old responses was lower than

that of the 9-months-olds (age group main effect): -0.08 μV vs. -0.28 μV, respectively. Post-hoc Tukey's pairwise comparisons for the significant anterior-posterior main effect revealed that frontal ($t(684) = -5.82$, $p < 0.0001$) and central ($t(684) = -6.45$, $p < 0.0001$) sites are characterized by significantly higher early negative responses compared to posterior sites (-0.306 μV and -0.386 μV vs. 0.09 μV, respectively). Post-hoc ANOVAs conducted for the significant anterior-posterior × age group interaction revealed that the groups differed only at the central line of electrodes ($F(1,232) = 36.71$, $p < 0.0001$, $\eta p2 = 0.14$) with 4-months-old infants producing significantly lower-amplitude (less negative) responses than 9-months-olds (-0.08 μV vs. -0.6 μV).

For noise segments, the amplitude measurement window for the peak of the early negativity was set from 362 to 412 ms at 4 months and from 352 to 400 ms at 9 months (Figure 2A). Results of the ANOVA revealed only a significant anterior-posterior main effect ($F(2,558) = 76.82$, $p < 0.0001$, $\eta p^2 = 0.22$). Post-hoc pairwise comparisons revealed that frontal ($t(558) = -10.48$, $p < 0.0001$) and central ($t(558) = -9.71$, $p < 0.0001$) sites are characterized by significantly higher early negativity responses compared to posterior sites (-0.44 μV and -0.38 μV vs. 0.39 μV, respectively).

Generators of the early negative difference. Distributions of the grand-average current source density (CSD) differences are shown on Figure 1B and Figure 2B, for tones and noise segments, respectively. Significant age group effects were found for tones in the right temporal cortices (right MTG, STG, and the bank of the STS) and in the IFG (left pars opercularis). Significant age group effects were found for noise segments in all left partitions of the IFG (left pars opercularis, triangularis, orbitalis) and in the right bank of the STS. All significant statistical findings are reported in Table 2.

Table 2. Testing age effects for source amplitudes in the early time window
ANOVA results comparing mean CSD values in the time window of the early negative difference, separately for the two stimulus types (tones and noise segments; columns) and ROIs (rows).

Cortical ROIs	TONES	NOISE SEGMENTS
Frontal cortex		
Right triIFG	-	-
Left triIFG	-	$F(1,62) = 6$; $p = 0.017$, $\eta p2 = 0.09$
Right opIFG	-	-

Left opIFG	$F(1,76) = 6.08$; $p = 0.015$, $\eta p^2 = 0.07$	$F(1,62) = 4.91$; $p = 0.03$, $\eta p2 = 0.07$
Right orbIFG	-	-
Left orbIFG	-	$F(1,62) = 5.07$; $p = 0.027$, $\eta p2 = 0.08$
Right medial OFG	-	-
Left medial OFG	-	-
Right lateral OFG	-	-
Left lateral OFG	-	-
Temporal cortex		
Right STG	$F(1,76) = 4.17$; $p = 0.044$, $\eta p2 = 0.05$	-
Left STG	-	-
Right MTG	$F(1,76) = 13.06$; $p = 0.005$, $\eta p2 = 0.15$	-
Left MTG	-	-
Right ITG	-	-
Left ITG	-	-
Right bank of STS	$F(1,76) = 9.09$; $p = 0.003$, $\eta p2 = 0.11$	$F(1,62) = 14.57$; $p = 0.003$, $\eta p2 = 0.19$
Left bank of STS	-	-

Right PAC	-	-
Left PAC	-	-
Parietal cortex		
Right SMG	-	-
Left SMG	-	-
Right IPG	-	-
Left IPG	-	-

3.2. Late negative difference

For tones, the amplitude measurement window for the late negativity was set from 592 to 640 ms (Figure 1A). Results of the ANOVA showed a significant main effect of age group ($F(1,684) = 11.41$, $p = 0.002$, $\eta p^2 = 0.02$), and a significant interaction between the anterior-posterior and age group ($F(2,684) = 11.09$, $p = 0.0001$, $\eta p^2 = 0.03$). The mean amplitude of the 4-months-old responses was positive in contrast to that of the 9-months-olds (age group main effect): 0.3 µV vs. -0.18 µV. The post-hoc ANOVA on the anterior-posterior and age group interaction revealed that the groups significantly differed only at frontal ($F(1,232) = 20.6$, $p < 0.0001$, $\eta p^2 = 0.08$) and central electrodes ($F(1,232) = 13.59$, $p = 0.0002$, $\eta p^2 = 0.06$), with the 4-months-old infants responses characterized by positive mean amplitudes (0.55 µV and 0.51 µV for frontal and central sites, respectively) compared to that of the 9-months-olds (-0.52 µV and -0.33 µV for frontal and central sites, respectively).

For noise segments, the amplitude measurement window for the late negativity was set from 588 to 636 ms (Figure 2A). Results of the ANOVA revealed a significant interaction between anterior-posterior and age group ($F(2,558) = 17.09$, $p < 0.0001$, $\eta p^2 = 0.06$). The post-hoc ANOVAs on this interaction revealed that the groups differed at all three electrode lines. At frontal ($F(1,190) = 8.37$, $p = 0.004$, $\eta p^2 = 0.04$) and central

(F(1,190) = 10.3, p = 0.001, ηp2 = 0.05) electrodes, 4-months-old infants (0.44 µV and 0.18 µV respectively) produced positive responses as opposed to 9-months-olds (-0.38 µV and -0.6 µV respectively). In contrast, 4-months-old infants were characterized by negative responses in the late negativity while 9-months-olds by positive ones (-0.65 µV vs. 0.48 µV) at posterior sites (F(1,190) = 16.15, p = 0.0008, ηp^2 = 0.08).

Generators of the late negative difference. Distributions of the grand-average current source density (CSD) differences are shown on Figure 1C and 2C. The effect of age group on the late negative difference to tones was significant in temporal areas (bilateral STG, and right PAC), in the right parietal SMG region, and in the left pars triangularis of the IFG. Significant effects of age group on the late negative difference to noise segments were observed in temporal areas (bilateral STG, and PAC, left bank of the STS, ITG, MTG), in the right SMG, in the left lateral OFG, in the right medial OFG, and in the left pars orbitalis of the IFG. All significant statistical findings are reported in Table 3.

Table 3. Testing age effects for source amplitudes in the late time window
ANOVA results comparing mean CSD values in the time window of the early negative difference, separately for the two stimulus types (tones and noise segments; columns) and ROIs (rows).

Cortical ROIs	TONES	NOISE SEGMENTS
Frontal cortex		
Right triIFG	-	-
Left triIFG	F(1,76) = 8.84; p = 0.003, ηp2 = 0.1	-
Right opIFG	-	-
Left opIFG	-	-
Right orbIFG	-	-
Left orbIFG		F(1,62) = 4.23; p = 0.043, ηp2 = 0.06
Right medial OFG	-	F(1,62) = 4.86; p = 0.031, ηp2 = 0.07
Left medial OFG	-	-

Right lateral OFG	-	-
Left lateral OFG	-	F(1,62) = 5.59; p = 0.021, ηp2 = 0.08
Temporal cortex		
Right STG	F(1,76) = 11.38; p = 0.0011, ηp2 = 0.13	F(1,62) = 8.21; p = 0.005, ηp2 = 0.12
Left STG	F(1,76) = 12.44; p = 0.0007, ηp2 = 0.14	F(1,62) = 5.95; p = 0.017, ηp2 = 0.09
Right MTG	-	
Left MTG	-	F(1,62) = 9.45; p = 0.0031, ηp2 = 0.13
Right ITG	-	-
Left ITG	-	F(1,62) = 6.8; p = 0.011, ηp2 = 0.1
Right bank of STS	-	-
Left bank of STS	-	F(1,62) = 14.52; p = 0.0003, ηp2 = 0.19
Right PAC	F(1,76) = 12; p = 0.0008, ηp2 = 0.14	F(1,62) = 12.89; p = 0.0006, ηp2 = 0.17
Left PAC	-	F(1,62) = 6.07; p = 0.016, ηp2 = 0.09
Parietal cortex		
Right SMG	F(1,76) = 9.62; p = 0.0026, ηp2 = 0.11	F(1,62) = 9.58; p = 0.0029, ηp2 = 0.13

Left SMG	-	-
Right IPG	-	-
Left IPG	-	-

4. Discussion

Results of the current study showed that the brain of 4- and 9-months-old infants is sensitive to sound duration, as reflected by ERPs elicited by short, isolated sounds. Effects of maturation were found for the scalp-recorded waveforms and the corresponding source activity. Both scalp-recorded ERPs and the related source activity became stronger and spatially better defined between 4 and 9 months of age. Generators of the event-related potentials sensitive to sound duration were located in the ventral frontotemporal auditory pathway. Details of the results are discussed below.

Two successive fronto-centrally negative sound-duration sensitive waveforms were identified in the ERPs elicited by 9-months old infants. In contrast, only the first of these two negativities were present in the responses of 4-months-olds with similar, 350-400 ms peak latencies in the two age groups. The late negativity peaked between 580 and 650 ms in the 9-months-olds over fronto-central scalp sites. The emergence of a second sound duration sensitive waveform between 4 and 9 months of age may reflect that this information is further processed in the latter age group. The ERP results are generally congruent with reports of waveforms related to the N250 and N450 in infant studies with duration manipulation (Čeponienė, 2002, 2004; Kushnerenko et al., 2001, 2002). While we did observe slightly longer latencies in the present study compared to some of the previous ones, this may be due to variability in the maturational stages of auditory brain networks (Adibpour, Lebenberg, Kabdebon, Dehaene-Lambertz, & Dubois, 2020). The fronto-centrally maximal scalp distributions of the current waveforms are also compatible with those reported in previous literature (Adibpour et al., 2020; Hämäläinen, Leppänen, Guttorm, & Lyytinen, 2008; Kushnerenko et al 2001). This fronto-central gradient has been suggested to be part of the developmental trajectory of auditory ERPs (van den Heuvel et al., 2015).

In contrast to some other studies, we did not observe lateralized effects for the early negativity. Different generators underlying the same ERP can have different maturational pathways (Adibpour et al., 2020). This may explain the mixed reports of age effects on lateralization (Cantiani et al., 2016; Ortiz-Mantilla et al., 2012). Thus, the contrasting age-

related differences in generator activity may be related to differential maturational trajectories (Lebenberg et al., 2019). This hypothesis is strengthened by previous reports highlighting that ERP components underlying duration perception are not unitary in nature and reflect dynamic processes involving multiple brain regions (Bendixen, Grimm, & Schröger, 2005).

The source-localized results show that the generators underlying the early negativity are distributed exclusively within the ventral auditory pathway consisting of bilateral temporal (PAC, MTG, STG, ITG) and inferior frontal cortices (see Figure 1B and 2B). The neural generators of the late negativity were located in overlapping areas (see. Figure 1C and 2C). However, they showed more focal distribution within primary and secondary auditory areas. The duration sensitive sources found in this study are consistent with those reported in the literature in response to both speech sounds and tones: the superior temporal, supramarginal, and inferior frontal gyri, and primary and secondary auditory cortical areas (Adibpour et al., 2020; Hämäläinen, Lohvansuu, Ervast, & Leppänen, 2015; Li et al., 2019; Mayhew et al., 2010; Parviainen, Helenius, & Salmelin, 2019; Ruhnau, Herrmann, Maess, & Schröger, 2011). This points towards the reliability of the use of anatomical templates in developmental samples, which represents an important turning point in studying electrical brain activity in infants.

A somewhat surprising result was the similarity between the scalp-recorded ERP morphologies of 4- and 9-months-old infants. Because the infant brain is subject to dramatic development during this period (Adibpour et al., 2020), one may expect the morphology of all ERP responses to change accordingly. However, similarity between the results in the two age groups sheds light on the central role the processing temporal information plays in sound processing: the processing of temporal features is more consistent than that of spectral features throughout development and even its neural substrate may not change dramatically from infancy to adulthood. Thus, whereas pitch discrimination initially requires very large separation between the sounds to elicit reliable discriminative brain activity (Novitski, Huotilainen, Tervaniemi, Näätänen, & Fellman, 2007), temporal features are reliably detected already at birth (Háden et al., 2015; Kushnerenko et al., 2013) and responses sensitive to different sound durations are quite similar between neonates and adults (Kushnerenko, Fellman et al., 2001).

The underlying generators show a more nuanced picture. As seen in Figure 1 and 2, the cortical distribution of the sound duration sensitive generators active during the early scalp-recorded negativity are almost identical in the two age groups. The exception is represented by the involvement of inferior frontal regions, suggesting the presence of brain maturational mechanisms. Supporting this assumption, we found statistically significant increments in current source density within the time window of the early negativity:

increased activity was observed in the right hemispheric part of the previously described fronto-temporal network. These mechanisms may be based on either higher efficiency of the primary and secondary auditory sensory cortices (temporal brain regions), and/or the increased involvement of additional regions such as the inferior frontal cortex.

The cortical distribution of the neural activity within the late negativity time window also strongly overlapped between age groups. However, substantial age-related differences were found in left primary and secondary auditory areas (bilateral STG, bilateral PAC, left bank of the STS, left ITG, and left MTG). In addition, we observed age related differences in the associative right SMG and in the OFG. These distributions indicate that maturation aids auditory network development in a quantitative manner at the later stages of duration processing. These developmental findings are especially interesting in light of our investigation of intersensory binding in Chapter 4, in which brain responses to multisensory binding were investigated in adults and HL-ASD infants. Indeed, we found surprisingly similar generators in adults and infants, but we direct the reader to Chapter 4 for a more in-depth discussion of the implications of this result.

Nonetheless, age-related effects on duration-sensitive ERPs and their generators may stem from multiple types of maturational processes (Kushnerenko et al., 2002). For example, myelination of auditory pathways plays a crucial role by enabling rapid coordination among neurons (Long, Wan, Roberts, & Corfas, 2018). The timeline of myelination in auditory pathways mirrors key developmental milestones in auditory perception such as the appearance of the P1-N1-P2-N2 (Long et al., 2018). Regarding the involvement of frontal regions, despite the widely accepted notion that prefrontal areas show prolonged development compared to sensory cortical regions (Lebenberg et al., 2019), some studies show that already at birth infants recruit frontal areas during processing speech sound (Dehaene-Lambertz, Dehaene, & Hertz-Pannier, 2002; Dehaene-Lambertz, et al., 2006; Gervain, Macagno, Cogoi, Peña, & Mehler, 2008). Moreover, temporal and frontal regions do not develop independently but show high correlations in their developmental course (Leroy et al.,2011). These findings suggest that the frontal cortices and their connections to auditory sensory regions (temporo-parietal areas) play a major role in language acquisition. In infants, frontal and temporo-parietal regions are already well connected, and these connections are hypothesized to underlie the early processing of speech (Dubois et al., 2016). Both bottom-up and top-down connections are necessary for language (Friederici, 2012). We hypothesize the first duration-sensitive component to reflect a first prodromal involvement of frontal cortices in auditory temporal processing (also related to language acquisition), the late component may then reflect feedback from frontal cortices to temporal areas. This hypothesis is supported by the involvement of primary auditory cortices in the generation of the later

duration-sensitive component. In this sense, the late component might be reflecting a maturational step of duration processing which might not yet be stable at 4 months.

This chapter represent a fundamental steppingstone to increase complexity and investigate higher-order unisensory and multisensory processes. Indeed, if we ignore the role played by, for example, associative regions in basic auditory mechanisms we may be tempted to misinterpret their function in response to the complex social stimuli presented in the following chapters. While this chapter lacks a direct testing of low-level processing in HL-ASD infants, we are currently in the middle testing the role of low-level auditory abilities in HL-ASD infants, infants at risk of developmental language disorder, and LR infants (Polver*, Cantiani*, Bulf, Piazza, Turati, Molteni, Riva, in preparation). Our preliminary results and other reports indicate that HL-ASD infants are characterized by fairly preserved low-level auditory skills (Xiao et al., 2022). Indeed, converging studies show that individuals with ASD are characterized by impaired auditory processing skills in tasks involving spectro-temporal complex and dynamic stimuli without impairments in basic acoustic processing (Riva et al., 2018; Lloyd Fox et al., 2013). However, results are non-univocal with some studies reporting functional abnormalities within low level primary sensory auditory pathways in adults with ASD (Edgar et al., 2015; Sperdin & Schaer, 2016). Further work is thus required to understand how these impairments in the early stages of auditory processing might impinge on the development of higher-order processes such as speech and language. Indeed, individuals with ASD may be characterized by subtle impairments in the processing of low-level acoustical features that may have cascading effects at higher hierarchical levels (Abrams et al., 2013). Interestingly, previous research has provided evidence in support of the sensory deficit by showing reduced activity in the STS (Gervais et al., 2004), a core region associated with both human voice processing and intersensory binding mechanisms (Belin, 2006; Stevenson et al., 2016). We will thus begin to investigate the role played by the human voice in typical and atypical development in the next chapter.

Chapter 2

How does it sound? Social saliency affects voice cortical tracking in low risk and HL-ASD infants.

In Chapter 1 we investigated the brain machinery involved in the perception of low-level acoustic stimuli. However, to understand auditory functioning and its underlying patterns of cortical activation in social situations, it is necessary to increase the level of complexity (Belin, 2006). Moreover, while HL-ASD infants may be characterized by fairly preserved low-level auditory processing, impairments may become progressively more apparent as stimuli complexity increases (de Boer-Schellekens, Keetels, Eussen, & Vroomen, 2013).

To capture the role of social cues in shaping auditory processing, we shall directly manipulate the social saliency of stimuli by presenting LR and HL-ASD infants with the same nursery rhyme narrated either by a synthesiser, an actress, or the infant own mother. As the mother's voice is one of the first and most important sounds in infants' environments, being perceived already in the womb (Abrams et al., 2013), we shall test the facilitating effects it might have on scaffolding socio-auditory processing in LR infants. On the contrary, considering the social impairments outlined in the Introduction, HL-ASD infants should fail in exploiting facilitation from social cues. They might however benefit from the lack of prosodic cues characterizing the synthesiser.

Our research questions will be answered through the novel perspective of voice cortical tracking, by assuming that brain oscillations are directly tuned to specific features of external auditory stimuli. Results will thus have both theoretical and methodological implications for the characterization of the developmental trajectories (and its deviations) of the first, and ontogenetically older, foundations of social processing in infancy: the human voice.

1. Introduction

The human voice is the most salient sound in the environment. Our ability to process information contained in voices plays a key role in social interactions (Belin, Fecteau, & Bedard, 2004). In this sense the voice can be conceptualized as a modulated carrier signal with specific spectral and temporal properties (Nallet & Gervain, 2022) conveying vital information such as the identity of the speaker, his/her affective state, and of course speech information (Adam-Darque et al., 2020; Belin et al., 2000, 2004; Sperdin & Schaer, 2016).

While the voice does carry speech, it cannot be reduced to it (Belin et al., 2004) For instance, both time-reversed speech, a signal devoid of linguistic content but with preserved vocal features, and normal speech cause similar patterns of activity in large portions of the auditory cortex (Belin, 2006). In fact, while speech is a more recent evolutionary breakthrough, vocalizations are much more phylogenetically and ontogenetically older (Belin et al., 2004; Pell et al., 2015). Indeed, voice-processing abilities precede speech perception in human development. For example, while phonemes discrimination appears at around 2 months of age (Cheng, Lee, Chen, Wang, & Decety 2012; Dehaene-Lambertz et al., 2002; Friederici, 2005), at birth infants already show well-developed voice perception abilities (Cheng et al., 2012; DeCasper and Spence, 1986; Grossman, 2010), enabling them to be socially responsive (Belin & Grosbras, 2010; Cheng et al., 2012; Vouloumanos & Curtin, 2014). Moreover, newborns can discriminate their mother's voice from the voice of another woman (Beauchemin et al., 2011; Cheng et al., 2012).

Taken together, these results suggest that the ability to process and discriminate the human voice emerge very early in development (Cheng et al., 2012; Davies-Thompson et al., 2019; Grossmann, 2010) and allow the preverbal infant to interact with the environment and learn from it (Abrams et al., 2013, 2016). Thus, focusing on the broader context of voice perception might give us more direct insights into the social components carried by acoustic stimuli (Belin et al., 2004). However, while speech processing has been extensively studied, far less is known about general voice processing and on how it develops (Grossmann, 2010; Grossmann et al., 2005; Vouloumanos & Curtin, 2014).

In adults' brain the processing of human voices is known to engage the so-called 'temporal voice areas' (TVA) located in auditory cortices and alongside the superior temporal sulcus (STS; Beauchemin et al., 2011; Belin, 2006; Belin et al., 2000; Belin & Grosbras, 2010; Capilla, Belin, & Gross, 2013; Fecteau, Armony, Joanette, & Belin, 2004). This finding mirrors the discovery of face selective areas in the visual cortex and suggests that face and voice processing could follow similar organizational principles (Belin et al. 2004, 2006; Fecteau et al., 2004). Indeed, TVA show greater activity in response to voices, in terms of both speech and non-speech vocalization such as laughs and coughs, compared to non-vocal sounds (Belin & Grosbras, 2010; Gervais et al., 2004; Latinus & Belin, 2011). Interestingly, some voice sensitive regions, such as the planum temporal and the superior temporal gyri (STG), have also been shown to encode speaker's identity (Maguinness et al., 2018). Given their long and preserved evolutionary history (Belin & Grosbras, 2010; Latinus & Belin, 2011), being also found in the macaque brain (Capilla et al., 2013; Petkov et al., 2008), the TVA have been shown to be already functionally specialized in the infant brain (Belin & Grosbas, 2010). This may represent a fertile

substrate for subsequent linguistic acquisitions (Sperdin & Schaer, 2016; Cheng et al., 2012). However, little is still known on the actual organization of TVA in the infant brain and of their interactions with other brain networks, such as the social brain (Grossman & Johnson, 2007).

Within this context, a specific class of vocal stimuli to which infants show preference is their mother's voice (Adam-Darque et al., 2020; Abrams et al., 2016). Indeed, the mother's voice is relevant for the emotional and social development of the infant and lays foundations for later developmental acquisitions, such as language (Abrams et al., 2016; Adam-Darque et al., 2020; Beauchemin et al., 2011; Zhao, Chronaki, Schiessl, Wan, & Abel, 2019). Moreover, caregivers' voices have been shown to provide critical emotional value and promote bonding between infants and their parents (Abrams et al., 2013, 2016). Despite the outlined role of the mother's voice, there is surprisingly little research on how it shapes brain development (Abrams et al., 2016). Some reports show that there are specific activations in response to the mother's voice in many brain structures encompassing both primary and secondary auditory cortices and voice selective temporal regions but also in reward, affective, face processing, salience detection and mnemonic circuits (Abrams et al., 2016). This observation is of particular interest since it might reflect a specific engagement of social brain in response to the mother's voice. Indeed, what strikes is the width of the systems involved, even in response to brief meaningless samples of the mother's voice (Abrams et al., 2016).

A fundamental question is whether voice processing is impaired in neurodevelopmental disorders whose core features regard dysfunctions in social interactions. Difficulties to speak and to appropriately interact with the social environment are central features for the diagnosis of autism spectrum disorder (ASD; American Psychological Association, 2013). Children with autism spectrum disorder (ASD) are often non-responsive to voices, and it has been hypothesized that voice processing impairments contribute to the pronounced social communication difficulties in ASD (Abrams et al 2019; Sperdin & Schaer, 2016). Specifically, children and adults with ASD tend to ignore human vocalizations but are responsive to other non-vocal stimuli (Sperdin & Schaer, 2016), suggesting specific impairments in the processing of the social cues inherently present in voices. Although, the exact etiology of such impairments is still not certain, there are reports showing anomalies within the TVA in both children and adults with ASD (Belin & Grosbras, 2010; Gervais et al., 2004; Sperdin & Schaer, 2016). Indeed, activity in both TVA and the extended voice processing networks including auditory, reward, and salience regions, has been shown to be positively related social functioning in children with ASD (Abrams et al., 2019). Moreover, children with ASD show a striking lack of preference for their mother's voice (Abrams et al., 2019).

As according to neuroconstructivism even impairments in low-level processes early in development can lead to marked dysfunction later on (Karmiloff-smith, 2009), early atypicalities in voice processing may be indicators of subsequent impairments in ASD (Sperdin & Schaer, 2016; Vouloumanos & Curtin, 2014). For this reason, infants at higher likelihood of developing ASD (HL-ASD), due to having a diagnosed sibling, are increasingly studied prospectively to understand risk factors (Sperdin & Schaer, 2016). From what we do know, HL-ASD infants show reduced sensitivity to human emotional vocalizations compared to LR infants (Blasi et al., 2015). Thus, aberrant functioning within the extended voice processing network may underlie insensitivity to biologically salient voices in ASD, which may subsequently affect key developmental processes associated with social and linguistic achievements (Abrams et al., 2019). Early impairments suggest aberrant wiring of cerebral regions and might underlie early clinical signs, such as unresponsiveness to name, lack of orientation to human voices and delay in the development of receptive and expressive language skills (Belin & Grosbras, 2010; Cheng et al., 2012; Elsabbagh & Johnson, 2007; Sperdin & Schaer, 2016). However, very little is known about the neurobiology of early voice processing impairments in ASD (Abrams et al., 2019).

In the present study we aimed to characterize voice brain processing in 18-months-old LR and HL-ASD infants along a socially salient continuum involving the voice of a synthesiser, an actress, and the infant's mother narrating a nursery rhyme. Hence, while the actress' and the mother's voice will probe voice processing with increasing social salience, the synthesiser, being devoid of them, will allow us to test the role of prosodic cues in the encoding of voices. Indeed, since emotional prosody consists of variations in the pitch contours, in the intensity and in the duration of voices, all conveying social information (Clumeck et al., 2014), LR infants should greatly benefit from the presence of prosodic cues, especially if coming from their mothers. On the other hand, prosodic processing seems to pose a greater challenge for individuals with ASD due to their impairments in social perception (Clumeck et al., 2014; Gebauer, Skewes, Hørlyck, & Vuust, 2014; Maslennikova et al., 2022). Thus, these infants should be facilitated in processing the synthesiser voice, since it lacks salient prosodic cues. We chose to investigate the hypotheses outlined above through voice-brain coherence. More in depth, voice-brain coherence measures the consistency of phase differences between two signals and thus synchronization between neural oscillations and the acoustic envelope at any given frequency (Assaneo et al., 2019; Attaheri, Choisdealbha, et al., 2022; Kalashnikova, Peter, Di Liberto, Lalor, & Burnham, 2018; Kolozsvári et al., 2021; Menn, Michel, Meyer, Hoehl, & Männel 2022; Nallet & Gervain, 2022; Casas et al., 2021). The higher the coherence between brain activity and the acoustic envelope, the better the

tracking of incoming auditory information (Kolozsvári et al., 2021). The processual benefit hypothesised in LR infants to their mother's and secondly to the actress voice, may thus take the form of higher coherence values in response to these stimuli compared to HL-ASD infants. In addition, previous studies in both adult and developmental samples have shown a prominent role played by delta and theta oscillations in encoding voice stimuli (Attaheri, Choisdealbha, et al., 2022). However little is still known about the role played, if any, by faster oscillations in cortical tracking across infancy. In this study we will therefore investigate all proposed ranges of oscillations, from the slow delta to the fast gamma oscillations, to disentangle developmental trajectories in frequency tracking. To our knowledge, this is the first study to directly address the issue of social salience in voice processing through naturalistic stimuli.

Due to the Covid-19 pandemic, data collection has significantly slowed down. Indeed, as data collection takes place in the IRCCS Eugenio Medea healthcare facility as a part of a broader longitudinal project, extra care was taken to avoid the spread of the virus in susceptible patients. For this reason, data collection is still ongoing. Here, we discuss preliminary results obtained through the application of our analysis pipelines on the samples we collected thus far.

2. Methods

2.1. Participants

Participants recruitment is part of an on-going longitudinal project carried out at the IRCCS Eugenio Medea, aimed at identifying early risk markers in ASD, and in the general population, focusing on infants at familiar risk for ASD (Riva et al., 2018, 2022). Infants were included if both parents were native Italian speakers, their gestational age was ≥36 weeks, their birth weight was ≥2500 g, and their Griffiths general quotient was ≥85 (Griffiths, 1984), if there were no known medical, genetic, or neurological conditions, and major complications in pregnancy and/or delivery likely to affect brain development. Our sample thus far consists of thirty-three 18-months-old infants, of which 23 LR infants and 10 HL-ASD infants. As we shall see in the next sections, after preprocessing 7 infants were excluded from analysis due to excessive artifacts (N = 4, all LR), technical issues (N = 1, all LR), and excessive fussiness (N = 3, HL-ASD = 1). The mean age of LR infants was 18.05 months (SD = 0.8), while the mean age of HL-ASD infants was 18.05 months (SD = 0.7). Written informed consent was obtained from all parents prior to testing. The experiment was performed in accordance with relevant guidelines and regulations and was approved by the Eugenio Medea Scientific Institute Ethical and Scientific Committees.

2.2. Stimuli

Our stimuli are represented by the nursery rhyme ‘Il maestro giusto’ by the writer Gianni Rodari narrated by an actress, a synthesiser, and the infant’s mother. Specifically, we first started by training an actress to narrate the nursery rhyme. Then, we proceeded to feed a text-to-speech phoneme-based synthesiser the text of the nursery rhyme. We chose this type of synthesiser due to the mechanical qualities the voice output has. Indeed, our aim was to obtain a stimulus devoid of prosodic qualities and of clearly non-human nature, to assess the role of the social actor in voice perception. Mothers were sent both the audio recordings of the actress and the text of nursery rhyme and were trained to match the tempo of the actress while imagining speaking to their child (Adam-Darque et al., 2020). The end results were audio recordings of approximately 43 s duration. We chose lengthy recordings to keep stimuli as naturalistic as possible and to preserve social cues. Stimuli were preprocessed in Audacity 3.0.3, where noise and clicks in the audio recording were reduced. Then stimuli were matched to the actress intensity and pitch range to make them comparable regarding low-level acoustical in Praat (v6.1.37). As the speech prosody is conveyed via several acoustical speech features, controlling further the acoustical differences between the stimuli, for example by exactly matching pitch contours, would have affected their emotional content, and made the stimuli unnatural and ecologically invalid (Lindström, Lepistö-paisley, Vanhala, Alén, & Kujala 2016). This procedure was repeated for each mother’s recording. Stimuli were presented free field at an intensity of 55 dB via speakers located on both sides and equidistant from the infant. Each stimulus was randomly presented 3 times. During stimuli presentation infants were shown the same engaging video to maintain attention. As the video was the same across stimuli, we expect its effect to be cancelled out during statistical analysis. Stimuli were presented using the E-Prime software v2.0 (Psychology Software Tools Inc., Pittsburgh, PA).

