amU
Aix Marseille Université

NNT : 2024AIXM0001

THÈSE DE DOCTORAT

Soutenue à AMU – Aix-Marseille Université
le 18 décembre 2024 par

Estelle HERVE

From speech perception in babies to vocabulary acquisition in children

Discipline
Sciences Cognitives

École doctorale
356 - Cognition, Langage, Education

Laboratoire/Partenaires de recherche
Laboratoire Parole et Langage
Centre National de la Recherche Scientifique (CNRS)
Institute of Language, Communication and the Brain (ILCB)

Composition du jury

Laurianne CABRERA Chargée de Recherche – CNRS, Université de Paris	Rapporteure
Judit GERVAIN Full professor – CNRS, Université de Paris & Università di Padova	Rapporteure
Claudia MÄNNEL Professor – Charité - Universitätsmedizin Berlin	Examinatrice
Ruth DE DIEGO-BALAGUER Research Professor – Universitat de Barcelona	Examinatrice
Johannes ZIEGLER Directeur de Recherche – CNRS, Aix-Marseille Université	Président du jury
Laurent PREVOT Professeur – CNRS, Aix-Marseille Université	Directeur de thèse
Clément FRANCOIS Chargé de Recherche – CNRS, Aix-Marseille Université	Co-directeur de thèse

ILCB Institute of Language, Communication and the Brain

UMR 7309 • CNRS • AMU PAROLE ET LANGAGE

cnrs

Affidavit

I, undersigned, Estelle Hervé, hereby declare that the work presented in this manuscript is my own work, carried out under the scientific supervision of Laurent Prévot and Clément François, in accordance with the principles of honesty, integrity and responsibility inherent to the research mission. The research work and the writing of this manuscript have been carried out in compliance with both the french national charter for Research Integrity and the Aix-Marseille University charter on the fight against plagiarism.

This work has not been submitted previously either in this country or in another country in the same or in a similar version to any other examination body.

Aix-en-Provence, October 20, 2024



This work is licensed under Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International Public License

Liste de publications et participation aux conférences

Liste des publications et/ou brevets réalisées dans le cadre du projet de thèse :

1. **Hervé, E.**, Mento, G., Desnous, B., and François, C. (2022). Challenges and new perspectives of developmental cognitive EEG studies. *NeuroImage*, 260, 119508.
2. **Hervé, E.**, François, C., and Prévot, L. (2024) Daily auditory environments in French-speaking infants : A longitudinal dataset. *ACL 2024 Cognitive Modeling and Computational Linguistics Workshop*.
3. **Hervé, E.**, Dubarry, A.S., Legou, T., Prévot, L., François, C. (in prep.) Development of cortical and subcortical responses to non-native speech sounds during infancy.
4. **Hervé, E.**, Geringswald, F., Zielinksi, C., François, C. (in prep.) Minimal pair associative word learning in 18- and 24-month-old infants.

Participation aux conférences et écoles d'été au cours de la période de thèse :

1. 2021 Summer School of the Institute for Language, Communication and the Brain (30 Aout 2021 - 3 Septembre 2021) Institute for Language, Communication and the Brain – Marseille.
2. CuttingEEG 2021 (4-7 Octobre 2021, Aix-en-Provence) – Le Cube, Aix-en-Provence. *Bénévole, membre de l'équipe d'organisation.*
3. 4th Summer School ISOLDE Infant Studies On Language Development in Europe (14 juin 2021) – University of Potsdam, Germany.
4. FFR Workshop 2022 (8-10 Juin 2022) – Université de Barcelone,Catalogne, Espagne. *Présentation d'un poster.*
5. PracticalMEEG 2nd Edition (14-16 Décembre) – Le Cube, Aix-en-Provence. *Bénévole, membre de l'équipe d'organisation.*

6. Multimodality in Social Interactions Workshop (3-4 Avril 2023) – Marseille. *Présentations orales.*
7. 15th Annual Meeting of the Society for the Neurobiology of Language (24-26 Octobre) – Palais du Pharo, Marseille. *Présentation d'un poster.*
8. Multimodality in Social Interactions Workshop (3-4 Avril 2023) – Marseille. *Présentations orales.*
9. Multimodality in Social Interactions Workshop (3-4 Avril 2023) – Marseille. *Présentations orales.*
10. audidev : Auditory Development Workshop (20-21 Novembre 2023) – Paris. *Présentation d'un poster.*
11. WILD : Workshop on Infant Language Development (18-21 Juin 2024) - Lisbonne. *Présentation d'un poster.*

Résumé et mots clés

L'acquisition du langage est une prouesse remarquable de la cognition humaine. Les enfants maîtrisent la complexité de leur langue maternelle en quelques années seulement, et semblent y parvenir sans effort. Malgré cette apparente facilité, la compréhension de ce processus est complexifiée par l'interaction qui existe entre de nombreux facteurs, notamment neuronaux, environnementaux et cognitifs. Dans ce contexte, les études multimodales intégratives sont essentielles pour mieux comprendre la manière dont ces facteurs interagissent. De plus, la mise en œuvre de paradigmes longitudinaux permet de suivre les changements intra-individuels au fil du temps et d'aborder les relations de causalité. Ma thèse adopte cette approche, en combinant des données électrophysiologiques, environnementales et comportementales pour explorer les corrélats neuronaux de la perception de la parole, le rôle de l'environnement sonore et les capacités d'apprentissage associatif de nouveaux mots dans la prédiction de la taille du vocabulaire au cours des deux premières années de vie. Grâce à cette approche intégrative, mon travail contribue à une meilleure compréhension des prédicteurs précoces du développement du langage.

La thèse est divisée en trois études. Dans la première étude, j'ai exploré les capacités de perception précoce de la parole au cours des deux premières années de vie. J'ai validé un paradigme innovant permettant d'enregistrer les réponses cérébrales sous-corticales et corticales précoces en une seule session. L'analyse des Frequency-Following Responses (FFRs) a révélé des trajectoires développementales de l'encodage spectro-temporel des sons de parole, en particulier en ce qui concerne la fréquence fondamentale et les harmoniques, suggérant des changements maturationnels sous-corticaux. Les réponses corticales (MMRs) ont révélé des différences dans l'encodage des sons de parole entre 10 et 18 mois, en accord avec la théorie de "perceptual narrowing" sur l'émergence des catégories phonologiques chez les jeunes enfants.

La seconde étude visait à caractériser les environnements auditifs quotidiens des enfants, en utilisant le système LENA (Language ENvironment Analysis). J'ai exploré un nouveau corpus LENA collecté dans des foyers français, enrichissant ainsi la littérature existante sur les jeux de données non anglophones. La fiabilité du logiciel d'annotation automatique de LENA sur les données françaises a été établie par le biais d'analyses

comparatives de différents logiciels d'annotation automatique. Les évaluations quantitatives ont révélé une augmentation de la parole des adultes et des enfants au fil du temps. J'ai également reproduit les résultats des observations de Warlaumont et al. (2014) concernant l'émergence d'une boucle sociale, soulignant le rôle crucial des facteurs sociaux dans le développement du langage.

Enfin, la dernière étude a examiné les interactions entre le traitement phonologique et l'apprentissage associatif de nouveaux mots pendant la période cruciale de l'acquisition de vocabulaire. Deux nouveaux mots constituant une paire minimale ont été présentés aux enfants dans une tâche de "Looking-While-Listening". Les temps de regard moyens ont été analysés à partir de mesures oculométriques, révélant que les enfants n'ont pas réussi à apprendre les nouveaux mots. Ce résultat inattendu a suscité une discussion sur la trajectoire développementale des mécanismes d'apprentissage de nouveaux mots, en accord avec les modèles antérieurs de développement du langage.

En résumé, cette thèse apporte des informations précieuses sur la nature multiple des processus impliqués dans l'acquisition précoce du vocabulaire, en mettant l'accent sur l'interaction entre le traitement neuronal, les facteurs environnementaux et le développement cognitif dans l'apprentissage du langage dans les deux premières années de vie.

Mots clés : développement du langage, longitudinal, perception de la parole, électrophysiologie, interactions parents-enfants, LENA, apprentissage de nouveaux mots, oculométrie.

Abstract and keywords

Language acquisition is one of the most remarkable feats of human cognition. Infants master the complexities of their native language in just a few years, in a seemingly effortless manner. Despite its apparent ease, understanding this process is challenging due to the interplay of numerous factors, including neural, environmental, and cognitive influences. In this context, multimodal integrative studies are essential to reach a more comprehensive understanding of how these factors interact. Additionally, the implementation of longitudinal designs allows for tracking intra-individual changes over time and tackling causal relationships. My thesis adopts this approach, combining electrophysiological, environmental, and behavioral data to explore the neural correlates of speech perception, the role of the auditory environment, and associative word learning abilities in predicting vocabulary size during the first two years of life. Through this integrated approach, my work contributes to a better understanding of the early predictors of language development.

The thesis is divided into three studies. In the first study, I explored early speech perception abilities within the first two years of life. I validated an innovative paradigm to record early subcortical and cortical brain responses in a single session. Exploratory analyses of the frequency-following response (FFR) revealed distinct developmental patterns in temporal and spectral encoding of speech sounds, particularly concerning fundamental frequency and higher harmonics, suggesting maturational changes at the subcortical level. The cortical responses were assessed through a multi-featured oddball paradigm, where mismatch negativity responses (MMRs) were observed, indicating automatic change detection in the auditory cortex. Notably, group comparisons indicated differences in phonetic encoding, aligning with the perceptual narrowing theory, while also considering the implications of phonological categories on infants' speech perception.

The second study aimed at characterizing the daily auditory environments of infants, using the LENA (Language ENvironment Analysis) system. I explored a new LENA corpus collected in French households, enhancing the existing literature on non-English datasets. The reliability of LENA's automatic annotation software on French data was established via comparative analyses of different automatic annotation softwares. Quan-

titative assessments revealed an increase in both adult and infant speech over time. I also replicated findings from Warlaumont et al. (2014)'s observations regarding the emergence of a social loop, underscoring the critical role of social factors in language development.

Finally, the last study investigated the interactions between phonological processing and associative word learning during the crucial vocabulary spurt period. Infants were presented with two new words constituting a minimal pair, in a Looking-While-Listening paradigm. Mean looking times were analyzed from eye-tracking measures, and revealed that infants did not successfully learn the new words. This unexpected result brought up a discussion on the developmental trajectory of word learning mechanisms, in line with previous models of language development.

Overall, this thesis contributes valuable insights into the multifaceted nature of early vocabulary acquisition, emphasizing the interplay between neural processing, environmental factors, and cognitive development in shaping language learning during infancy.

Keywords: language development, longitudinal, speech perception, electrophysiology, parents-children interactions, LENA, words learning, eye-tracking.

Remerciements

First, I would like to express my sincere thanks to the members of the jury, Dr. Lauriane Cabrera, Prof. Ruth de Diego, Prof. Judit Gervain, Prof. Claudia Männel, and Dr. Johannes Ziegler, for their valuable time and feedback on my thesis. I am very grateful for their willingness to engage with my research and for bringing their expertise to foster constructive and enriching discussions around my work.

Merci au Centre National de la Recherche Scientifique (CNRS) d'avoir financé mes trois premières années de thèse. Merci à l'Institute of Language, Communication and the Brain (ILCB) d'avoir financé ma quatrième année ainsi qu'un contrat d'assistante de recherche qui m'ont permis de terminer ce travail dans de bonnes conditions. Merci au Laboratoire Parole et Langage de m'avoir accueillie dans une atmosphère propice à la recherche et aux connexions humaines.

Naturellement, j'aimerais exprimer toute ma gratitude à mes directeurs de thèse, Clément François et Laurent Prévot. Ce projet est né dans l'esprit de Clément, et a pris un nouveau chemin grâce à ses échanges avec Laurent, puis à travers ma contribution durant ces 4 dernières années.
Clément, merci de m'avoir fait confiance pour prendre les rênes de ce projet qui te tenait tant à cœur. Merci d'avoir été présent autant scientifiquement qu'humainement, d'avoir supporté mes marées hautes comme mes marées basses, de m'avoir aussi bien comprise et accompagnée. Merci pour tout ce que tu m'as appris et pour tout le temps que tu m'as accordé. Merci de m'avoir sortie de ma zone de confort régulièrement pour me faire progresser et grandir. Ton honnêteté, ton empathie et ton soutien m'ont vraiment aidée à ne pas lâcher, alors merci pour tout cela.
Laurent, merci de m'avoir envoyé ce mail, en mai 2020, qui contenait l'annonce du projet de thèse. Merci d'avoir vu en moi la candidate idéale, et d'y avoir cru jusqu'au bout. Merci d'avoir assuré un suivi régulier malgré tes nombreux autres engagements scientifiques et institutionnels. Et merci pour ton immense bienveillance et ton soutien. Avant de commencer ma thèse, on m'avait dit qu'il valait mieux choisir des bons superviseurs plutôt qu'un sujet intéressant : moi, j'ai eu le combo gagnant.

Un grand merci aux crèches Babilou de la région Sud-Est. Sans elles, ce projet n'aurait

pas pu exister tel qu'il est. Je remercie particulièrement les directeurs et directrices pour leur aide inestimable, leur adaptation aux imprévus et leur investissement dans ce projet : Mickael, Jennifer, Christine, Claire. Ce fut un vrai plaisir de collaborer avec vous. Merci, bien sûr, au personnel des crèches sans qui les nombreux enregistrements n'auraient pas été possibles. Pardonnez-moi de ne pas citer tous les noms, mais sachez que votre participation à ce travail a été extrêmement précieuse.

Merci aux familles pour leurs participations nombreuses et essentielles. Merci aux enfants d'avoir accepté tous ces enregistrements dans le jeu et la bonne humeur. Merci aux parents pour votre confiance, votre investissement, et les échanges enrichissants que nous avons eus ensemble.