2.3. EEG data collection and preprocessing

The infant sat comfortably on his/her mother lap in a sound-attenuated and electrically shielded room during EEG data collection. Infants were always accompanied by their mothers. The EEG signal was recorded using a 128-electrode HydroCel Geodesic Sensor Net (Electrical Geodesic In., Eugene, OR). The vertex was used as online reference and the signal was sampled at 500 Hz. Data were preprocessed using the EEGLab toolbox (Delorme & Makeig, 2004) in Matlab (ver. R2021a). We first removed the external ocular channels as they are not present in all the caps measures, then data were high-pass filtered at 0.1 Hz and low-pass filtered at 100 Hz. Cut-off frequencies were 0.05 and 100.05 Hz (-6 dB). The 50Hz line noise was removed by using the CleanLineNoise function incorporated in the PREP pipeline (Bigdely-Shamlo, Mullen, Kothe, Su, & Robbins, 2015). Data were then cleaned using Clean Rawdata EEGLab plugin (Kothe &

Makeig, 2013; Plechawska-Wojcik et al., 2019): channels were deemed 'bad' and removed if either flat for longer than 5 seconds, or through computation of each channel's correlation to its RANdom SAmple Consensus (RANSAC) reconstruction for each window. The correlation threshold for channel removal was set at 0.8. The maximum number of channels rejected was 13. Then, to all recordings, a PCA (Principal Component Analysis) decomposition was applied to identify artifactual PCs (defined by a comparison against the data's own cleanest parts) in order to reject them and reconstruct activity from the remaining components. Windows' Standard Deviation (SD) was set at 20. After these steps, removed channels were spherically interpolated by referencing to the full-rank channel matrix and data were re-referenced to average reference. To further remove artifacts, the Independent Component Analysis (ICA) was performed on PCA reduced components. Indeed, re-referencing and channel interpolation resulted in an uneven number of ranks. As the ICA is heavily affected by ranks variations and produces unreliable ICs under such conditions (Schneider, 2010), we applied the PCA as a data reduction technique to avoid issues. The resulting 80 components were subjected to manual rejection of eye, heart line or muscle artifacts. The maximum number of automatically rejected components was set to 9. Data were then divided into 2000 ms epochs. For the final epoch rejection, a threshold of +/- 250 µV was set to remove abrupt changes in amplitude, while two thresholds of 150 µV for slope across the whole epoch and 0.5 for r^2 were set to remove step-like artifacts and linear trends. In the LR group the mean number of trials for the synthesiser was 36.82 (SD = 18.24), for the actress voice was 35.82 (SD = 18.30), for the mother voice 38.41 (SD = 17.79). In the HL-ASD group the mean number of trials for the synthesiser was 30.37 (SD = 12.07), for the actress voice was 25.75 (SD = 13.54), for the mother voice 30.5 (SD = 13.9). To check for any disparity between groups and voices we performed a set of t-tests: no significant differences were found between groups ($p > 0.14$).

2.4. Voice-brain coherence computation and statistical analyses

To match the EEG signal we first resampled our audio recordings to 500 Hz and extracted 2000 ms epochs. Then, we computed the average of both the EEG and the sound epochs. Afterwards we computed the absolute values of the Hilbert transform on the broadband audio recording to obtain the amplitude envelope of the signal. To avoid spurious EEG values due to volume conduction, we computed the scalp current source density (CSD) using the second-order derivative (the surface Laplacian) of the potential distribution through a spherical spline method. We then computed the magnitude-squared coherence between the audio and the EEG signal for each frequency band. The frequency bands considered were the delta (1 – 3 Hz), theta (3 – 5 Hz), alpha (6 – 9 Hz), beta (10 – 29 Hz), lower gamma (30 – 49 Hz), higher gamma (51 – 80 Hz). Where possible we used

published guidelines to choose infant-appropriate frequency ranges (Saby & Marshall, 2012). If that was not possible, such as in the case of higher frequency bands, we referenced to the adults OHBM COBIDAS guidelines (Pernet et al., 2020). A graphical representation of the analysis performed is depicted in Figure 1. We then computed t-tests comparing coherence values in response to the synthesiser, actress and mother's voices in the two groups separately and across groups for each frequency band. However, once our samples will be completed, we plan to synthesise noise having the same power spectrum and duration as the original signal by randomizing the phase of the Fourier spectrum. The implementation of the same procedure described above on the derived noise, will provide surrogate data to which our original data will be compared through the cluster-based statistic (CBS) in order to control for multiple comparisons (Maris & Oostenveld, 2007). More in depth, surrogate and original data will be compared by means of a t-value to create pairwise group contrasts matrices. Selected samples will then be clustered in connected sets, on the basis of spatial and spectral adjacency, and cluster level-statistics will be calculated by taking the sum of the t-values for every cluster. Neighbouring sensors will be identified through a triangulation algorithm, thus irrespectively of their relative distance. The significance probability for the CBS will be then calculated via the Monte Carlo permutations method. Within this framework, p-values will be the proportion of random partitions that resulted in a larger statistic than the observed one for each cluster. Thus, our results will report the sensors significantly different from the surrogate data, showing only activity in topographical scalp regions and at Hz ranges not due to artifacts or other sensory confounds. This, besides making coherence computations more reliable, will enable us to roughly interpret the distribution of coherence values in term of broad CSD-reconstructed brain areas. All analyses are or will be completed using FieldTrip (Oostenveld, Fries, Maris & Schoffelen, 2011) and in-house Matlab scripts.

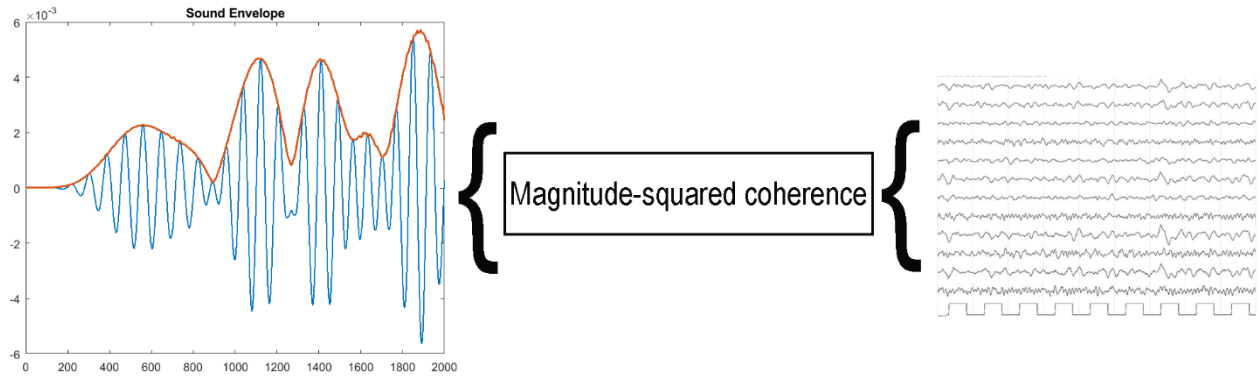


Figure 1. *Graphical representation of the analysis pipeline. The figure shows the steps followed to relate the sound envelope to the EEG signal through the magnitude-squared coherence.*

3. Results

For the delta band the mean value of the coherence in the LR group was 0.35 (SD = 0.12) in response to the synthesiser voice, 0.38 (SD = 0.12) in response to the actress voice, 0.32 (SD = 0.12) in response to the mother's voice. In the HL-ASD group it was 0.39 (SD = 0.11) in response to the synthesiser voice, 0.34 (SD = 0.12) in response to the actress voice, and 0.38 (SD = 0.11) in response to the mother's voice (Figure 2).

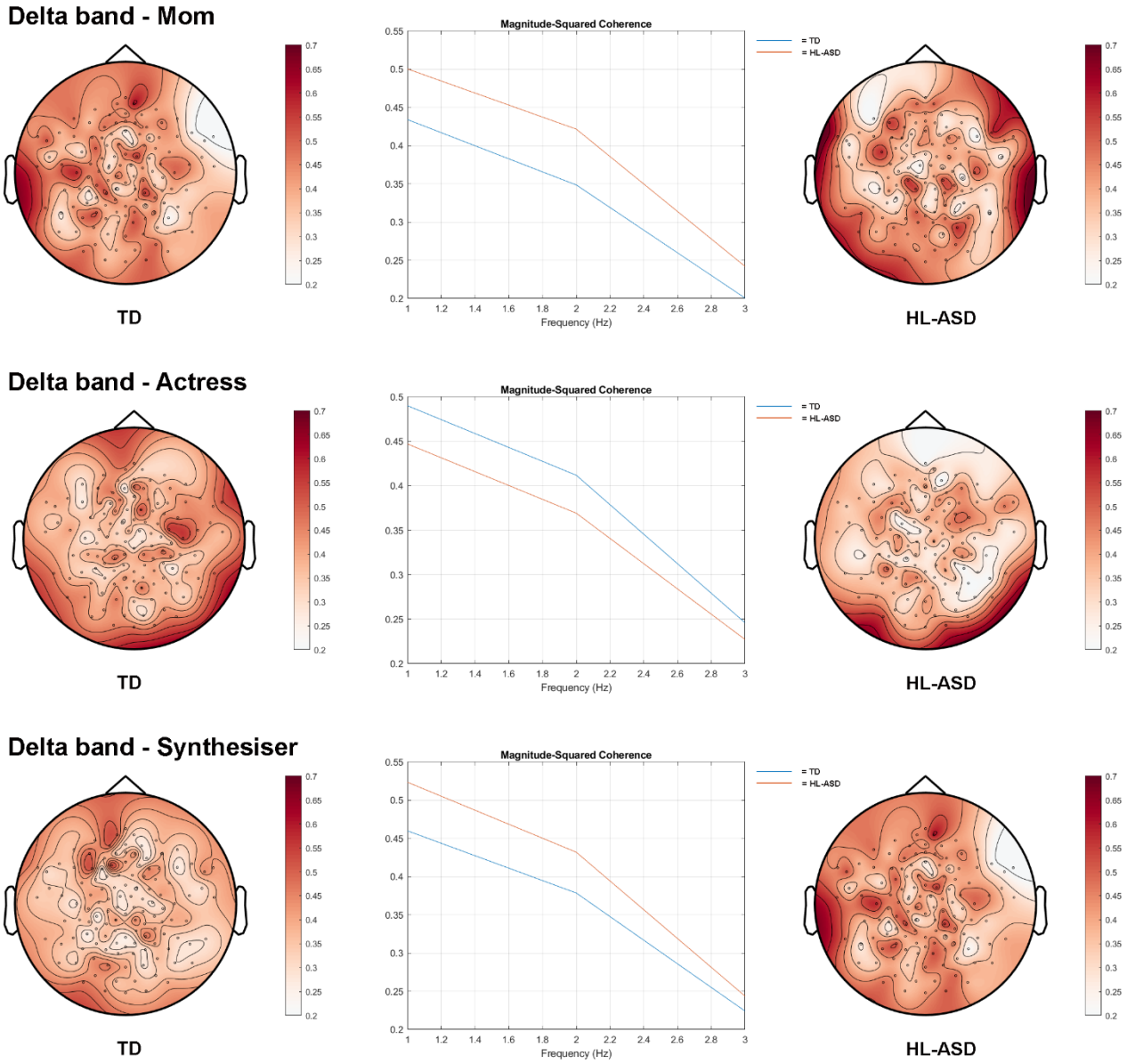


***Figure 2**. Magnitude-squared coherence values between the sound envelop and the EEG across stimuli. On the sides magnitude-squared coherence values are represented in their topological distribution on the scalp on both LR (left) and HL-ASD infants (right). In the middle the distribution of the magnitude-squared coherence values is represented across the Hz of the delta range.*

For the theta band the mean value of the coherence in the LR group was 0.14 (SD = 0.03) in response to the synthesiser voice, 0.19 (SD = 0.04) in response to the actress voice, 0.15 (SD = 0.05) in response to the mother's voice. In the HL-ASD group it was 0.16 (SD = 0.05) in response to the synthesiser voice, 0.18 (SD = 0.04) in response to the actress voice, and 0.15 (SD = 0.05) in response to the mother's voice (Figure 3).

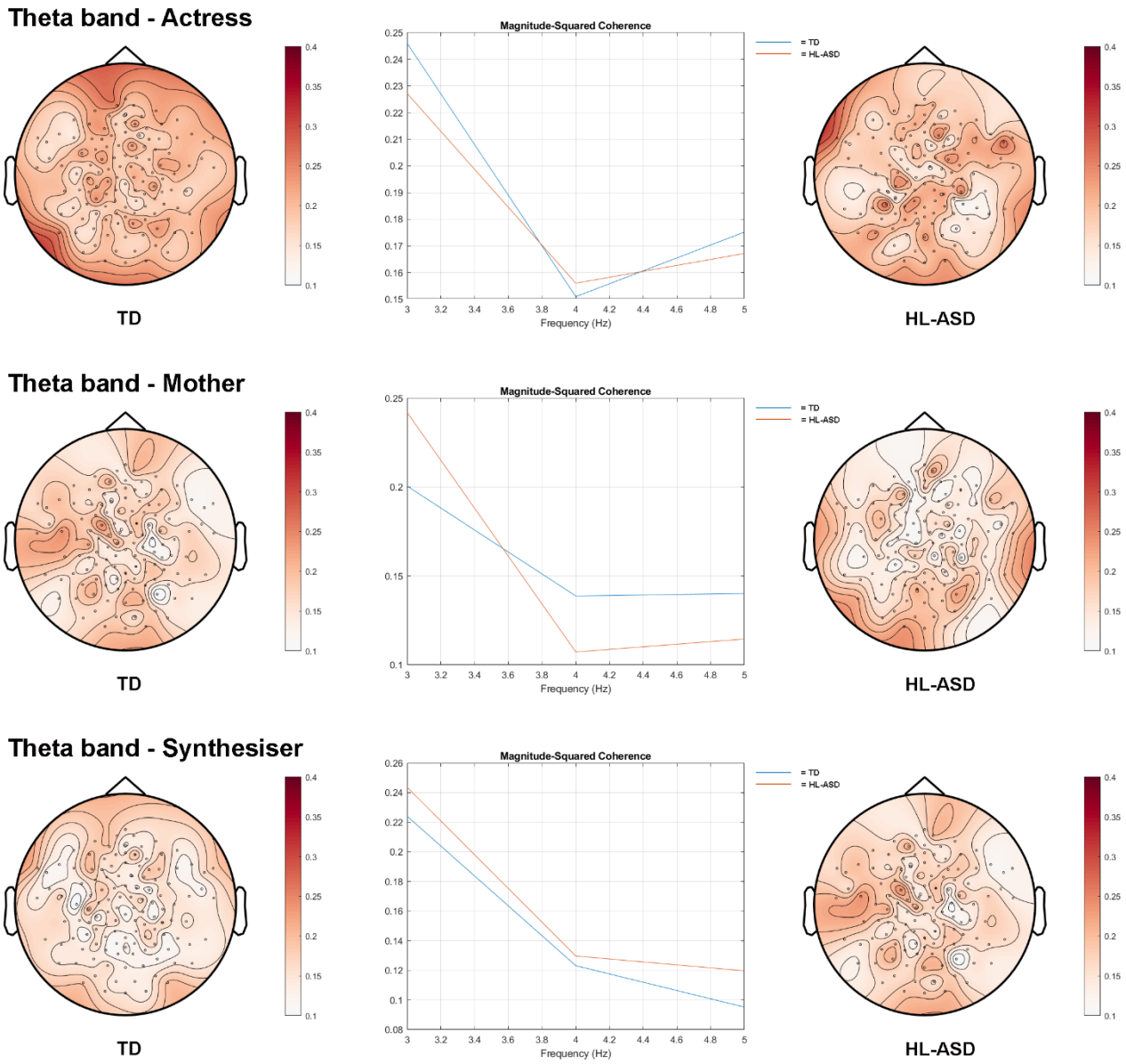


Figure 3*. Magnitude-squared coherence values between the sound envelop and the EEG across stimuli. On the sides magnitude-squared coherence values are represented in their topological distribution on the scalp on both LR (left) and HL-ASD infants (right). In the middle the distribution of the magnitude-squared coherence values is represented across the Hz of the theta range.*

For the alpha band the mean value of the coherence in the LR group was 0.13 (SD = 0.03) in response to the synthesiser voice, 0.12 (SD = 0.02) in response to the actress voice, 0.13 (SD = 0.03) in response to the mother's voice. In the HL-ASD group it was 0.12 (SD = 0.03) in response to the synthesiser voice, 0.11 (SD = 0.02) in response to the actress voice, and 0.11 (SD = 0.04) in response to the mother's voice (Figure 4).

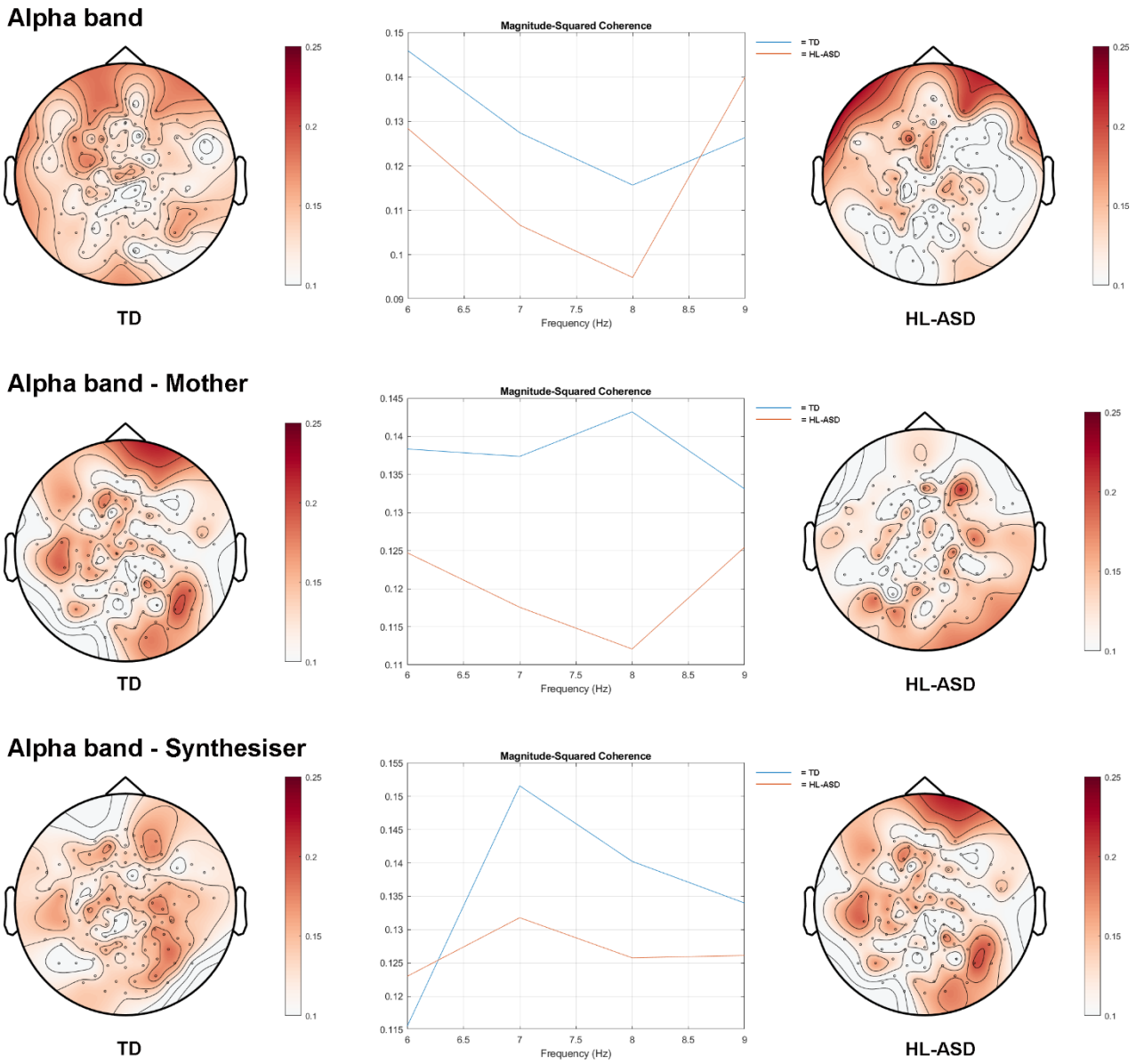


Figure 4. *Magnitude-squared coherence values between the sound envelop and the EEG across stimuli. On the sides magnitude-squared coherence values are represented in their topological distribution on the scalp on both LR (left) and HL-ASD infants (right). In the middle the distribution of the magnitude-squared coherence values is represented across the Hz of the alpha range.*

For the beta band the mean value of the coherence in the LR group was 0.13 (SD = 0.01) in response to the synthesiser voice, 0.12 (SD = 0.01) in response to the actress voice, 0.13 (SD = 0.01) in response to the mother's voice. In the HL-ASD group it was 0.13 (SD = 0.01) in response to the synthesiser voice, 0.13 (SD = 0.01) in response to the actress voice, and 0.13 (SD = 0.01) in response to the mother's voice (Figure 5).

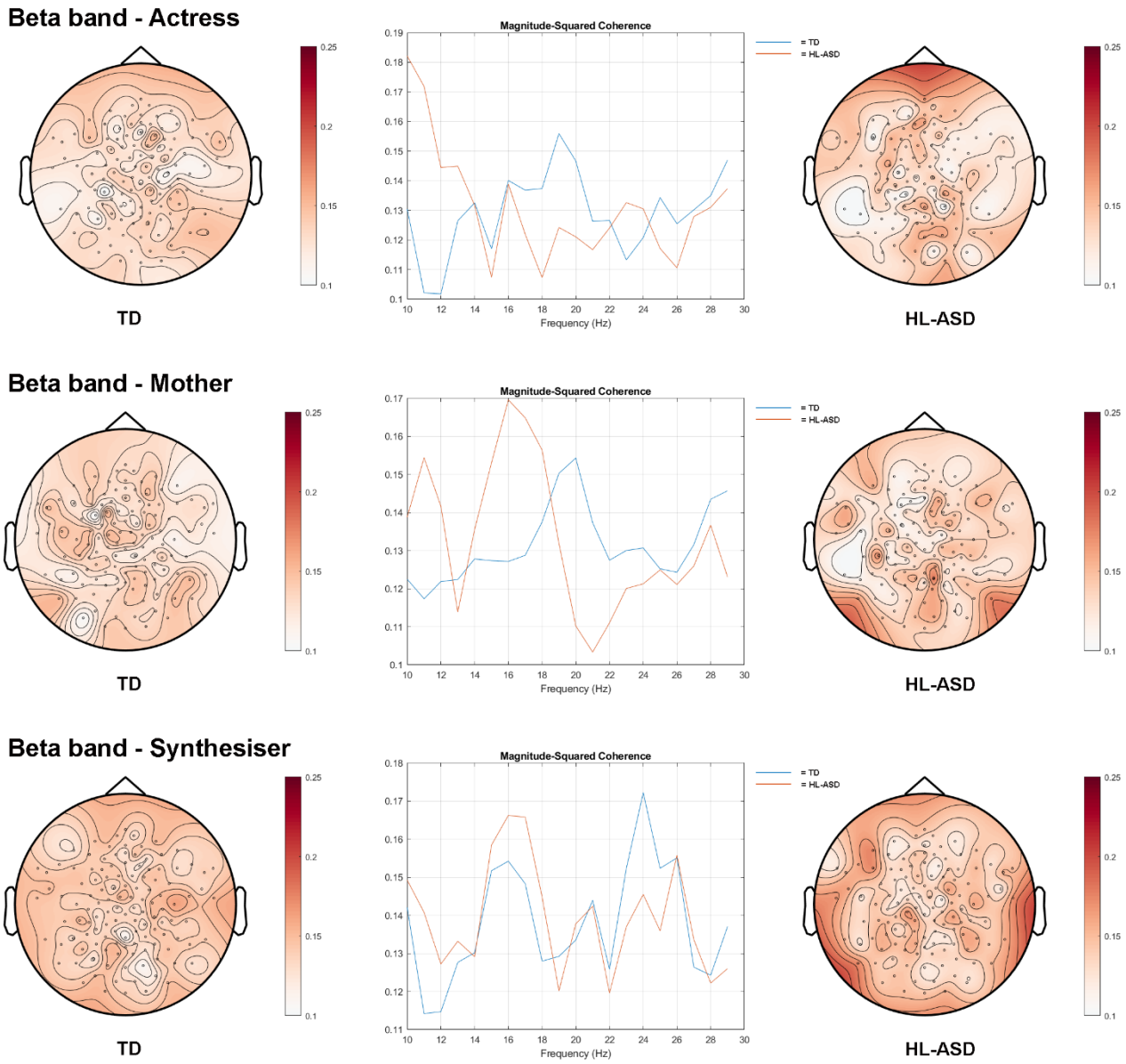


Figure 5*. Magnitude-squared coherence values between the sound envelop and the EEG across stimuli. On the sides magnitude-squared coherence values are represented in their topological distribution on the scalp on both LR (left) and HL-ASD infants (right). In the middle the distribution of the magnitude-squared coherence values is represented across the Hz of the beta range.*

For the lower gamma band the mean value of the coherence in the LR group was 0.13 (SD = 0.01) in response to the synthesiser voice, 0.13 (SD = 0.01) in response to the actress voice, 0.13 (SD = 0.01) in response to the mother's voice. In the HL-ASD group it was 0.12 (SD = 0.01) in response to the synthesiser voice, 0.12(SD = 0.01) in response to the actress voice, and 0.13 (SD = 0.01) in response to the mother's voice (Figure 6).

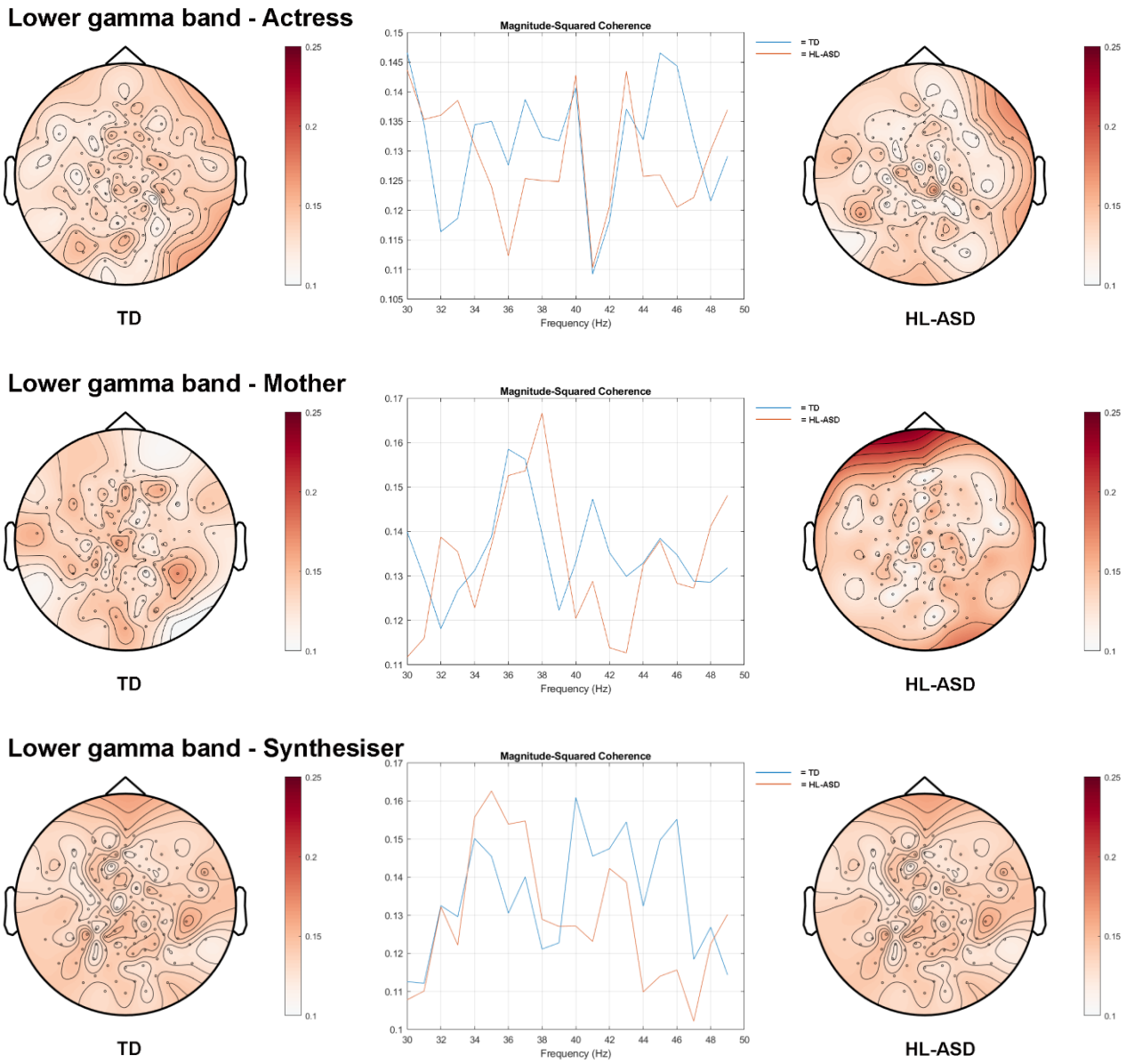


Figure 6. *Magnitude-squared coherence values between the sound envelop and the EEG across stimuli. On the sides magnitude-squared coherence values are represented in their topological distribution on the scalp on both LR (left) and HL-ASD infants (right). In the middle the distribution of the magnitude-squared coherence values is represented across the Hz of the lower gamma range.*

For the higher gamma band the mean value of the coherence in the LR group was 0.13 (SD = 0.009) in response to the synthesiser voice, 0.13 (SD = 0.01) in response to the actress voice, 0.13 (SD = 0.01) in response to the mother's voice. In the HL-ASD group it was 0.13 (SD = 0.01) in response to the synthesiser voice, 0.12 (SD = 0.01) in response to the actress voice, and 0.13 (SD = 0.008) in response to the mother's voice (Figure 7).

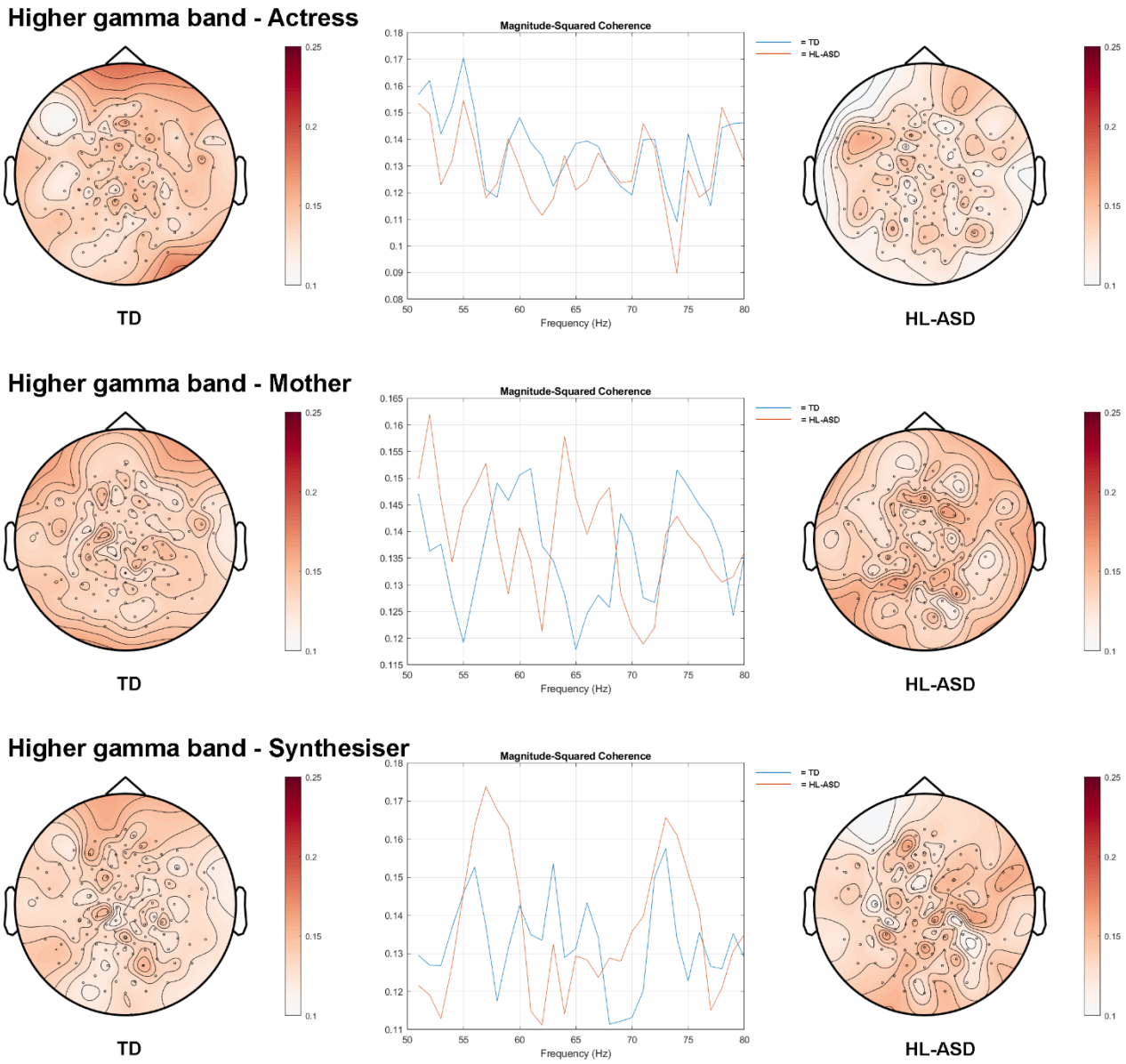


Figure 7. *Magnitude-squared coherence values between the sound envelop and the EEG across stimuli. On the sides magnitude-squared coherence values are represented in their topological distribution on the scalp on both LR (left) and HL-ASD infants (right). In the middle the distribution of the magnitude-squared coherence values is represented across the Hz of the higher gamma range.*

As we observed varying peaks within frequency bands, once our sample will be completed, we shall focus on finding peaks differences through the CBS as outlined in the methods section.

From the t-tests we observed that in the LR group coherence values in response to the actress voice are higher compared to the synthesiser voice ($t(16) = 3.3$, $p = 0.004$). No other comparison was found to be significant (all $p > 0.08$). We propose that the absence of other significant results is due to the reduced statistical power in our samples.

4. Discussion

In this chapter we set out to investigate the role of social cues in scaffolding brain processing of voices in both LR and HL-ASD infants. We hypothesised that social salience

may play a prominent role in facilitating voice processing in LR infants, due to the greater involvement of brain areas mediated by activity in the social brain (Grossman & Johnson, 2007). Conversely, considering the impairments in social processing characterizing ASD and more specifically their dysfunctions in prosody encoding, we expected HL-ASD infants to fail to track highly salient stimuli such as their mother's voice, and to succeed in tracking the synthesiser voice, due to the absence of social cues.

Before going in depth regarding social saliency manipulations, since this is a fairly novel area of research, it is important to describe the overall distribution across frequencies bands of coherence values in our samples. Indeed, as we can appreciate from these preliminary results, coherence values tended to be higher for the slower bands and smaller for faster frequencies. We thus confirm reports in the literature highlighting the role of slower frequency bands in cortical tracking across infancy, especially in the delta range (Attaheri, Choisdealbha, et al., 2022), where we find fairly high coherence values. Delta cortical tracking has been hypothesised to reflect higher-order attentional and linguistic mechanisms modulating voice processing (Molinaro & Lizarazu, 2018), while the role of theta oscillations is more debated. It has been proposed that theta tracking may reflect a passive synchronization of auditory cortices to acoustic regularities (Molinaro & Lizarazu, 2018). However, our results argue against this view as we observed significantly different coherence values in the LR group in response to stimuli varying only in their social salience and prosodic features, as low-level characteristics were as controlled as possible. Moreover, theta tracking increases in strength in adulthood, possibly reflecting neural maturation processes (Attaheri, Panayiotou, et al., 2022; Kolozsvári et al., 2021; Power et al., 2010). The lack of results on higher frequency bands could be either due to the fact that, for example, the beta and gamma bands are more sensitive to cross-modal stimuli compared to unimodal ones (Doesburg, Emberson, Rahi, Cameron, & Ward, 2008; Senkowski, Talsma, Grigutsch, Herrmann, & Woldorff, 2007), or to an actual less prominent role of those frequencies in cortical tracking of voices. In addition, it is important to note that we observe overall higher coherence values in slower compared to faster frequencies even in the HL-ASD group, which to our knowledge has never occurred before. This points toward a somewhat preserved voice cortical tracking in this group (Menn et al., 2022). However, this result must be interpreted with care as our HL-ASD sample is especially small.

Regarding social salience manipulations, we observed higher coherence in response to the actress voice compared to the synthesiser in the LR group. This result confirms that the presence of social cues guides tracking of auditory stimuli (Kalashinkova et al., 2019). This finding highlights the role that prosody plays in voice processing, a role that it is unfortunately understudied in most speech tacking studies (Myers et al., 2019).

Contrary to our initial hypothesis we do not observe results in response to the mother's voice in LR infants. However, it is interesting to note this effect may descend from novelty processing (Adam-Darque et al., 2020). This is in line with reports of a greater activation of the STS and frontal regions in adults in response to an unknown voice (Adam-darque et al., 2020; Kriegstein & Giraud, 2004). While these results are to be interpreted with care since we did not yet collect our full sample and our statistical power is thus low, our results thus far confirm the usefulness of neural tracking analysis in samples characterized by reduced data quality such as infants (Jessen, Obleser, & Tune, 2021).

Interestingly, we did not observe effects in the HL-ASD group. Since we do observe preserved cortical tracking of acoustic stimuli in HL-ASD infants in accordance with other reports (Menn et al., 2022), it seems plausible to hypothesise, according to the literature, that HL-ASD infants tend to treat voices as merely an auditory stimulus and do not differentiate among them. Indeed, while LR newborns are rapidly attracted by human voices within the first days of life (Cheng et al., 2012; Sperdin & Schaer, 2016), young children and toddlers with ASD are characterized by atypical preferences. For example, 5-year-old children with ASD do not reliably orient to their name (Vouloumanos & Curtin, 2014), and 3–4-year-olds with ASD prefer superimposed voices over their mother's voice (Vouloumanos & Curtin 2014). This is in stark contrast with reports showing that LR infants attend to their name by 4.5 months of age (Mandel, Jusczyk, & Pisoni, 1995; Vouloumanos & Curtin, 2014) and prefer listening to their mother's voice from birth (DeCasper & Fifer, 1980; Vouloumanos & Curtin, 2014). This suggests that acoustic stimuli rich in social features, such as voices, might not be adequately processed, most likely due to the abnormal brain development (Gervais et al., 2004; Sperdin & Schaer, 2016). Indeed, it has been proposed that the dysfunction that HL-ASD infants show in reward, affective, face processing, salience detection and mnemonic circuits prevent them from engaging with social acoustical stimuli (Abrams et al., 2013). In support of this hypothesis, HL-ASD infants and adults with ASD reportedly do not show early specialization in the processing of emotional voices in TVA (Belin, 2006; Blasi et al., 2015; Gervais et al. 2004; Sperdin & Schaer, 2016). This impairment has been shown to be specific for vocal stimuli, since responses proved to be normal to non-vocal sounds (Belin, 2006; Gervais et al. 2004). In other words, for the auditory areas of ASD individuals, voices are just another sound category (Belin, 2006; Gervais et al. 2004). Such a dysfunction might not only undermine early caregiver - infant relationship but also therapeutic efforts to socially engage toddlers with ASD (Xiao et al., 2022).

Finally, from a topological point of view, we mainly observed higher coherence values in temporal and frontal areas in our subjects. While these results are to be interpreted with care, there are reports showing that top-down signals, coming from frontal

areas, increase the synchronization between the auditory cortex and the stimulus envelope (Assaneo et al., 2019). Similarly to what we observed in the previous chapter in response to simpler stimuli, frontal areas may modulate oscillations in voice sensitive regions on the basis of predictive cues, allowing a more flexible tracking (Assaneo et al., 2019; Park et al., 2015). For example, there are reports of extensive networks oscillating in the delta and theta bands, comprising the middle and inferior temporal gyri, the superior parietal lobules, the middle and inferior frontal gyri, auditory cortices involved in the tracking of speech (Park et al., 2015). As we shall see in Chapter 4, these areas are strikingly similar to those involved in multimodal integration, highlighting the additive role that sensory information plays in brain processing.

In sum, the study of the neural correlates of voice perception in ASD is a promising area of research which deserves as much attention as the domain of face processing (Belin, 2006), coupled with specific scrutiny of cross-modality impairments (Stevenson et al., 2016), as we shall see in the next chapters. As our focus was voice processing, this study specifically manipulated social salience in voices without confounds related to the presence of other sensory modalities. Indeed, most studies of cortical tracking present their stimuli in the form of audiovisual recordings (Menn et al., 2022). This may represent an issue in results interpretation as ASD individuals tend to show specific impairments in the processing of cross-modal stimuli (Hyde et al., 2017). Moreover, accurate tracking of the speech envelope has been shown to be affected by intersensory integration mechanisms (Crosse, Di Liberto, & Lalor, 2016; Kalashnikova et al., 2018; Menn et al., 2022; Power, Fair, Schlaggar, & Petersen, 2010), either by enhancing acoustic processing itself or by providing additional information such as lip movements (Park et al., 2015; Menn et al., 2022). This facilitation has been shown to be especially strong in preverbal infants (Hillairet de Boisferon, Tift, Minar, & Lewkowicz, 2017) as we shall describe more in depth in Chapter 4.

Chapter 3

Emotional faces guide sensory and higher-order processing during development. [2]

As we have seen in the Introduction, to properly perceive the social environment two of the most fundamental cues are human faces and human vocalizations. Indeed, the face provides a wealth of socially relevant information often co-occurring with vocal cues (Grossman & Johnson, 2007), and the ability to recognize faces is considered a fundamental social adaptation appearing very early in development (Grossman & Johnson, 2007). In the previous chapter we discussed in depth the role that the human voice has in guiding development and in describing ASD, in which voices lose their salience. In this chapter we shall do the same with human emotional faces.

1. Introduction

As briefly outlined, investigating face processing in infants allows us to understand the shape and state of development of brain topological organization at specific developmental time points. This is in line with evidence that social processes involving interactions with others play a vital role in network shaping during development (van der Velde et al., 2021). In fact, facial expressions provide a remarkable source of information, allowing to infer others' internal dispositions and mental states. This is especially true for infants, who, through non-verbal social communication, learn about their environment and derive expectations about others' behaviours (Quadrelli, Conte, Macchi Cassia, & Turati 2019; Rotem-Kohavi, Oberlander, & Virji-babul, 2017). Indeed, detecting others' emotions through facial expressions is specifically important in social interactions (Grossman & Johnson, 2007). Recognizing facial expression allows us to detect another person's emotional state and provides cues on how to respond in social interactions (Grossman & Johnson, 2007). This explains why most brain structures involved in emotional processing are reportedly also fundamental for social behaviours (Grossman & Johnson, 2007).

In this context, research investigating the early development of emotional processing has mostly focused on static emotional expressions. Quadrelli, Conte and colleagues (2019) report that static happy faces entail faster allocation of attention in 7-month-old infants compared to angry faces, in the form of an earlier negative central (Nc) component. However, when facial expressions are presented dynamically, happy, and angry faces elicit similar neural activation while entailing greater allocation of attention

[2] Parts of this chapter are included in Polver, S., Quadrelli, E., Turati, C., & Bulf, H. (2022). Decoding functional brain networks through graph measures in infancy: The case of emotional faces. Biological Psychology, 170, 108292.

compared to neutral faces, in the form of a larger Nc. In interpreting their results, the authors suggest that 7-month-old infants are sensitive to facial dynamics and posit that the dynamicity of emotional expressions may play a role in shaping the neural processing of emotional expressions. While these results provide a considerable wealth of information regarding the processing of facial expression of emotion, they fail to address the possible topological variations across frequency bands and to discern between brain activities related to differential processing of static and dynamic emotional faces. An effective way to address the latter issue is to refer to decoding techniques by which it is possible to directly differentiate specific brain activities related to, for example, stimuli categories (Bae & Luck, 2018).

Capitalizing on existing evidence suggesting that dynamic emotional expressions are differentially processed compared to static displays of emotions in adults (Ambadar, Schooler, & Cohn, 2005; Recio, Sommer, & Schacht, 2011) and infants (e.g., Quadrelli, Conte, et al., 2019), the present study investigated brain topological properties of 7-month-old infants in response to static and dynamic emotional faces across different EEG frequency bands. More in depth, starting from the underlying structure of infants' brain activity in the alpha and theta bands, we sought to find the components specific to presentation modalities by decoding whether an individual was viewing either static or dynamic facial expressions of emotions. We hypothesized to find a differential pattern of brain topological organization in infants' processing of static and dynamic emotional faces, as a function of the involvement of, respectively, ventral and dorsal visual neural pathways (Pitcher et al., 2014), and due to the engagement of sensorimotor areas. Indeed, some influential models of adults' face perception (Calder & Young, 2005; Pitcher et al., 2014) posit that functionally distinct but integrated cortical pathways process different facial aspects. The ventral pathway, which includes the fusiform face area, preferentially responds to invariant facial aspects, while more lateral and dorsal pathways, including the posterior temporal sulcus, preferentially respond to changeable facial aspects, such as emotional expression. Interestingly, the literature suggests that the processing of dynamic and static faces in perceptual tasks might be functionally dissociable and exploit different but integrated pathways originating in the early visual cortex (Pitcher et al., 2014). Indeed, some studies have shown that, even if responses to faces in the ventral stream keeps on developing until late childhood, a gradual tuning to stimuli in sensitive cortical areas may occur quite early during development (Dekker, Mareschal, Sereno, & Johnson, 2011). In fact, it has been reported that 4-month-old infants already show cortical specialization in regions involved in the perception of facial communication cues, while the recognition of faces' invariant aspects (such as identity) is thought to specialize more gradually and require more widespread processing (Cohen Kadosh & Johnson, 2007; Grossmann et al.,

2008; Simic & Rovet, 2017). Hence, due to their different developmental trajectories, the perception of dynamic social communication cues and static invariant face characteristics have been suggested to be two distinct processes (Grossman et al., 2008). Furthermore, for what concerns sensorimotor areas, a substantial amount of evidence points towards the existence of a complex network of brain structures involved in the processing of facial emotional expressions, encompassing, besides extrastriate areas, structures involved in visual perception (Eimer & Holmes, 2007). These results are indeed not surprising, since facial expressions contain both emotional and motor components (Dapretto et al., 2006; van der Gaag, Minderaa, & Keysers, 2007). Thus, to investigate the presence of functionally specialized modules, we focused on the Within Module Degree Z Score (WMDZ), a topological measure derived from graph theory. Nodes highly connected to other nodes within the same functional module facilitate functional specialization (Bullmore, 2016), thus WMDZ values were expected to represent circumscribed activity patterns, related to the simultaneous activity of functional modules subtending specific processing of stimuli characteristics.

In the current study, we applied graph measures to analyse infants' functional brain activity across the alpha and theta EEG frequency bands in response to facial expressions of emotions, investigating whether these measures can represent useful tools to probe the underlying functional networks' organization during emotional perception tasks and the network configurations in response to emotional faces perception during development. We focused on the alpha and theta bands, since their configurations during development are well described. Regarding the alpha band, most research focused on its central configurations, encompassing the mu rhythm, and have found it to be consistently associated with sensorimotor activity even during development (e.g., Quadrelli, Geangu, et al., 2019; Saby & Marshall, 2012). The theta band in infancy has been mostly linked to the expression of emotions (Michel et al., 2015; Saby & Marshall, 2012) and to attentional modulation (Orekhova, Stroganova, & Posikera, 1999). We hypothesised that the topological structure of functional brain components would change among the alpha and theta bands, according to their differential functional properties. Regarding the alpha band we hypothesised to find specific activations related to motion in our dynamic condition, in accordance with the sensorimotor sensitivity of the alpha band (Quadrelli, Geangu, et al., 2019). Conversely for the theta band, we expected to find activations related to the emotional content of stimuli, in accordance with reports that find theta activity able to discriminate, for example, between social and non-social stimuli (van der Velde et al., 2021).

2. Methods

2.1. Participants

Participants were forty-three 7-month-old infants (Mean age = 217 days, SD = 13 days, range = 201-233 days), born at term (37-42 weeks of gestation) with normal birth weight (>2500 g), that did not suffer of any neurological or other medical conditions, and had normal vision and hearing for their age. The original participants sample consisted of 52 participants; however, 9 infants were excluded from analyses due to fussiness (N = 4), excessive artifacts (N = 4), and technical problems during data collection (N = 1). Infants were randomly assigned to one of two experimental conditions: 22 infants were presented with dynamic stimuli, while 21 infants were presented with static stimuli. The procedure followed the ethical standards (the Declaration of Helsinki, BMJ 1991; 302:1194) and was approved by the ethical committee of the University of Milano-Bicocca. All infants' caregivers gave their informed consent prior to study inclusion.

2.2. Stimuli

Dynamic stimuli consisted of short 1000 ms color videos of 9 female Caucasian actresses portraying angry, happy, and neutral facial expressions while facing forward. Static stimuli consisted of photographs depicting the full emotional expression extracted from videos used in the dynamic condition. Identities did not overlap since a different one was used for each emotion portrayed (Figure 1). While dynamic happy and angry faces were extracted from videos of the Binghampton University 4D Facial Expression database (BU-4DFE; Yin, Chen, Sun, Worm, & Reale, 2008), neutral faces were extracted from videos recorded at our laboratory. Using Adobe Photoshop, all images were cropped to remove hair and external features in order to emphasize and facilitate the processing of cues differentiating emotions. Stimuli were then equalized for luminance, which did not differ between emotion categories (Kruskal–Wallis H test, $\chi^2(2) = 5.60$; $p = 0.08$; $\eta^2 = 0.60$). Faces were pasted against a grey background. Then, stimuli were screened and selected for their emotional valence. Wilcoxon Signed-ranks tests computed on intensity scores for each emotion showed that both happy ($M = 7.20$; $SD = 0.81$), $Z = 3.84$; $p < 0.001$, $\eta^2 = 1.64$, and angry expressions ($M = -6.56$; $SD = 1.08$), $Z = -3.84$; $p < 0.001$, $\eta^2 = 1.64$, were perceived as equally different from neutral expressions, which instead was properly perceived as being non-emotional ($M = 0.42$; $SD = 0.89$), $Z = 1.814$; $p = 0.07$, $\eta^2 = 0.38$. The unfolding of each emotional expression (i.e., neutral to 100% intensity) lasted 500 ms, and the full expression remained on the screen until the end of the video. Static stimuli were presented for 1000 ms.

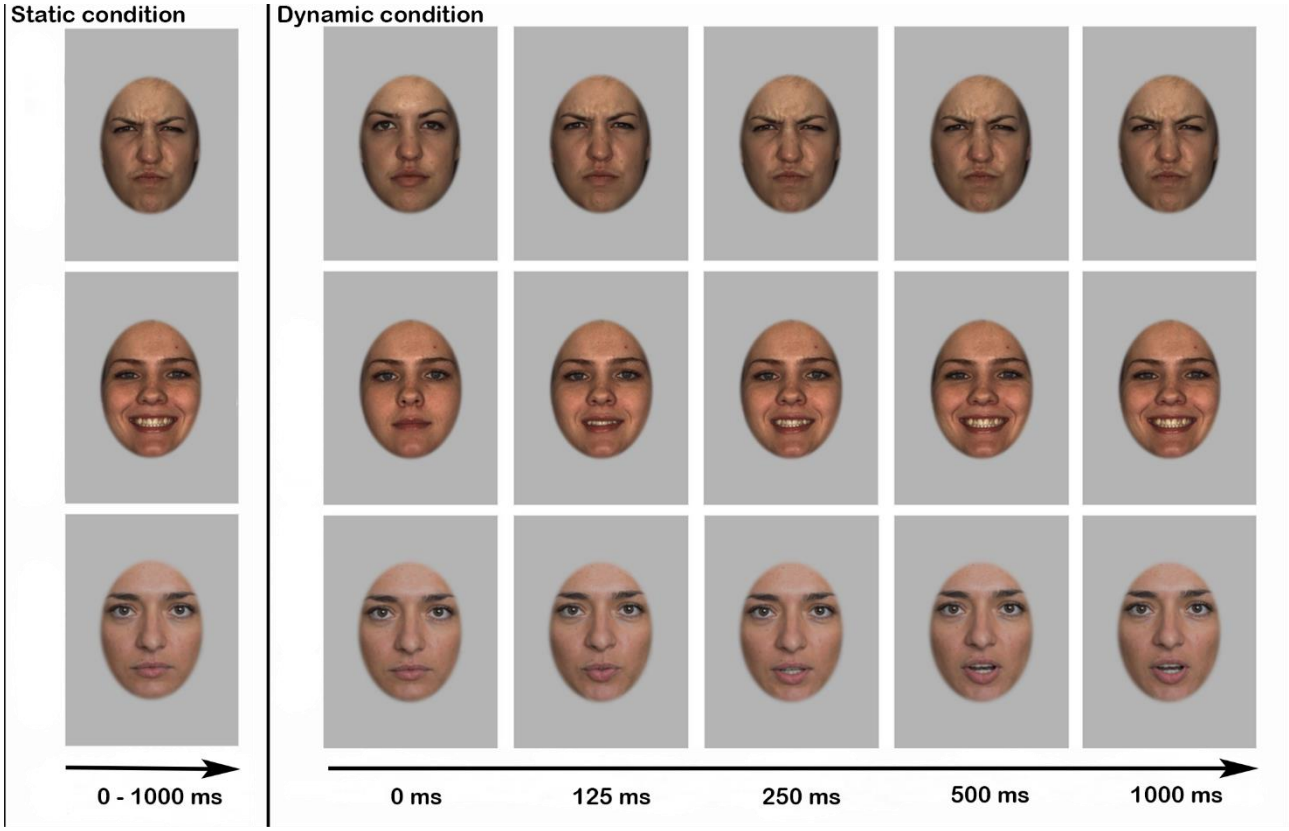


Figure 1. *Examples of the stimuli used in both static and dynamic conditions.*

2.3. Procedure

The experiment took place in a dimly lit, audiometric, and electrically shielded room, where participants were seated on their mother's lap, at approximately 60 cm from a 24-inch monitor, in quiet alertness. Stimuli were presented using the E-Prime software v2.0 (Psychology Software Tools Inc., Pittsburgh, PA). Mothers were instructed to remain as still as possible and keep silent during the experimental session to avoid acoustic interferences. The data acquisition computer displayed the live image of the infants' face and body to allow the experimenter to pause or terminate the session when the infant became too fussy. Each infant was presented with all the face identities (n = 9) randomly, with the only constraint that models expressing the same emotion could not occur more than three times in a row. The experimental session was terminated when infants attended to the maximum number of trials (N = 270) or got tired of the experiment. A trial consisted of 1000 ms stimulus presented centrally followed by an interstimulus interval, characterized by a black fixation cross at the centre of the screen against a grey background, which varied randomly in duration between 900 and 1100 ms. Whenever necessary the experimenter presented a looming fixation point between trials to reorient the infant's attention to the monitor.

2.4. EEG recordings

The EEG signal was recorded using a 128-electrode HydroCel Geodesic Sensor Net (Electrical Geodesic In., Eugene, OR) and sampled at 500 Hz through an EGI NetAmps 300 amplifier. The signal was recorded with respect to the vertex electrode. A

bandpass filter of 0.1 to 100 Hz was applied online, and impedances were kept below 50 KΩ. EEG data were further high-pass filtered offline (0.3 Hz), low-pass filtered (30 Hz) and divided into 2400 ms segments, beginning 1000 ms before and ending 1400 ms after stimulus onset. Trials in which the infant did not look at the screen or moved excessively were manually rejected after visual inspection of experimental video recordings.

2.5. Preprocessing and graph metrics computations

Data were preprocessed using the EEGLab toolbox (Delorme & Makeig, 2004) in Matlab (ver. R2020b). First, the epochs in which the infant did not pay attention to the stimuli were rejected. Then, non-functioning EOG channels were removed, and data were high-pass filtered at 0.1 Hz and low-pass filtered at 30 Hz. The 50 Hz line noise was removed via subtraction through a sliding window estimating sine wave amplitude (Bigdely-Shamlo et al., 2015). Data were then cleaned using the Artifact Subspace Reconstruction (ASR) algorithms in EEGLab (Plechawska-Wójcik et al., 2019): channels were deemed 'bad' and removed if either flat for longer than 5 seconds, or through computation of each channel's correlation to its RANdom SAmple Consensus (RANSAC) reconstruction for each window. The correlation threshold for channel removal was set at 0.8. Then, to all recordings, a Principal Component Analysis (PCA) decomposition was applied to identify artifactual PCs (defined by a comparison against the data's own cleanest parts) and activity was reconstructed from remaining components. More in depth, data were corrected through a 5 s sliding window PCA, classifying PCs in high variance (10 SD from clean data) or normal variance. High variance parts of data, which correspond to artifacts, were then reconstructed based on normal variance subspace components. After these steps, removed channels were spherically interpolated by referencing to the full-rank channel matrix and data were re-referenced to average reference. To further remove artifacts, the Independent Component Analysis (ICA) was performed on PCA reduced components. Usually, the number of data rank is the same as the number of channels, however due re-referencing and adding the previous reference back to the data our ranks exceeded what expected, while the opposite happened when we removed noisy channels. The ICA is heavily affected by ranks and tends to produce unreliable ICs under such conditions (Schneider, 2010). As we used the ICA to reject artifacts, we applied the PCA as a data reduction technique to avoid issues. Then, the resulting 64 components were subjected to automatic flagging of artifactual components through the ICLabel algorithm (Pion-Tonachini, Kreutz-Delgado & Makeig 2019). Components were rejected if had a minimum 0.7 probability of being either eye, heart, line, channel, muscle, or other artifacts. For the final epoch rejection, a threshold of +/- 150 µV was set to remove abrupt changes in amplitude, while two thresholds of 100 for slope and 0.5 for r^2 were set to remove step-like artifacts and linear trends. The mean number of artifact-free trials was

72.41 (SD = 30.91) for the dynamic condition (anger: 23.91, SD = 10.11; happiness: 24.27, SD = 10.40; neutral: 24.23, SD = 11.37) and 77.95 (SD = 30.99) for the static condition (anger: 26.48, SD = 11.17; happiness: 26.43, SD = 10.17; neutral: 25.05, SD = 10.82). To check for any disparity between conditions we performed a set of t-tests: no significant differences were found between the two conditions and emotions portrayed (all ps > .43). Data were then passed to FieldTrip (Oostenveld et al., 2011) for adjacency matrices computation. To avoid issues caused by volume conduction, we computed the scalp current source density (CSD) using the second-order derivative (the surface Laplacian) of the EEG potential distribution thorough a spherical spline method. On CSD data, we computed the imaginary part of the coherence (ICOH; Nolte et al., 2004) as a connectivity measure. The adjacency matrices thus calculated were left weighted. Frequencies of interest were the alpha (6-9 Hz) and theta (3-5 Hz) bands. One weighted adjacency matrix based on ICOH values was derived per subject, per condition and per emotion, for each frequency band. We considered electrodes as nodes and ICOH values as edges for WMDZ computation. To avoid arbitrary threshold choices to remove spurious connections, the WMDZ was computed on unthresholded adjacency matrices and then normalized by their random counterparts. To achieve this, we rewired ICOH values in the adjacency matrices to create random graphs that maintain the same weight, degree, and strength distributions as the original ones. Random graphs were thus computed by rewiring ICOH values every 10th step until the whole network was rewired to be completely random. On such null models, we again computed the WMDZ, and we normalized WMDZ values computed on the original graphs by WMDZ values computed on the random graphs. Specifically, the WMDZ was estimated for a node, *i*, as

$$z_i = \frac{k_i(m_i) - \bar{k}_i\,(m_i)}{\sigma_{k(m_i)}}$$

where $k_i(m_i)$ is the within-module degree (i.e., number of connections linking node *i* to other nodes in the same module, m_i), $\bar{k}_i\,(m_i)$ is the mean withinmodule degree of nodes in the same module as node *i*, and $\sigma_{k(m_i)}$ is the standard deviation of $k_i(m_i)$ values across all nodes in module m_i. A node with a within module degree z-score equal to zero has an average level of intramodule connectivity relative to other nodes in the same module. A positive z-score indicates above average intramodule connectivity and a negative score indicates below average intramodule connectivity (Bullmore, 2016). All steps were completed using the Brain Connectivity Toolbox (BCT) (Rubinov & Sporns, 2010) and ad-hoc scripts in Matlab.