Merci à toutes les personnes qui m'ont aidée sur les plans technique et logistique :
» Thierry Legou : j'en ai passé du temps dans ton atelier ! Malgré toutes les péripéties avec ce fichu câble, tu n'as pas lâché l'affaire et m'a aidée jusqu'au bout : un grand merci pour ça. J'ai également découvert en toi une personne profondément gentille et positive, alors merci pour tous ces moments passés ensemble.
» Anne-Sophie Dubarry : avec toi, c'était l'aventure Matlab / GitHub ! J'ai adoré apprendre et travailler avec toi, merci de m'avoir partagé autant de connaissances et de savoir-faire. Merci pour ton aide et ton investissement sans faille malgré ton changement de labo en cours de projet et ton emploi du temps bien chargé. Merci pour toutes ces heures passées devant le code, jusqu'à en devenir zinzines parfois !
» Franziska Geringswald : sans toi il n'y avait pas d'analyses eye-tracking. Merci pour ton aide précieuse. Merci d'avoir pris le temps de m'expliquer toutes ces notions, surtout en stats, sans jamais avoir jugé mon niveau de connaissances ridicule dans le domaine !
» Christelle Zielinski : c'est grâce à toi que les premières analyses de mes données eye-tracking ont vu le jour. Tu as passé la main à Franziska en cours, mais c'est toi qui as initié ce travail et as permis à ces données d'en arriver là où elles sont. Un grand merci pour ça. Et merci pour ton sens de l'humour inégalable !
» Stéphane Rauzy : merci d'avoir contribué à relever mon niveau en statistiques, c'était vraiment nécessaire ! Et merci pour ta bonne humeur et ton soutien à chaque fois que l'on se croisait (c'est-à-dire très souvent).
» Fabrice Cauchard : merci pour ton aide lors de mon problème technique avec l'eye-tracker, tu m'as vraiment sauvé la mise !
» Merci à Seb, Cyril, Sabrina, Carine, Alain, Gilles et Claudia pour votre support technique (et moral) tout au long de ma thèse, qui m'ont beaucoup aidée.

Merci aux stagiaires, Lise Tonellotto et Héloïse Didier, pour leur participation active à mon projet. Lise, merci pour la création de la tâche eye-tracking, et pour les premières passations de la tâche. Héloïse, merci pour ton investissement et ta motivation admirables.
Merci à Mathilde Gaujard : tu as tout simplement sauvé ma dernière année de thèse. Ta présence sur ce projet a été un grand soulagement pour moi à une période où tout s'accumulait. Un grand merci pour ton aide précieuse !

Merci à tous les membres du laboratoire avec qui j'ai eu l'occasion d'échanger, que ce soit sur des sujets scientifiques ou du quotidien. Merci à Serge Pinto, Sophie Herment, Christine Meunier, Amandine Michelas, Philippe Blache, Béatrice Priego-Valverde, Roxane Bertrand, Chotiga Patamadilok, avec qui j'ai eu de nombreuses interactions, et qui ont contribué à me sentir intégrée et à ma place au LPL. Merci également à Nadéra Bureau, Stéphanie Lhuillier-Desous, et Annelise Coquillon : au-delà de votre aide administrative, ce sont surtout des super moments partagés avec vous.

J'aimerais adresser un mot tout particulier à Cristel Portes, qui a été une personne déterminante dans mon parcours universitaire. Cristel, c'est toi qui m'as appris l'existence du MasCo et qui m'a fait prendre un chemin complètement différent de celui auquel j'avais pensé. En L3, lors d'une discussion remplie de bienveillance et de respect mutuel, tu m'as permis d'agrandir mon champ des possibles, et c'est en partie grâce à toi que j'en suis arrivée ici. Alors un grand merci pour cela.

Merci également à tous les chercheurs, doctorants et posdoctorants rencontrés lors d'évènements scientifiques. A particular thought to Carles Escera and all his Barecelona team, who I met several times and with whom I spent great moments!

Evidemment, merci aux doctorants et postdocs du LPL, sans qui cette aventure aurait été bien compliquée. Je vais essayer de n'oublier personne, mais après 4 ans passés ici, il y a du monde à remercier ! Merci à l'équipe des anciens de m'avoir accueillie et intégrée : Rémi, Axel, Giusy, Chiara, Francesca, Mary, Marie, Marie-Charlotte. Merci aux un peu moins anciens, qui ont participé à l'énergie du groupe et aux after-work bien mérités : Dasha, Kübra, Lena, Sneza, Xenia, YenLin, Salvatore, Lise. Et merci à la relève, qui continue de faire vivre l'open space et les repas du midi : Marine, Giulia, Leela, Clara.

Mentions spéciales :

» Salomé, une rencontre incroyable. Merci pour ton énergie et ta folie.
» Airelle et Emilie, autrement dit Tic et Tac : merci pour votre bonté, vos oreilles bienveillantes et votre coolitude.
» Lydia, une personne en or, tu as égayé un bon nombre de mes journées.
» Anastasia, une personne adorable et attachante : je passe bientôt te voir en Espagne !
» Auriane et Outhmane, aka la team Karim : sans vous, cette thèse n'aurait pas la même saveur. Vous avez été mes compagnons de route, mon soutien moral, mes acolytes de folie, mes partenaires de potins. Les copains, merci pour ces années passées ensemble !

Enfin, j'aimerais remercier mes proches. Ceux qui ne comprennent pas forcément ce que je fais et qui pourtant m'ont soutenue tout au long de ces 4 années, et depuis bien plus longtemps encore.
Merci à ma team du collège/lycée qui est là depuis toujours : Thomas, promis tu n'entendras plus jamais le mot *passation* !! Merci pour ton soutien infaillible depuis toutes ces années, peu importe ce que je fais. Mathis, merci de m'avoir aidée à surmonter ce challenge en partageant nos expériences de doctorants (et en allant se perdre dans la montagne aussi). Du cours d'allemand en 5ème à docteurs, qui l'aurait cru ? Louise, merci pour nos soirées mamies et nos discussions sans fins, les meilleur anti-stress ! Elise, Tiffany, merci de m'avoir écoutée me plaindre pendant des heures et d'avoir été un énorme soutien moral.
Merci à Amandine d'être restée dans ma vie malgré mon emploi du temps de ministre : promis après la thèse on se voit un peu plus qu'une fois tous les 3 mois !

Merci à ma famille, et particulièrement Papa, Maman, et Elise. Grâce à vous, j'ai pu m'entrainer à vulgariser mon travail. Mais surtout, j'ai trouvé du soutien quand j'en avais besoin, et même quand je ne pensais pas en avoir besoin. Merci de créer ce cocon familial depuis toujours, qui a été d'autant plus nécessaire ces dernières années.

Et pour finir, merci à mon compagnon de vie, Gaël. Merci de m'avoir supportée (dans tous les sens du terme), d'avoir accepté mon emploi du temps parfois absurde, de m'avoir fait prendre du recul quand j'en avais besoin, d'avoir géré mes périodes de stress, d'avoir essayé de comprendre ce que je faisais, d'avoir célébré mes petites victoires, enfin, la liste est longue. Tout simplement, merci d'avoir été et d'être toujours là.

Estelle.

Contents

Affidavit . . . 2
Liste de publications et participation aux conférences . . . 3
Résumé et mots clés . . . 5
Abstract and keywords . . . 7
Remerciements . . . 9
List of Figures . . . 18
List of Tables . . . 20
List of acronyms . . . 21
Glossary . . . 25
General introduction . . . 27
1 Language development in infancy: history . . . 28
1.1 First reports: diary studies and influences of behaviorism . . . 28
1.2 The nature vs. nurture debate . . . 29
1.3 The emergence of cognitive science . . . 30
2 Developmental perspective . . . 30
3 Predicting language abilities: implications . . . 31
4 General thesis overview . . . 32
Chapter 1 Theoretical frameworks . . . 35
1 Early speech perception . . . 37
1.1 Maturation of the auditory system . . . 37
1.1.1 The human auditory system: development . . . 37
1.1.2 Critical periods . . . 38
1.1.3 The first hearing experiences . . . 39
1.2 From universal to language-specific speech perception . . . 40
1.2.1 An initial bias for speech . . . 40
1.2.2 The citizen of the world . . . 40
1.2.3 Vowel perception in infancy . . . 41
1.2.4 Change in consonant perception . . . 43
1.2.5 Perceptual narrowing . . . 44

1.3 Theories of categorical speech perception . . . 45
1.3.1 Motor theories . . . 46
1.3.2 The Perceptual Assimilation Model . . . 48
1.3.3 The Native Language Magnet theories . . . 49
1.4 Electrophysiological responses to speech sounds . . . 52
1.4.1 Cortical brain responses to speech sounds . . . 52
1.4.2 Subcortical brain responses to speech sounds . . . 55
2 Auditory environment and language development . . . 59
2.1 Auditory inputs in early infancy . . . 59
2.1.1 Input from the caregivers . . . 59
2.1.2 Infant-Directed Speech . . . 60
2.1.3 Non linguistic inputs . . . 61
2.2 Daylong recordings in infancy . . . 62
2.2.1 A technological advance . . . 62
2.2.2 Daylong recording studies . . . 64
2.3 The daily auditory environment as a predictor for language skills 66
3 Vocabulary acquisition in early infancy . . . 68
3.1 Word learning in infancy: challenges . . . 68
3.1.1 Speech segmentation . . . 68
3.1.2 Word-to-world mapping . . . 69
3.1.3 Generalization and categorization . . . 70
3.2 First word acquisition: overcoming the challenges . . . 71
3.2.1 Internal factors . . . 71
3.2.2 Capitalizing on external cues: the Emergentist Coalition Model . . . 74
3.3 How to study vocabulary acquisition in infancy? . . . 77
3.3.1 Focusing on associative word learning . . . 77
3.3.2 Lexical-phonological aspect of word learning . . . 79
3.3.3 Paradigms and standardized scales . . . 82
4 Chapter's summary . . . 89
Chapter 2 General aims and hypotheses . . . 91
1 Objectives of the research . . . 91
2 Main hypotheses . . . 93
2.1 Cortical and subcortical brain responses to speech sounds in early infancy . . . 93
2.2 Daily auditory environment . . . 94

2.3 Minimal pair associative word learning . . . 94
Chapter 3 Working with infants, a challenge . . . 97
1 Recruitment and logistical aspects . . . 97
1.1 Collaboration with the Babilou group . . . 97
1.1.1 The beginning of the collaboration . . . 97
1.1.2 Description of the daycare centers . . . 98
1.1.3 Recruitment procedure . . . 99
1.1.4 Benefits and difficulties . . . 100
1.2 Le Labo des Minots: a new Babylab in Marseille . . . 100
2 Electrophysiological recordings in infants: challenges and perspectives 102
Chapter 4 Development of cortical and subcortical responses to non-native speech sounds during infancy . . . 103
1 Introduction . . . 104
2 Methods . . . 106
2.1 Participants . . . 106
2.2 Stimuli . . . 106
2.3 Procedure . . . 107
2.4 EEG data processing . . . 109
2.4.1 Subcortical FFRs . . . 109
2.4.2 Cortical ERPs . . . 112
2.5 Statistical analyses . . . 113
2.5.1 Subcortical FFRs . . . 114
2.5.2 Cortical ERPs . . . 115
3 Results . . . 116
3.1 Subcortical FFRs . . . 116
3.1.1 Group comparison . . . 116
3.1.2 Longitudinal sample . . . 117
3.1.3 Age-based regressions . . . 119
3.2 Cortical ERPs . . . 123
3.2.1 Cluster-based permutations. . . . 123
3.2.2 Group comparison . . . 123
3.2.3 Longitudinal sample . . . 127
4 Discussion . . . 133
4.1 Recording simultaneous cortical and subcortical responses in infants: feasibility . . . 133
4.2 Subcortical encoding of speech sounds . . . 134

4.3 Cortical responses to speech sounds . . . 138
4.4 Limitations and future directions . . . 142
5 Conclusion . . . 142
Chapter 5 Investigation daily auditory environment in infants with LENA . 145
1 Introduction . . . 146
2 Related Work . . . 147
2.1 Experimental studies . . . 147
2.2 Validation studies . . . 148
3 Rationale . . . 149
3.1 Participants . . . 149
3.2 Procedure . . . 150
4 Tools and Methods . . . 150
4.1 LENA . . . 151
4.2 VTC and VCM . . . 154
5 Dataset . . . 154
6 Investigating the dataset . . . 155
6.1 Testing age . . . 155
6.2 Interactional metrics . . . 157
7 Discussion . . . 159
7.1 Dataset investigation: linguistic input and contingencies . . . 159
8 Limitations . . . 161
9 Conclusion . . . 162
Chapter 6 Minimal pair associative word learning in 18- and 24-month-olds 163
1 Introduction . . . 163
2 Methods . . . 167
2.1 Participants . . . 167
2.2 Associative word learning task . . . 167
2.3 Vocabulary size and language and cognitive development assessment . . . 171
2.4 Data analysis . . . 172
3 Results . . . 174
4 Discussion . . . 184
4.1 Dataset investigation . . . 184
4.2 Focus on the consonant contrast . . . 187
4.3 The vocabulary spurt: a critical period for word-learning mechanisms? . . . 189

4.4 Links with language outcomes 191
4.5 Limitations and future directions 192
General discussion 195
1 Focus on the main results 195
2 From speech perception to vocabulary acquisition 197
3 Limitations and future directions 199
Bibliography 201

ANNEXES **244**

A Recruitment documents 245
B List of data collected for each participant 257
C Challenges and new perspectives of developmental cognitive EEG studies: full text 259
D Daily auditory environments in French-speaking infants: A longitudinal dataset: full text 275

List of Figures

0.1 Universal developmental trajectory of receptive and expressive language abilities in the first year of life . . . 27

1.1 The human auditory system: main structures . . . 38
1.2 Roles of experience in perceptual narrowing and associated theories . . . 45
1.3 Illustration of the motor theory . . . 47
1.4 Representation of the magnet effect (Kuhl, 2000) . . . 50
1.5 Mismatch response in infancy as a function of change type and age 55
1.6 FFRs to different syllables . . . 56
1.7 Pauses in speech do not mark word boundaries . . . 69
1.8 The constraints/principles of word learning . . . 72
1.9 The WOLVES model (Samuelson, 2021) . . . 73
1.10 The emergentist coallition model . . . 75
1.11 Eye-tracking set up in a nursery. . . . 86