2.6. PCA and PCR

To obtain the hypothesised differential motifs of infants' brain functional organization in response to static and dynamic facial expressions of emotions, we applied the PCA decomposition to detect structured patterns of brain activity in topological EEG data (Leonardi et al., 2013). The PCA as input of the PCR was applied on the computed WMDZ values per channel. On extracted components, we performed the Principal Component Regression (PCR). The PCR uses PCA outputs as new regressors (Kendall, 1957), in this way, the independent variables are rendered orthogonal, to prevent errors caused by dependencies across variables (Jolliffe, 1982). One major advantage of this technique is that it can be performed on datasets with a high number of predictor variables (Hadi & Ling, 1998), such as in our case, since all extracted components were considered as predictors. More in depth, to identify the appropriate number of features to extract for each frequency band and to avoid overfitting, data were first subjected to parallel analysis. Parallel analysis represents one way to determine the appropriate number of components to extract and is performed by comparing the components' solution to random data with the same properties as the real data set. If the input is a data matrix, the comparison includes random samples from the real data, as well as normally distributed random data with the same number of subjects and variables. Then the PCA was carried out with the varimax orthogonal rotation (selected after correlation matrices exploration), to increase components' explanatory power. Extracted components were then reverted to their total score matrices and such matrices were used for statistical analysis. PCA scores index the contribution of each node to the overall component, after the linear transformation applied by the PCA (Jolliffe & Cadima, 2016). A PCR was then used, in the form of a binomial logistic regression, in which condition membership (static vs. dynamic) was considered as the dependent variable, while extracted components were considered independent variables. All statistical tests were conducted on a 0.05 level of significance (two-tailed) and results underwent False Discovery Rate (FDR) corrections (Benjamini & Hochberg, 1995). The aim of these analyses was to determine if components derived from brain activation were strong enough to predict condition membership in a data driven way (Mihalik et al., 2020). Since the dynamic condition was set as reference, a negative estimate indicated that participants were more likely to have seen dynamic stimuli. Thus, the higher the component's eingevalues were, the higher the probability the subject had to belong to one of the two conditions depending on coefficients signs. Finally, to control for effects of the emotions portrayed we computed a multinomial logistic regression with emotional category as predictor.

All analyses were performed in R (ver. 4.0.3), with the aid of the psych (Revelle, 2016) and nnet (Venables & Ripley, 2002) packages, while circular graphs for components visualization were computed in Matlab.

3. Results

Inspection of parallel analysis results led to the extraction of 22 principal components respectively for the alpha and theta bands. The high number of components extracted was due to the fact that the WMDZ identifies patterns of intra-modular connectivity, and in fact, the spatial configuration of components was coherently circumscribed. The scalp configurations of components were inspected referring to electrodes' loadings. Indeed, higher loadings define the relative weight of that electrode in the component.

Results of the PCR indicated that there was a significant association between experimental condition (static vs. dynamic) and PCA WMDZ components in all frequency bands. In the alpha band, $\chi^2(22) = 49.26$, $p < .001$) the eleventh component proved to be significant ($b = -1.05$, $z = -3.49$, $p = .01$). This component encompassed mainly right temporal, frontal, central and occipital electrodes, and its activity indexed dynamic stimuli presentation (Figure 2). No other component attained statistical significance (all ps > .13).

In the theta band, $\chi^2(22) = 50.14$, $p < .001$) the sixth ($b = 0.77$, $z = 3.13$, $p = .01$), the eighth ($b = 0.98$, $z = 3.54$, $p = .009$) and the thirteenth ($b = -0.75$, $z = -2.76$, $p = .04$) components were significant. The sixth component mainly encompassed centro-parietal and temporal electrodes, while the eighth component comprised frontal, left temporal and electrodes; both components indexed activity in the static condition. The thirteenth component comprised temporo-parieto-occipital and central electrodes and indexed activity in the dynamic condition (Figure 2). No other component attained statistical significance (all ps > .20).

Finally, results for the PCR were never significant across the emotional categories and frequency bands (all ps > .56), ruling out the possibility of a differential effect of portrayed emotions.

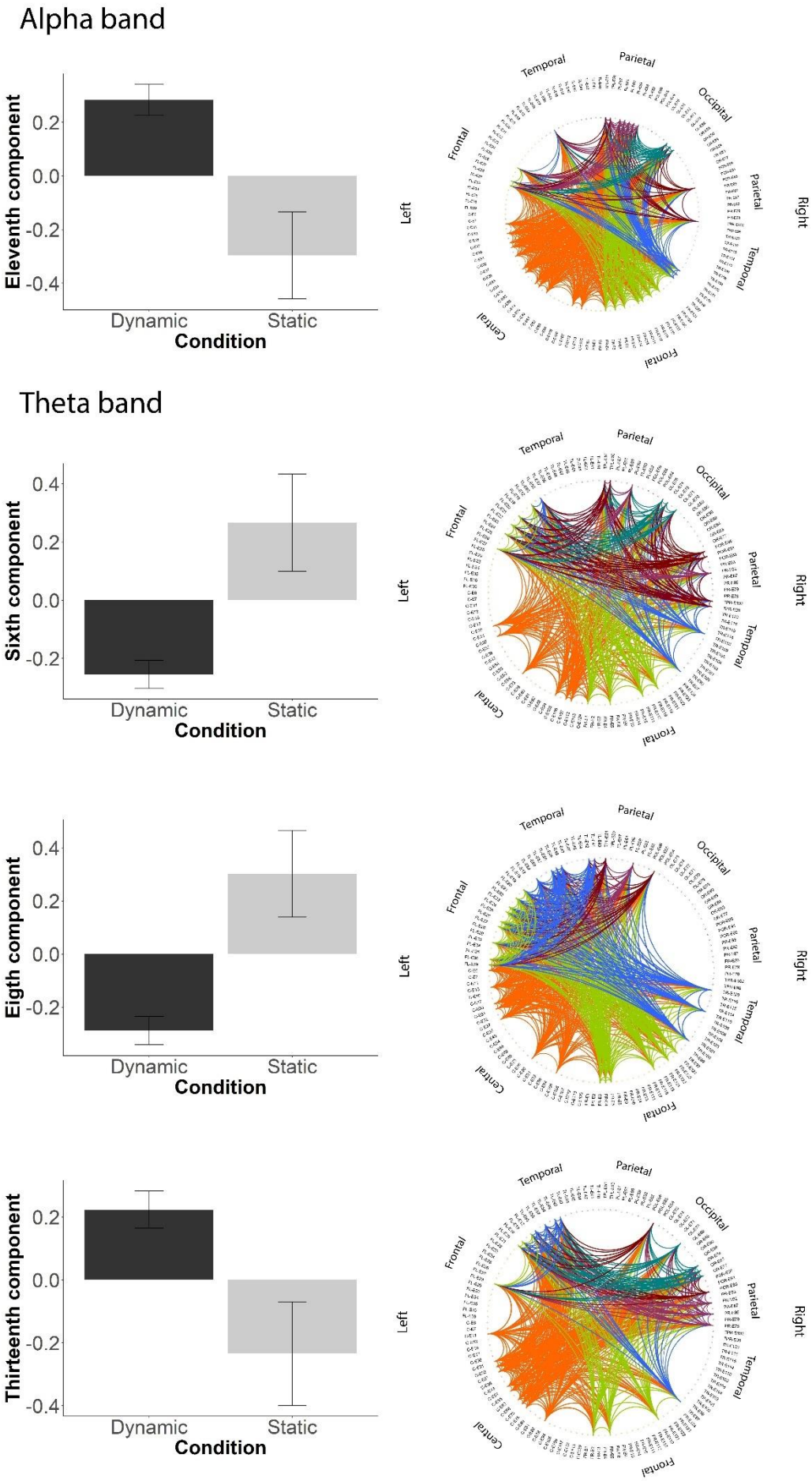


Figure 2. *Visualization of all significant components for WMDZ results across frequency bands. The circular graphs represent the topological organization of significant components indexing activity in either static or dynamic stimuli as per PCR results. The components' functional structure*

is plotted against the barplot showing the direction of PCR results. The y axis represents components' scores. Error bars represent the standard error of the mean.

4. Discussion

This study aimed to disambiguate differential sets of brain activity in response to static and dynamic facial expressions of emotion in a sample of 7-month-old infants. We applied graph theoretical measures to assess the topological structure of functional networks. We thus decomposed EEG data with the PCA and applied the PCR to evaluate the components specific to the static vs dynamic facial expressions of emotion in each EEG frequency band.

Results for the alpha band show the presence of a specific modular structure indexing activity in response to dynamic emotional faces. From a developmental perspective, we may assume that, as in other domains such as language, the processing of emotional faces requires the recruitment of broader cortical areas within the first months of life as compared to adulthood (Johnson, Grossmann & Kadosh, 2009; Johnson, 2011). This finding is in accordance with the hypothesis that infants require the involvement of wider sets of brain areas for face processing compared to adults (Grossman et al., 2008). As the alpha band has been implied in perceptual and cross-modal binding in the context of large-scale integration processes (Hummel & Gerloff, 2005), alpha-mediated communication between distant brain areas may provide the basis for integration of complex information. In this sense, activity in sensorimotor networks may scaffold more advanced processes, such as those triggered by dynamic facial expressions of emotion (Fransson, Åden, Blennow, & Lagercrantz, 2011). Finally, our results further support the specific tuning of alpha band activity in response to dynamic emotional faces presentation across techniques (Rayson, Bonaiuto, Ferrari, & Murray, 2016).

Results for the theta band were coherent with the differentiation between visual streams: we observed that activity in response to static emotional faces was more ventral, while activity in response to dynamic emotional faces was more dorsal. Indeed, components indexing activity in the static condition encompassed temporal electrodes, while the component indexing activity in the dynamic condition encompassed central and parietal electrodes. However, besides differentiations we observed invariances in topological organizations. WMDZ components across static and dynamic emotional faces implicated mostly the same topological areas, albeit with different electrode compositions. Such activities may index an invariance of face processing in infants' brains, going beyond presentation modalities. Our results are in line with findings in older children showing increased theta activity in both frontal and temporal areas in response to social stimulation (Bazhenova et al., 2007). In addition, since theta activity occurs within the suitable time range of Hebbian learning processes, high levels of theta activity in infants may reflect a

brain state optimal for synaptic plasticity, related to visuomotor learning (Saby & Marshall, 2012). Overall, results for the theta band are in line with the hypothesis of a specific functional role of theta oscillations in the development of brain areas subtending social processing, otherwise known as the social brain network (van der Velde et al., 2021).

While several ERPs studies report differential effects in infants' neural processing of emotional expressions (Missana, Rajhans, Atkinson & Grossman, 2014; Missana, Atkinson & Grossman, 2015; Quadrelli, Conte et al., 2019), there are many methodological differences that might explain observed discrepancies between the existing literature and our results. For one thing, the temporal accuracy of ERP responses may render them more sensitive to capture fast network configurations, thus shedding light on the differential neural responses to emotional categories. Conversely, through graph measures, we gain insights into networks' global organizations (Bullmore & Sporns, 2009). Furthermore, the lack of differentiations between emotional categories highlighted by our analyses is consistent with literature using other brain imaging techniques, such as the fNIRS (Bayet et al., 2021). In this vein, the absence of differences between emotional categories might be related to immature neural network organizations related to face processing in the first year of life (Grossman, 2010). It is then likely that, throughout development, we may observe the emergence of more specific network configurations related to the maturation of emotion processing abilities (Quadrelli, Roberti, Polver, Bulf, & Turati, 2021). In this sense, differences across emotional categories may have not surfaced because the stimuli were not salient enough to elicit detectable changes in the whole brain's functional architecture. Through the use of more realistic stimuli involving multiple sensory systems, we may then be able to detect differences.

Altogether, our results are coherent with previous findings using graph theoretical measures to assess brain electrical activity in newborns, pointing toward differential patterns of functional network topologies across frequency bands (Tóth et al., 2017). In fact, it has been proposed that the emergence of discrete EEG frequency bands in infancy and childhood may reflect the maturation of the cortical generators producing such rhythms (Tóth et al., 2017). As a conclusion, we can assert that brain activity does contain enough information to differentiate between the processing of static and dynamic facial expressions of emotions. The invariances in brain responses we observed may then reflect the existence of a rudimental network tuned to face processing already at 7-months of age. In fact, we found coherence between WMDZ components, indexing organized brain activity in early infancy. However, such organizations, characterized by differential dorsal and ventral activities, coupled with activations in sensorimotor areas, might not yet be fully mature and, as the infant grows, we may observe a progressively less central role for sensory areas, associated with the emergence of more advanced modular

organizations around higher-order association areas, such as parietal and frontal cortices. This is especially interesting since the prevalence of cortical hubs in primary brain regions may provide scaffold for higher-order behaviours maturing during development (Fransson et al., 2011; Santamaria et al., 2020). Thus, as the infant grows, the orientation toward specific emotional categories might become progressively more fine-tuned. Indeed, we may hypothesise that as the infant brain develops, the ability to roughly differentiate between emotional stimuli based on their motion will turn into a more fine-tuned specialization for emotional processing. Such specialization, even if not yet strongly embedded in the brain's organization, appears to be underway at 7 months as indexed, for example, by coherent topological organizations in response to emotional faces irrespectively of presentation modality. This hypothesis should, however, be better assessed in future studies.

While we were not be able to isolate the contribution of single brain areas in this study, it is nonetheless noteworthy to heed that the spatial locations of the topological areas we observe are remarkably similar to those of areas such as the fusiform and occipital gyri (Grossman & Johnson, 2007; Maguinness et al., 2018). Interestingly, face processing in infancy also recruits areas such as the inferior frontal gyrus and the superior temporal sulcus, which are usually identified as associative areas part of the voice processing network in adults (Grossman & Johnson, 2007; Maguinness et al., 2018; Belin et al., 2004). This finding highlights the extensive interdependencies of brain networks involved in social processing (Grossman & Johnson, 2007; Maguinness et al 2018). However, as we have consistently reported in these chapters, emotional processing has been mainly studied in each sensory modality separately (Grossman & Johnson. 2007). Thus, the main question that should arise is: how emotional information from voices and faces are integrated? Finally, while some reports on individuals with ASD have shown dysfunctions in the processing of both emotional facial expressions (Adolphs, Sears, & Piven, 2000) and voice prosody (Myers et al., 2019), other show preserved skills (Grossman & Tager-Flusberg, 2008). This may indicate a potentially significant effect of perceptual and/or social complexity on socio-emotional processing in ASD (Shic et al., 2013; Grossman & Tager-Flusberg, 2012). Indeed, 6-months-old infants that would be later diagnosed with ASD, showed reduced attention to key facial features only in response to talking faces (Shic et al., 2013). This again suggest that ASD may be a disorder of complex intersensory binding (Shic et al., 2013). We shall specifically address this statement in the next chapter

Chapter 4

Multimodal temporal binding of emotional vocalizations: typical and atypical brain development.

As we have uncovered brain responses to unimodal visual and acoustic social stimuli in the previous chapters, we will now combine visual and auditory cues to characterize both typical and atypical social processing in early infancy. Indeed, combining information from different modalities is necessary to perceive and interact with the complex social environment surrounding us (Bahrick, McNew, Pruden, & Castellanos, 2019). Moreover, the binding of stimuli from different sensory sources is especially important in infancy as it aids social and communicative development (Hyde et al., 2017). Given its importance in social interactions, it is no wonder that intersensory binding is found to be specifically impaired in ASD (de Boer Schellekens, Keetels et al., 2013).

To this end, we will shed light on intersensory binding in adults, low risk (LR) and infants at higher likelihood for ASD (HL-ASD) by presenting positive, negative, and neutral audiovisual (AV) recordings. By manipulating the synchrony of the acoustic and visual components we will address the brain mechanisms underlying intersensory binding. Indeed, through our analyses, on one hand we will be able to characterize brain networks in a low risk sample and relating it to the developmental end point represented by adulthood. On the other, our analyses may help answer an ever-present question in the literature regarding the source of observed impairments in ASD. By addressing these questions from state-of-the-art neuroscientific methods, we hope to aid the search of early developmental markers of the disorder before full-blown behavioural deficits are established.

1. Introduction

A fundamental feature that guides attentional allocation in early development is intersensory redundancy (Curtindale, Bahrick, Lickliter, & Colombo 2019; Lewkowicz, Schmuckler, & Agrawal, 2021). Intersensory redundancy can be defined as the simultaneous availability of the same information across two or more sensory systems (e.g., auditory, visual; Curtindale et al., 2019; Bahrick et al., 2019). Indeed, in everyday life interactions are complex and, for their processing, require the involvement of multiple sensory systems (Cecere, Gross, Willis, & Thut, 2017; Coutrot & Guyader, 2014). Redundant and temporally synchronous information guides selective attention, aiding perception, learning of events, and memory (Balk et al., 2010; Calvert, Hansen, Iversen, & Brammer, 2001; Curtindale et al., 2019; Feldman et al., 2018; Lewkowicz et al., 2021; Patten, Watson, & Baranek, 2014).

It has been proposed that infants, like adults, can integrate redundant multisensory information early in sensory processing (Hyde et al., 2017). Within this context, social stimuli are highly redundant compared to non-social ones. For example, speech is characterized by lips movements that are highly coherent in time with audible information, naturally guiding attention and aiding learning in the infant (Coutrot & Guyader, 2014; Lewkowicz et al., 2021; Patten et al., 2014; Hillairet de Boisferon et al., 2017). In addition, changes in the tempo and the rhythm of audiovisual (AV) events are detected more readily by the infant when acoustic and visual information are presented together rather than when they are presented either in just one sensory modality or out of synchrony (Bahrick et al., 2019). Moreover, redundancy across synchronous facial and vocal cues facilitate the detection of emotional cues in both infants and adults (Bahrick et al., 2019). Consequently, for infants to benefit from various forms of AV redundancy, the integration of cross-modal features into a single coherent percept is needed (de Boer-Schellekens, Eussen, et al., 2013; de Boer-Schellekens, Keetels, et al., 2013; Hillairet de Boisferon et al., 2017; Feldman et al., 2018; Stevenson et al., 2016).

Intersensory temporal synchrony, by which two stimuli originating from the same source will have highly similar temporal structures, is arguably the simplest and developmentally earliest basis of multisensory binding (Calvert et al., 2001; Lewkowicz, 2003; Lewkowicz et al., 2021; Stevenson et al., 2016). Indeed, the perception of the temporal relationships between stimuli is especially important in guiding the binding of two stimuli originating from different sensory sources (Stevenson et al., 2016; Zhou, Cheung, & Chan, 2020). However, two stimuli don't have to be perfectly aligned for integration to happen (Kopp, 2014; Stevenson et al., 2016). Rather, stimuli must appear within a specific temporal window to be integrated (Stevenson et al., 2016). This window is commonly known as Temporal Binding Window (TBW; (Colonius & Diederich, 2004; Stevenson et al., 2016). The temporal range of the TBW shrinks across the lifespan: multisensory abilities improve and responsiveness to cross-modal stimuli increase accordingly (Kopp, 2014). The late maturation of the TBW may thus coincide with protracted maturation of multisensory brain networks (Zhou et al., 2020). Moreover, the width of the window partially depends on the nature of the stimulus being processed (Cook, Van Valkenburg, & Badcock, 2011). It's relatively short for simple stimuli and widens at increasing stimulus complexity (Cook et al., 2011; Zhou et al., 2020; Lewkowicz et al., 2021). In this sense, the ability of the perceptual system to flexibly adapt and reconcile small temporal asynchronies highlights its refinement (Navarra et al., 2005). However, there is no agreement on which aspects of stimuli complexity guide changes in the TBW (Cook et al., 2011).

While the ability to form unified percepts of the multisensory world from unisensory inputs is a fundamental brain function (Balk et al., 2010; Feldman et al., 2018), there is

surprisingly little research on how cross-modal integration occurs, its developmental changes, and its neural substrates (Hyde, Jones, Flom, & Porter, 2011; Kopp, 2014; Kopp & Dietrich, 2013). Regarding brain areas involved in the integration of percepts, it has been proposed that cross-modal interactions start in regions mostly thought to be unimodal and proceeds into multisensory ones (Bonath et al., 2014; Macaluso, George, Dolan, Spence, & Driver, 2004), with reciprocal feedback among regions influencing and modulating activity (Bonath et al., 2014; Calvert et al., 2001: Cecere et al., 2017; Macaluso et al., 2004). Interestingly, this reciprocal feedback mechanism resonates with what we uncovered in Chapter 1 between temporal and frontal regions. One of the higher-order regions thought to be central in cross-modal binding is the superior temporal sulcus (STS; Balk et al., 2010). Indeed, the STS is involved in low-level auditory processing (Chapter 1), voice processing (Chapter 2; Belin & Grosbas, 2010), and also face processing (Chapter 3; Maguinness et al., 2018). However, no brain area works in isolation and large-scale networks have been repeatedly reported to be involved in intersensory processing (Calvert et al., 2001; Adhikari, Goshorn, Lamichhane, & Dhamala, 2013). Regions thought to comprise these networks, besides the STS and primary sensory cortices, are the superior occipital cortices, the inferior parietal sulci and lobules, the ventromedial prefrontal cortex, the superior frontal gyri, the insula, and the planum temporale (Calvert et al., 2001; Lewis & Noppeney, 2010; Macaluso et al., 2004).

Networks interactions happen very early in processing stages, as early as 50 ms (Cecere et al., 2017). Therefore, the temporal resolution of the EEG is especially useful in this context (Zhou et al., 2020), more so if coupled with source reconstruction and data analysis techniques allowing us to get time-resolved insights into functional networks configurations, partly overcoming the spatial limits of the EEG (Cecere et al., 2017; Kopp, 2014; O'Sullivan, Crosse, Di Liberto, de Cheveigné, & Lalor, 2021). A putative mechanism to bind unimodal sensory features in a single percept is oscillatory brain activity (Cecere et al., 2017; Doesburg et al., 2008). Indeed, inputs in one sensory modality can influence stimulus processing in a different modality by synchronization of ongoing oscillations (Cecere et al., 2017; Covic, Keitel, Porcu, Schröger, & Müller, 2017; Doesburg et al., 2008; van Driel, Knapen, van Es, & Cohen, 2014). Thus, long-range phase synchronization between neural groups provides a mechanism for information transfer and may explain how distributed networks are able to exhibit transient enhancements of functional connectivity (Fries, 2005). In this context, neural populations oscillate at a specific frequency in a synchronous manner binding together percepts (Doesburg et al., 2008). Almost all frequency bands have been proposed to be involved in this mechanism, from the delta up to the gamma band. However, since most studies focus on a single band in isolation their role in multisensory integration has remained elusive.

Besides through which mechanisms the brain achieves perceptual integration, another open question regards what happens when the brain is impaired in its cross-modal binding function. A prime example is autism spectrum disorder (ASD). ASD is characterized by deficits in social interactions, communication, and by restricted interests and/or repetitive behaviours (American Psychological Association, 2013). A key finding in ASD is disruptions in intersensory integration (Feldman et al., 2018), which may take the form of a wider TBW (de Boer-schellekens, Eussen, et al., 2013; de Boer-schellekens, Keetels, et al., 2013; Feldman et al., 2018; Stevenson et al., 2016; Turi, Karaminis, Pellicano, & Burr, 2016). Indeed, poor temporal acuity across the visual and auditory domains is reflected in a decreased ability to bond sensory inputs together, leading to asynchrony in behavioural responses to such stimuli (de Boer-schellekens, Keetels, et al., 2013; Murat Baldwin, Xiao, & Murray, 2021; Stevenson et al., 2016). One hypothesis is that a widened TBW may reflect issues in combining irrelevant sensory information, resulting in a confusing and unpredictable environment (Zhou et al., 2020). For example, the failure to utilize temporal cues to match facial information with auditory signals may cause ASD individuals to have trouble in communicating effectively with others (Zhou et al., 2020). Nevertheless, the exact nature of such impairments remains elusive (Righi et al., 2018). For example, ASD individuals tend to not show impairments with simpler stimuli such as clicks and flashes (de Boer Schellekens, Keetels et al., 2013), pointing to toward ASD being a disorder of complex information processing (de Boer Schellekens, Keetles et al., 2013). However, there is no agreement on the exact facets of complexity resulting in impairments. More in depth, do impairments regard complexity per se, so that we may expect to find disruptions at increasing perceptual complexity of stimuli regardless of their contents, or impairments derive from social aspects naturally embedded in complex social stimuli? In other words, what guides impairments are higher-order perceptual features or are social features making the stimulus too complex for ASD individuals? Indeed, if complexity does not impact stimuli binding per se, attention orienting in ASD may not be tuned to prioritize highly social stimuli (Abrams et al., 2013; Patten et al., 2014; Shic, Wang, Macari, & Chawarska, 2020). In this sense, studies using stimuli high in sensory complexity with and without social cues are needed to disentangle such mechanisms (Crosse et al., 2016; Eg & Behne, 2015; Patten et al., 2014). One first step is to refer to non-speech emotional interjections, such as laughs, cries, screams, which are inherently social but don’t imply linguistic processes that can be affected by developmental delays (Belin et al., 2004). Moreover, during emotion processing vocalizations are encoded very early in the brain (Pell et al., 2015). This makes emotional vocalization main candidates to probe the nature of impairments in multisensory temporal binding in ASD.

While symptoms of ASD, including sensory difficulties, occur early in life, with an 82% stability of ASD diagnosis by the age of 2 years (Ozonoff et al., 2015; Riva et al., 2022), a diagnosis is typically not made before the age of 3 years and early detection of biomarkers of the disorder represents a significant challenge (Riva et al., 2018). For this reason, one of the most relevant approaches for the search of early markers is focusing on infants at higher likelihood of developing ASD (HL-ASD; Riva et al 2018, 2022). Indeed, evidence shows that HL-ASD show atypical sensory processing in a number of sensory modalities as early as the first months of life. For example, 12-months-old HL-ASD infants fail to show AV integration even with preserved visual processing, pointing toward a specific deficit in cross-modal integration, mirroring results with older diagnosed children (Riva et al., 2022). In this sense, studies on siblings provide the unique opportunity to identify very early biomarkers of the disorder aiding early diagnosis and interventions, prior to any overt behavioural changes (Riva et al., 2018, 2022).

The aim of this study was to address multisensory binding in 9-months-old infants and in their developmental end point represented by adulthood, with a particular focus on the deviations from the typical pathways characterizing ASD. Capitalizing on the knowledge derived from investigations of both low-level and higher-order unisensory information reported in the previous chapters, we here focus on the temporal binding of AV emotional vocalizations. To address both typical and impaired emotional processing we used AV recordings of two actresses crying, laughing, and coughing thus comprising positive, negative, and neutral categories. We want to highlight that we chose to focus on emotional signals that both infants and adults spontaneously produce, increasing the possibility of cross-sectional results interpretation (Blasi et al., 2011). In this context, some studies report that both adults and infants allocate more resources to the processing of negative compared to positive emotional displays, suggesting an evolutionary propensity to react more strongly to negative compared to either positive or neutral stimuli (Grossman et al., 2005). We expected HL-ASD to fail in the appropriate salience attribution to stimuli, resulting in impairments in attentional allocations to negative emotional stimuli. Considering the earlier development and advantage of auditory over visual systems (Grossman, 2010), we investigated sensory binding by anticipating the audio over the video and by applying a complexity-scaled and appropriately large TBW of 1.2 s. These manipulations allow us to gather insights on the mechanisms in place when the brain tries to bind stimuli expected to belong together. Indeed, notwithstanding the intentional width of the TBW used, we expect HL-ASD to fail in their attempts at integration. These failures may take the form of reduced engagement of higher-order regions and in the absence of feedback across networks.

To acquire as realistic as possible insights into brain network configurations, their interactions and feedbacks across time, and their developmental pathways, we source-reconstructed EEG activity in both infants and adults. By leveraging the time resolution of the EEG, we characterized brain areas interactions across time by referring to time-varying connectivity analysis. To our knowledge the time-resolved characterization of EEG functional networks was never done before in developmental samples. This is also one of the reasons why we chose to collect an adult sample, to be able to reliably describe our results without task confounds.

By combining state-of-the-art analysis pipelines with theoretical insights from the developmental literature we hope, once our samples will be completed, to characterize binding mechanisms in the infant brain in response to highly complex socials stimuli and their deviations from the norm. Indeed, as noted in Chapter 2, due to the Covid-19 pandemic, data collection has significantly slowed down. Indeed, as data collection takes place in the IRCCS Eugenio Medea healthcare facility as part of a broader longitudinal project, extra care was taken to avoid the spread of the virus in susceptible patients. For this reason, data collection is still ongoing. Here, we shall discuss preliminary results obtained through the application of our analysis pipelines on the samples we collected thus far. Due to the health-care primary function of the IRCCS, the number of LR infants specifically collected for research projects is lower compared to the number of HL-ASD infants accessing the structure. Consequently, as our LR sample is composed thus far by only 2 infants, we chose to exclude them from our analyses for now.

2. Methods

2.1. Participants

The infant sample is part of an on-going longitudinal project aiming at identifying early risk markers in ASD, and in the general population, focusing on infants at higher likelihood for ASD (Riva et al., 2018;,2022) conducted at the IRCCS Eugenio Medea. Inclusion criteria were both parents being native Italian speakers, gestational age ≥36 weeks, birth weight ≥2500 g, and Griffiths general quotient ≥85 (Griffiths, 1984), the absence of known medical, genetic, or neurological conditions, and major complications in pregnancy and/or delivery likely to affect brain development. To address the developmental end point of networks' organizations in response to intersensory binding we collected 40 healthy, native Italian speakers, young adults (13 males) with mean age of 22.5 years (SD = 3.53). As we shall see in the subsequent sessions, after preprocessing 6 adults were excluded from analyses leaving an overall sample of 34 participants. Participants were recruited by advertising at the University of Milano-Bicocca and at the Eugenio Medea Scientific Institute. Moreover, this study comprises preliminary investigation of 12 HL-ASD infants (M = 9.87 months, SD = 1.03). Again, as we shall see

in the next sessions, after preprocessing 3 HL-ASD infants were excluded from analyses. Written informed consent was obtained from parents or from the participant, in the case of the adult sample, prior to testing. The experiment was performed in accordance with relevant guidelines and regulations and was approved by the Eugenio Medea Scientific Institute Ethical and Scientific Committees.

2.2. Stimuli

Stimuli were the same for both infants and adults as we were interested in addressing the configurations of brain network within development and at the end of growth trajectories. By presenting the same stimuli to all samples we ensured that the same perceptual and integration systems would be at play, ruling out experimental design confounds from our analyses. Specifically, stimuli were AV recordings of two Caucasian actress laughing, crying, and coughing. While laughing represented a positive emotional cue, and crying represented a negative emotional cue, coughing represented the neutral condition. The actresses were trained to behave as natural as they could while trying to match the duration of their vocalizations. Moreover, the actresses wore the same white t-shirt, their hair were tied, and were recorded half bust against a black backdrop (Figure 1). We chose to keep such details in order to maintain stimuli as naturalistic as possible. Nonetheless to avoid confounds related to low-level stimuli characteristics, recordings were matched for luminance using in-house Matlab (ver. R2021a) scripts. Stimuli were presented in a synchronous or asynchronous way, with the audio anticipated by 1.2 s. Regardless of condition, after 1 s of still face the audio started. Only if the stimulus was asynchronous face movement began after 1.2 s compared to the audio (Figure 1). Stimuli duration was 3.48 s. All videos processing was done using Adobe Premiere. Presentation was randomised across 4 blocks, each block comprised 3 repetitions of each stimulus, controlling for actress' identity. The experiment was interrupted when the infant became too fussy to continue. Adults completed all blocks. While stimuli were presented using E-Prime 3.0 (Psychology Software Tools Inc., Pittsburgh, PA), participants were positioned approximately 95 cm from the monitor and audio intensity was set at 55 dB free field.

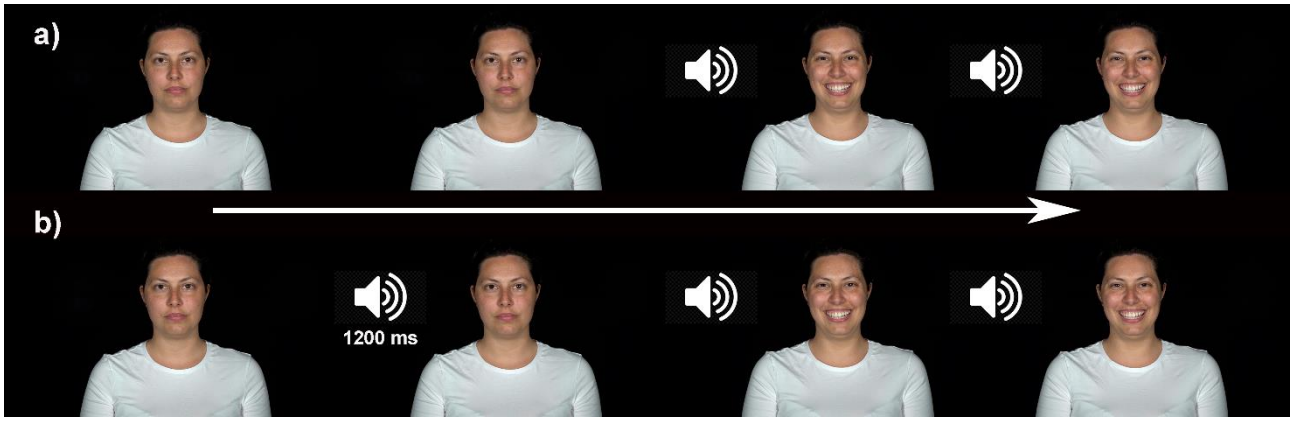


Figure 1. *Graphical depiction of synchronous a) and asynchronous stimuli b).*

2.3. EEG data collection and preprocessing

The infant sat comfortably on his/her caregiver lap in a sound-attenuated and electrically shielded room during EEG data collection. Adults likewise sat on a chair in the same room. The EEG signal was recorded using a 128-electrode HydroCel Geodesic Sensor Net (Electrical Geodesic In., Eugene, OR) with age-appropriate caps. The vertex was used as online reference and the signal was sampled at 500 Hz. Prior to preprocessing, infant data were subjected to the manual flagging of events in which the infant did not look at the screen. Data were then preprocessed using the EEGLab toolbox (Delorme & Makeig, 2004) in Matlab (ver. R2021a). For both infants and adults, data were high-pass filtered at 0.1 Hz and low-pass filtered at 100 Hz. Cut-off frequencies were 0.05 and 100.05 Hz (-6 dB). The 50Hz line noise was removed by using the CleanLineNoise function incorporated in the PREP pipeline (Bigdely-Shamlo et al., 2015). Data were then cleaned using Clean Rawdata EEGLab plugin (Kothe & Makeig, 2013; Plechawska-Wojcik et al., 2019): channels were deemed 'bad' and removed if either flat for longer than 5 seconds, or through computation of each channel's correlation to its RANdom SAmple Consensus (RANSAC) reconstruction for each window. The correlation threshold for channel removal was set at 0.8. The maximum number of channels rejected was 13. Then, to all recordings, a PCA (Principal Component Analysis) decomposition was applied to identify artifactual PCs (defined by comparison against the data's own cleanest parts) in order to reject them and reconstruct activity from the remaining components. Windows' standard deviation was set at 20. After these steps, removed channels were spherically interpolated by referencing to the full-rank channel matrix and data were re-referenced to average reference. To further remove artifacts, the Independent Component Analysis (ICA) was performed on PCA reduced components. The resulting 80 components were subjected to manual rejection of eye, heart, line or muscle artifacts. The maximum number of rejected components was set to 8 for the infant sample and to 20 for adults. The difference is number is due to the increased similarity in the spectro-temporal characteristics of artifacts in adults, allowing them to be better isolated by the ICA and thus conveniently removed. Event markers during recording were in conjunction with stimulus onset, we thus shifted events by 1 s to match the audio onset. Data were then divided into 2500 ms epochs, beginning 1 s before the audio onset and ending 1500 ms after. For the final epoch rejection, a threshold of +/- 200 µV for adults and of +/- 300 µV for infants were set to remove abrupt changes in amplitude, while thresholds of 160 µV slope for adults and 170 µV slope for infants and 0.5 for adults and 0.8 for infants for r^2 were set to remove step-like artifacts and linear trends. In this step we also rejected the events in which the infant was not looking at the screen. We did not do so before as some preprocessing steps, for example the ICA, perform better with continuous data. In the HL-ASD group the

mean number of trials for the asynchronous laughter was 6.66 (SD = 2.39), for the synchronous laughter was 7.44 (SD = 3.8), for the asynchronous cough was 7.77 (SD = 3.59), for the synchronous cough was 7.11 (SD = 2.75), for the asynchronous cry was 7.66 (SD = 3.27), for the synchronous cry was 7.11 (SD = 2.31). To check for any disparity between conditions we performed a set of t-tests: no significant differences were found ($p > 0.055$). Our developmental sample thus far consists of nine 9-months-old HL-ASD infants. Three infants were excluded from analysis due to excessive artifacts (N = 1) and excessive fussiness (N = 2). In the adult sample the mean number of trials for the asynchronous laughter was 20.67 (SD = 4.06), for the synchronous laughter was 19.73 (SD = 5.96), for the asynchronous cough was 20.82 (SD = 3.35), for the synchronous cough was 20.23 (SD = 4.97), for the asynchronous cry was 20.38 (SD = 4.7), for the synchronous cry was 20.64 (SD = 4.44). To check for any disparity between conditions we performed a set of t-tests, resulting in no significant differences ($p > 0.35$). Our sample thus consists of thirty-four adults. Six adults were excluded from analysis due to excessive artifacts (N = 4) and technical problems (N = 2). While our adult sample is completed, we highlight that data collection for the infant sample is still ongoing, explaining the low numbers in the sample and lenient parameters used in the preprocessing. Once our samples will be completed, we plan to re-run all using more stringent parameters.

2.4. Source reconstruction

For the infants samples we selected the 9-months-old anatomical template (O'Reilly et al., 2021). For adults, the ICBM 152 (Mazziotta et al., 2001) template was used. Default anatomical regions and default electrode locations were entered into the forward boundary element head model (BEM) provided by the openMEEG algorithm (Gramfort et al., 2011). The tissue conductivity values for infants were those used in O'Reilly et al. (2021) for 7-months-old infants, as in Chapter 1: 0.33 S/m for the gray matter, 0.0041 S/m for the skull, and 0.33 S/m for the scalp. For adults we used 0.33 S/m for the gray matter, 0.0043 S/m for the skull, and 0.33 S/m for the scalp (Thorpe, Cannon, & Fox, 2016).

For the modelling of time-varying source signals (current density) of all cortical voxels, a minimum norm estimate inverse solution was applied after noise covariance estimation. Matrices were then scaled through dynamical Statistical Parametric Mapping normalization (Dale et al., 2000). Averaging current density across voxels yielded time series for 66 cortical areas (region of interest ROIs) for both infants an aduts in the Desikan and Killiany (2006) parcellation scheme. For time-varying connectivity measures we computed the time-resolved amplitude envelope correlation (Hipp, Hawellek, Corbetta, Siegel, & Engel 2012), every 50 ms from audio onset to 1150 ms for every EEG frequency band (delta 0.1 - 4; theta 5, 8; alpha 9 - 13; beta 14 - 30; lower gamma 31 - 59; higher gamma 60 – 80 for adults; delta 1 – 3 Hz, theta 3 – 5 Hz, alpha 6 – 9 Hz, beta 10 – 29 Hz,

lower gamma 30 – 49 Hz, higher gamma 51 – 80 Hz for infants; Pernet et al., 2020; Saby & Marshall, 2012). All computations were done using Brainstorm (Tadel et al., 2011) and ad-hoc Matlab scripts.

2.5. Time-varying connectivity analysis

In order to quantify relationships between brain areas we referred to graph theory (Bullmore et al., 2009). Through graph theory brain areas are reduced to nodes and every connection between them to an edge. Within this approach, brain functions are considered to emerge from synchronized activity (i.e., edges) of several information processing nodes (Quadrelli et al., 2021). In a graph, the backbone structure underlying brain activations can be evaluated by the minimum spanning tree (MST). MST is an unbiased method to represent the essential features of brain networks, as it identifies the strongest connections avoiding loops (Stam et al., 2014). The result is a network comprising all the major qualitative properties of connectivity while allowing comparison across conditions (Stam et al., 2014). Previous reports have shown that brain networks measured through MSTs become progressively more connected with age, mirroring myelination processes and providing evidence for the usefulness of MSTs to describe networks configurations (Stam et al., 2014). A tree was derived for each time-resolved adjacency matrix containing amplitude envelope correlations values. Then, the edge betweenness centrality was computed to quantify relevant connections in the network at each time point. The edge betweenness centrality represents the fraction of all shortest paths in the network that contain a given edge. Edges with high values represent fundamental hubs controlling information flow (Brandes, 2001; Rubinov & Sporns, 2010). To test differences across brain areas and time-points between conditions within each sample (HL-ASD, adults) we applied non-parametric permutation testing. Data were permuted 5000 times and significant t-values were extracted for each cluster. Following the literature reviewed above, a priori one-tailed planned comparisons focused on asynchronous stimuli compared to synchronous ones, where asynchronous stimuli were hypothesised to involve greater connectivity values compared to synchronous ones, were conducted. The same approach was used to test synchronous cry compared to synchronous cough stimuli and synchronous laughter compared to synchronous cough stimuli. This was done for each frequency band across groups. All computations were performed using the Brain Connectivity Toolbox (Rubinov & Sporns, 2010) and ad-hoc Matlab scripts.

3. Results

Adults

Regarding the asynchronous compared to synchronous comparison, where we expected greater connectivity in response to asynchronous stimuli, in the delta band we observed significant hubs at 250 ms after audio onset in the bilateral pars opercularis and

in the left pars orbitalis (t sum = 5.9, p = 0.006); at 750 ms in the bilateral STS (t sum= 4.26; p = 0.02).

In the theta band we observed significant hubs at 150 ms in the bilateral pars orbitalis and in the left pars triangularis (t = 5.19, p = 0.0092); at 200 ms in the right pars orbitalis and left pars triangularis (t = 4.19; p = 0.024); at 300 ms in the right temporal pole and left fusiform gyrus (t = 4.74, p = 0.014) and in the bilateral posterior cingulate cortex (PCC; t = 4.16, p = 0.036); at 700 ms in the bilateral insula (t = 5.21; p = 0.0082); at 950 ms in the bilateral caudate anterior cingulate cortex (ACC; t = 4.3, p = 0.025); at 1000 ms in the bilateral lingual gyrus and in the left pericalcarine cortex (t = 5.56, p = 0.006) in response to asynchronous compared to synchronous stimuli.

In the alpha band significant hubs were distributed at 300 ms in the bilateral STS and in the left Superior Frontal Gyrus (SFG) (t = 5.83, p = 0.0042); at 400 ms in the right rostral part of the middle frontal gyrus (MFG), in the bilateral caudal part of the MFG (t = 6.87, p = 0.001), in the right frontal pole and in the bilateral precentral gyrus (t = 5.82, t = 0.004); at 500 ms in the bilateral parahippocampal gyrus (t = 4.17, p = 0.02); at 700 ms in the bilateral medial part of the orbitofrontal cortex (OFC; t = 4.61, p = 0.01); at 750 ms in the bilateral SFG and in the left rostral MFG (t = 5.63, p = 0.004); at 1050 ms in the right superior parietal lobule and in the bilateral inferior parietal lobule (t = 5.19, p = 0.007); at 1100 ms in the right superior parietal lobule, the bilateral inferior parietal lobule and in the left precuneus (t = 7.29, p = 0.0008), all in response to asynchronous compared to synchronous stimuli.

In the beta band, in response to asynchronous vs. synchronous stimuli, we observed significant hubs at 250 ms in the right rostral MFG and in the left caudal MFG (t = 4.01, p = 0.047); at 650 ms in the bilateral precentral gyrus and in the left paracentral lobule (t = 5.49. p = 0.006).

Finally, while in the lower gamma band we observed significant hubs at 600 ms in the right inferior parietal lobule and bilateral precuneus (t = 6.19, p = 0.002); at 700 ms in the bilateral PCC (t = 4.51, p = 0.017); at 850 ms in the right precentral gyrus and in the left paracentral lobule (t = 4.09, p = 0.042); at 1150 ms in the right PCC, and in the left ACC (t = 4.94, p = 0.009), in the higher gamma we observed significant activity at audio onset in the right cuneus, in the left lateral occipital gyrus (LOG; t = 4.44, p = 0.02); at 50 ms in the bilateral cuneus (t = 4.22, p = 0.02); at 250 ms in the right pars triangularis and in the bilateral lateral OFC (t = 5.89, p = 0.0036); at 350 ms in the bilateral cuneus (t = 4.19, p = 0.02); at 400 ms in the bilateral temporal pole (t = 4.35, p = 0.02); at 800 ms in the right STG and in the bilateral MTG (t = 6.39, p = 0.0008); at 850 ms in the bilateral pericalcarine cortex (t = 4.03, p = 0.04).

The areas most active in asynchronous compared to synchronous stimuli in adults are depicted across time in Figure 2.

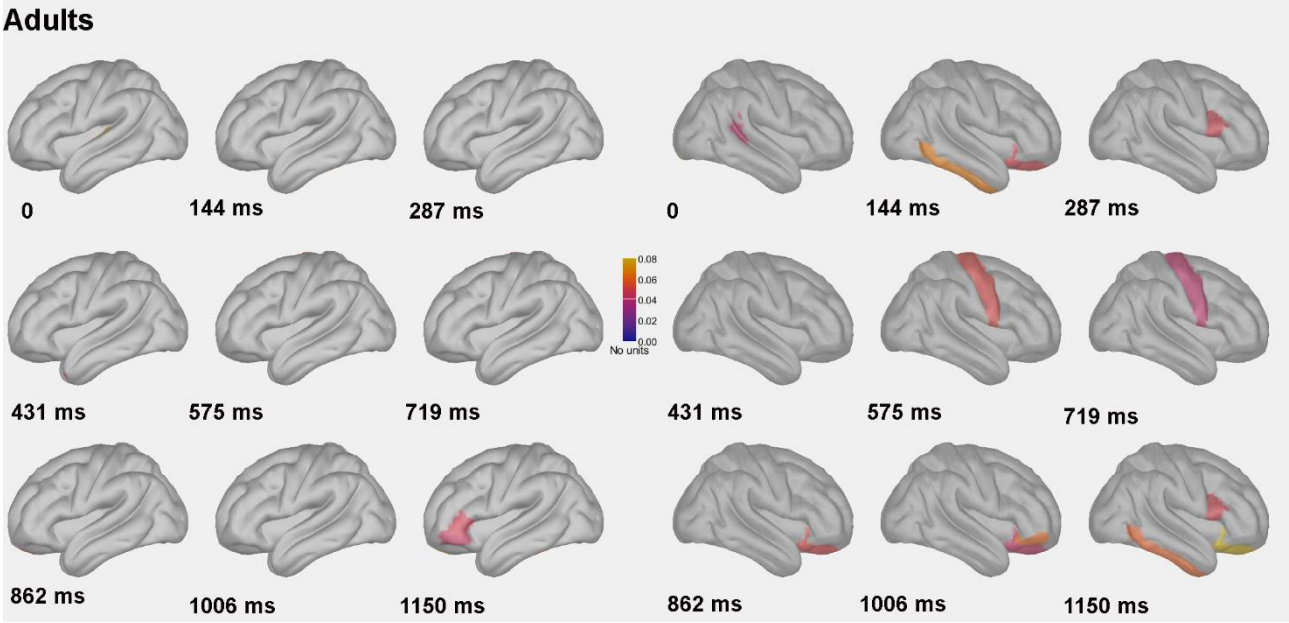


Figure 2. *Current source density differential activity in response to asynchronous compared to synchronous stimuli in adults across time*

Regarding the synchronous cry vs. cough comparison, where we expected increased connectivity in response to stimuli involving crying, in the delta band we observed significant hubs at 100 ms in the right parahippocampal gyrus and in the left temporal pole (t = 4.38, p = 0.015); at 200 ms in the bilateral ITG (t = 4.84; p = 0.008); at 450 ms in the bilateral ITG and in the primary auditory cortex (PAC; t = 5.69, p = 0.0038); at 950 ms in the bilateral supramarginal gyrus (t = 4.38, p = 0.014).

In the theta band we observed activity in response to cry vocalizations starting from 50 ms in the right LOG and in the left rostral ACC (t = 4.44, p = 0.01); then from 300 ms in the right MTG and in the left ITG (t = 4.29, p = 0.01); from 500 ms in the bilateral temporal pole (t= 4.9, p = 0.007); and at 550 ms in the bilateral caudate MFG (t = 4.67, p = 0.008); at 1050 ms in the bilateral MTG (t = 3.93, p = 0.045).

In the alpha band hubs sensitive to cry vocalizations were observed at audio onset in the bilateral lingual gyrus (t = 3.95, p = 0.04) and at 450 ms in the bilateral pars orbitalis (t = 4.14, p = 0.03).

In the beta band significant hubs in response to crying were observed at 100 ms in the right pericalcarine cortex and left cuneus (t = 4.05, p = 0.03), at 450 ms in the bilateral pars orbitalis (t = 4.06, p = 0.03); at 1050 ms in the bilateral ACC (t = 4.18, p = 0.039).

Finally, in the lower gamma band we observed significant hubs at 250 ms in the right MTG and left ITG (t= 4.36, p = 0.018) while in the higher gamma band at 400 ms in the right MTG and in the bilateral ITG (t = 5.87, p = 0.0042); at 650 ms in the bilateral precuneus and in the right lingual gyrus (t = 5.22, p = 0.005).

The areas most active in synchronous cry compared to synchronous cough stimuli in adults are depicted across time in Figure 3.

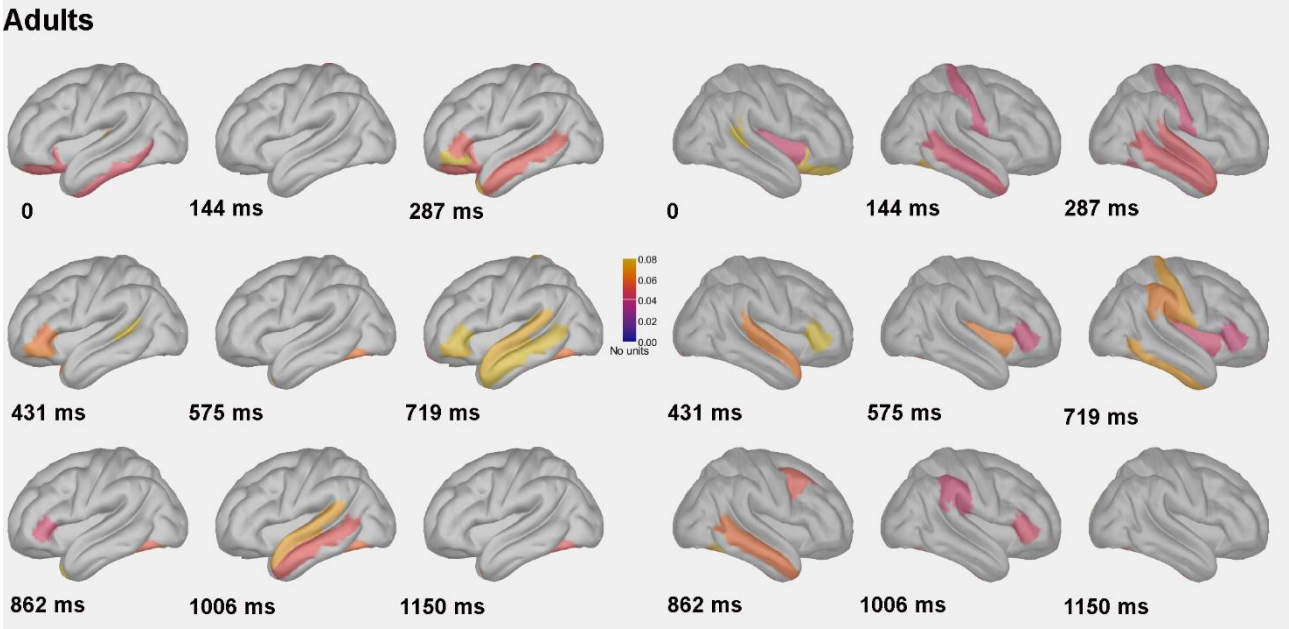


Figure 3. *Current source density differential activity in response to synchronous cry compared to cough stimuli in adults across time*

Regarding the synchronous laughter vs. cough comparison, in which we expected increased connectivity in response to laughter, in the alpha band we observed significant hubs at 500 ms in the bilateral pericalcarine cortex and in the left cuneus ($t = 5.8$, $p = 0.002$); at 750 ms in the bilateral PAC ($t = 4$, $p = 0.04$).

In the beta band activity from hubs sensitive to laughter started at 250 ms in the bilateral ACC ($t = 4.04$, $p = 0.03$); then at 500 ms was in the bilateral caudal MFG and in the left pars opercularis ($t = 6.03$, $p = 0.0024$); at 600 ms in the bilateral MFG ($t = 4.76$, $p = 0.005$); at 800 ms in the right supramarginal gyrus and in the left superior parietal lobule ($t = 3.92$,$p = 0.046$); and at 1100 was in the bilateral cuneus and in the left LOG ($t = 5.47$, $p = 0.005$).

In the lower gamma band we observed activity in significant hubs in response to laughter vocalization at 1100 ms in the bilateral PCC ($t = 4.27$, $p = 0.02$) while in the higher gamma band at 550 ms in the bilateral pars orbitalis ($t = 3.49$, $p = 0.04$); at 600 ms in the right pericalcarine cortex and in the left cuneus ($t = 4.13$, $p = 0.03$) at 750 ms in the bilateral cuneus ($t = 3.94$, $p = 0.04$) and at 900 ms in the bilateral paracentral lobule ($t = 4.06$, $p = 0.04$). We do not observe significant hubs in the delta and theta bands.

The areas most active in synchronous laughter compared to synchronous cough in adults are depicted across time in Figure 4.

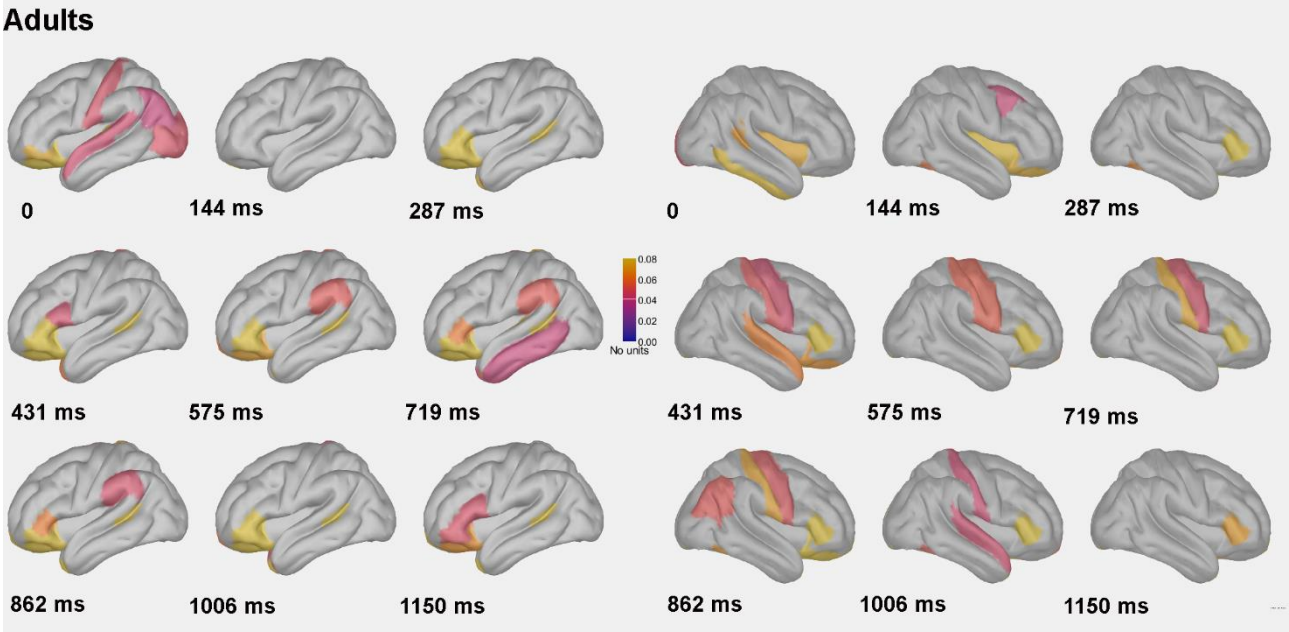


Figure 4. *Current source density differential activity in response to synchronous laughter compared to cough stimuli in adults across time*

Infants

While preliminary, in the asynchronous vs. synchronous comparison in the HL-ASD group in the delta band we observed significant hubs at audio onset in the bilateral LOG and in the right rostral ACC (t = 5.46, p = 0.005); at 150 ms in the bilateral postcentral gyrus (t = 5.46, p = 0.03); at 200 ms in the bilateral insula (t = 4.27, p = 0.01); at 350 ms in the bilateral PCC (t = 4.17, p = 0.02).

In the theta band significant hubs in response to asynchronous stimuli were distributed at audio onset in the bilateral pericalcarine cortex (t = 4.15, p = 0.025) and at 100 ms in the bilateral insula (t = 2.27, p = 0.021).

In the alpha band we observed significant hubs in response to asynchronous compared to synchronous stimuli at 50 ms in the bilateral STS (t = 4.18, p = 0.02); at 100 ms in the bilateral rostral ACC (t = 4.2, p = 0.03); at 150 ms in the bilateral PAC (t = 4.25, p = 0.03); at 200 ms in the bilateral PCC (t = 4.2, p = 0.02); at 250 ms in the right pars opercularis and in the left pars orbitalis (t = 4.36, p = 0.019) and at 300 ms in the bilateral PCC (t = 5.11, p = 0.009).

In the beta band significant hubs in response to asynchronous stimuli were distributed at audio onset in the bilateral insula (t = 4.6, p = 0.015); at 100 ms in the bilateral rostral and caudal ACC (t = 8.67, p = 0.0004), while in the lower gamma were present at 50 s in the bilateral precuneus (t = 4.32, p = 0.028) and at 100 ms in the bilateral PAC and left STS (t = 5.34, p = 0.008). No other frequency band was found to be significant (ps > .92).

The areas most active in asynchronous compared to synchronous stimuli in HL-ASD infants are depicted across time in Figure 5.

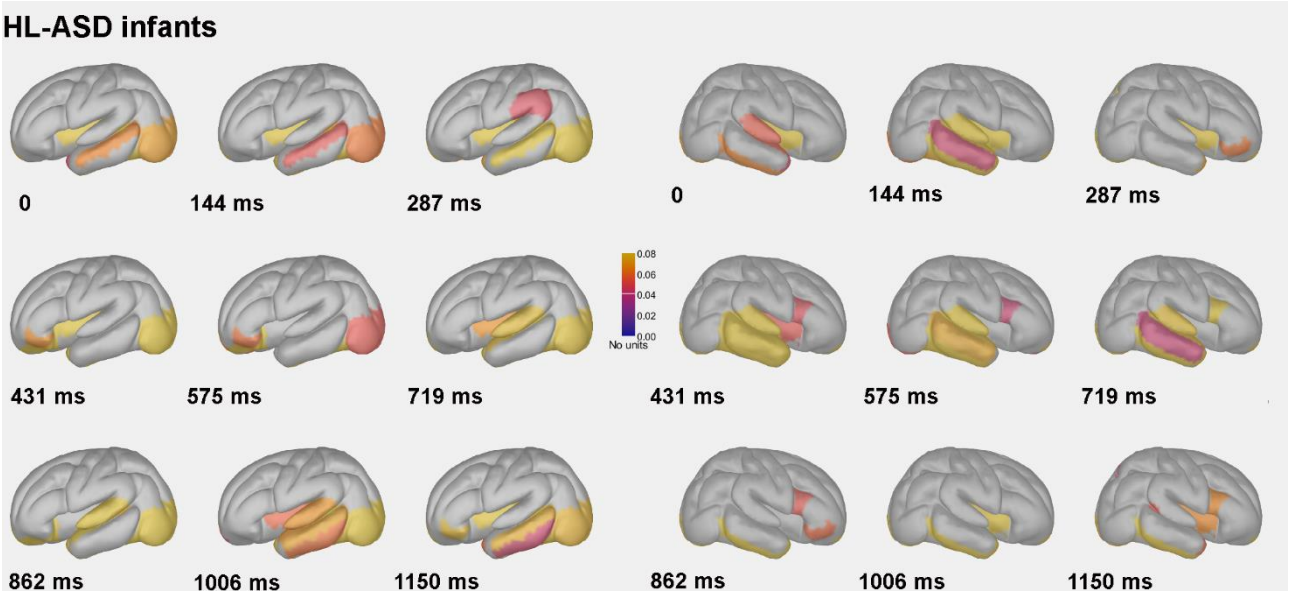


Figure 5*. Current source density differential activity in response to asynchronous compared synchronous stimuli HL-ASD infants across time*

Regarding the synchronous cry vs. cough comparison in the HL-ASD group we found significant hubs in the delta band at 350 ms in the bilateral temporal pole and in the left fusiform gyrus (t = 5.2, p = 0.004) and in the alpha band at 250 ms in the right STS and in the left SFG (t = 4.29, p = 0.016).