4.1 Brain responses to speech sounds: Stimuli and experimental design . . . 108
4.2 FFR: Temporal and spectral representations by group . . . 117
4.3 FFR: Violin plots of the SNR_{FFR} by group . . . 118
4.4 FFR: Temporal displays at five assessment points . . . 120
4.5 FFR: Temporal display at five assessment points . . . 121
4.6 FFR: Correlation matrix . . . 122
4.7 FFR: Correlation between pitch tracking and age . . . 122
4.8 EEG: Grand averaged responses to standard /da/ . . . 124
4.9 ERPs: Difference waves to both deviants, group comparison . . . 125
4.10 EEG: Grand averaged responses to standard /da/ . . . 128
4.11 ERPs: Difference waves to both deviants, all assessment points . 129
4.12 ERPs: Patterns of the mean amplitude of the difference waves, all assessment points . . . 132

5.1 LENA: Recording materials . . . 152
5.2 LENA: Mean speaking ratios . . . 155

5.3 LENA: Speaking time as a function of gender, SES, linguistic background, and age . . . 156
5.4 LENA: Speech vs. Non-speech (VCM) . . . 157
5.5 LENA: Child/Adult speech contingencies . . . 158
5.6 LENA: Speech contingencies (VTC) . . . 158
5.7 LENA: Replication of Warlaumont's (2014) results . . . 160

6.1 Word learning: visual stimuli . . . 168
6.2 Word learning: time courses . . . 170
6.3 Word learning results: Mean proportion of fixation on screen during familiarization phase . . . 175
6.4 Word learning results: Habituation criterion . . . 176
6.5 Word learning results: Look at screen during test . . . 177
6.6 Word learning results: Mean proportion of fixation to target during test phase . . . 178
6.7 Word learning results: Object preference during test . . . 179
6.8 Word learning results: Mean proportion of fixation to target by word occurrence . . . 180
6.9 Word learning results: Mean proportion of fixation to target by word type . . . 181
6.10 Word learning results: Timecourses . . . 183
6.11 Word learning results: Correlations . . . 184

List of Tables

4.1 FFR: Statistics for time and frequency domain measures 119
4.2 ERPs: Statistics for standard response, group comparison 124
4.3 ERPs: Difference wave statistics, group comparison 126
4.4 ERPs: Statistics for standard response, all assessment points . . . 128
4.5 ERPs: Difference wave statistics, all assessment points 130

5.1 LENA: Data set description 154
5.2 LENA: Selected children 155

6.1 Word learning: summary statistics of mean looking times. 179

List of Acronyms

ABR

Auditory Brainstem Response. 39, 55, 56

ADEX

Advanced Data Extractor. 153

ADS

Adult-Directed Speech. 60

AOI

area of interest. 170, 172–174, 177, 181

ASD

autism spectrum disorder. 64, 65

AWC

Adult Word Count. 64–67, 152, 153

BSID

Bayley Scales of Infant Development. 87, 171, 174, 182, 184, 191, 197

CHN

CHild Near. 152, 153

CTC

Conversational Turn Count. 64–67, 153

CVC

Child Vocalization Count. 64, 67, 152, 153

CXN

Other Child Near. 152, 153

DLP

digital language processor. 151, 152

DRT

Direct Realist Theory. 47, 48

EEG

electroencephalography / electroencephalogram. 52, 55, 56, 101, 102

ERP

Event-related potential. 43, 52, 104–106, 108, 112, 113, 115, 123, 138, 196

F0

Fundamental frequency. 54, 56, 57, 105, 111, 112, 116–118, 121, 135, 136

FAN

Female Adult Near. 152, 153

FFR

Frequency-Following Response. 56–58, 89, 105–110, 116, 117, 119, 134, 135, 196

FFT

Fast-Fourrier Transform. 111

FSR

Fast Stimulation Rate. 109, 110

FUZ

Fuzzy. 153

IDS

Infant-Directed Speech. 60, 61, 67, 78, 168, 171, 189, 190

IPLP

Intermodal Preferential Looking Paradigm. 83, 169

IQR

interquartile range. 116

ISI

Inter-Stimulus Interval. 133, 142

LENA

Language ENvironment Analysis. 63–67, 89, 98, 101, 145, 151–153

LWL

Looking-While-Listening. 83, 84, 165, 184, 185

MAN

Male Adult Near. 152, 153

MCDI

Mac-Arthur Bates Communicative Development Inventories. 86, 87, 171, 174, 182, 184, 191, 197

MMN

Mismatch Negativity. 52, 53

MMR

Mismatch Response. 53–55, 89, 104–106, 138–142, 196

MT

Motor Theory. 46–48

NLM

Native Language Magnet. 49, 50, 141

NLM-e

Native Language Magnet extended. 51

NLNC

Native Language Neural Commitment. 51

NON

NOise Near. 153

OLN

OverLap Near. 153

PAM

Perceptual Assimilation Model. 48, 141

Q1

first quartile. 116

RMS

Root Mean Square. 110, 111

SES

socio-economic status. 60, 66, 67, 198

SIL

SILence. 153

SNR

Signal-to-Noise Ratio. 111, 134–136

SOA

Stimulus Onset Asyncrony. 109

SSR

Slow Stimulation Rate. 109, 112, 113

TVN

TV Near. 153

VOT

Voice Onset Time. 41, 43, 44, 53, 54, 105, 107, 166, 188

WL

Word Learning. 169, 171, 173, 174, 176–178, 180–187, 191

WOLVES

Word Object Learning via Visual Exploration in Space. 73, 74

WR

Word Recognition. 169, 171, 173, 174, 176, 177, 180, 182, 184–186, 191

Glossary

allophone

Phonetic variation of a phoneme. Two allophones are two different pronunciations of the same phoneme: in a given language, allophones can be used interchangeably without changing the meaning of a word. 41

Auditory Brainstem Response

Electrophysiological evoked potential in response to sound that originates from the brainstem. 39

category goodness

"Category goodness is a matter of degree, where some members are perceived as better exemplars, more representative or prototypic, than others (Rosch, 1975)." (Kuhl, 1991). 49

distractor

In a Looking-While-Listening procedure, the distractor is the object that is not named during a test trial. 83, 165

ERP

A "depiction of the changes in scalp-recorded voltage over time that reflect the sensory, cognitive, affective, and motor processes elicited by a stimulus." (Kappenman & Luck, 2011). 43

eye-tracking

A technological tool and process that detects the human's pupil to measure the position and movements of the eyes on a screen. 84, 85, 165

familiarization

In the context of a Looking-While-Listening task, a familiarization is a learning phase that involves a fixed number of trials for all participants. 84, 85, 168, 169, 171, 172, 174, 175, 185

Fast-Fourier Transform

A mathematical transformation of a signal from the time domain to the frequency domain. The Fast-Fourier Transform decomposes the signal into a sequence of its constituing frequencies. 56

formant

A peak of high acoustic energy at specific frequencies in speech sounds. The value and trajectory of the formants determine the nature of the speech sound, particularly for vowels. 42, 46, 53, 54, 105, 107, 108

habituation

In the context of a Looking-While-Listening task, a habituation is a learning phase in which the number of trials depends on the participant's looking time. It relies on the idea that a decrease in the proportion of fixation to the screen indicates that the participant has learned. 78, 82, 84, 85, 164, 169, 185

minimal pair

Two words constitute a minimal pair when they differ by only one feature (most of the time, by one phoneme). 79–82, 88, 89, 164–166

Mismatch Response

In the context of an oddball paradigm, a Mismatch Response is evoked by an automatic change detection in the series of auditory stimuli. 42, 44

oddball paradigm

A paradigm that involves the auditory presentation of a deviant stimuli in series of identical stimuli. The oddball paradigm has been extensively used in developmental research to explore native and nonnative speech sound discrimination. 43, 52, 55, 141

perceptual narrowing

Refers to the loss of sensitivity to nonnative contrasts and increase in native speech sound discrimination in infancy. 44, 45, 48, 49, 104, 143, 196

phoneme

Abstract distinctive unit of a given language. 40–42, 46, 48, 164

target

In a Looking-While-Listening procedure, the target is the object that is named during a test trial. 83, 165

vocabulary spurt

A period of sudden increase in the pace of vocabulary growth in infants' development, that occur between 18 and 20 months on average. 72, 79, 165

General introduction

Human language is part of the most complex and elaborated communication systems. It constitutes a direct window to our minds, allowing us to describe our environment, share ideas, express emotions, and most importantly, create social bonds with our peers. As social beings, mastering our native(s) language(s) is a crucial step in life but also very challenging. Indeed, young learners must acquire all aspects of their native language(s) including the phonological repertoire, the lexicon, the morphosyntactic constraints, the prosodic features, the pragmatic rules, and the social conventions, while also developing at the motor, social, cognitive, and emotional levels. Nevertheless, infants possess incredible resources that enable them to produce their first words, typically between the first and the second year of life. Although a universal developmental trajectory of vocabulary acquisition has been described in early infancy (see Figure 0.1), a great variability between infants is reported (Bates et al., 1994, 1995; Frank et al., 2021).

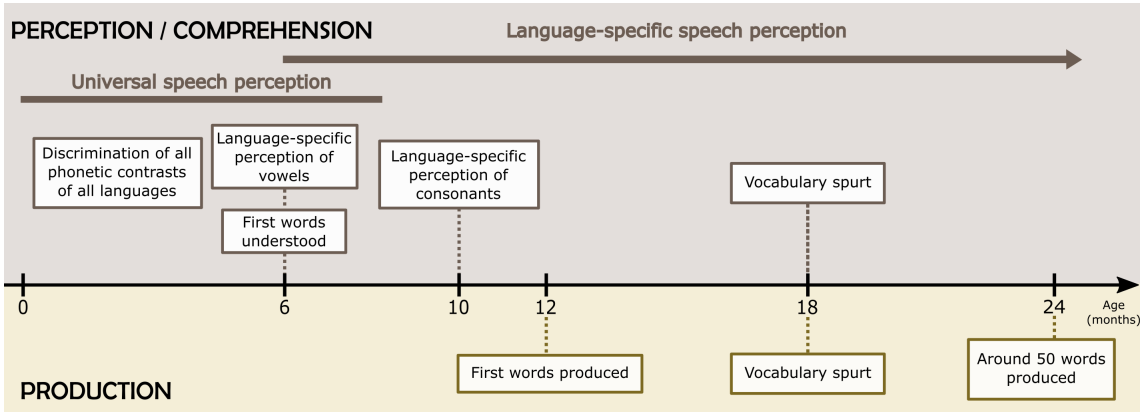


Figure 0.1. – *inspired by Kuhl, 2004.* Universal developmental trajectory of receptive and expressive language abilities in the first year of life.

This inter-individual variability has long been set aside by the many cross-sectional studies that revealed averaged developmental patterns and critical periods of early language development. Longitudinal studies have recently gained interest in the developmental community, enabling for grasping intra-individual changes through time. After decades of research, the methodologies and theoretical insights have evolved into more multimodal studies, involving interdisciplinary approaches that came along with the advent of cognitive sciences. My PhD work comes within this current scientific

approach, at the crossroads of linguistics, neurosciences, and developmental psychology. Specifically, I created a new corpus of multimodal data gathered longitudinally, involving electrophysiological measures of speech perception, daylong recordings of auditory and linguistic environments, and behavioral measures of associative word learning abilities using eye-tracking. This data set was completed by neurodevelopmental scores of language and cognitive development at 24 months, to assess the complex interactions between early predictors of language development and vocabulary size in the first two years of life.

1. Language development in infancy: history

1.1. First reports: diary studies and influences of behaviorism

Language acquisition in babies is one of the most fascinating processes of human cognition, yet very complex and tricky to study. The first reports on early language development can be traced back to the 19th century, when scientists started observing their own infants and noting everything in diaries (Darwin, 1877; Leopold, 1939; Stern & Stern, 1907). These first observations constituted detailed and rich datasets that are now considered valuable single case studies. During the first half of the 20th century, the emergence of behaviorism paved the way for more systematic studies involving empirical methods, larger samples of participants, and detailed data analyses. Behaviorist theories postulated that learning occurs through repeated behaviors conditioned or reinforced by experience and external input. This hypothesis came from the Russian theorist Ivan Petrovich Pavlov who noticed that his dogs salivated even before seeing their food, triggered by the sound of his footsteps (Pavlov, 1927). He concluded that they had been conditioned to salivate in response to specific inputs that they had associated with the food. Later, systematic studies in other species (Eibl-Eibesfeldt & Kramer, 1958; Guthrie, 1930; Skinner, 1938), then in humans (Skinner, 1953; Watson, 1913), reinforced this idea that experience with the environment is sufficient to understand and explain learning processes.

Behaviorism influenced language development research at both methodological and theoretical levels. Indeed, behavioral measures were dominant and studies used large cross-sectional populations to highlight the role of infants' environment on language acquisition (Miller & Dollard, 1941; Skinner, 1957). To the same extent, the first large-scale normative studies appeared in the mid-20th century. The main goal was to describe observable behaviors such as speech sound production or discrimination abilities, vo-

cabulary size, or sentence length (Agheyisi & Fishman, 1970; McCarthy, 1946; Ridgeway et al., 1985; Templin, 1957). At that time, there were no precise theoretical questions about language development. Rather, the main goal was to establish norms of linguistic abilities in early infancy.