The areas most active in synchronous cry compared to synchronous cough stimuli in HL-ASD infants are depicted across time in Figure 6

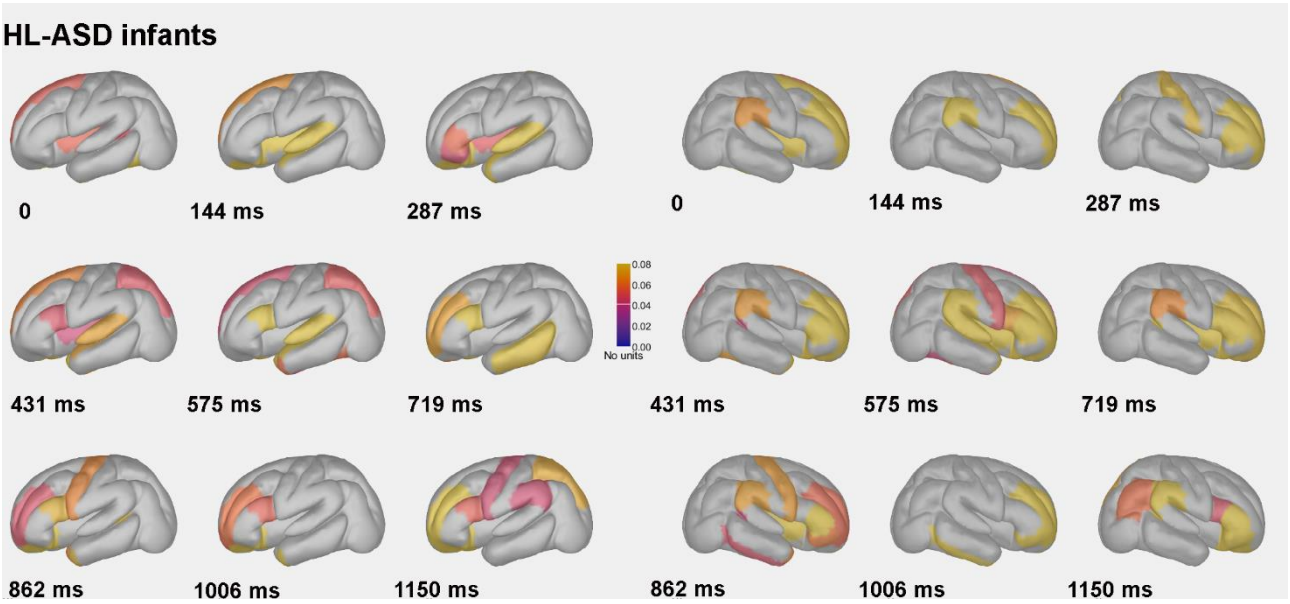


Figure 6. *Current source density differential activity in response to synchronous cry compared to cough stimuli in HL-ASD infants across time*

Regarding the synchronous laughter vs. cough comparison in the HL-ASD group, we found significant hubs to be located in the delta band at 300 ms in the right paracentral lobule and in the left postcentral gyrus (t = 4.35, p = 0.018); in the theta band at 150 ms in the right STS and in the bilateral SFG (t = 5.03, p = 0.004); in the alpha band at 50 ms in the bilateral ACC (t = 3.9, p = 0.041) and at 200 ms in the bilateral caudal MFG (t = 4.34, p = 0.011); in the beta band at 250 ms in the bilateral PCC (t = 4.13, p = 0.03); in the lower

gamma band at audio onset in the bilateral caudal MFG (t = 4.23, p = 0.02); in the higher gamma band at 50 ms in the bilateral postcentral gyrus (t = 4.05, p = 0.04).The areas most active in synchronous laughter compared to synchronous cough in HL-ASD are depicted across time in Figure 7.

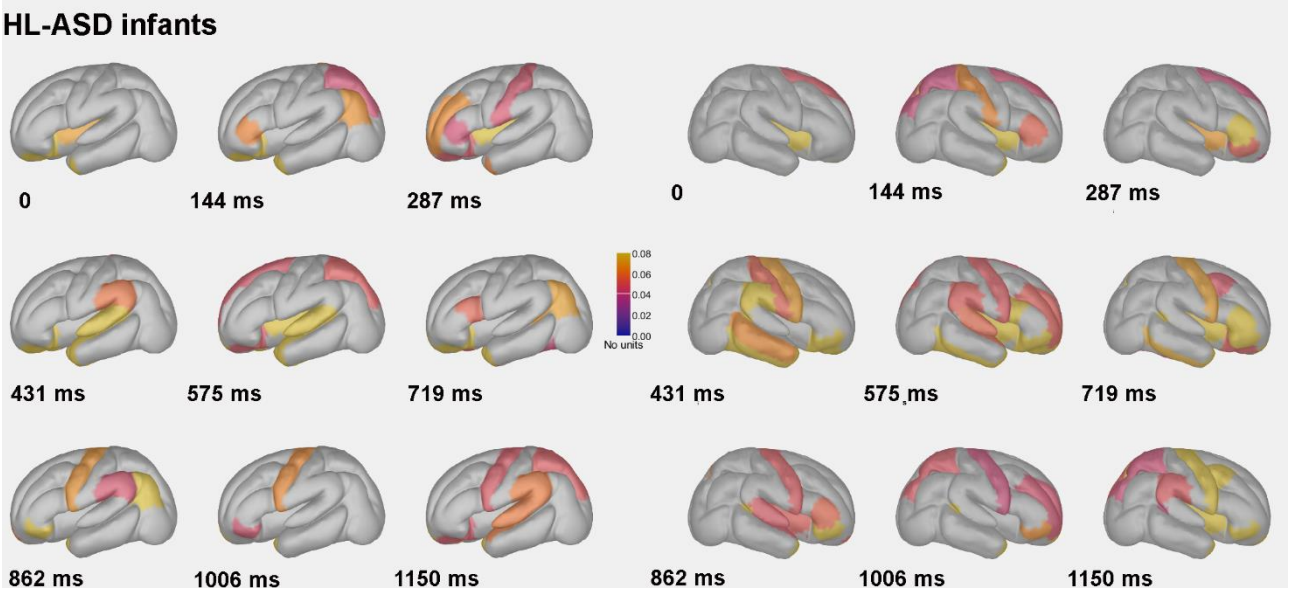


Figure 7. *Current source density differential activity in response to synchronous laughter compared to cough stimuli in HL-ASD infants across time*

4. Discussion

In this study we set out to characterize the brain machinery underlying intersensory binding of emotional stimuli. More in depth, by first identifying the configuration of functional brain networks involved in intersensory binding at its developmental end point represented by adulthood, we wanted to address the role of socio-emotional complexity in guiding impairments in multimodal integration in HL-ASD infants. Indeed, due to their reported social and perceptual impairments, we hypothesised HL-ASD infants to lack integration across sensory and higher-order regions in response to inherently complex social stimuli.

Regarding multisensory binding in adults, we observe a very complex interplay between brain regions and frequency bands. In the delta band we observe a central role in the control of information flow played by the IFG followed by activity in the STS, highlighting the role of the STS in multisensory binding (Stevenson et al., 2016). In the theta band we observe activity in the IFG followed by activity in the fusiform gyrus, ACC and temporal pole. Next, we discovered activity in the insula, followed by the ACC and then by activity in visual cortices. This activations highlight the feedback from higher-order areas involved in multisensory integration such as the temporal pole and the STS (Pascual et al., 2015; Stevenson et al., 2016), from areas involved in conflict monitoring triggered by the lack of temporal synchrony in our stimuli such as the ACC (Botvinick, Cohen, & Carter, 2004), and from areas involved in salience detection such as the insula (Menon & Uddin,

2010), to unimodal sensory regions. In the alpha band we again observe primary roles in controlling information flow first in the STS and IFG, followed by activity in the MFG, primary motor cortices, and OFC, feeding back into the MFG and SFG followed by activity in the superior and inferior parietal lobules. In this context, activity in medial frontal regions most likely relate to short-term memory processes and stimulus mapping (Bonte, Ley, Scharke, & Formisano, 2016). Interestingly shifting from slower to faster frequency bands we coherently observe earlier onsets of hub activities, in line with the role of faster oscillations in binding of percepts (Doesburg et al., 2008). Nonetheless, even in faster frequencies we observe the continuous feedback and exchange between sensory and associative regions as stimuli unfold. Indeed, while oscillations in the delta range are reportedly involved in the integration of cross-modal stimuli over a broad and relatively late time window (Crosse et al., 2016) and alpha phase synchrony mediates large-scale communication between distant brain regions (Hummel & Gerloff, 2005; Ikumi, Torralba, Ruzzoli, & Soto-Faraco, 2019; van Driel et al., 2014), the gamma band has been proposed to mediate fast network associations regarding the binding of AV stimuli (Doesburg et al., 2008; Senkowski et al., 2007). Coherently, it is likely that diverse windows of integration characterized by oscillations at different frequencies reflect neural integration at multiple stages of the integration processing hierarchy, with different frequencies encoding different aspects (Crosse et al., 2016; van Driel et al., 2014). Overall, our results confirm the presence of a feedforward and feedback loop comprising sensory and associative regions. More in depth, after travelling across modality-specific sensory cortices information arrives in higher-order areas for further processing. Then, information travels back to sensory areas. We propose these mechanisms to be involved in the building of coherent and informed percepts of the complex world surrounding us.

While our results on the infant sample are extremely preliminary, they nonetheless show that HL-ASD are still characterized by a similar pattern of cross-talks between unimodal and associative regions as adults in higher frequencies (Adam-Darque et al., 2020; Blasi et al., 2011), highlighting the early ontogenesis of binding mechanisms and confirming their putative role in later acquisitions (Blasi et al., 2011; Hillairet de Boisferon et al., 2017). However, in the alpha and theta band they first show activity in unisensory regions followed by activity in higher-order areas. If confirmed, this result would be particularly noteworthy as it would suggest the absence of feedbacks from higher-order to sensory regions, resulting in failures in intersensory binding, especially in the frequencies involved in social processing across infancy (Saby & Marshall, 2012; van der Velde et al., 2021). For example, theta power is reportedly stronger to social stimuli compared to non-social ones in infancy (Jones et al., 2015). This would thus suggest that social complexity leads to greater disruptions in intersensory binding in ASD. Thus, impaired feedback

between higher-order associative regions such as the STS and sensory areas could reflect reduced perceived salience of social stimuli independently of low-level sensory impairments (Aglieri, Chaminade, Takerkart, & Belin, 2018). However, before embracing this hypothesis it is necessary to repeat our analysis on the full samples and to add the insights on LR infants.

Beyond differences, it is interesting to note that the temporal onsets of significant hubs activity were about the same across HL-ASD infants and adults. This result confirms previous reports of the timing of integration of emotional facial and vocal information (Adam-Darque et al., 2020; Grossmann & Johnson, 2007.; Kushnerenko et al., 2013; Schweinberger, Kloth, & Robertson, 2010), and of the timing of effects we observed and discussed in Chapter 1 in response to tones and noises designed to mimic speech duration characteristics. Moreover, although the frontal lobe is generally viewed as a late-maturing area, ours and other reports show that these higher-order areas are already operative in early infancy and their functional role, besides fine-tuning, is preserved up to adulthood (Adam-Darque et al., 2020). This once again confirms the fact that broad brain mechanisms are not stimuli specific but rather invariant (Park & Friston, 2013). Within this context, even minor changes in the overall architecture can lead to very different behavioural outcomes.

Regarding activity in response to emotional synchronous stimuli compared to neutral stimuli, in adults we coherently observe increased processing of negative compared to positive emotions (Blasi et al., 2011; Grossmann, Oberecker, Koch, & Friederici, 2010; Zhao et al., 2019) Indeed, the areas involved in the processing of crying are more numerous across frequency bands. Remarkably, we observe a reversal in HL-ASD infants. Indeed, HL-ASD infants are characterized by increased processing of laughter compared to cry stimuli. This effect, while baffling, may be related to the more reinforcing value of positive stimuli (Bahrick et al., 2019), which may aid the attentional engaging of HL-ASD infants. Indeed, it has been reported that ASD individuals show increased processing of positive affective prosody (Gebauer et al., 2014). This increased activation was found, among others, in the MFG and in the SFG (Gebauer et al., 2014), areas we also report to be active during laughter processing. Thus, these increased activations might support increased attentional demands during positive emotion recognition in ASD, considering that both the MFG and SFG are implicated in executive functions, attention, and working memory (Gebauer et al., 2014). This may indicate that positive vocalizations undergo more explicit evaluations compared to negative ones (Grossman et al., 2010). Nonetheless, besides impairments, in HL-ASD infants we observe the involvement of similar areas to those reported in previous investigations of emotional vocalizations (Blasi et al., 2011; 2015), such as TVA, inferior prefrontal cortex,

premotor cortical regions, the OFC, and the insula (Blasi et al., 2011). Of interest is also the involvement of sensorimotor regions in both infants and adults, suggesting that the involvement of sensorimotor areas during facial emotional processing we observed in Chapter 3 is fairly preserved across development. Overall, our findings confirm that emotional vocalizations are associated with differential activity in emotion-processing networks from very early in development (Blasi et al., 2011). However, as our infants' samples are still very small, we postpone a deeper discussion of these results until our samples are completed.

Finally, we do not observe reliable lateralization of activity across our samples. While there are many reports of increased right-lateralized auditory encoding in response to both simple acoustic stimuli (Chapter 1), voices (Chapter 2; Blasi et al., 2011; Fecteau et al., 2004; Myers et al., 2019; Grossman et al., 2010; Belin & Grosbras, 2010) and emotional vocalizations in the literature, the activity we found in this study was mostly bilateral. This can be due to mainly three reasons. First this may be a low statistical power artefact. However, the fact that we observe bilateral hubs in our adult sample argues against this conclusion. On the other hand, this effect could be attributed to the changing results that source and scalp level analysis provide (Kolozsvári et al., 2021). However, while we do observe right-lateralization in Chapter 1 we do not observe strong lateralization in response to intersensory integration in this study, even though source reconstruction analysis was performed for both experiments. This observation argues against the second interpretation. As a third interpretation we propose that right-lateralization happens mainly in response to simpler acoustic stimuli but at increasing complexity processing becomes necessarily more bilateral. While promising, further studies are needed to confirm this hypothesis.

In conclusion, in the present study we were able to characterize multisensory binding mechanisms across time in adults and have begun to do so in 9-months-old LR and HL-ASD infants. We showed the presence of feedforward and feedback loops across multisensory and unisensory regions enabling the processing of complex social stimuli. Once our developmental samples will be completed, we hope to confirm the absence of feedback mechanisms from higher-order to sensory regions in HL-ASD infants and provide evidence for their role in observed behavioural impairments. Indeed, if confirmed these results, coupled with the reversal of the negative bias in emotion perception, suggest the presence of higher-order social processing impairments in ASD not related to sensory impairments. This confirms the role of social complexity impairments in the genesis of dysfunctions in ASD. Finally, while we were not able to discern causal network connections in our analysis, future studies will benefit greatly from the use of directed

connectivity measures to causally address the directionality of the feedbacks we described.

General Discussion

In the first year of life HL-ASD infants are characterized by widespread atypicalities across multiple brain systems, both in the social brain and in perceptual and cognitive networks (Elsabbagh & Johnson, 2016). It is thus not possible to ascribe all the interweaving of impairments reported in the literature to dysfunctions in a single set of brain areas (Elsabbagh & Johnson, 2016; Righi et al., 2014; Abrams et al., 2013). In this thesis we set out to investigate social brain development at increasing perceptual, but most of all, social complexity, and to assess brain networks and their interactions in both typical and atypical development. In Chapter 1 we showed that the brain networks involved in auditory processing narrow across development, suggesting an increased proficiency as the infant grows. In Chapter 2 we suggested that the social cues inherent in the human voice actively modify the tuning of cortical areas, and consequently both perceptual and social processing. We also observed that HL-ASD infants appear to be unaffected by social saliency carried by vocal stimuli. In Chapter 3 we uncovered an early involvement of dorsal and ventral visual streams in response to dynamic and static emotional faces respectively, suggesting a specific encoding of faces as salient social cues in the infant brain. Finally, in Chapter 4 we described the ontogenetically preserved spatial and temporal organizations of networks as the brain attempts to integrate multimodal social stimuli.

Overall, perceiving socially relevant stimuli, such as human faces and voices, is fundamental for social interactions and it is achieved without explicit effort in our everyday lives. However, as we saw in the studies reported here, this task requires different neural processing stages. For example, low-level acoustic cues are first processed in primary auditory areas, then trough feedback loops involving higher-order temporal and frontal regions, information travels back to sensory regions informing subsequent percepts (Chapter 1). Higher-level information, such as vocal or facial cues, is then processed capitalizing on the same mechanisms, involving functionally dissociable but interacting brain regions, coupled with wider involvement of associative areas (Chapter 2 and 3). This has also been confirmed in studies regarding face perception in which, besides face sensitive areas, processing involves non-specific regions such as anterior temporal lobes and prefrontal areas (Davies-Thompson et al., 2019; Haxby, Hoffman, & Gobbini, 2000). Similarly, while the presence of temporal voice areas (TVA; Belin et al., 2000) is well documented, other areas have been found to be sensitive to voices (Davies-Thompson et al., 2019; Maguinness et al., 2018; Pernet et al., 2015). This opens the interesting possibility of a specific tuning to socially salient cues in the brain during development (Davies-Thompson et al., 2019; Liu et al., 2012). Indeed, we may hypothesise that after

mandatory modality specific sensory processing, social information widely engages mostly the same associative areas irrespectively of whether it's faces or voices.

This hypothesis is confirmed by our results on multimodal integration (Chapter 4). In fact, we observe the engagement of approximately the same brain areas across synchronous emotional categories. Moreover, these mechanisms seem to be preserved across development as we observe them, besides in adults, in both younger and older infants (Chapter 1). Indeed, while networks undergo fine-tuning as the infant develops and into adulthood, we do not observe changes in their composition. This is a striking result, suggesting the presence of hard-wired mechanisms in the brain for the processing of salient social cues irrespectively of their sensory origins (Grossman & Johnson, 2007). More in depth, while the presence of two largely distinct networks of face and voice perception (Chapter 2 and 3) is widely acknowledged, there are reports describing a shared 'multisensory convergence zone' for the integration of information across sensory modalities (Davies-Thompson et al., 2019; Pourtois, de Gelder, Bol, & Crommelinck 2005). This zone includes areas we find to form a network in Chapter 4, such as the orbitofrontal cortex, the superior temporal sulcus (STS) and the inferior frontal gyrus (IFG; Davies-Thompson et al., 2019; Maguinness et al., 2018). Thus, face and voice regions bound together may form a general social processing network, underlying the visual and auditory processing of human cues (Davies-Thompson et al., 2019; Staib & Frühholz, 2021). Indeed, functional flexibility and broader sensitivity to stimuli represents a more evolutionarily plausible model compared to strict functional specialization across large cortical areas (Staib & Frühholz 2021, 2022). Moreover, despite the functional specialization of face and voice sensitive networks, there is a growing number of studies highlighting commonalities in the neural coding of faces and voices due to their frequent co-occurrence in the environment (Young, Frühholz, & Schweinberger, 2020). Taken together our results point toward the presence of a sequence of increasingly complex processes in response to social stimuli, with both feedforward and feedback communication between primary sensory regions and higher-order associative areas (Liu et al., 2012). While we propose this to be a sensible model, it is nonetheless important to highlight that our samples are not completed. Thus, while we expect to confirm and strengthen our results as we collect participants, these conclusions are still to be treated with care.

Either way, in this context, a lingering question regards the role of the loops we proposed in disorders with prominent social impairments such as ASD. In Chapter 2 we observe a lack of differentiation between voices in HL-ASD infants. Thus, irrespectively whether the infant was listening to, for example his/her mother voice or a synthesiser, cortical tracking remained the same. These results are in accordance with reports showing

a diminished response and reduced sensitivity of TVA in HL-ASD infants (Lloyd Fox et al., 2013). In Chapter 4 we showed that HL-ASD infants seemingly lack the feedforward and feedback loops that allow adults to integrate multisensory information during emotional stimulation. While further studies are needed to verify this, a plausible hypothesis is that a lack of synchronization between broad social processing networks and perceptual and cognitive systems may explain both our results and the wider array of perceptual and social impairments reported in ASD (Righi et al., 2014). For example, as complex social interactions (i.e., conversations) require fine perceptual tuning and attentional orienting, the absence of coordinated activity between perceptual systems and the social brain may impair the automatic orienting of attention to salient social cues and may result in a confusing and unpredictable environment from both a sensory and interactional point of view (Clumeck et al., 2014; Elsabbagh & Johnson, 2016). Indeed, as we discussed in Chapter 4, individuals with ASD are characterized by impairments in contexts in which intersensory binding is necessary (Hyde et al., 2017).

In this context, studying HL-ASD infants is fundamental to track developmental trajectories and to relate them to the emergence of symptoms (Righi et al., 2014). Indeed, together with recent findings of aberrant wiring of white matters tracts in ASD (Barnea-Goraly, Lotspeich, & Reiss, 2010; Righi et al., 2014), our results across chapters provide evidence for disrupted neural architecture already in early developmental stages in individuals that may develop ASD (Righi et al., 2014). This disruptions in brain architecture may in turn explain the lack of cortical specialization in response to social stimuli reported in ASD, even in early infancy (Lloyd Fox et al., 2013). Moreover, since our focus were 9- and 18-months-old HL-ASD infants, we are able to assert that these impairments if not present at birth, emerge early in development and remain stable at least across the first year of life. It is important to note that we did not observe strikingly different brain organizations in HL-ASD infants compared to either LR peers or adults. However, we observed slight differences in the magnitude of activity and in the composition of brain systems involved. This result is unsurprising, since even mild impairments in the complexity characterizing brain architecture can result in destructive impairments in the processing of complex, dynamic, less predictable stimuli, such as social ones (Elsabbagh & Johnson, 2016).

Finally, our results are in accordance with the interactive specialization model (Johnson, 2001). Indeed, across development, cortical areas transition from responding to wide arrays of stimuli in an undifferentiated manner to having increasingly fine-tuned responses to stimuli (Grossman & Johnson, 2007; Johnson, 2001). The direct consequences of this are increasingly focal pattern of functional activation across cortical areas in response to specific stimuli (Grossman & Johnson, 2007). In conformity with this,

in Chapter 1 we observed more focal activities in 9- compared to 4-months-olds in response to simple auditory stimuli, suggesting a narrowing of brain areas involved in perceptual processing due to proficiency. The same was observed in Chapter 4, with adults showing more focal activities compared to 9-months-old infants in response to complex multimodal social stimuli, providing further support for the IS. Indeed, as cortical regions become progressively more specialized with experience (Grossman & Johnson, 2007), social stimuli may have a prominent role in shaping this tuning, since one of the most fundamental tasks the infants face is learning about the social world (Grossman, 2015). Specifically, brain areas may show undifferentiated responses in the beginning, but with social experience their function may flexibly specialize to specific social features (Grossman & Johnson, 2007). Indeed, we hypothesise that the loops we described between sensory and higher-order areas would shape the progressive specialization of networks involved in social and perceptual processing. Within this context, ASD may be characterized by failures or delays within networks' differentiation processes (Grossman & Johnson, 2007) due to the absence of the feedforward and feedback loops between sensory and higher-order regions we previously described. However, the completion of our samples and further studies are needed to confirm this hypothesis.

In conclusion, we provide evidence regarding the development of social processing in the brain, both in typical and atypical infants, in response to some of the most important social cues for humans: faces, voices, and their combinations. While we recognize that due to Covid-19 our samples are still preliminary, we highlight the methodological and theoretical insights that this thesis hopes to provide. Indeed, we chose to adopt a multidisciplinary approach for these investigations, trying to test perspectives coming from developmental literature with cutting-edge analytical pipelines. This choice was made to capture the underlying organizations of brain processing across development in response to some of the most salient stimuli the infant can find in his/her environment. By the investigation of at-risk populations, we hope to aid the search of early developmental markers of ASD. Indeed, their identification is fundamental to anticipate diagnosis and consequently interventions (Riva et al 2018, 2022). Late identification of neurodevelopmental disorders entails considerable social and economic costs. These costs could be partly, if not considerably, prevented by focusing interventions to appropriate developmental time windows, offering the potential for altering long-term outcomes in a clinically meaningful manner. Besides interventions, our results have implications for future studies addressing higher-order social interactions. Indeed, at the highest levels of complexity the impairments we report, for example, in multimodal integration, may manifest as dysfunctions in interpersonal synchrony, such as in conversation turn taking (Honisch, Mane, Golan, & Chakrabarti, 2021; Murat Baldwin et

al., 2021). In fact, poor temporal acuity across the visual and auditory domains is reflected in a decreased ability to bond sensory inputs together, leading to asynchrony in behavioural responses to such stimuli (de Boer Schellekens, Keetels et al., 2013; Murat Baldwin et al., 2021; Stevenson et al., 2016). Thus, future studies may focus on brain responses to complex social scenes involving turn-taking. This will reduce even further the distance between laboratory studies and real-world experiences.

Our investigation is not free from limitations. First, as previously stated, our investigations were greatly impacted by the Covid-19 pandemic. In order to guarantee maximal safety to our participants, data collections in the health care facility we conduct our research in was restarted slowly. Nonetheless, we hope to confirm and strengthen our results with the data we are currently collecting. Another possible criticism may regard an overestimation of the role played by frontal cortices in the loops we propose. However, while previous reports implied that frontal cortices are silent during infancy due to delayed maturation (Grossman, 2015; Zelazo & Müller, 2002), there is now considerable evidence suggesting that frontal cortices are already active early in infancy. Indeed, long-range connections between posterior and frontal areas emerge throughout the first year of life, with top-down signals coming from frontal cortices modulating activity in, for example, sensory cortices (Assaneo et al., 2019; Davies-Thompson et al., 2019; Fransson et al., 2011; Grossman, 2015; Jones et al., 2015; Park et al., 2015). Moreover, it is important to highlight that the feedback and feedforward loops we propose between sensory and associative regions are not backed by white matter investigations in our samples. Nonetheless, while previous studies have shown little white matter connections between face and voice-selective networks (Davies-Thompson et al., 2019; Gschwind, Pourtois, Schwartz, Van De Ville, & Vuilleumier, 2012), the functional connections we propose may not be a direct proxy for anatomical connectivity. As such, connections may not be direct but may reflect the overall functional architecture of the brain (Davies-Thompson et al., 2019). Future studies would benefit greatly from the combination of structural and functional measurements, in order to shed light on the complex connections between brain areas during development. This is of particular importance since all the outstanding developmental milestones the infant achieves, descend from the incredibly complex interweaving of activity happening inside our brain.

References

Abrams, D. A., Chen, T., Odriozola, P., Cheng, K. M., Baker, A. E., Padmanabhan, A., ... & Menon, V. (2016). Neural circuits underlying mother's voice perception predict social communication abilities in children. *Proceedings of the National Academy of Sciences*, *113*(22), 6295-6300. https://doi.org/10.1073/pnas.1602948113

Abrams, D. A., Lynch, C. J., Cheng, K. M., Phillips, J., Supekar, K., Ryali, S., ... & Menon, V. (2013). Underconnectivity between voice-selective cortex and reward circuitry in children with autism. *Proceedings of the National Academy of Sciences*, *110*(29), 12060-12065. https://doi.org/10.1073/pnas.1302982110

Abrams, D. A., Padmanabhan, A., Chen, T., Odriozola, P., Baker, A. E., Kochalka, J., ... & Menon, V. (2019). Impaired voice processing in reward and salience circuits predicts social communication in children with autism. *elife*, *8*, e39906. https://doi.org/10.7554/eLife.39906

Adam-Darque, A., Pittet, M. P., Grouiller, F., Rihs, T. A., Leuchter, R. H. V., Lazeyras, F., ... & Hüppi, P. S. (2020). Neural correlates of voice perception in newborns and the influence of preterm birth. *Cerebral Cortex*, *30*(11), 5717-5730. https://doi.org/10.1093/cercor/bhaa144

Adhikari, B. M., Goshorn, E. S., Lamichhane, B., & Dhamala, M. (2013). Temporal-order judgment of audiovisual events involves network activity between parietal and prefrontal cortices. *Brain connectivity*, *3*(5), 536-545. https://doi.org/10.1089/brain.2013.0163

Adolphs, R., Sears, L., & Piven, J. (2001). Abnormal processing of social information from faces in autism. *Journal of cognitive neuroscience*, *13*(2), 232-240. 10.1162/089892901564289

Adibpour, P., Lebenberg, J., Kabdebon, C., Dehaene-Lambertz, G., & Dubois, J. (2020). Anatomo-functional correlates of auditory development in infancy. *Developmental cognitive neuroscience*, *42*, 100752. https://doi.org/10.1016/j.dcn.2019.100752

Aglieri, V., Chaminade, T., Takerkart, S., & Belin, P. (2018). Functional connectivity within the voice perception network and its behavioural relevance. *Neuroimage*, *183*, 356-365. https://doi.org/10.1016/j.neuroimage.2018.08.011

Ambadar, Z., Schooler, J. W., & Cohn, J. F. (2005). Deciphering the enigmatic face: The importance of facial dynamics in interpreting subtle facial expressions. *Psychological science*, *16*(5), 403-410. https://doi.org/10.1111/j.0956-7976.2005.01548.x

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th ed.)