1.2. The nature vs. nurture debate

The second half of the 20th century marks the beginnings of major theoretical conflicts, with the advent of innate and then constructivist ideologies. The *nature versus nurture* debate started in the 1960s when Noam Chomsky shared his innate theory of language. As opposed to behaviorism which emphasizes the role of experience and external factors, Chomsky's theory holds that all humans are born with an innate and biologically determined ability for language. Although he did not reject the influence of environment on cognitive development, he believed that infants could learn their native language(s) even without any external input. He introduced the notion of innate grammar, a genetically predetermined set of rules and learning capacities that allow infants to acquire language. Chomsky revolutionized linguistics by claiming that human behavior observations are no longer relevant to understanding linguistic processes. Instead, he defended the idea that mental grammar, namely what is happening in speakers' minds, was the only way to explain linguistic behaviors. Moreover, he asserted that the linguistic inputs infants are exposed to are not sufficient to learn their native language in such a short period of time: this is the *poverty of the stimulus* argument. He noticed that infants were exposed to only a small and finite subset of all possible syntactic combinations. Yet, at some point, infants were able to produce complex grammatical sentences that they never heard before. To account for this phenomenon, Chomsky assessed the existence of a *universal grammar*, a set of innate grammatical constraints that every speaker possesses. Chomsky's position opened up new theoretical considerations, especially on infants' development of grammatical knowledge (Brown, 1973).

In a more nuanced view, constructivist (or interactionist) theories also posit the existence of innate characteristics that allow infants to develop language rapidly, but emphasize the co-construction of linguistic knowledge through interactions with the environment. Associated with Jean Piaget's conception of development, *constructivism* argues that language acquisition is a co-construction process in which infants actively engage with social and linguistic factors, which bootstrap the preexisting internal dispositions.

1.3. The emergence of cognitive science

Chomsky's work in psychology and linguistics has influenced the cognitivist view of human research. Indeed, in parallel with the nature vs nurture debate, the 1950s saw the emergence of a new interdisciplinary field: cognitive science. At the crossroads of psychology, linguistics, neuroscience, anthropology, artificial intelligence, and philosophy, cognitive science aims at explaining how the human mind works. Studies on language development played a major role in this "cognitive revolution", as it moved the focus of the debate to understanding how the human brain works and generates the observed behaviors. Although it might not be the goal for many researchers who study language development, this new focus on the human brain has led to the emergence of new tools and methodological approaches. Specifically, neurophysiological procedures such as electroencephalography (EEG) and event-related potentials (ERPs) have emerged in developmental populations (Dehaene-Lambertz, 2000a; M. H. Johnson, 1998; Kutas & Hillyard, 1980). These tools have shown to be particularly reliable for capturing the brain's real-time processing of linguistic input like speech sounds or semantic properties, thus offering valuable insights into the neural mechanisms underlying language acquisition. From these investigations on infants' brain responses, different patterns of behavioral and electrophysiological data have been observed, as well as differences between clinical and typical populations (Bates, 1999; Dawson, 1991; Friedrich & Friederici, 2010; Kuhl, 2010). These results oriented the researchers' interest toward the exploration of links between multimodal data, and especially relations between early biomarkers and later language outcomes. Thus, cognitive science, influenced by evolving perspectives on language development, has not only benefited from these advancements but has also contributed significantly to shaping new, integrative approaches for studying language development.

2. Developmental perspective

Along with the emergence of interdisciplinary work on language development and efforts to predict later language outcomes, the question of inter-individual variability arose. To address it, longitudinal designs are often preferred because they enable tracking developmental changes over time. This method implies collecting data in a single cohort of participants with several follow-up recordings, thus eliminating the multiple-cohort effect. Although reliable and particularly relevant for developmental studies, longitudinal designs are not predominant in the literature. Indeed, compared to cross-sectional studies, longitudinal designs are difficult to implement because they are

expensive, very time-consuming, and defy the current "publish or perish" publication system. Despite these constraints, the present research has been conducted from a developmental perspective, involving longitudinal follow-up of infants in their first two years of life.

This work can be associated with the *developmental* approach of language development, although it does not ignore the *learnability approach*. The *developmental approach* focuses on how language learning abilities develop and evolve over time and aims at providing a developmental trajectory of language acquisition, from the first stage of speech perception and production to the production of the first words and sentences. This approach can be associated with the *ontogenic* question of Tinbergen (1963), which contributes to understanding animal behaviors by questioning its changes throughout an individual's life. This also relates to the interactionist/constructivist view of language development described above. Specifically, it questions the interactions between biological predispositions, environmental and social factors, and cognitive development. This approach investigates how and when children acquire specific language skills (phonological, syntactic, semantic), relying on naturalistic observation, longitudinal studies, and neurophysiological techniques (EEG, MEG, fNIRS). On the other hand, researchers from the *learnability approach* are rather interested in the theoretical aspects of language development. Closely related to *nativist theories*, this approach supports the idea that language is "learned" thanks to an innate grammatical structure or cognitive constraints. The main research questions focus on understanding the intrinsic mechanisms and the nature of the representations that allow infants to learn their native language(s) despite the constraints and poverty of the input. The perspective of my PhD works is thus embedded in the *developmental approach* via its design, research interest, and methods.

3. Predicting language abilities: implications

By investigating the interactions between the multimodal factors involved in language development, the current work further explores early predictors of vocabulary acquisition. Predicting language abilities is recent but not new, and has several implications that should be mentioned.

From a clinical perspective, early identification of language abilities predictors can help detect language delays or disorders at a time when interventions may be most

effective. Accurate and well-timed assessments are essential for children with known risks, such as those with familial history of language disorders, but also for children who may not present obvious risk factors. Indeed, previous studies have shown strong relations between early electrophysiological (Kuhl et al., 2008), environmental (Cristia et al., 2021a), or behavioral (Tsao et al., 2004) markers and later language outcomes. Targeting the predictors and deciphering those interactions allow for tailored interventions, supporting children's language development and helping clinicians evaluate the efficacy of treatments.
Beyond clinical applications, predicting language abilities in infancy contributes to a better understanding of human cognition. Specifically, it explores the mechanisms underlying language acquisition, distinguishing language-specific processes from broader domain-general cognitive functions. Additionally, this approach can provide insights into how infants acquire language and how linguistic abilities intersect with other cognitive functions such as memory and attention.
Finally, early predictions offer valuable feedback for parents, providing them with robust knowledge for adapting their infant's auditory environment during a critical period for brain development. Thus, early prediction serves not only to improve the efficiency of clinical interventions but also to enrich our understanding of cognitive development and offer practical guidance for caregivers.

My PhD work fits within the context outlined in this general introduction. Next, I provide an overview of the chapters' content to offer a comprehensive guide through the manuscript.

4. General thesis overview

Chapter 1 provides the theoretical background of this thesis and introduces key notions for the subsequent chapters. The chapter is divided into three sections that correspond to the three main studies developed in chapters 4, 5, and 6. The first section describes the developmental milestones of the auditory system's maturation and speech perception abilities. Key theoretical models of categorical perception and methodological paradigms involving electrophysiological data are reviewed. The second section is dedicated to the auditory environment of infants' first months of life. The main inputs of infants' soundscape are characterized, and daylong recording studies' main findings are reported. Finally, the third section encompasses theories and methodological frameworks from previous works on vocabulary acquisition. The main challenges of

early word learning are exposed, before digging into some models that attempted to explain the interactions between external cues and internal processes. The section ends with a short review of the main approaches and paradigms used in early word learning research.

Chapter 2 introduces the general aims and hypotheses of the project. Specifically, the main goal of this project was to investigate the neural biomarkers of speech perception, the daily auditory environment, and the associative word-learning abilities of infants as predictors of vocabulary size in the first two years of life.

Chapter 3 discusses methodological aspects related to developmental populations. Specifically, I introduce the challenges faced during the thesis and the choices I made to overcome them. I provide details about the recruitment procedures and the collaboration with daycare centers.

Chapter 4 addresses the question of early speech perception abilities in the first two years of life. An innovative paradigm that allowed recording simultaneous subcortical and cortical brain responses to speech sounds is described and tested in infants aged 6 to 24 months. I show substantial changes in the spectro-temporal encoding of speech sounds between 6 and 24 months of age at the subcortical level, revealing different developmental patterns in the time and frequency domains. Cortical responses gathered during the multi-featured oddball paradigm reveal a change in the discrimination of synthetic non-native speech sounds, in line with the *perceptual narrowing* theory. Finally, I explore the relation between brain responses to speech sounds and later language outcomes.

Chapter 5 explores the daily auditory environment of infants. I introduce the corpus of daylong auditory data gathered during this PhD using the LENA system. I show an interaction between the overall proportion of speech and infants' age, in line with previous literature. Importantly, I replicate results from previous work on speech contingencies, providing further evidence for the existence of a social loop between infants and caregivers. These results further emphasize the role of external input on infants' language development.

Chapter 6 investigates minimal pair associative word learning abilities at 18 and 24 months. A Looking-While-Listening paradigm is coupled with eye-tracking to explore

overall looking times as well as fine-grained looking behaviors during the task. I obtained unexpected results indicating that infants did not learn the two new words at both ages. The implications for the developmental trajectory of word learning mechanisms are discussed.

Chapter 7 summarizes the main results and contributions of this four-year work. A discussion of the implications of our findings is provided. Finally, I propose some improvements and future directions.

Chapter 1. Theoretical frameworks

How do infants, starting as mere passive listeners, evolve into masters of language acquisition in a few months, achieving one of the most profound transformations in human development?
Although this question remains unresolved, developmental research has brought some evidence and theoretical insights that contribute to solving this mystery. This chapter aims to provide an overview of the main findings and remaining questions that motivated this PhD work. As mentioned in the General introduction, this thesis falls within a developmental perspective of vocabulary acquisition, from early stages of speech perception to first words' understanding and production. During this doctoral work, I encompassed internal and external factors involved in vocabulary acquisition. Specifically, I focused on (1) neural correlates of speech perception, (2) infants' daily auditory environment, and (3) associative word learning abilities and neurodevelopmental outcomes related to language and cognitive development. The chapter has three sections covering each main topic: 1. the neurodevelopmental stages of early speech perception; 2. the characteristics of infants' daily auditory environment and its role in language development; 3. vocabulary acquisition's main challenges and current theories. Fundamental notions and pioneering studies are introduced to provide a theoretical framework for the following chapters.

Contents

1 Early speech perception 37
1.1 Maturation of the auditory system 37
1.2 From universal to language-specific speech perception 40
1.3 Theories of categorical speech perception 45
1.4 Electrophysiological responses to speech sounds 52
2 Auditory environment and language development 59
2.1 Auditory inputs in early infancy 59
2.2 Daylong recordings in infancy 62
2.3 The daily auditory environment as a predictor for language skills 66
3 Vocabulary acquisition in early infancy 68

3.1 Word learning in infancy: challenges 68
3.2 First word acquisition: overcoming the challenges 71
3.3 How to study vocabulary acquisition in infancy? 77
4 Chapter's summary 89

1. Early speech perception

Language acquisition involves the development of many skills, from domain-general to domain-specific [1] and from low-level perceptual to high-level cognitive processes. Although recruiting common neural bases, speech perception abilities differ from simple auditory perception. Indeed, "perceiving speech is detecting and discerning language-specific kinds of biologically significant events: those generated by communicative intentions of fellow human talkers" (O'Callaghan, 2015, p. 494). Understanding the neural substrates and associated developmental changes of early speech perception is pivotal to unraveling the complex process of language acquisition. In this first section, I describe the development of speech perception abilities in infancy, from the first hearing experiences to later language-specific speech perception. I introduce substantial work and theories that explored the mechanisms of speech perception, and I report recent studies on cortical and subcortical brain responses to speech sounds.

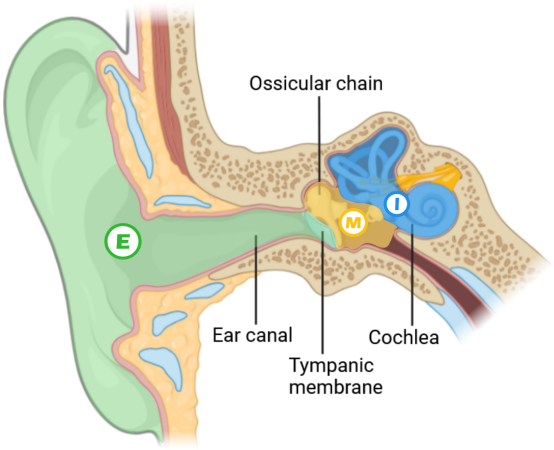
1.1. Maturation of the auditory system

1.1.1. The human auditory system: development

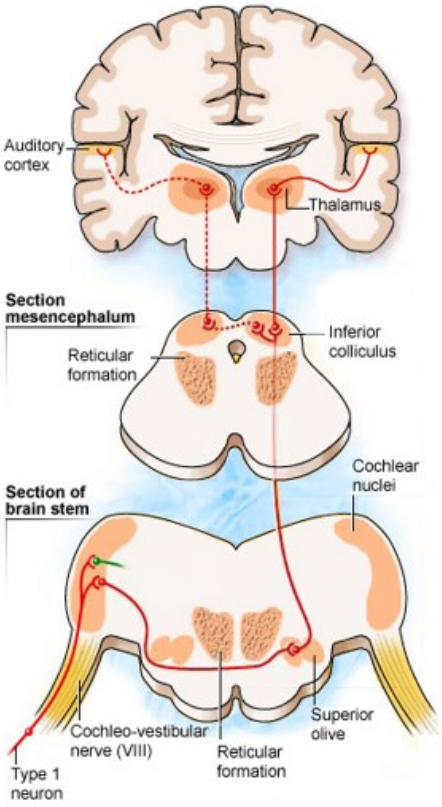
The human auditory system involves several brain structures that operate within a complex hierarchical organization. Sound waves arrive through the peripheral structures, which include the outer, middle, and inner ear (Figure 1.1a). Once the air vibrations have reached the cochlea, they are transformed into bioelectric information that is transferred up to the auditory cortex through the central structures (Figure 1.1b). The brainstem plays a crucial role in the auditory pathway and comprises a set of important relays for neural conduction of auditory information.

The fetal auditory system is the first sensory system to develop in the mother's womb. It is fully functional between the 25th and 29th gestational weeks (Moore & Linthicum, 2007) where the perception of sounds begins through liquidian conduction. During the following weeks and months, intense maturational changes occur in the auditory system. The fast myelination of the auditory pathway is essential to provide the high temporal resolution required for accurate spatial localization and fine-grained speech processing (Litovsky, 2015; Long et al., 2018; Moore & Linthicum, 2007; Ponton et al., 2002). Indeed, impaired temporal processing in infancy and childhood can induce language disorders such as dyslexia (Abrams et al., 2009; Lehongre et al., 2011). The beginning of myelination along the auditory pathway is concurrent with the functional

1. here, "domain" refers to *language*



(a) The auditory system: peripheral structures. E = external (outer) ear, M = middle ear, I = inner ear.



(b) Auditory pathway: the relays of the central structures. Figure from *http://www.cochlea.eu/cerveau-auditif*

Figure 1.1. – Schematic view of the human auditory system: main structures.

maturation of the auditory system at the end of the second trimester of intrauterine life (Litovsky, 2015; Moore & Linthicum, 2007; Ponton et al., 2002). While myelination in the brainstem starts prenatally and increases until the end of the first year (Moore et al., 1995), it takes place later in the auditory cortex and continues until adulthood (Su et al., 2008). In parallel, intense synaptogenesis allows grey matter volume to reach a maximum at the end of the second year of life (Knickmeyer et al., 2008).

1.1.2. Critical periods

The developmental windows of these maturational changes constitute critical periods, defined as *"window[s], typically in early development, during which a system is open to structuring or restructuring based on input from the environment"* (Werker & Hensch, 2015). Changes in environments might negatively impact these critical periods.

A striking example is the case of premature babies: during gestation, the maternal tissue and fluid act as filters for high-frequency sounds, protecting the cochlea during its maturation. However, preterm newborns are deprived of this natural protection too early and exposed to different frequency bands compared to full-terms. This premature exposure to the external world impairs the maturation of the auditory pathway (McMahon et al., 2012). Specifically, neonatal Auditory Brainstem Responses in preterms showed retarded myelination of the central auditory pathway when compared to full-term babies (Pasman et al., 1992).

1.1.3. The first hearing experiences

The earliest behavioral responses to sound were recorded through ultrasound imaging between the 24th and the 25th gestational weeks, and consistent responses were shown around the 28th week (Birnholz & Benacerraf, 1983; Kuhlman et al., 1988). One classical protocol to evaluate subcortical brain responses in humans and animals relies on the presentation of simple auditory clicks. Animal studies showed evidence of click-evoked Auditory Brainstem Responses (ABRs) in ewes' fetuses between 117 and 125 days gestation (Plessinger & Woods, 1986; Woods et al., 1984), and constituted the first electrophysiological evidence of prenatal sound processing at the subcortical level. Although animal models constitute interesting comparative tools to study language's phylogeny and ontogeny, they do not provide direct information about human perception and cognition (Cartmill, 2023). For this reason, electrophysiological measures obtained in human newborns and infants are essential. Click-evoked ABRs consist of seven consecutive waves, each one informing about the integrity of neural transmission at each relay of the auditory pathway (see Figure 1.1b) (Intartaglia, 2016). Neonatal ABRs shows functional subcortical processing of sounds at birth: they have been used as a hearing screening test since the mid-90s (Meyer et al., 1999; Oudesluys-Murphy et al., 1996) and still constitute a reliable and efficient evaluation of newborns' auditory system functioning (Acke et al., 2022).

Although newborns can hear, their auditory system continues to mature during the postnatal months, allowing for the development of refined speech perception abilities. Next, I will address the following question: **What are the developmental milestones of early speech perception?**

1.2. From universal to language-specific speech perception

1.2.1. An initial bias for speech

In their first days of life, babies already have a bias for speech sounds (Vouloumanos et al., 2010). Indeed, behavioral studies on newborns' auditory perception show a preference for speech sounds over non-speech sounds that share similar spectro-temporal features (Butterfield & Siperstein, 1970; Vouloumanos & Werker, 2007). When exposed to filtered (Spence & DeCasper, 1987) or reversed speech (Pena et al., 2003), neonates still exhibit a preference for natural speech. Previous research also shows that newborns prefer their mothers' voices over other voices (DeCasper & Fifer, 1980), that they can recognize their native language over an unfamiliar language based on prosodic cues only (Mehler et al., 1988), and that they prefer listening to their native language over nonnative speech sequences (Moon et al., 1993). Moreover, evidence for face-like patterns preference in neonates (Goren et al., 1975; M. H. Johnson et al., 1991; Kleiner, 1987; Valenza et al., 1996) indicates that visual perception also tends to orient babies' attention toward social cues at birth. In addition to these perceptual biases, neonates are already equipped with speech discrimination abilities that allow them to process speech sounds even with missing temporal information (Cabrera & Gervain, 2020). Overall, these results suggest that humans are born with perceptive systems that are prepared and biased toward relevant environmental information for language acquisition (Kuhl et al., 2005).

1.2.2. The citizen of the world

The critical periods mentioned in the previous section (see Section 1.1) apply to various aspects of development, including speech perception. Accordingly, a universal developmental trajectory of speech perception has been identified and described in the literature (Kuhl, 2004; Kuhl et al., 2008; Werker & Tees, 1984b). From birth and until six to ten months of age, infants discriminate all phonetic contrasts of the world's languages, independently of their native language (Dehaene-Lambertz & Pena, 2001; Eimas et al., 1971; Ferguson, 1994): they are born "citizens of the world" (Kuhl, 1999). This universal speech perception diminishes through development and is lost in adults (but see Sections 1.2.4 and1.2.5). Indeed, adults' speech perception is shaped by exposure to their native language and its phonemic repertoire (see Box 1.2.1). Miyawaki et al. (1975) showed that adult native speakers of Japanese do not discriminate between /r/ and /l/, two speech sounds that do not constitute a pair of phonemes in Japanese. However, American adults tested on the same contrast correctly categorized the two

speech sounds, which do belong to different phonemic categories in American English. When presented with the /p/-/b/ nonnative contrast[2], however, American English adults hardly heard the difference between the two stimuli (Abramson & Lisker, 1973). Then, **when do humans lose this universal speech perception ability?**

More about: Phonemic repertoires of the world's languages.

The distinction between speech sounds and phonemes is crucial for understanding the notion of phonemic repertoire and the difference between phonetic (or acoustic) versus phonemic (or phonological) processing. Speech sounds refer to all the sounds a human can anatomically produce, whereas phonemes are abstract distinctive units. Speech sounds relate to the physical manifestation of speech and can be phonemic or non-phonemic sounds in a given language. In contrast, phonemes are always defined in the context of a specific language and are used to distinguish between words. For instance, the speech sounds /r/ and /l/ are two different phonemes in English, but constitute allophones of the same phoneme in Japanese. Thus, they can be used to differentiate between the words /leɪt/ ("late") and /reɪt/ ("rate") in English, but not between /arigato/ and /aligato/ ("arigatō"), that both mean *thank you* in Japanese.
Each language has its own set of phonemes or ***phonemic repertoire***. Some languages share common phonemes, others have unique ones. Overall, there is great variability in the sizes and contents of the phonemic repertoires of the world.

Box 1.2.1 – Phonemic repertoires of the world's languages

1.2.3. Vowel perception in infancy

Around six months of age, a decline in nonnative vowel discrimination is observed at both behavioral and electrophysiological levels. In a seminal behavioral study, Polka and Werker (1994) found that 4-month-olds English-learning perceived the nonnative /y/-/u/ German contrast, whereas 6-month-olds and older infants did not. Kuhl et al. (1992) obtained similar results using American English and Swedish synthesized vowel prototypes (/i/ and /y/ respectively) and several variants of each vowel. The variants were created to form a continuum based on the acoustic distance from the prototype. English- and Swedish-learning infants were tested at 6 months old on their perception of the similarity between each prototype (native and nonnative) and its variants. Results showed that the variants of the native vowel were perceived as similar to the prototype

2. English also has the phonemes /p/ and /b/, but the Voice Onset Time (VOT) values that allow the distinction between the two categories are different in Spanish and in English.

independently of the acoustic distance for both groups. However, as their acoustic distance from the nonnative prototype increased, variants were perceived as more different. These findings were one of the first evidence in favor of the *perceptual magnet effect*, Kuhl's theory of early speech perception that I will develop in Section 1.3.3. More broadly, these results suggest that early exposure to the native language shapes the perception of speech sounds .

Electrophysiological studies brought consistent results with these findings, providing additional information about the intensity and velocity of the neural responses involved in speech sound discrimination. Cheour, Ceponiene, et al. (1998)'s study is the first electrophysiological investigation of native versus nonnative vowel perception in infants using the Mismatch Response (see Section 1.4.1). Estonian and Finnish infants were presented with two contrasts: (1) /e/-/ö/, two phonemes that exist in both Estonian and Finnish; and (2) /e/-/õ/, which involves the phoneme/õ/ that exists only in Estonian and that is further from /e/ in the F1-F2 space [3]. At 6 months of age, Finns were able to discriminate both the native and nonnative contrasts, with amplitudes of the evoked responses reflecting the acoustic-phonetic distance between vowels. However, when presented with the same contrasts at 12 months of age, the same infants showed a reversed pattern: the response amplitude increased for the native contrast (/e/-/ö/) but decreased in the nonnative contrast (/e/-/õ/). In addition, the encoding of the acoustic-phonetic distance between vowels disappeared with age. When tested on the same contrasts, 12-month-old Estonians showed the same response patterns as 6-month-old Finns, with overall higher amplitudes that reflect stronger speech sound encoding. These results indicate that Estonian infants' discrimination abilities increased with age on both contrasts, whereas Finns showed this increased response only for the native contrast. Moreover, the acoustic-phonetic distance encoded in the neural response at 6 months old has been lost at 12 months old in Finns but not in Estonians. This suggests that categorical speech perception emerged in both groups, reinforcing native contrast discrimination and diminishing nonnative contrast discrimination abilities. Cheour et al.'s study constitutes compelling evidence of the switch from acoustic to phonological processing of vowels around 6 months of age.

3. The F1-F2 space is a 2D representation of vowels as a function of their first (F1) and second (F2) formants in Hertz. The F1 value is related to the opening of the mouth: the more open, the higher the F1 value. F2 relates to the degree of "backness" of the tongue in the mouth: the more at the front, the higher the F2 value.

1.2.4. Change in consonant perception

The same switch is observed for consonants, but later in the developmental trajectory. Consonant perception becomes language-specific between ten and twelve months of age (Best & McRoberts, 2003; Tsushima et al., 1994; Weikum et al., 2012; Werker & Tees, 1984b). Early consonant discrimination was first explored by Eimas et al. (1971), who showed that 1-to-4-month-olds raised in an English-speaking environment discriminated native VOT-based contrasts such as /b/-/p/. A few years later, Trehub (1976) showed that infants from the same type of population (1-to-4-month-olds from English households) discriminated two pairs of nonnative contrasts from Czech and French languages which they never heard before. When tested with the same stimuli, English-speaking adults were at chance in the discrimination task but showed perfect accuracy on a control English contrast. Similar results were obtained in English- and Hindi-speaking populations by Werker et al. (1981): whereas English adults did not discriminate the two Hindi consonant contrasts that they were presented, English infants aged 6-8 months of age and Hindi adults did. Then, the pivotal study of Werker and Tees (1984b) brought evidence for the critical period where native language attunement occurs for consonants. They compared discrimination abilities in infants aged 6-8, 8-10, 10-12 months, and adults. Results showed a clear switch in the 10-12 group: infants aged 6-8 and 8-10 discriminated the nonnative contrasts, and infants aged 10-12 and adults did not. This study constitutes the first evidence of the consonant perception switch towards the native language phonemic repertoire between 10 and 12 months.

All works on consonant perception mentioned above used behavioral data. However, as seen previously for vowel perception (subsection 1.2.3), changes in speech perception can be explored using electrophysiological data. Accordingly, several studies focused on early consonant perception using cortical ERPs. For instance, Rivera-Gaxiola et al. (2005) explored developmental changes in brain responses to native and nonnative consonants using a multi-feature oddball paradigm (see Section 1.4.1 for an explanation of the paradigm). American infants aged 7 and 11 months were presented with native and nonnative contrasts involving changes in VOT. The group-level analysis revealed a similar developmental pattern as found in behavioral studies: at 7 months of age, infants discriminated both contrasts; at 11 months of age, the same infants discriminated only the native contrast. However, individual-level analyses revealed different patterns of response, indicating that 11-month-olds still encoded the nonnative consonant contrast although less strongly than the native contrast. This was evidenced by the smaller Event-related potentials (ERPs) amplitude to the nonnative contrast compared to the

native contrast, reflecting the "*neural commitment*" to the native language, as argued by Rivera-Gaxiola et al. (2005). This work is in line with other studies showing that different components of the Mismatch Response in passive listening oddball tasks might reflect distinct aspects of speech sound encoding (see Section 1.4).
Electrophysiological responses to different consonant contrasts revealed different developmental time windows and different neural components' characteristics depending on the phonetic features involved in the discrimination. Indeed, changes in voicing (VOT), place of articulation, and manner of articulation led to different encoding of the contrasts reflected in different patterns of brain responses (Cheour, Alho, et al. (1998), Cheour, Ceponiene, et al. (1998), Dehaene-Lambertz (2000b), François et al. (2021), and Pang et al. (1998); see Narayan (2019) for a review).

1.2.5. Perceptual narrowing

The change in speech perception described above is called "perceptual narrowing" (Maurer & Werker, 2014). This phenomenon refers to the parallel loss of sensitivity to nonnative contrasts and increase in native speech sound discrimination. The studies mentioned in the previous paragraphs suggest that perceptual narrowing results from an interaction between three main factors: an initial perceptual sensitivity to speech sounds, an increasing exposure to the native language, and maturational changes involving critical periods (see Section 1.1). To account for perceptual narrowing, Aslin (1980) proposed three possible scenarios that depend on (1) the developmental state of perceptual abilities at birth and (2) the effects of experience on perceptual abilities (see Figure1.2). The three alternatives are universal theory, attunement theory, and perceptual learning theory. The first one supposes fully developed perceptual abilities at birth and maintenance/loss effects of the native language exposure during the postnatal period. The second one, the attunement theory, postulates that speech perception abilities are partially developed at birth and are either facilitated, maintained, or lost through native language exposure. The last theory involves an underdeveloped perception of speech sounds that is then fully shaped by experience.