Assaneo, M. F., Rimmele, J. M., Orpella, J., Ripollés, P., de Diego-Balaguer, R., & Poeppel, D. (2019). The lateralization of speech-brain coupling is differentially modulated by intrinsic auditory and top-down mechanisms. *Frontiers in Integrative Neuroscience*, *13*, 28. https://doi.org/10.3389/fnint.2019.00028

Attaheri, A., Choisdealbha, Á. N., Di Liberto, G. M., Rocha, S., Brusini, P., Mead, N., ... & Goswami, U. (2022). Delta-and theta-band cortical tracking and phase-amplitude coupling to sung speech by infants. *NeuroImage*, *247*, 118698. https://doi.org/10.1016/j.neuroimage.2021.118698

Attaheri, A., Panayiotou, D., Phillips, A., Choisdealbha, Á. N., Di Liberto, G. M., Rocha, S., ... & Goswami, U. (2022). Cortical tracking of sung speech in adults vs infants: A developmental analysis. *Frontiers in Neuroscience*, *16*. https://doi.org/10.3389/fnins.2022.842447

Bae, G. Y., & Luck, S. J. (2018). Dissociable decoding of spatial attention and working memory from EEG oscillations and sustained potentials. *Journal of Neuroscience*, *38*(2), 409–422. https://doi.org/10.1523/JNEUROSCI.2860-17.2017

Bahrick, L. E., McNew, M. E., Pruden, S. M., & Castellanos, I. (2019). Intersensory redundancy promotes infant detection of prosody in infant-directed speech. *Journal of experimental child psychology*, *183*, 295-309. https://doi.org/10.1016/j.jecp.2019.02.008

Balk, M. H., Ojanen, V., Pekkola, J., Autti, T., Sams, M., & Jääskeläinen, I. P. (2010). Synchrony of audio–visual speech stimuli modulates left superior temporal sulcus. *Neuroreport*, *21*(12), 822-826. https://doi.org/10.1097/WNR.0b013e32833d138f

Barnea-Goraly, N., Lotspeich, L. J., & Reiss, A. L. (2010). Similar white matter aberrations in children with autism and their unaffected siblings: a diffusion tensor imaging study using tract-based spatial statistics. *Archives of general psychiatry*, *67*(10), 1052-1060. doi:10.1001/archgenpsychiatry.2010.123

Bayet, L., Perdue, K. L., Behrendt, H. F., Richards, J. E., Westerlund, A., Cataldo, J. K., & Nelson III, C. A. (2021). Neural responses to happy, fearful and angry faces of varying identities in 5-and 7-month-old infants. *Developmental cognitive neuroscience*, *47*, 100882. https://doi.org/10.1016/j.dcn.2020.100882

Bazhenova, O. V., Stroganova, T. A., Doussard-Roosevelt, J. A., Posikera, I. A., & Porges, S. W. (2007). Physiological responses of 5-month-old infants to smiling and blank faces. *International journal of psychophysiology*, *63*(1), 64-76. https://doi.org/10.1016/j.ijpsycho.2006.08.008

Beauchemin, M., Gonzalez-Frankenberger, B., Tremblay, J., Vannasing, P., Martínez-Montes, E., Belin, P., ... & Lassonde, M. (2011). Mother and stranger: an

electrophysiological study of voice processing in newborns. *Cerebral cortex*, *21*(8), 1705-1711. https://doi.org/10.1093/cercor/bhq242

Belin, P. (2006). Voice processing in human and non-human primates. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *361*(1476), 2091-2107. https://doi.org/10.1098/rstb.2006.1933

Belin, P., Fecteau, S., & Bedard, C. (2004). Thinking the voice: neural correlates of voice perception. *Trends in cognitive sciences*, *8*(3), 129-135. https://doi.org/10.1016/j.tics.2004.01.008

Belin, P., & Grosbras, M. H. (2010). Before speech: cerebral voice processing in infants. *Neuron*, *65*(6), 733-735. https://doi.org/10.1016/j.neuron.2010.03.018

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). *Voice-selective areas in human auditory cortex*. 309–312.

Benasich, A. A., Choudhury, N., Friedman, J. T., Realpe-Bonilla, T., Chojnowska, C., & Gou, Z. (2006). The infant as a prelinguistic model for language learning impairments: predicting from event-related potentials to behavior. *Neuropsychologia*, *44*(3), 396-411. https://doi.org/10.1016/j.neuropsychologia.2005.06.004

Benasich, A. A., Thomas, J. J., Choudhury, N., & Leppänen, P. H. (2002). The importance of rapid auditory processing abilities to early language development: evidence from converging methodologies. *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology*, *40*(3), 278-292. https://doi.org/10.1002/dev.10032

Bendixen, A., Grimm, S., & Schröger, E. (2005). Human auditory event-related potentials predict duration judgments. *Neuroscience letters*, 383(3), 284-288. https://doi.org/10.1016/j.neulet.2005.04.034

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bigdely-Shamlo, N., Mullen, T., Kothe, C., Su, K. M., & Robbins, K. A. (2015). The PREP pipeline: standardized preprocessing for large-scale EEG analysis. Frontiers in neuroinformatics, 9, 16. https://doi.org/10.3389/fninf.2015.00016

Bishop, D. V., Hardiman, M., Uwer, R., & Von Suchodoletz, W. (2007). Maturation of the long-latency auditory ERP: step function changes at start and end of adolescence. *Developmental Science*, 10(5), 565-575. https://doi.org/10.1111/j.1467-7687.2007.00619.x

Blasi, A., Lloyd-Fox, S., Sethna, V., Brammer, M. J., Mercure, E., Murray, L., Williams, S. C. R., Simmons, A., Murphy, D. G. M., & Johnson, M. H. (2015). Atypical processing

of voice sounds in infants at risk for autism spectrum disorder. *Cortex*, *71*, 122–133. https://doi.org/10.1016/j.cortex.2015.06.015

Blasi, A., Mercure, E., Lloyd-Fox, S., Thomson, A., Brammer, M., Sauter, D., Deeley, Q., Barker, G. J., Renvall, V., Deoni, S., Gasston, D., Williams, S. C. R., Johnson, M. H., Simmons, A., & Murphy, D. G. M. (2011). Early specialization for voice and emotion processing in the infant brain. *Current Biology*, *21*(14), 1220–1224. https://doi.org/10.1016/j.cub.2011.06.009

Boersma, P., & Weenink, D. (2022). Praat: A system for doing phonetics by computer, 2000. *Software available at www. praat. org*, *4*(2).

Bonath, B., Noesselt, T., Krauel, K., Tyll, S., Tempelmann, C., & Hillyard, S. A. (2014). NeuroImage Audio-visual synchrony modulates the ventriloquist illusion and its neural / spatial representation in the auditory cortex. *NeuroImage*, *98*, 425–434. https://doi.org/10.1016/j.neuroimage.2014.04.077

Bonte, M., Ley, A., Scharke, W., & Formisano, E. (2016). NeuroImage Developmental re fi nement of cortical systems for speech and voice processing. *NeuroImage*, *128*, 373–384. https://doi.org/10.1016/j.neuroimage.2016.01.015

Botvinick, M. M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: an update. *Trends in cognitive sciences*, *8*(12), 539-546. https://doi.org/10.1016/j.tics.2004.10.003

Bradshaw, J., McCracken, C., Pileggi, M., Brane, N., Delehanty, A., Day, T., ... & Wetherby, A. (2021). Early social communication development in infants with autism spectrum disorder. *Child development*, *92*(6), 2224-2234.https://doi.org/10.1111/cdev.13683

Brandes, U. (2001). A faster algorithm for betweenness centrality. *Journal of mathematical sociology*, *25*(2), 163-177. https://doi.org/10.1080/0022250X.2001.9990249

Bullmore, E., Barnes, A., Bassett, D. S., Fornito, A., Kitzbichler, M., Meunier, D., & Suckling, J. (2009). Generic aspects of complexity in brain imaging data and other biological systems. *NeuroImage*, *47*(3), 1125–1134. https://doi.org/10.1016/j.neuroimage.2009.05.032

Bullmore, E., & Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature reviews neuroscience*, *10*(3), 186-198. https://doi.org/10.1038/nrn2575

Calder, A. J., & Young, A. W. (2005). Understanding the recognition of facial identity and facial expression. *Nature Reviews Neuroscience*, *6*(8), 641–651. https://doi.org/10.1038/nrn1724

Calvert, G. A., Hansen, P. C., Iversen, S. D., & Brammer, M. J. (2001). Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the

BOLD effect. *Neuroimage*, *14*(2), 427-438. https://doi.org/10.1006/nimg.2001.0812

Cantiani, C., Riva, V., Piazza, C., Bettoni, R., Molteni, M., Choudhury, N., Marino, C., & Benasich, A. A. (2016). Auditory discrimination predicts linguistic outcome in Italian infants with and without familial risk for language learning impairment. *Developmental Cognitive Neuroscience*, *20*, 23–34. https://doi.org/10.1016/j.dcn.2016.03.002

Capilla, A., Belin, P., & Gross, J. (2013). The early spatio-temporal correlates and task independence of cerebral voice processing studied with MEG. *Cerebral Cortex*, *23*(6), 1388-1395. https://doi.org/10.1093/cercor/bhs119

Casas, A. S. H., Lajnef, T., Pascarella, A., Guiraud-Vinatea, H., Laaksonen, H., Bayle, D., ... & Boulenger, V. (2021). Neural oscillations track natural but not artificial fast speech: Novel insights from speech-brain coupling using MEG. *Neuroimage*, *244*, 118577. https://doi.org/10.1016/j.neuroimage.2021.118577

Cecere, R., Gross, J., Willis, A., & Thut, G. (2017). Being first matters: topographical representational similarity analysis of ERP signals reveals separate networks for audiovisual temporal binding depending on the leading sense. *Journal of Neuroscience*, *37*(21), 5274-5287. https://doi.org/10.1523/JNEUROSCI.2926-16.2017

Čeponienė, R., Kushnerenko, E., Fellman, V., Renlund, M., Suominen, K., & Näätänen, R. (2002). Event-related potential features indexing central auditory discrimination by newborns. *Cognitive Brain Research*, 13(1), 101-113. https://doi.org/10.1016/S0926-6410(01)00093-3

Čeponienė, R., Lepistö, T., Soininen, M., Aronen, E., Alku, P., & Näätänen, R. (2004). Event-related potentials associated with sound discrimination versus novelty detection in children. *Psychophysiology*, 41(1), 130-141. https://doi.org/10.1111/j.1469-8986.2003.00138.x

Cheng, Y., Lee, S. Y., Chen, H. Y., Wang, P. Y., & Decety, J. (2012). Voice and emotion processing in the human neonatal brain. *Journal of cognitive neuroscience*, *24*(6), 1411-1419. https://doi.org/10.1162/jocn_a_00214

Choudhury, N. & Benasich, A. A. Clinical Neurophysiology Maturation of auditory evoked potentials from 6 to 48 months : Prediction to 3 and 4 year language and cognitive abilities. *Clin. Neurophysiol.* 122, 320–338 (2011). https://doi.org/10.1016/j.clinph.2010.05.035

Clumeck, C., Suarez Garcia, S., Bourguignon, M., Wens, V., Op de Beeck, M., Marty, B., ... & De Tiege, X. (2014). Preserved coupling between the reader's voice and the listener's cortical activity in autism spectrum disorders. *Plos one*, *9*(3), e92329. https://doi.org/10.1371/journal.pone.0092329

Colonius, H., & Diederich, A. (2004). Multisensory interaction in saccadic reaction time: a time-window-of-integration model. *Journal of cognitive neuroscience*, *16*(6), 1000-

1009. 10.1162/0898929041502733

Conte, S., & Richards, J. E. (2022). Cortical source analysis of event-related potentials: A developmental approach. *Developmental Cognitive Neuroscience*, *54*, 101092. https://doi.org/10.1016/j.dcn.2022.101092

Cook, L. A., Van Valkenburg, D. L., & Badcock, D. R. (2011). Predictability affects the perception of audiovisual synchrony in complex sequences. *Attention, Perception, & Psychophysics*, *73*, 2286-2297. https://doi.org/10.3758/s13414-011-0185-8

Coutrot, A., & Guyader, N. (2014). How saliency, faces, and sound influence gaze in dynamic social scenes. *Journal of vision*, *14*(8), 5-5. https://doi.org/10.1167/14.8.5.doi

Covic, A., Keitel, C., Porcu, E., Schröger, E., & Müller, M. M. (2017). Audio-visual synchrony and spatial attention enhance processing of dynamic visual stimulation independently and in parallel: a frequency-tagging study. *Neuroimage*, *161*, 32-42. https://doi.org/10.1016/j.neuroimage.2017.08.022

Crosse, M. J., Di Liberto, G. M., & Lalor, E. C. (2016). Eye can hear clearly now: inverse effectiveness in natural audiovisual speech processing relies on long-term crossmodal temporal integration. *Journal of Neuroscience*, *36*(38), 9888-9895. https://doi.org/10.1523/JNEUROSCI.1396-16.2016

Curtindale, L. M., Bahrick, L. E., Lickliter, R., & Colombo, J. (2019). Journal of Experimental Child Effects of multimodal synchrony on infant attention and heart rate during events with social and nonsocial stimuli. *Journal of Experimental Child Psychology*, *178*, 283–294. https://doi.org/10.1016/j.jecp.2018.10.006

Dale, A. M., Liu, A. K., Fischl, B. R., Buckner, R. L., Belliveau, J. W., Lewine, J. D., & Halgren, E. (2000). Dynamic statistical parametric mapping: combining fMRI and MEG for high-resolution imaging of cortical activity. *neuron*, *26*(1), 55-67. https://doi.org/10.1016/S0896-6273(00)81138-1

Dapretto, M., Davies, M. S., Pfeifer, J. H., Scott, A. A., Sigman, M., Bookheimer, S. Y., & Iacoboni, M. (2006). Understanding emotions in others: mirror neuron dysfunction in children with autism spectrum disorders. *Nature neuroscience*, 9(1), 28.

Davies-Thompson, J., Elli, G. V., Rezk, M., Benetti, S., van Ackeren, M., & Collignon, O. (2019). Hierarchical brain network for face and voice integration of emotion expression. *Cerebral Cortex*, *29*(9), 3590-3605. https://doi.org/10.1093/cercor/bhy240

de Boer-Schellekens, L., Eussen, M., & Vroomen, J. (2013). Diminished sensitivity of audiovisual temporal order in autism spectrum disorder. *Frontiers in integrative neuroscience*, *7*, 8. https://doi.org/10.3389/fnint.2013.00008

de Boer-Schellekens, L. De, Keetels, M., Eussen, M., & Vroomen, J. (2013). No evidence for impaired multisensory integration of low-level audiovisual stimuli in adolescents and young adults with autism spectrum disorders. *Neuropsychologia*, *51*(14), 3004-

3013. https://doi.org/10.1016/j.neuropsychologia.2013.10.005

DeCasper, A. J., & Spence, M. J. (1986). Prenatal maternal speech influences newborns' perception of speech sounds. *Infant behavior and Development*, *9*(2), 133-150. https://doi.org/10.1016/0163-6383(86)90025-1

Dehaene-Lambertz, G., Dehaene, S., & Hertz-Pannier, L. (2002). Functional neuroimaging of speech perception in infants. *Science*, 298(5600), 2013-2015 10.1126/science.1077066

Dehaene-Lambertz, G., Hertz-Pannier, L., Dubois, J., Mériaux, S., Roche, A., Sigman, M., & Dehaene, S. (2006). Functional organization of perisylvian activation during presentation of sentences in preverbal infants. *Proceedings of the National Academy of Sciences*, 103(38), 14240-14245. https://doi.org/10.1073/pnas.0606302103

Dekker, T., Mareschal, D., Sereno, M. I., & Johnson, M. H. (2011). Dorsal and ventral stream activation and object recognition performance in school-age children. *NeuroImage*, *57*(3), 659–670. https://doi.org/10.1016/j.neuroimage.2010.11.005

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., ... & Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage*, *31*(3), 968-980. https://doi.org/10.1016/j.neuroimage.2006.01.021

Doesburg, S. M., Emberson, L. L., Rahi, A., Cameron, D., & Ward, L. M. (2008). Asynchrony from synchrony: long-range gamma-band neural synchrony accompanies perception of audiovisual speech asynchrony. *Experimental Brain Research*, *185*, 11-20. https://doi.org/10.1007/s00221-007-1127-5

Dubois, J., Poupon, C., Thirion, B., Simonnet, H., Kulikova, S., & Leroy, F. & Dehaene-Lambertz, G.(2016). Exploring the early organization and maturation of linguistic pathways in the human infant brain. *Cerebral Cortex*, 26(5), 2283-2298. https://doi.org/10.1093/cercor/bhv082

Edgar, J. C., Murray, R., Kuschner, E. S., Pratt, K., Paulson, D. N., Dell, J., ... & Roberts, T. P. (2015). The maturation of auditory responses in infants and young children: a cross-sectional study from 6 to 59 months. *Frontiers in Neuroanatomy*, *9*, 131. https://doi.org/10.3389/fnana.2015.00131

Eg, R., & Behne, D. M. (2015). Perceived synchrony for realistic and dynamic audiovisual events. *Frontiers in psychology*, *6*, 736. https://doi.org/10.3389/fpsyg.2015.00736

Eimer, M., & Holmes, A. (2007). Event-related brain potential correlates of emotional face processing. *Neuropsychologia*, 45(1), 15-31.

https://doi.org/10.1016/j.neuropsychologia.2006.04.022

Elsabbagh, M., & Johnson, M. H. (2016). Review Autism and the Social Brain : The First-Year Puzzle. *Biological Psychiatry*, *80*(2), 94–99. https://doi.org/10.1016/j.biopsych.2016.02.019

Fecteau, S., Armony, J. L., Joanette, Y., & Belin, P. (2004). Is voice processing species-specific in human auditory cortex? An fMRI study. *Neuroimage*, *23*(3), 840-848. https://doi.org/10.1016/j.neuroimage.2004.09.019

Feldman, J. I., Dunham, K., Cassidy, M., Wallace, M. T., Liu, Y., & Woynaroski, T. G. (2018). Audiovisual multisensory integration in individuals with autism spectrum disorder: A systematic review and meta-analysis. *Neuroscience & Biobehavioral Reviews*, *95*, 220-234. https://doi.org/10.1016/j.neubiorev.2018.09.020

Fellman, V., Kushnerenko, E., Mikkola, K., Ceponiene, R., Leipälä, J., & Näätänen, R. (2004). Atypical auditory event-related potentials in preterm infants during the first year of life: a possible sign of cognitive dysfunction?. *Pediatric research*, 56(2), 291-297.

Fornito, A., Zalesky, A., & Bullmore, E. (2016). *Fundamentals of brain network analysis*. Academic Press.

Fransson, P., Åden, U., Blennow, M., & Lagercrantz, H. (2011). The functional architecture of the infant brain as revealed by resting-state fMRI. *Cerebral Cortex*, *21*(1), 145–154. https://doi.org/10.1093/cercor/bhq071

Friederici, A. D. (2005). Neurophysiological markers of early language acquisition: from syllables to sentences. *Trends in cognitive sciences*, *9*(10), 481-488. https://doi.org/10.1016/j.tics.2005.08.008

Friederici, A. D. (2012). The cortical language circuit: from auditory perception to sentence comprehension. *Trends in cognitive sciences*, *16*(5), 262-268. https://doi.org/10.1016/j.tics.2012.04.001

Fries, P. (2005). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends in cognitive sciences*, *9*(10), 474-480. https://doi.org/10.1016/j.tics.2005.08.011

Gebauer, L., Skewes, J., Hørlyck, L., & Vuust, P. (2014). Atypical perception of affective prosody in Autism Spectrum Disorder. *NeuroImage: Clinical*, *6*, 370-378. https://doi.org/10.1016/j.nicl.2014.08.025

Gervain, J., & Geffen, M. N. (2019). Efficient Neural Coding in Auditory and Speech Perception. *Trends in Neurosciences*, *42*(1), 56–65. https://doi.org/10.1016/j.tins.2018.09.004

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

14222-14227 https://doi.org/10.1073/pnas.0806530105

Gervais, H., Belin, P., Boddaert, N., Leboyer, M., Coez, A., Sfaello, I., Barthélémy, C., Brunelle, F., Samson, Y., & Zilbovicius, M. (2004). Abnormal cortical voice processing in autism. *Nature Neuroscience*, *7*(8), 801–802. https://doi.org/10.1038/nn1291

Gramfort, A., Papadopoulo, T., Olivi, E., & Clerc, M. (2011). Forward field computation with OpenMEEG. *Computational intelligence and neuroscience*, *2011*. https://doi.org/10.1155/2011/923703

Griffiths, R. (1984). *Griffiths mental development scales*. Test Angency.

Grossmann, T. (2010). The development of emotion perception in face and voice during infancy. *Restorative neurology and neuroscience*, *28*(2), 219-236. https://doi.org/10.3233/RNN-2010-0499

Grossmann, T. (2015). The development of social brain functions in infancy. *Psychological bulletin*, *141*(6), 1266. https://doi.org/http://dx.doi.org/10.1037/bul0000002

Grossmann, T., & Johnson, M. H. (2007). The development of the social brain in human infancy. *European Journal of Neuroscience*, *25*(4), 909-919. https://doi.org/10.1111/j.1460-9568.2007.05379.x

Grossmann, T., Johnson, M. H., Lloyd-Fox, S., Blasi, A., Deligianni, F., Elwell, C., & Csibra, G. (2008). Early cortical specialization for face-to-face communication in human infants. *Proceedings of the Royal Society B: Biological Sciences*, 275(1653), 2803-2811. https://doi.org/10.1098/rspb.2008.0986

Grossmann, T., Oberecker, R., Koch, S. P., & Friederici, A. D. (2010). The Developmental Origins of Voice Processing in the Human Brain. *Neuron*, *65*(6), 852–858. https://doi.org/10.1016/j.neuron.2010.03.001

Grossmann, T., Striano, T., & Friederici, A. D. (2005). Infants' electric brain responses to emotional prosody. *Neuroreport*, *16*(16), 1825-1828. 10.1097/01.wnr.0000185964.34336.b1

Grossman, R. B., & Tager-Flusberg, H. (2008). Reading faces for information about words and emotions in adolescents with autism. *Research in autism spectrum disorders*, *2*(4), 681-695. https://doi.org/10.1016/j.rasd.2008.02.004ù

Grossman, R. B., & Tager-Flusberg, H. (2012). “Who said that?” Matching of low-and high-intensity emotional prosody to facial expressions by adolescents with ASD. *Journal of autism and developmental disorders*, *42*, 2546-2557.

Gschwind, M., Pourtois, G., Schwartz, S., Van De Ville, D., & Vuilleumier, P. (2012). White-matter connectivity between face-responsive regions in the human brain. *Cerebral cortex*, *22*(7), 1564-1576. https://doi.org/10.1093/cercor/bhr226

Háden, G. P., Honing, H., Török, M. & Winkler, I. Detecting the temporal structure of sound sequences in newborn infants. *Int. J. Psychophysiol.* 96, 23–28 (2015).

https://doi.org/10.1016/j.ijpsycho.2015.02.024

Hadi, A. S., & Ling, R. F. (1998). Some cautionary notes on the use of principal components regression. *American Statistician*, *52*(1), 15–19. https://doi.org/10.1080/00031305.1998.10480530

Hämäläinen, J. A., Leppänen, P. H., Guttorm, T. K., & Lyytinen, H. (2008). Event-related potentials to pitch and rise time change in children with reading disabilities and typically reading children. *Clinical Neurophysiology*, 119(1), 100-115. ps://doi.org/10.1016/j.clinph.2007.09.064

Hämäläinen, J. A., Ortiz-Mantilla, S., & Benasich, A. A. (2011). Source localization of event-related potentials to pitch change mapped onto age-appropriate MRIs at 6 months of age. *Neuroimage*, *54*(3), 1910-1918. https://doi.org/10.1016/j.neuroimage.2010.10.016

Hautus, M. J., & Johnson, B. W. (2005). Object-related brain potentials associated with the perceptual segregation of a dichotically embedded pitch. *The Journal of the Acoustical Society of America*, *117*(1), 275-280. https://doi.org/10.1121/1.1828499

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in cognitive sciences*, *4*(6), 223-233. https://doi.org/10.1016/S1364-6613(00)01482-0

Hillairet de Boisferon, A., Tift, A. H., Minar, N. J., & Lewkowicz, D. J. (2017). Selective attention to a talker's mouth in infancy: Role of audiovisual temporal synchrony and linguistic experience. *Developmental science*, *20*(3), e12381. https://doi.org/10.1111/desc.12381

Hipp, J. F., Hawellek, D. J., Corbetta, M., Siegel, M., & Engel, A. K. (2012). Large-scale cortical correlation structure of spontaneous oscillatory activity. *Nature neuroscience*, *15*(6), 884-890. https://doi.org/10.1038/nn.3101

Honisch, J. J., Mane, P., Golan, O., & Chakrabarti, B. (2021). Keeping in time with social and non-social stimuli: synchronisation with auditory, visual, and audio-visual cues. *Scientific reports*, *11*(1), 1-11. https://doi.org/10.1038/s41598-021-88112-y

Huang, H., Shu, N., Mishra, V., Jeon, T., Chalak, L., Wang, Z. J., ... & He, Y. (2015). Development of human brain structural networks through infancy and childhood. *Cerebral Cortex*, *25*(5), 1389-1404. https://doi.org/10.1093/cercor/bht335

Hummel, F., & Gerloff, C. (2005). Larger interregional synchrony is associated with greater behavioral success in a complex sensory integration task in humans. *Cerebral Cortex*, *15*(5), 670–678. https://doi.org/10.1093/cercor/bhh170

Huttenlocher, P. R., & Dabholkar, A. S. (1997). Regional differences in synaptogenesis in human cerebral cortex. *Journal of comparative Neurology*, *387*(2), 167-178. https://doi.org/10.1002/(SICI)1096-9861(19971020)387:2<167::AID-CNE1>3.0.CO;2-

Z

Hyde, D. C., Flom, R., Porter, C. L., Hyde, D. C., Flom, R., & Behavioral, C. L. P. (2017). Behavioral and Neural Foundations of Multisensory Face-Voice Perception in Infancy Behavioral and Neural Foundations of Multisensory Face-Voice. *Developmental Neuropsychology*, *41*(5–8), 273–292. https://doi.org/10.1080/87565641.2016.1255744

Hyde, D. C., Jones, B. L., Flom, R., & Porter, C. L. (2011). Neural signatures of face–voice synchrony in 5-month-old human infants. *Developmental Psychobiology*, *53*(4), 359-370. https://doi.org/10.1002/dev.20525

Ikumi, N., Torralba, M., Ruzzoli, M., & Soto-Faraco, S. (2019). The phase of pre-stimulus brain oscillations correlates with cross-modal synchrony perception. *European Journal of Neuroscience*, *49*(2), 150-164. https://doi.org/10.1111/ejn.14186

Jessen, S., Obleser, J., & Tune, S. (2021). Developmental Cognitive Neuroscience Neural tracking in infants – An analytical tool for multisensory social processing in development. *Developmental Cognitive Neuroscience*, *52*, 101034. https://doi.org/10.1016/j.dcn.2021.101034

Jing, H., & Benasich, A. A. (2006). Brain responses to tonal changes in the first two years of life. *Brain and Development*, 28(4), 247-256. https://doi.org/10.1016/j.braindev.2005.09.002

Johnson, B. W., & Hautus, M. J. (2010). Neuropsychologia Processing of binaural spatial information in human auditory cortex : Neuromagnetic responses to interaural timing and level differences. *Neuropsychologia*, *48*(9), 2610–2619. https://doi.org/10.1016/j.neuropsychologia.2010.05.008

Johnson, M. H. (2001). Functional brain development in humans. *Nature Reviews Neuroscience*, *2*(7), 475–483. https://doi.org/10.1038/35081509

Johnson, M. H. (2011). Interactive specialization: a domain-general framework for human functional brain development?. *Developmental cognitive neuroscience*, *1*(1), 7-21. https://doi.org/10.1016/j.dcn.2010.07.003

Johnson, M. H., Grossmann, T., & Kadosh, K. C. (2009). Mapping functional brain development: Building a social brain through interactive specialization. *Developmental psychology*, 45(1), 151. http://dx.doi.org/10.1037/a0014548

Johnson, M. H., Griffin, R., Csibra, G., Halit, H., Farroni, T., De Haan, M., Tucker, L. A., Baron-Cohen, S., & Richards, J. (2005). The emergence of the social brain network: Evidence from typical and atypical development. *Development and Psychopathology*, *17*(3), 599–619. https://doi.org/10.1136/bmj.38519.678148.8F

Jolliffe, I. T. (1982). A Note on the Use of Principal Components in Regression. *Applied Statistics*, *31*(3), 300. https://doi.org/10.2307/2348005

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent

developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2065), 20150202. https://doi.org/10.1098/rsta.2015.0202

Jones, E. J. H., Venema, K., Lowy, R., Earl, R. K., & Webb, S. J. (2015). Developmental changes in infant brain activity during naturalistic social experiences. *Developmental Psychobiology*, *57*(7), 842–853. https://doi.org/10.1002/dev.21336

Kadosh, K. C., & Johnson, M. H. (2007). Developing a cortex specialized for face perception. *Trends in cognitive sciences*, *11*(9), 367-369. https://doi.org/10.1016/j.tics.2007.06.007

Kalashnikova, M., Peter, V., Di Liberto, G. M., Lalor, E. C., & Burnham, D. (2018). Infant-directed speech facilitates seven-month-old infants' cortical tracking of speech. *Scientific reports*, *8*(1), 1-8. https://doi.org/10.1038/s41598-018-32150-6

Karmiloff-Smith, A. (2009). Preaching to the converted? From constructivism to neuroconstructivism. *Child development perspectives*, *3*(2), 99-102. https://doi.org/10.1111/j.1750-8606.2009.00086.x

Karmiloff-Smith, A., Scerif, G., & Thomas, M. (2002). Different approaches to relating genotype to phenotype in developmental disorders. *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology*, *40*(3), 311-322. https://doi.org/10.1002/dev.10035

Kendall, M. G. (1957). *A course in multivariate analysis* (No. 311.23 K4).

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3?.

Klin, A., & Jones, W. (2015). Measurement and mismeasurement of social development in infants later diagnosed with autism spectrum disorder. *International Journal of Statistics in Medical Research*, *4*(2), 180.

Kolozsvári, O. B., Xu, W., Gerike, G., Parviainen, T., Nieminen, L., Noiray, A., & Hämäläinen, J. A. (2021). Coherence between brain activation and speech envelope at word and sentence levels showed age-related differences in low frequency bands. *Neurobiology of Language*, *2*(2), 226-253. https://doi.org/10.1162/nol_a_00033

Kopp, F. (2014). Audiovisual temporal fusion in 6-month-old infants. *Developmental Cognitive Neuroscience*, *9*, 56-67. https://doi.org/10.1016/j.dcn.2014.01.001

Kopp, F., & Dietrich, C. (2013). Neural dynamics of audiovisual synchrony and asynchrony perception in 6-month-old infants. *Frontiers in psychology*, *4*, 2. https://doi.org/10.3389/fpsyg.2013.00002

Korpilahti, P., Jansson-Verkasalo, E., Mattila, M. L., Kuusikko, S., Suominen, K., Rytky, S., ... & Moilanen, I. (2007). Processing of affective speech prosody is impaired in Asperger syndrome. *Journal of Autism and Developmental Disorders*, *37*, 1539-1549. https://doi.org/10.1007/s10803-006-0271-2

Kothe, C. A., & Makeig, S. (2013). BCILAB: a platform for brain–computer interface development. *Journal of neural engineering, 10*(5), 056014. https://doi.org/10.1088/1741-2560/10/5/056014

Kriegstein, K. V., & Giraud, A. L. (2004). Distinct functional substrates along the right superior temporal sulcus for the processing of voices. *Neuroimage*, *22*(2), 948-955. https://doi.org/10.1016/j.neuroimage.2004.02.020

Kudo, N., Nonaka, Y., Mizuno, N., Mizuno, K., & Okanoya, K. (2011). On-line statistical segmentation of a non-speech auditory stream in neonates as demonstrated by event-related brain potentials. *Developmental science*, *14*(5), 1100-1106. https://doi.org/10.1111/j.1467-7687.2011.01056.x

Kushnerenko, E., Ceponiene, R., Balan, P., Fellman, V., Huotilainen, M., & Näätänen, R. (2002). Maturation of the auditory event-related potentials during the first year of life. *Neuroreport*, *13*(1), 47-51. 13(1):p 47-51, January 21, 2002.

Kushnerenko, E., Cheour, M., Ceponiene, R., Fellman, V., Renlund, M., Soininen, K., ... & Naatanen, R. (2001). Central auditory processing of durational changes in complex speech patterns by newborns: an event-related brain potential study. *Developmental neuropsychology*, *19*(1), 83-97. https://doi.org/10.1207/S15326942DN1901_6

Kushnerenko, E., Fellman, V., Huotilainen, M., & Winkler, I. (2001). Event-related potential correlates of sound duration: similar pattern from birth to adulthood. *NeuroReport*, **12**(17), 3777-3781.

Kushnerenko, E. V., Van den Bergh, B. R., & Winkler, I. (2013). Separating acoustic deviance from novelty during the first year of life: a review of event-related potential evidence. *Frontiers in psychology*, *4*, 595. https://doi.org/10.3389/fpsyg.2013.00595

Latinus, M., & Belin, P. (2011). Human voice perception. *Current Biology*, *21*(4), R143-R145. https://doi.org/10.1016/j.cub.2010.12.033

Lebenberg, J., Mangin, J. F., Thirion, B., Poupon, C., Hertz-Pannier, L., Leroy, F., ... & Dubois, J. (2019). Mapping the asynchrony of cortical maturation in the infant brain: a MRI multi-parametric clustering approach. *Neuroimage*, *185*, 641-653. https://doi.org/10.1016/j.neuroimage.2018.07.022

Leonardi, N., Richiardi, J., Gschwind, M., Simioni, S., Annoni, J. M., Schluep, M., Vuilleumier, P., & Van De Ville, D. (2013). Principal components of functional connectivity: A new approach to study dynamic brain connectivity during rest. *NeuroImage*, *83*, 937–950. https://doi.org/10.1016/j.neuroimage.2013.07.019

Leroy, F., Glasel, H., Dubois, J., Hertz-Pannier, L., Thirion, B., Mangin, J. F., & Dehaene-Lambertz, G. (2011). Early maturation of the linguistic dorsal pathway in human infants. *Journal of Neuroscience*, 31(4), 1500-1506. https://doi.org/10.1523/JNEUROSCI.4141-10.2011

Lewis, R., & Noppeney, U. (2010). Audiovisual synchrony improves motion discrimination via enhanced connectivity between early visual and auditory areas. *Journal of Neuroscience*, *30*(37), 12329-12339. https://doi.org/10.1523/JNEUROSCI.5745-09.2010

Lewkowicz, D. J. (2003). Learning and discrimination of audiovisual events in human infants: the hierarchical relation between intersensory temporal synchrony and rhythmic pattern cues. *Developmental psychology*, *39*(5), 795. https://doi.org/10.1037/0012-1649.39.5.795

Lewkowicz, D. J., Schmuckler, M., & Agrawal, V. (2021). The multisensory cocktail party problem in adults: Perceptual segregation of talking faces on the basis of audiovisual temporal synchrony. *Cognition*, *214*, 104743. https://doi.org/10.1016/j.cognition.2021.104743

Li, Q., Liu, G., Wei, D., Guo, J., Yuan, G., & Wu, S. (2019). The spatiotemporal pattern of pure tone processing: a single-trial EEG-fMRI study. *Neuroimage*, *187*, 184-191. https://doi.org/10.1016/j.neuroimage.2017.11.059

Lindström, R., Lepistö-paisley, T., Vanhala, R., Alén, R., & Kujala, T. (2016). Neuroscience Letters Impaired neural discrimination of emotional speech prosody in children with autism spectrum disorder and language impairment. *Neuroscience Letters*, *628*, 47–51. https://doi.org/10.1016/j.neulet.2016.06.016

Liu, T. Y., Hsieh, J. C., Chen, Y. S., Tu, P. C., Su, T. P., & Chen, L. F. (2012). Different patterns of abnormal gamma oscillatory activity in unipolar and bipolar disorder patients during an implicit emotion task. *Neuropsychologia*, *50*(7), 1514–1520. https://doi.org/10.1016/j.neuropsychologia.2012.03.004

Lloyd-Fox, S., Blasi, A., Elwell, C. E., Charman, T., Murphy, D., & Johnson, M. H. (2013). Reduced neural sensitivity to social stimuli in infants at risk for autism. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1758). https://doi.org/10.1098/rspb.2012.3026

Long, P., Wan, G., Roberts, M. T., & Corfas, G. (2018). Myelin development, plasticity, and pathology in the auditory system. *Developmental neurobiology*, 78(2), 80-92. https://doi.org/10.1002/dneu.22538

Macaluso, E., George, N., Dolan, R., Spence, C., & Driver, J. (2004). Spatial and temporal factors during processing of audiovisual speech: a PET study. *Neuroimage*, *21*(2), 725-732. https://doi.org/10.1016/j.neuroimage.2003.09.049

Maguinness, C., Roswandowitz, C., & Kriegstein, K. Von. (2018). Neuropsychologia Understanding the mechanisms of familiar voice-identity recognition in the. *Neuropsychologia*, *116*(May 2017), 179–193. https://doi.org/10.1016/j.neuropsychologia.2018.03.039

Mandel, D. R., Jusczyk, P. W., & Pisoni, D. B. (1995). Infants' recognition of the sound patterns of their own names. *Psychological Science,* 6(5), 314-317.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Maslennikova, A. V., Portnova, G. V., & Martynova, O. V. (2022). Brain oscillatory patterns of affective prosody perception in children with autism spectrum disorder. *Research in Autism Spectrum Disorders*, *96*, 101993. https://doi.org/10.1016/j.rasd.2022.101993

Mayhew, S. D., Dirckx, S. G., Niazy, R. K., Iannetti, G. D., & Wise, R. G. (2010). EEG signatures of auditory activity correlate with simultaneously recorded fMRI responses in humans. *Neuroimage*, *49*(1), 849-864. https://doi.org/10.1016/j.neuroimage.2009.06.080

Mazziotta, J., Toga, A., Evans, A., Fox, P., Lancaster, J., Zilles, K., ... & Mazoyer, B. (2001). A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *356*(1412), 1293-1322. https://doi.org/10.1098/rstb.2001.0915

Menn, K. H., Michel, C., Meyer, L., Hoehl, S., & Männel, C. (2022). Natural infant-directed speech facilitates neural tracking of prosody. *NeuroImage*, *251*, 118991. https://doi.org/10.1016/j.neuroimage.2022.118991

Menon, V., & Uddin, L. Q. (2010). Saliency, switching, attention and control: a network model of insula function. *Brain structure and function*, *214*, 655-667.

Michel, C., Stets, M., Parise, E., Reid, V. M., Striano, T., & Hoehl, S. (2015). Theta- and alpha-band EEG activity in response to eye gaze cues in early infancy. *NeuroImage*, *118*, 576–583. https://doi.org/10.1016/j.neuroimage.2015.06.042

Mihalik, A., Ferreira, F. S., Moutoussis, M., Ziegler, G., Adams, R. A., Rosa, M. J., Prabhu, G., de Oliveira, L., Pereira, M., Bullmore, E. T., Fonagy, P., Goodyer, I. M., Jones, P. B., Hauser, T., Neufeld, S., Romero-Garcia, R., St Clair, M., Vértes, P. E., Whitaker, K., … Mourão-Miranda, J. (2020). Multiple Holdouts With Stability: Improving the Generalizability of Machine Learning Analyses of Brain–Behavior Relationships. *Biological Psychiatry*, *87*(4), 368–376. https://doi.org/10.1016/j.biopsych.2019.12.001

Missana, M., Atkinson, A. P., & Grossmann, T. (2015). Tuning the developing brain to emotional body expressions. *Developmental Science*, *18*(2), 243-253. https://doi.org/10.1111/desc.12209

Missana, M., Rajhans, P., Atkinson, A. P., & Grossmann, T. (2014). Discrimination of fearful and happy body postures in 8-month-old infants: an event-related potential study. Frontiers in Human Neuroscience, 8, 531.

https://doi.org/10.3389/fnhum.2014.00531

Molinaro, N., & Lizarazu, M. (2018). Delta (but not theta)-band cortical entrainment involves speech-specific processing. *European Journal of Neuroscience*, *48*(7), 2642-2650.https://doi.org/10.1111/ejn.13811

Murat Baldwin, M., Xiao, Z., & Murray, A. (2021). Temporal synchrony in Autism: a systematic review. *Review Journal of Autism and Developmental Disorders*, 1-22.

Myers, B. R., Lense, M. D., & Gordon, R. L. (2019). Pushing the envelope: Developments in neural entrainment to speech and the biological underpinnings of prosody perception. *Brain sciences*, *9*(3), 70. https://doi.org/10.3390/brainsci9030070

Nallet, C., & Gervain, J. (2022). Atypical neural oscillations in response to speech in infants and children with speech and language impairments : a systematic review. *Hearing, Balance and Communication*, *20*(3), 145–154. https://doi.org/10.1080/21695717.2022.2084864

Navarra, J., Vatakis, A., Zampini, M., Soto-Faraco, S., Humphreys, W., & Spence, C. (2005). Exposure to asynchronous audiovisual speech extends the temporal window for audiovisual integration. *Cognitive Brain Research*, *25*(2), 499-507. https://doi.org/10.1016/j.cogbrainres.2005.07.009

Nolte, G., Bai, O., Wheaton, L., Mari, Z., Vorbach, S., & Hallett, M. (2004). Identifying true brain interaction from EEG data using the imaginary part of coherency. *Clinical Neurophysiology*, *115*(10), 2292–2307. https://doi.org/10.1016/j.clinph.2004.04.029

Novitski, N., Huotilainen, M., Tervaniemi, M., Näätänen, R., & Fellman, V. (2007). Neonatal frequency discrimination in 250–4000-Hz range: Electrophysiological evidence. *Clinical Neurophysiology*, 118(2), 412-419. https://doi.org/10.1016/j.clinph.2006.10.008

Obrig, H., Rossi, S., Telkemeyer, S., & Wartenburger, I. (2010). From acoustic segmentation to language processing: evidence from optical imaging. *Frontiers in Neuroenergetics*, 13. https://doi.org/10.3389/fnene.2010.00013

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational intelligence and neuroscience*, *2011*. https://doi.org/10.1155/2011/156869

Orekhova, E. V., Stroganova, T. A., & Posikera, I. N. (1999). Theta synchronization during sustained anticipatory attention in infants over the second half of the first year of life. *International Journal of Psychophysiology*, *32*(2), 151–172. https://doi.org/10.1016/S0167-8760(99)00011-2

Ortiz-Mantilla, S., Hämäläinen, J. A., & Benasich, A. A. (2012). Time course of ERP generators to syllables in infants: a source localization study using age-appropriate

brain templates. *Neuroimage*, *59*(4), 3275-3287. https://doi.org/10.1016/j.neuroimage.2011.11.048

Ozonoff, S., Young, G. S., Carter, A., Messinger, D., Yirmiya, N., Zwaigenbaum, L., ... & Stone, W. L. (2011). Recurrence risk for autism spectrum disorders: a Baby Siblings Research Consortium study. *Pediatrics*, *128*(3), e488-e495. https://doi.org/10.1542/peds.2010-2825

Ozonoff, S., Young, G. S., Landa, R. J., Brian, J., Bryson, S., Charman, T., ... & Iosif, A. M. (2015). Diagnostic stability in young children at risk for autism spectrum disorder: a baby siblings research consortium study. *Journal of Child Psychology and Psychiatry*, *56*(9), 988-998. https://doi.org/10.1111/jcpp.12421

Park, H. J., & Friston, K. (2013). Nonlinear multivariate analysis of neurophysiological signals. *Science*, *6158*(1238411), 3. https://doi.org/10.1126/science.1238411

Park, H., Ince, R. A. A., Schyns, P. G., Thut, G., Park, H., Ince, R. A. A., Schyns, P. G., Thut, G., & Gross, J. (2015). Frontal Top-Down Signals Increase Coupling of Auditory Low-Frequency Oscillations to Continuous Speech in Human Listeners Report Frontal Top-Down Signals Increase Coupling of Auditory Low-Frequency Oscillations to Continuous Speech in Human Listeners. *Current Biology*, *25*(12), 1649–1653. https://doi.org/10.1016/j.cub.2015.04.049

Parviainen, T., Helenius, P., & Salmelin, R. (2019). Children show hemispheric differences in the basic auditory response properties. *Human Brain Mapping*, 40(9), 2699-2710. https://doi.org/10.1002/hbm.24553

Pascual, B., Masdeu, J. C., Hollenbeck, M., Makris, N., Insausti, R., Ding, S. L., & Dickerson, B. C. (2015). Large-scale brain networks of the human left temporal pole: a functional connectivity MRI study. *Cerebral cortex*, *25*(3), 680-702. https://doi.org/10.1093/cercor/bht260

Patten, E., Watson, L. R., & Baranek, G. T. (2014). Temporal synchrony detection and associations with language in young children with ASD. *Autism Research and Treatment*, *2014*. https://doi.org/10.1155/2014/678346

Pell, M. D., Rothermich, K., Liu, P., Paulmann, S., Sethi, S., & Rigoulot, S. (2015). Preferential decoding of emotion from human non-linguistic vocalizations versus speech prosody. *Biological Psychology*, *111*, 14–25. https://doi.org/10.1016/j.biopsycho.2015.08.008

Pernet, C., Garrido, M. I., Gramfort, A., Maurits, N., Michel, C. M., Pang, E., ... & Puce, A. (2020). Issues and recommendations from the OHBM COBIDAS MEEG committee for reproducible EEG and MEG research. *Nature neuroscience*, *23*(12), 1473-1483. https://doi.org/10.1038/s41593-020-00709-0

Pernet, C. R., Mcaleer, P., Latinus, M., Gorgolewski, K. J., Charest, I., Bestelmeyer, P. E.

G., Watson, R. H., Fleming, D., Crabbe, F., Valdes-sosa, M., & Belin, P. (2015). NeuroImage The human voice areas : Spatial organization and inter-individual variability in temporal and extra-temporal cortices. *NeuroImage*, *119*, 164–174. https://doi.org/10.1016/j.neuroimage.2015.06.050

Petkov, C. I., Kayser, C., Steudel, T., Whittingstall, K., Augath, M., & Logothetis, N. K. (2008). A voice region in the monkey brain. *Nature neuroscience*, *11*(3), 367-374. https://doi.org/10.1038/nn2043

Pihko, E., Leppänen, P. H., Eklund, K. M., Cheour, M., Guttorm, T. K., & Lyytinen, H. (1999). Cortical responses of infants with and without a genetic risk for dyslexia: I. Age effects. *Neuroreport*, 10(5), 901-905. 10(5):p 901-905, April 6, 1999.

Pion-Tonachini, L., Kreutz-Delgado, K., & Makeig, S. (2019). ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. *NeuroImage*, *198*, 181-197. https://doi.org/10.1016/j.neuroimage.2019.05.026

Pitcher, D., Duchaine, B., & Walsh, V. (2014). Combined TMS and fMRI reveal dissociable cortical pathways for dynamic and static face perception. *Current Biology*, *24*(17), 2066–2070. https://doi.org/10.1016/j.cub.2014.07.060

Plechawska-Wójcik, M., Tokovarov, M., Kaczorowska, M., & Zapała, D. (2019). A three-class classification of cognitive workload based on EEG spectral data. *Applied Sciences*, *9*(24), 5340. https://doi.org/10.3390/app9245340

Ponton, C. W., Eggermont, J. J., Kwong, B., & Don, M. (2000). Maturation of human central auditory system activity: evidence from multi-channel evoked potentials. *Clinical neurophysiology*, 111(2), 220-236. https://doi.org/10.1016/S1388-2457(99)00236-9

Pourtois, G., de Gelder, B., Bol, A., & Crommelinck, M. (2005). Perception of facial expressions and voices and of their combination in the human brain. *Cortex*, *41*(1), 49-59. https://doi.org/10.1016/S0010-9452(08)70177-1

Power, J. D., Fair, D. A., Schlaggar, B. L., & Petersen, S. E. (2010). Review The Development of Human Functional Brain Networks. *Neuron*, *67*(5), 735–748. https://doi.org/10.1016/j.neuron.2010.08.017

Quadrelli, E., Conte, S., Macchi Cassia, V., & Turati, C. (2019). Emotion in motion: Facial dynamics affect infants' neural processing of emotions. *Developmental Psychobiology*, *61*(6), 843–858. https://doi.org/10.1002/dev.21860

Quadrelli, E., Geangu, E., & Turati, C. (2019). Human action sounds elicit sensorimotor activation early in life. Cortex, 117, 323-335. https://doi.org/10.1016/j.cortex.2019.05.009

Quadrelli, E., Roberti, E., Polver, S., Bulf, H., & Turati, C. (2021). *brain sciences*

Sensorimotor Activity and Network Connectivity to Dynamic and Static Emotional Faces in 7-Month-Old Infants.

Rayson, H., Bonaiuto, J. J., Ferrari, P. F., & Murray, L. (2016). Mu desynchronization during observation and execution of facial expressions in 30-month-old children. *Developmental Cognitive Neuroscience*, 19, 279-287. https://doi.org/10.1016/j.dcn.2016.05.003

Recio, G., Sommer, W., & Schacht, A. (2011). Electrophysiological correlates of perceiving and evaluating static and dynamic facial emotional expressions. *Brain research*, *1376*, 66-75. https://doi.org/10.1016/j.brainres.2010.12.041

Reilly, C. O., Larson, E., Richards, J. E., & Elsabbagh, M. (2021). NeuroImage Structural templates for imaging EEG cortical sources in infants. *NeuroImage*, *227*(December 2020), 117682. https://doi.org/10.1016/j.neuroimage.2020.117682

Revelle, W. (2016) psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA, http://CRAN.R-project.org/package=psych

Righi, G., Tenenbaum, E. J., McCormick, C., Blossom, M., Amso, D., & Sheinkopf, S. J. (2018). Sensitivity to audio-visual synchrony and its relation to language abilities in children with and without ASD. *Autism Research*, *11*(4), 645-653. https://doi.org/10.1002/aur.1918

Righi, G., Tierney, A. L., Tager-Flusberg, H., & Nelson, C. A. (2014). Functional connectivity in the first year of life in infants at risk for autism spectrum disorder: an EEG study. *PLoS One*, *9*(8), e105176.https://doi.org/10.1371/journal.pone.0105176

Riva, V., Cantiani, C., Mornati, G., Gallo, M., Villa, L., Mani, E., Saviozzi, I., Marino, C., & Molteni, M. (2018). Distinct ERP profiles for auditory processing in infants at-risk for autism and language impairment. *Scientific Reports*, *8*(1), 1–11. https://doi.org/10.1038/s41598-017-19009-y

Riva, V., Riboldi, E. M., Dondena, C., Piazza, C., Molteni, M., & Cantiani, C. (2022). Atypical ERP responses to audiovisual speech integration and sensory responsiveness in infants at risk for autism spectrum disorder. *Infancy*, *27*(2), 369-388. https://doi.org/10.1111/infa.12456

Ronconi, L., Molteni, M., & Casartelli, L. (2016). Building blocks of others' understanding: a perspective shift in investigating social-communicative deficit in autism. *Frontiers in Human Neuroscience*, *10*, 144 https://doi.org/10.3389/fnhum.2016.00144

Rotem-kohavi, N., Oberlander, T. F., & Virji-babul, N. (2017). Neuroscience Letters Infants and adults have similar regional functional brain organization for the perception of emotions. *Neuroscience Letters*, *650*, 118–125. https://doi.org/10.1016/j.neulet.2017.04.031

Rubinov, M., & Sporns, O. (2010). Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage*, *52*(3), 1059–1069. https://doi.org/10.1016/j.neuroimage.2009.10.003

Ruhnau, P., Herrmann, B., Maess, B. & Schröger, E. NeuroImage Maturation of obligatory auditory responses and their neural sources : Evidence from EEG and MEG. *Neuroimage* 58, 630–639 (2011). https://doi.org/10.1016/j.neuroimage.2011.06.050

Saby, J. N., & Marshall, P. J. (2012). The utility of EEG band power analysis in the study of infancy and early childhood. *Developmental Neuropsychology*, *37*(3), 253–273. https://doi.org/10.1080/87565641.2011.614663

Sanchez, C. E., Richards, J. E., & Almli, C. R. (2012). Neurodevelopmental MRI brain templates for children from 2 weeks to 4 years of age. *Developmental psychobiology*, *54*(1), 77-91. https://doi.org/10.1002/dev.20579

Santamaria, L., Noreika, V., Georgieva, S., Clackson, K., Wass, S., & Leong, V. (2020). Emotional valence modulates the topology of the parent-infant inter-brain network. *NeuroImage*, *207*(October 2019), 116341. https://doi.org/10.1016/j.neuroimage.2019.116341

Schweinberger, S. R., Kloth, N., & Robertson, D. M. (2011). Hearing facial identities: Brain correlates of face–voice integration in person identification. *Cortex*, *47*(9), 1026-1037. https://doi.org/10.1016/j.cortex.2010.11.011

Senkowski, D., Talsma, D., Grigutsch, M., Herrmann, C. S., & Woldorff, M. G. (2007). Good times for multisensory integration: effects of the precision of temporal synchrony as revealed by gamma-band oscillations. *Neuropsychologia*, *45*(3), 561-571. https://doi.org/10.1016/j.neuropsychologia.2006.01.013

Shic, F., Macari, S., & Chawarska, K. (2013). Speech Disturbs Face Scanning in 6-Month-Old. *Biological Psychiatry*, *75*(3), 231–237. https://doi.org/10.1016/j.biopsych.2013.07.009

Shic, F., Wang, Q., Macari, S. L., & Chawarska, K. (2020). The role of limited salience of speech in selective attention to faces in toddlers with autism spectrum disorders. *Journal of Child Psychology and Psychiatry*, *61*(4), 459-469. https://doi.org/10.1111/jcpp.13118

Simic, N., & Rovet, J. (2017). Dorsal and ventral visual streams: Typical and atypical development. *Child Neuropsychology*, *23*(6), 678-691. https://doi.org/10.1080/09297049.2016.1186616

Snyder, J. S., Alain, C., & Picton, T. W. (2006). Effects of attention on neuroelectric correlates of auditory stream segregation. *Journal of cognitive neuroscience*, *18*(1), 1-13. 10.1162/089892906775250021

Sperdin, H. F., & Schaer, M. (2016). Aberrant development of speech processing in young

children with autism: New insights from neuroimaging biomarkers. *Frontiers in neuroscience*, *10*, 393. https://doi.org/10.3389/fnins.2016.00393

Staib, M., & Frühholz, S. (2021). Progress in Neurobiology Cortical voice processing is grounded in elementary sound analyses for vocalization relevant sound patterns. *Progress in Neurobiology*, *200*, 101982. https://doi.org/10.1016/j.pneurobio.2020.101982

Staib, M., & Frühholz, S. (2022). Distinct functional levels of human voice processing in the auditory cortex. *Cerebral Cortex*. https://doi.org/10.1093/cercor/bhac128

Stam, C. J., Tewarie, P., Van Dellen, E., van Straaten, E. C. W., Hillebrand, A., & Van Mieghem, P. (2014). The trees and the forest: Characterization of complex brain networks with minimum spanning trees. *International Journal of Psychophysiology*, *92*(3), 129–138. https://doi.org/10.1016/j.ijpsycho.2014.04.001

Stevenson, R. A., Segers, M., Ferber, S., Barense, M. D., Camarata, S., & Wallace, M. T. (2016). Keeping time in the brain: Autism spectrum disorder and audiovisual temporal processing. *Autism Research*, *9*(7), 720-738. https://doi.org/10.1002/aur.1566

Stropahl, M., Bauer, A. K. R., Debener, S., & Bleichner, M. G. (2018). Source-modeling auditory processes of EEG data using EEGLAB and brainstorm. *Frontiers in neuroscience*, *12*, 309. https://doi.org/10.3389/fnins.2018.00309

Sullivan, A. E. O., Crosse, M. J., Di Liberto, G. M., de Cheveigné, A., & Lalor, E. C. (2021). Neurophysiological indices of audiovisual speech processing reveal a hierarchy of multisensory integration effects. Journal of Neuroscience, 41(23), 4991-5003. https://doi.org/10.1523/JNEUROSCI.0906-20.2021

Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D., & Leahy, R. M. (2011). Brainstorm: A user-friendly application for MEG/EEG analysis. *Computational Intelligence and Neuroscience*, *2011*. https://doi.org/10.1155/2011/879716

Tallal, P., & Gaab, N. (2006). Dynamic auditory processing, musical experience and language development. *Trends in Neurosciences*, *29*(7), 382–390. https://doi.org/10.1016/j.tins.2006.06.003

Thorpe, S. G., Cannon, E. N., & Fox, N. A. (2016). Spectral and source structural development of mu and alpha rhythms from infancy through adulthood. *Clinical Neurophysiology*, *127*(1), 254–269. https://doi.org/10.1016/j.clinph.2015.03.004

Tóth, B., Urbán, G., Háden, G. P., Márk, M., Török, M., Stam, C. J., & Winkler, I. (2017). Large-scale network organization of EEG functional connectivity in newborn infants. *Human Brain Mapping*, *38*(8), 4019–4033. https://doi.org/10.1002/hbm.23645

Turi, M., Karaminis, T., Pellicano, E., & Burr, D. (2016). No rapid audiovisual recalibration in adults on the autism spectrum. *Scientific reports*, *6*(1), 1-7. https://doi.org/10.1038/srep21756

van den Heuvel, M. I., Otte, R. A., Braeken, M. A. K. A., Winkler, I., Kushnerenko, E., & Bergh, B. R. H. Van Den. (2015). Differences between human auditory event-related potentials (AERPs) measured at 2 and 4 months after birth. *International Journal of Psychophysiology*, *97*(1), 75–83. https://doi.org/10.1016/j.ijpsycho.2015.04.003

van der Gaag, C., Minderaa, R. B., & Keysers, C. (2007). Facial expressions: what the mirror neuron system can and cannot tell us. *Social neuroscience*, 2(3-4), 179-222. https://doi.org/10.1080/17470910701376878

van der Velde, B., White, T., & Kemner, P. C. (2021). The emergence of a theta social brain network during infancy. *NeuroImage*, *240*(May), 118298. https://doi.org/10.1016/j.neuroimage.2021.118298

van Driel, J., Knapen, T., van Es, D. M., & Cohen, M. X. (2014). Interregional alpha-band synchrony supports temporal cross-modal integration. *Neuroimage*, *101*, 404-415. https://doi.org/10.1016/j.neuroimage.2014.07.022

Viola, F. C., Debener, S., Thorne, J., & Schneider, T. R. (2010). Using ICA for the analysis of multi-channel EEG data. *Simultaneous EEG and fMRI: Recording, Analysis, and Application: Recording, Analysis, and Application*, 121-133.

Venables WN, Ripley BD (2002). *Modern Applied Statistics with S*, Fourth edition. Springer, New York. ISBN 0-387-95457-0, https://www.stats.ox.ac.uk/pub/MASS4/.

Vouloumanos, A., & Curtin, S. (2014). Foundational tuning: How infants' attention to speech predicts language development. *Cognitive Science*, *38*(8), 1675-1686. https://doi.org/10.1111/cogs.12128

Xiao, Y., Wen, T. H., Kupis, L., Eyler, L. T., Goel, D., Vaux, K., ... & Courchesne, E. (2022). Neural responses to affective speech, including motherese, map onto clinical and social eye tracking profiles in toddlers with ASD. *Nature Human Behaviour*, *6*(3), 443-454. https://doi.org/10.1038/s41562-021-01237-y

Yin, L.; Chen, X.; Sun, Y.; Worm, T.; Reale, M. A high-resolution 3D dynamic facial expression database. In Proceedings of the 8th International Conference on Automatic Face and Gesture Recognition, Amsterdam, The Netherlands, 17–19 September 2008.

Young, A. W., Frühholz, S., & Schweinberger, S. R. (2020). Face and voice perception: Understanding commonalities and differences. *Trends in cognitive sciences*, *24*(5), 398-410. https://doi.org/10.1016/j.tics.2020.02.001

Zelazo, P. D., & Müller, U. (2010). Executive functioning in typical and atypical children. *U. Goswami, Blackwell Handbook of childhood cognitive development (2nd ed., pp. 574–603)*.

Zhao, C., Chronaki, G., Schiessl, I., Wan, M. W., & Abel, K. M. (2019). Is infant neural sensitivity to vocal emotion associated with mother-infant relational

experience?. *PLoS One*, *14*(2), e0212205. https://doi.org/10.1371/journal.pone.0212205

Zhou, H. Y., Cheung, E. F., & Chan, R. C. (2020). Audiovisual temporal integration: Cognitive processing, neural mechanisms, developmental trajectory and potential interventions. *Neuropsychologia*, *140*, 107396. https://doi.org/10.1016/j.neuropsychologia.2020.107396