As argued by Maurer and Werker (2014), the term "perceptual narrowing" conveys the idea of a decline in the perception of the underrepresented speech contrasts from the environment. However, the attunement and perceptual learning theories of Aslin (1980) involve enhancement mechanisms (facilitation and induction) that account for the increase in native speech perception, hence the term "perceptual attunement" proposed by Gottlieb (1976) and sometimes used in the literature. The attunement theory is particularly interesting because it accounts for data showing that the ability to

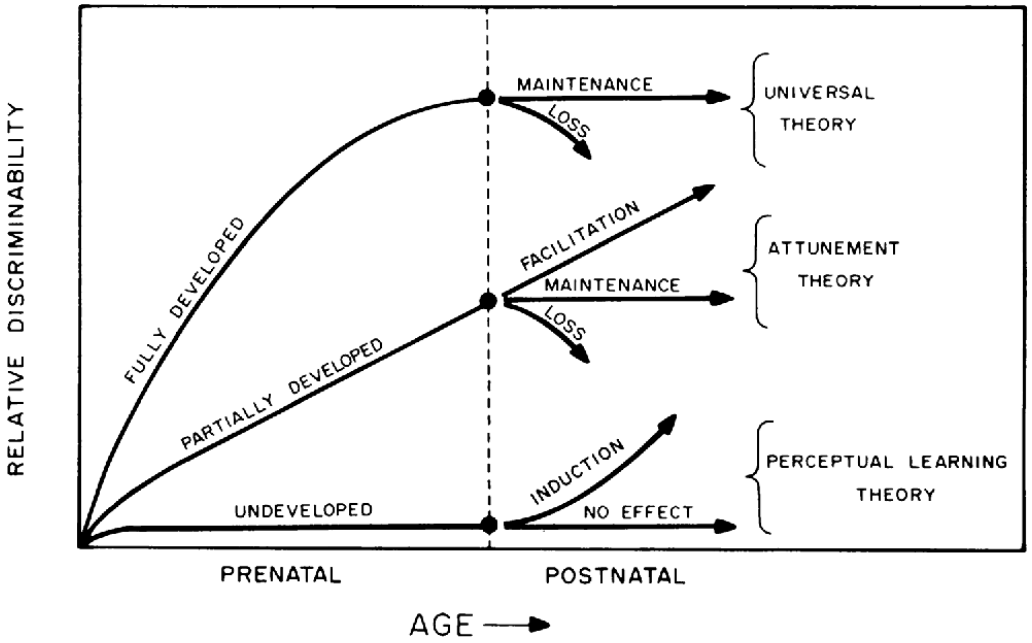


Figure 1.2. – Figure from Aslin (1980): Four possible mechanisms involved in perceptual narrowing (maintenance, facilitation, loss, induction) and the three classes of theories (universal, attunement, and perceptual learning) that depends on prenatal perceptual state and subsequent mechanisms implied.

discriminate nonnative speech sounds is not fully lost (Rivera-Gaxiola, Johnson, et al., 2000; Rivera-Gaxiola, Csibra, et al., 2000) or can be retrieved with training (Bradlow et al., 1997, 1999; Chandrasekaran, Sampath, & Wong, 2010; Lively et al., 1994; Perrachione et al., 2011).

Overall, electrophysiological evidence of native versus nonnative speech sound discrimination supports the idea of a developmental shift from phonetic to phonological speech perception (Cheour, Ceponiene, et al., 1998; Rivera-Gaxiola et al., 2005). However, the emergence of phonological categories is a complex process that still needs to be disentangled. For several decades, this question has been addressed, yielding different theories that will be developed in the next section.

1.3. Theories of categorical speech perception

In the previous section, I provided an overview of the literature describing a shift from universal to language-specific speech sound discrimination. Specifically, the time window between 6 and 10 months of age may constitute a critical period wherein perceptual narrowing occurs for vowels and consonants, respectively. This change in perceptual

abilities corresponds to the emergence of phonological representations, involving a shift from phonetic to phonological processing of speech sounds. In the present section, I review some of the theories that attempted to explain this developmental phenomenon.

1.3.1. Motor theories

In the early 1950s, researchers attempted to explain the mechanisms of speech perception by focusing on the mapping between the acoustic properties of the auditory signal and linguistic units (phonemes) and features. Because of the complexity of such mapping, A. Liberman and his colleagues at the Haskins Laboratories changed their perspective and focused on the link with the articulatory movements that generate the speech sounds instead of the resulting acoustic elements. This approach gave rise to the Motor Theory (MT) of speech perception (Liberman & Mattingly, 1985, 1989; Liberman et al., 1967; Mattingly & Liberman, 1988). The main claim of MT is that what we perceive when we process speech is articulatory events rather than acoustic or auditory events. This implies that the speech-motor system is involved during speech perception. This idea came from a series of experimental studies on the perception of synthetic consonants when presented in consonant-vowel syllables. Indeed, Delattre et al. (1955) noticed that different formant transitions in consonants could lead to the same phonemic percept. Figure 1.3 illustrates this phenomenon with the syllables /di/ and /du/. Whereas the first formant (F1) is identical in both syllables (F1 indicates the manner of articulation, here a voiced stop), the second formant (F2) displays different patterns (because of the coarticulation effect). Knowing that F2 informs about the place of articulation, the two consonants should be perceived as distinct phonemic units. However, both consonants are perceived as /d/, suggesting that "the second-formant transitions rather directly represent the articulatory movements from the place of production of the consonant to the position for the following vowel" (Delattre et al., 1955, p. 769). According to motor theorists, this stands for evidence that the information processed has an articulatory rather than acoustic nature.

In a revised version of the theory, Liberman and Mattingly (1985) highlighted that speech perception ability depends on a speech-specific *module* defined as "special neural structure[...], designed to take advantage of a systematic but unique relation between a proximal display at the sense organ and some property of a distal object." (Liberman & Mattingly, 1985, p. 7). The role of this *module* is to decode the articulatory gestures [4] perceived to convert them into phonemic units. In this view, speech percep-

4. As noted by Galantucci et al. (2006), [i]t is important to note that [...] gestures are not movements of individual articulators. For example, a bilabial consonant such as /b/ corresponds to one gesture —the

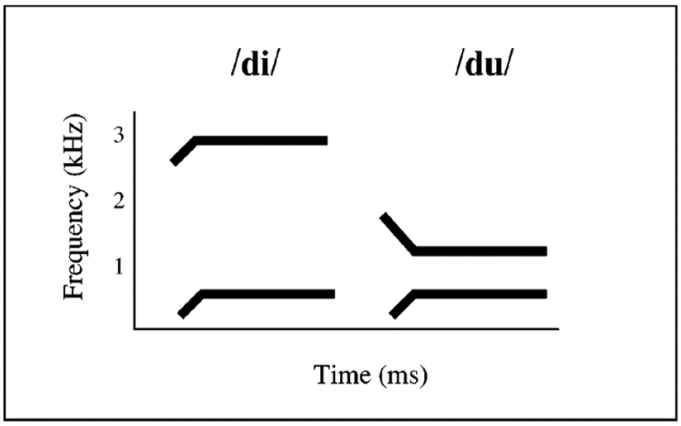


Figure 1.3. – Figure from Diehl et al. (2004): "Formant patterns for simplified versions of /di/ and /du/. Note that the transition of the second formant (i.e., the one higher in frequency) differs dramatically for the two syllables. Nonetheless, the consonant in both cases is perceived as /d/. The first formant trajectory, which is equivalent in both syllables, is not informative about place of articulation and would be the same for /b/ and /g/ initial syllables. (Adapted from Delattre et al. 1952.)"

tion and speech production rely on the same sequence of mechanisms, but in opposite orders. In the context of language development, MT implies that "infants initially detect all phonetically relevant gestures" (Kuhl et al., 2008, p.980). Moreover, the idea of a perception-production link suggests that as infants start to produce canonical sounds, they should improve their speech perception abilities. This perception-production link has later been explored in experimental studies (de Boysson-Bardies, 1993; Dehaene-Lambertz et al., 2006; Imada et al., 2006) and developed in another theory of speech perception (see 1.3.3).

A crucial aspect of the MT is that the units of speech perception are not the articulatory gestures themselves, but the *intended gestures*, "represented in the brain as invariant motor commands that call for movements of the articulators through certain linguistically significant configurations" (Liberman & Mattingly, 1985, p. 2). The Direct Realist Theory (DRT) proposed by Fowler (1996) agrees with the idea that the objects of speech perception are articulatory, nevertheless, it contrasts with the MT regarding the nature of these articulatory objects. For DRT, "the articulatory objects of perception are actual, phonetically structured, vocal tract movements, or gestures, and not events that are causally antecedent to these movements, such as neuromotor commands or

complete closure of the lips— achieved by the coordination of the movements of three articulators: the jaw and the two lips (Kelso et al., 1984).

intended gestures" (Diehl et al., 2004, p. 4). In this view, speech perception involves a *direct* access to the *real* articulatory movements without mental representation or conversion from a specialized neural network. Importantly, Fowler (1996) refutes the idea of *invariants control structures* and claims that because of coarticulation effects, vocal tract gestures might be variables. In the case of coarticulated phonemic units, DRT predicts that each articulatory gesture will be recognized and that the residual variability will be anticipated and attributed to the gestural properties of the neighboring speech sounds. This could also explain the perceptual phenomenon illustrated in Figure 1.3.

Although MT and DRT disagree on certain points, both theories consider the objects of speech perception as gestures (intended for the first, real for the latter). In the following subsections, I introduce two other theories that do not consider this motor aspect.

1.3.2. The Perceptual Assimilation Model

The Perceptual Assimilation Model (PAM) is a theory on cross-language speech perception proposed by C. Best (Best, 1993, 1994, 1995a, 1995b). The PAM aims to account for perceptual narrowing and the variability found in adult native/nonnative speech sound discrimination studies. According to PAM, the increased exposure to the native language attunes the perceptual system toward native gestural patterns. Thus, phoneme identification requires the detection of higher-order invariants in the speech sound properties, called *distinctive features*. These are higher-order features that are unique to each language and do not correspond to the lower-order phonetic features that characterize speech sounds, although it relies on them. Indeed, in a given language, only a set of lower-order phonetic features will be used as higher-order invariants. Thus, adults' perception of speech sounds consists of attending to the distinctive features from their native phonological repertoire to decide which phoneme is being heard. Adult listeners are constantly seeking similarities and dissimilarities with the gestural patterns of their native language to either *assimilate* a speech sound to a native phonemic category or decide that it is a nonnative or nonspeech sound. However, because languages of the world share lower-order phonetic features, some nonnative contrasts might be easier to discriminate than others. Through the scope of the native language repertoire, nonnative speech sounds can be assimilated to native ones, as "weak" or "non-prototypical" forms of phonemes. Best (1994) proposed several types of assimilation as a function of the similarities between the native gestural patterns and the low-level features of nonnative contrasts. The model predicts that if listeners can detect invariants from their native language in two nonnative categories, they will easily discriminate them: this is

the Two-Category (TC) type. Discrimination might be moderate when both nonnative sounds are assimilated to the same native category but differ in the level of similarity to this native category (Category-Goodness Difference or CG). However, poor discrimination is expected when both nonnative sounds are assimilated to the same native category and equally fit to the native prototype (Single-Category or SC), or when no native distinctive feature is detected in both nonnative sounds (Both Uncategorizable or UU). In this case, a low-level phonetic processing should be observed (L. Liu et al., 2023). Finally, good discrimination is expected when both nonnative sounds are not assimilated to a native category (Uncategorized vs. Categorized or UC) or are not perceived as speech sounds (Nonassimilable or NA).
This nuanced view of the perceptual narrowing of native/nonnative discrimination has been tested using electrophysiological data. Vowel (Näätänen et al., 2007; Tyler et al., 2014) and consonant (L. Liu et al., 2023) contrasts have been explored, showing that specific patterns of neural response might correspond to the different levels of assimilation proposed by Best (1994). These results might also be explained from P. Kuhl's perspective and her *Native Language Magnet theory*, which I present in the next section.

1.3.3. The Native Language Magnet theories

The Native Language Magnet (NLM) theory was first introduced in Kuhl (1991)'s study on the characterization of speech *prototypes*. Based on previous results suggesting that "speech categories may be organized with reference to a good exemplar, a prototype of the category" (Kuhl, 1991), P. Kuhl asked English-speaking adults to rate the category goodness of several instances of the vowel /i/. Results showed that rating was consistent across participants and decreased symmetrically as a function of the stimulus' distance from the prototype. Next, they assessed category perception using behavioral measures in adults (pressing a button) and 6-month-old infants (conditioned head turn procedure), showing that the perception of vowel categories was affected by the distance to the prototype in both groups. These results suggested that the prototype of the phonological category acted as a "*perceptual magnet*" for the other non-prototypical members of the category. This study also brought evidence that this structural influence of prototypes on categorical perception did not apply to non-human primates, suggesting a domain-specific mechanism.
Soon after, P. Kuhl showed that exposure to the native language had an impact on this prototypical sensitivity by comparing native and nonnative categorical perception (Kuhl, 2000; Kuhl et al., 1992). Kuhl argued that speech perception shifted from phonetic/a-

coustic to phonological/categorical, through the magnet effect. Figure 1.4 (from (Kuhl, 2000) illustrates this magnet effect for the two American English categories /r/ and /l/. Figure 1.4 A represents the actual physical distance between the two speech sounds in the F2/F3 space; figure 1.4 B shows the phonological distance perceived by native English speakers after the magnet effect.

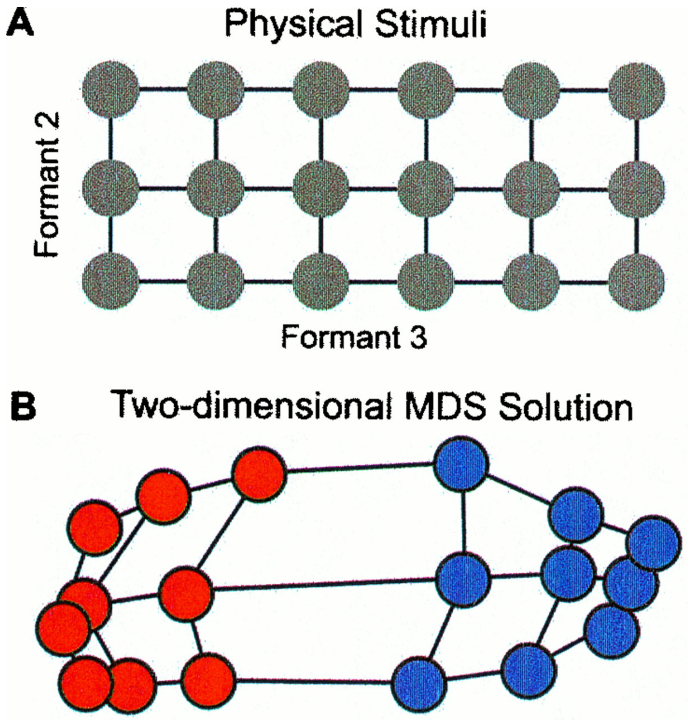


Figure 1.4. – The magnet effect: how is the perceptual system modified by prototypes of two categories. Here, the phonetic distance between /r/ and /l/ is represented (A) phonetically, and (B) phonologically (as perceived by American English speakers): the prototypes of the categories act as magnets for other members of the category. *From Kuhl (2000).*

NLM describes three main steps in speech perception development:

1) In the initial state, general auditory (rather than speech-specific) processing mechanisms allow infants to discriminate all speech sounds.

2) The native linguistic input contains distributional properties that organize the speech sounds into "modes". Infants' perceptual sensitivity is modulated by experience such that these modes become more salient. With increasing exposure, the most recurrent acoustic forms become *prototypes* that act as *perceptual magnets* for the speech sounds of the same category.

3) Perception is distorted: the *perceptual magnet effect* facilitates native contrast discrimination and impairs nonnative contrast discrimination abilities.

More recently, Kuhl proposed an extended version of the model, the Native Language Magnet extended (NLM-e) (Kuhl et al., 2008). This revised version incorporates additional factors such as neural commitment, social factors, and the link between perception and production. Accordingly, NLM-e comprises four main stages of development:

1) Initial stage of universal speech perception (identical to the NLM:) with the possible influence of acoustic salience.

2) As they get increasingly involved in social interactions and more sensitive to the exaggerated cues of Infant-Directed Speech (see section 2), infants become more sensitive to the distributional patterns of phonetic units. Infants also start to imitate the vocal gestures of their caregivers, creating new connections between perception and production. Native speech perception is facilitated by these factors (social interactions and perception-production link), and the increasing native language exposure induces Native Language Neural Commitment (NLNC). Kuhl et al. (2008) defines NLNC as "a process in which initial language exposure causes physical changes in neural tissue and circuitry that reflect the statistical and perceptual properties of language input". This neural commitment reinforces the detection of native higher-order patterns and diminishes the sensitivity to nonnative patterns (uncommitted circuitry).

3) In the third phase, NLNC has modified speech perception abilities, which has a cascading effect on other linguistic skills. In particular, three skills related to word-learning are improved: the detection of phonotactic patterns, the detection of word-like units, and the association between word forms and objects. At this stage, speech perception and word-learning abilities have *bidirectional* facilitatory effects, which reinforce native language development processes.

4) In phase 4, neural commitment is stable but still not at an adult state. Learning skills are affected by native pattern attunement, and the strength of the neural commitment predicts future learning.

Importantly, the NLM-e predicts a link between speech perception abilities and (1) the neural commitment for native-language patterns; (2) social factors, interactions with the environment, and linguistic input; and (3) later language development and learning skills. This model is of particular interest to the present work because I explored these three factors both longitudinally and cross-sectionally. Each of them is

addressed in three different chapters: Chapter 4 is dedicated to neural responses to speech sounds at different levels of the auditory pathway; Chapter 5 explores the daily linguistic environment in infancy; and Chapter 6 investigates associative word-learning abilities at 18 and 24 months of age. The next section introduces the main research on electrophysiological responses to speech sounds, which constitute the main focus of the Chapter 4.

1.4. Electrophysiological responses to speech sounds

Electrophysiology is the study of the electric and electrochemical phenomena of living cells and tissues, with a particular interest in neurons. In neuroscience, electroencephalography / electroencephalogram (EEG) is typically used to record the electric activity conveyed through brain neurons. Although the collection of electrophysiological measures in infants comes with its sets of challenges (see Hervé et al. (2022) and Section 2 from Chapter 3), the literature showed that neural responses to speech sounds provide crucial evidence that behavioral measures cannot (see Section 1.3). Importantly, pre-attentive responses can be observed with very high temporal resolution and used to decipher the perceptual mechanisms involved in early speech perception.
In the present section, I will focus on the methods and paradigms involving EEG data to study the development of speech perception abilities.

1.4.1. Cortical brain responses to speech sounds

Because of its high temporal resolution, EEG constitutes a reliable and efficient tool for studying the temporal dynamics of cognitive processes (Hervé et al., 2022). Early speech perception has been predominantly studied at the cortical level, using auditory paradigms that elicit cortical event-related brain potentials (ERPs). One particularly relevant evoked potential is the Mismatch Negativity (MMN), an automatic brain response elicited by automatic change detection of deviant auditory stimuli in a series of repeated standard stimuli (Näätänen et al., 2007). The MMN can be observed in a particular experimental paradigm called the oddball paradigm, a passive listening task involving the alternate presentation of at least two different auditory stimuli. One stimulus has a high probability of occurrence (the standard), and the other has a low probability of occurrence and should not be predictable (the deviant). The number of subsequent repeated standards should be high enough to create a short-term memory trace that diminishes as the number of presentations increases (Haenschel et al., 2005; Imada et al., 1993; Javitt et al., 1998; Sams et al., 1984). There is no consensus on the

minimum number of consecutive standards required, although recent work showed that two repeated standards before a deviant are enough to elicit an MMN (Y.-Y. Cheng et al., 2015; He et al., 2009; Virtala et al., 2022). The MMN's peak latency can be influenced by the stimulus rate, the magnitude of change, and the type of change (Kekoni et al., 1997; Schröger & Winkler, 1995; Takegata et al., 1999). In general, peak latency reflects the degree of myelination and the distance between the brain generator and the electrode; amplitude variations are related to neural synchronization, synaptogenesis, and the volume of cerebral cortex activated (in a sense, the intensity of the brain response) (Coch & Gullick, 2012; Picton & Taylor, 2007).

The MMN has been interpreted as a response that reflects the violation of the neural short-term memory representation of the standard stimulus (Näätänen et al., 2007). Dehaene-Lambertz and Gliga (2004) used the notion of *repetition suppression* to account for this phenomenon at the level of neural networks. According to the authors, when the standard stimuli are presented, specific neural networks fire in response to the standard stimulus, with different sets of neurons coding for different properties of the stimulus. As the number of repetitions increases, the intensity of the response decreases. When a deviant is presented, a new set of neurons will fire in response to the new feature (the feature that distinguishes the standard from the deviant). In contrast to the other neural networks already (decreasedly) activated from the standard presentation, the newly activated network will have a greater intensity. This difference of activity between neural networks is retrieved when subtracting the response to the standard from the response to the deviant and corresponds to the MMN. Thus, "by manipulating what elicits a mismatch response, that is, what counts or not as a repetition for a given neural network, it is possible to infer the format of representation that is computed by the network" (Dehaene-Lambertz & Gliga, 2004, p. 1376). Accordingly, changes at the acoustic level (speaker, loudness, etc.) will not elicit the same response as changes at the phonemic level (VOT, formant, etc.). Changes in phonemic features are particularly interesting to study more complex representations of speech sounds, especially in developmental studies.

The MMN has been observed in a 150-250 ms time window after deviant onset in adults, at frontocentral and central scalp electrodes (Näätänen et al., 2007). However, positive deflections in response to change have been observed in infancy, at various latencies between 150 and 400 ms. Thus, the term Mismatch Response (MMR) is preferred in developmental research to account for this variation in polarity. It has been

suggested that early MMRs might be divided into two distinct components, a positive and a negative peak, that appear at various latencies depending on their polarity and the infant's age at recording (He et al., 2007, 2009; Rivera-Gaxiola et al., 2005; L. Trainor et al., 2003; Werwach et al., 2022). Positive MMRs have been associated with immature perceptual systems, in longitudinal studies showing a shift from more positive to more negative responses with age (He et al., 2009; Rivera-Gaxiola et al., 2005; Virtala et al., 2022; Werwach et al., 2022). The MMR has been used to explore various change detection: sound duration (Čeponien, Kushnerenko, et al., 2002; Clarkson et al., 1989), intensity (Kushnerenko, Ceponiene, Balan, Fellman, & Näätänen, 2002; Partanen et al., 2013), frequency (Alho, Sainio, et al., 1990; Alho, Sajaniemi, et al., 1990; Bisiacchi et al., 2009; Kushnerenko, Ceponiene, Balan, Fellman, & Näätänen, 2002; Leppänen et al., 2010; Mento et al., 2010), loudness (Oceák et al., 2006; Wagner et al., 2023), phonetic features of speech sounds (Cheour, Alho, et al., 1998; Cheour, Ceponiene, et al., 1998; Dehaene-Lambertz, 2000b; François et al., 2021; Pang et al., 1998), and complex musical or speech sequences (François et al., 2017; Háden, Honing, et al., 2015; Háden, Németh, et al., 2015; Mueller et al., 2012). Importantly, MMRs to phonetic features such as VOT or formants give insights on consonants and vowels processing in the auditory cortex (see Section 1.2). For example, Werwach et al. (2022) explored the developmental trajectory of mismatch responses to consonant, vowel, vowel length, and frequency changes. The standard stimulus was the syllable /ba/ (Fundamental frequency (F0) = 198 Hz; duration = 170 ms), and the deviants were /ga/, /bu/, /ba:/ (270ms), and /ba+/ (F0 = 214Hz) for each deviant respectively. Figure 1.5 shows the difference wave obtained for each deviant at 2, 6, and 10 months of age. This longitudinal study shows that different phonemic features not only elicit different MMR patterns, but also show different developmental trajectories.

The MMR seems to constitute a valuable biomarker of language development abilities in infancy. Indeed, Kuhl et al. (2008) showed that consonant discrimination abilities at 7.5 months of age predicted productive vocabulary size at 18 and 24 months of age, and sentence complexity at 24 months of age. More precisely, greater MMR amplitudes to native contrasts were associated with more words produced and higher sentence complexity; conversely, greater MMR amplitudes to nonnative contrasts were associated with less word produced and lower sentence complexity. Other MMR studies showed links between speech perception abilities and linguistic outcomes in both typical (Rivera-Gaxiola et al., 2005; Schaadt et al., 2023; Von Holzen et al., 2018) and clinical populations (Bonte et al., 2007; Hämäläinen et al., 2018; Linnavalli et al., 2017; van der Leij et al., 2013). However, Kailaheimo-Lönnqvist et al. (2020) obtained more nuanced

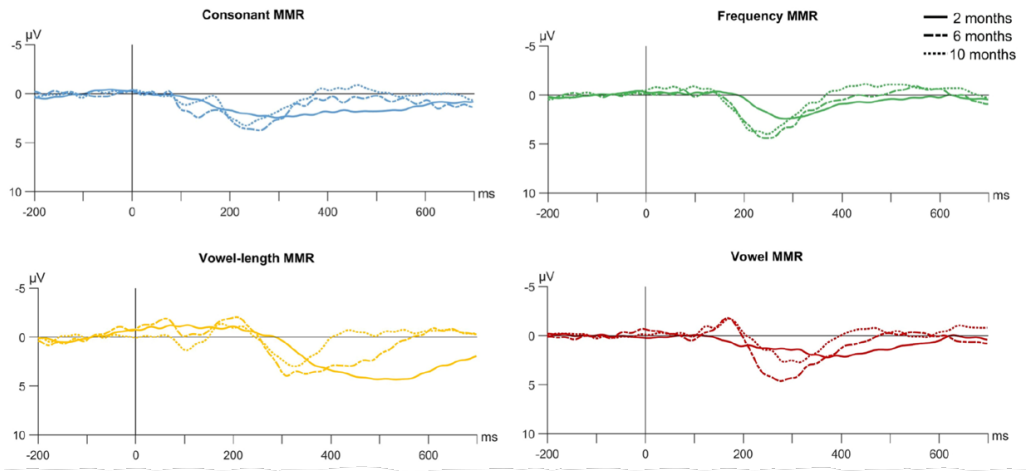


Figure 1.5. – Different mismatch responses in infancy as a function of deviant type and age. *From Werwach et al. (2022).*

results using a multi-feature oddball paradigm where a pseudo-word was presented as standard with three different deviants: duration, frequency, and vowel change. EEG data were recorded at 6 months of age, and a questionnaire filled by the parents at 6 and 12 months of age served as prelinguistic skills assessment. Results showed that the P1 component in the averaged response to standards predicted the prelinguistic skills improvement between 6 and 12 months of age. However, no correlations were found between the MMRs to all deviants and the prelinguistic skills improvement. The MMR to frequency deviant at 6 months old was positively correlated with the prelinguistic skill score at 6 months of age but not with the score improvement. Thus, more studies are needed to explore this link between early speech perception abilities assessed via MMR and later linguistic outcomes.

Cortical brain responses to speech sounds provide crucial biomarkers for studying the emergence of phonological categories. However, capturing brain activity at earlier stages of the auditory pathway offers complementary insights into early speech perception. In the next section, I shift my focus to subcortical brain responses, exploring the unique information they provide and their role in developmental research.

1.4.2. Subcortical brain responses to speech sounds

Section 1.1 showed that the brainstem contains important relays for auditory stimuli processing. In the same section, we learned that click-evoked Auditory Brainstem Responses (ABRs) were used as a hearing screening test in newborns. Indeed, this auditory

evoked potential consists of a series of seven waves that peak between 1.4 and 10 ms (Intartaglia, 2016). Each wave has been associated with one subcortical relay, allowing for the assessment of the functional integrity of the auditory pathway at multiple levels of the brainstem. ABRs can be recorded via a three-electrodes montage on the scalp using a classical EEG device with specific recording parameters. The Frequency-Following Response (FFR) is a particular variant of the ABR that starts between 3 and 10 ms after stimulus onset (Kraus et al., 2017; Krizman & Kraus, 2019) and reflects sustained, phase-locked activity of the auditory system. It can be recorded with the same montage as ABRs, using complex auditory stimuli such as speech or music. The FFR has recently gained interest in developmental auditory physiology because, in contrast to other auditory evoked potentials, it faithfully mimics the temporal and spectral features of complex auditory stimuli (Figure 1.6). This unique and powerful characteristic provides crucial information about the accuracy of fine-grained speech-sound neural encoding along the auditory pathway (Coffey et al., 2019; Gorina-Careta et al., 2021). This brain response has been considered to originate mainly from the brainstem, although recent findings showed that cortical generators may also contribute to the scalp-recorded FFR (G. Bidelman & Powers, 2018; Coffey et al., 2016, 2019, 2021).

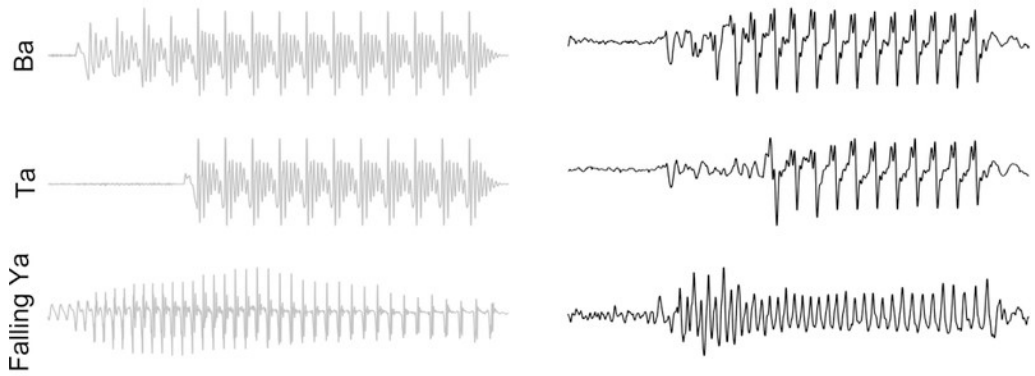


Figure 1.6. – Examples of Frequency-Following Responses to different syllables: on the left, the stimulus presented; on the right, the corresponding subcortical response. *From Krizman and Kraus (2019).*

The FFR can be analyzed in the time or frequency domain, using the same methods as any oscillatory signal (like a sound wave for example). We can thus display a temporal visualization, get peak latencies and amplitudes, extract an F0, or compute a Fast-Fourier Transform to analyze the frequency content (harmonics). Overall, the closer the spectro-temporal indices are to the stimulus' features, the better is the encoding (see Chapter 4 for more).

FFRs to speech sounds have been recorded in neonates (Arenillas-Alcón et al., 2021; Lemos et al., 2021; Ribas-Prats et al., 2019; Ribas-Prats et al., 2022) and infants (S. Anderson et al., 2015; Jeng et al., 2010, 2016; Ribas-Prats et al., 2023), in both typically developing and clinical populations. Ribas-Prats et al. (2019)'s normative study in healthy newborns provided spectral and temporal measures of /da/ and /ga/ syllables. Importantly, they showed accurate encoding of the F0 but not the higher harmonics, and no significant differences between the two syllables at both spectral and temporal levels. Following this work, important changes in subcortical auditory processing have been evidenced between birth and six months of age. Ribas-Prats et al. (2023) compared FFR responses to /da/ syllable in 0-, 1- and 6-month-olds. They found that the encoding of the /da/ was significantly faster between 0 and 1 month of age, and also between 0-1 month and 6 months of age. At the spectral level, F0 encoding did not significantly differ between 0 and 1 month of age, however, spectral amplitude did increase between 0-1 month and 6 months of age, consistent with Jeng et al. (2016)'s results. Regarding the fine structure encoding, a significant increase was observed both between 0 and 1 month of age and between 0-1 month and 6 months of age, as indicated by the comparison of the spectral amplitude at F1 value.
When comparing typically developing younger (3-5 months of age) and older (6-10 months of age) infants, S. Anderson et al. (2015) also showed that the amplitude of the FFR harmonics increased with age, reflecting a better encoding of speech sounds' fine structure (here the syllable /da/) in older infants. FFR peak latencies significantly decreased with age, as expected by the authors, considering the increasing myelination allowing faster neural conduction. However, no significant difference was found in the F0 amplitude between the two age groups.
Finally, work on newborns exposed to musical (Arenillas-Alcón et al., 2023) or bilingual (Gorina-Careta et al., 2024) environments using FFR recordings showed how prenatal environmental exposure influences the neural processing of speech. Specifically, Arenillas-Alcón et al. (2023) found enhanced spectro-temporal encoding of speech sounds in newborns with prenatal musical exposure, suggesting that musical experiences fine-tune the auditory system. Gorina-Careta et al. (2024) demonstrated that bilingual maternal speech exposure during pregnancy leads to different FFR patterns in neonates, highlighting enhanced neural processing of speech contrasts in both languages.

Overall, these studies indicate that important maturational changes occur at the subcortical level in the first months of life, that might have an impact on speech processing.

These changes are modulated by the auditory environment and might impact later language skills. Finally, spectro-temporal measures of the FFR to speech sounds might constitute a valuable tool for investigating early speech sound encoding.

1. Early speech perception: Summary

The first section described the main stages of infants' speech perception development, focusing on the neural responses to speech sounds. Importantly, a shift from phonetic to phonological speech perception is observed between 6 and 10 months of age, following critical periods' constraints. Several models and theories have attempted to explain this developmental trajectory through motor representations, assimilation of distinctive features, and prototypes' magnet effect. Electrophysiological methods are particularly efficient in deciphering the neural correlates of early speech perception, from subcortical relays in the brainstem to cortical auditory structures of the brain. In particular, the simultaneous recording of subcortical and cortical brain responses to speech sounds in infants has not been tested, although successful in adults. This new procedure might constitute a relevant tool for better understanding the relative contribution of these neural structures to speech sound encoding at different levels of the auditory pathway.

This first section showed that the development of speech perception abilities during the first two years of life is essential for later language outcomes. But during this period of high plasticity, the environment plays a crucial role in building robust linguistic skills. In the next section, I will focus on the role of the auditory environment in language development.

2. Auditory environment and language development

Current developmental theories claim that both internal and external factors play a role at all stages of language development. Section 1.1 focused on internal factors: we saw that humans start to hear in the mother's womb, and that the auditory system continues to mature after birth following a specific developmental trajectory. Nevertheless, regular input from the external environment is essential to trigger internal neurobiological changes that participate in the development of multiple linguistic skills. In this section, I emphasize the role of the auditory environment in language development.

2.1. Auditory inputs in early infancy

When acquiring their native language(s), infants receive multimodal input from their environment. In the present work, I focused on auditory input only, but the reader should keep in mind that language involves other modalities that are not addressed here. Some central questions regarding infants' auditory environment that are still under investigation are: what are the characteristics of this input? Does it change with age? If yes, how and when? How does it influence linguistic abilities? Which elements are essential to language development?

2.1.1. Input from the caregivers

The auditory inputs infants receive in their first months of life mainly come from their caregivers. The first auditory experience starts prenatally when infants hear their mother's voice filtered by the amniotic fluid (see Section 1.1). Then, full-term newborns are mainly exposed to their parent's voices daily, from birth until a few months of life. Therefore, studies on auditory environment in infants have primarily focused on caregivers' productions and infants-caregivers interactions. Indeed, even before infants can produce intelligible speech, alternations of caregivers and infants vocalizations are observed (Bateson, 1975; Trevarthen, 1977), yielding conversation-like patterns sometimes referred to as "protoconversations" (Bateson, 1979; Bruner, 1974). The first reports focused on caregiver-infant interactions using manually annotated audio recordings, although data collection methods were later considered to lack some ecological validity. (see section 3.1). Hart and Risley (1992a) is one of the pioneering studies examining infants' daily auditory environment by analyzing the number of parents-infants interactions per hour and the quality of the parents' utterances addressed to their infants. This work was more focused on the social aspects of those interactions but still provided a

description of the speech input infants received, showing a great variability between families. They found that the quality, content, and quantity of parent's speech input and child-caregiver interactions were influenced by the socio-economic status (SES) of the families. Specifically, lower SES families tended to use more prohibitive language, while higher SES families used more encouraging words, questions, and elaborations. However, this link between auditory input and SES was later called into question (Bergelson et al., 2023; Rindermann & Baumeister, 2015). In subsequent work, Hart and Risley (1992b) showed the importance of infant's linguistic stimulation from the parents and constituted one of the first studies to explore links between linguistic input and later language outcomes in children (see Section 2.2 for more).
Overall, caregivers tend to adapt their speech to young infants and modify acoustic, lexical, and pragmatic features (see 2.1.2). However, infants are also exposed to linguistic inputs that are neither addressed nor adapted to them and might be considered unhelpful for their language development. Yet, in some cultural groups, adults rarely or never address speech to infants before they start to produce intelligible speech (Casillas et al., 2020; Cristia et al., 2019; Shneidman & Goldin-Meadow, 2012). Indeed, both overheard speech and child-directed speech have been shown to facilitate language development (Akhtar, 2005; Au et al., 2002; Golinkoff et al., 2019; Rowe & Snow, 2020). Further, child-directed speech has also been associated with better linguistic abilities (Bergelson et al., 2019; Weisleder & Fernald, 2013a). Next, I focus on this very specific way of addressing speech to infants and its impact on the development of language and communicative skills.

2.1.2. Infant-Directed Speech

When talking to infants, adults modify their speech to adjust to their limited linguistic abilities. This singular way of talking to young children is called Infant-Directed Speech (IDS) and has been well documented at many levels (form, function, cross-cultural comparisons), mostly in English-speaking and occidental families. IDS is characterized by specific acoustic properties that differ from Adult-Directed Speech (ADS): increased pitch, greater pitch variations, lower speech rate, more rhythmical structure, more prosodic repetitions, and acoustically exaggerated vowels (Andruski & Kuhl, 1996; Burnham et al., 2002; Ferguson, 1964; Fernald, 1989; Fernald & Simon, 1984; Grieser & Kuhl, 1988; Kuhl et al., 1997). The exaggeration of vowel productions has led to the emergence of the hyper-articulation hypothesis: the idea that hyper-articulated vowels might help infants in vowel categorization. Interestingly, studies that tested this hypothesis did not obtain the expected results and tended to show that infants rely on prosodic rather

than acoustic features (Cristia & Seidl, 2014; Martin et al., 2015; Wong & Ng, 2018). IDS is also characterized by positive vocal affect, simplified grammar, a large number of questions, and easier vocabulary (Newman et al., 2016; Newport et al., 1977; Penman et al., 1983; Soderstrom et al., 2008). All these modifications in the adult's speech have been related to three main functions: linguistic, attentional, and socio-emotional. At the linguistic level, simplified and exaggerated speech features have been shown to clarify the speech input (Kalashnikova & Burnham, 2018; Kuhl et al., 1997; H.-M. Liu et al., 2003), helping infants understand the dominant and relevant acoustic features of their native language(s). In addition, the high variations in pitch and exaggerated acoustic features contribute to engaging and maintaining infants' attention. Finally, increased engagement in the interaction and positive vocal affect reinforce the emotional aspect of language and play an important role in social development (Golinkoff et al., 2015; Punamäki et al., 2024; Takada, 2005). These modifications of the speech input and their related functions have been observed in various linguistic communities and cultural groups, suggesting a cross-cultural phenomenon. However, some studies revealed a high degree of variability in some acoustic properties and in the frequency of IDS use, evoking the idea of a cross-cultural continuum (Bryant, 2022).

2.1.3. Non linguistic inputs

In addition to speech input, infants might be exposed to other non-linguistic sounds such as TV, other media, or music.

Indeed, TV, tablets, and smartphones are part of most families from industrialized countries nowadays, thus included in infants' auditory environment (Leblanc, 2017). Several studies reported negative impacts of children's screen exposure on outcomes such as attentional problems, obesity, aggression, inactivity, and sleep issues (D. R. Anderson et al., 2001; Christakis et al., 2004; MacBeth, 1996; Robinson, 1999). Importantly, negative impacts on language and cognitive outcomes have been found (Linebarger & Vaala, 2010; Linebarger & Walker, 2005; Zimmerman & Christakis, 2005), but some studies showed that in some contexts educational videos might be beneficial for children (Crawley et al., 1999; Hall et al., 1990; Linebarger & Walker, 2005). Also, whether TV is actively watched by young infants or overheard in the background seems to have different effects (D. R. Anderson & Pempek, 2005). Because TV necessarily involves both visual and auditory modalities, it is hard to disentangle the contribution of the auditory input from TV on infants' development.

Musical input, in contrast, is universal across cultures (Hilton et al., 2022; Mehr et al., 2019), and shares many acoustic features with language. Indeed, linguistic and musical auditory signals are both characterized by combinations of acoustic elements that vary in rhythm, melody (pitch), frequency, and duration through time. Although not consensual, Brandt et al. (2012) argue that speech is a "special type of music". Infants' musical environments have been mainly explored through questionnaires filled out by parents about their musical habits, revealing that the main source of music came from mother's songs and lullabies (Custodero et al., 2003; De Vries, 2021; Ilari, 2021; Young, 2008). In an attempt to provide more detailed and objective descriptions of these musical inputs, Politimou et al. (2018) developed an elaborated questionnaire for assessing the daily musical environment in families' homes. Politimou et al. (2018)'s Music@Home instrument comprises 67 items organized in 12 dimensions that covers both caregivers' habits, believes, and attitudes towards music, and children's productions, preferences, and reactions to music. Although very informative, at-home recordings of the daily auditory environment could yield more ecological and detailed data. These works are still scarce, but the few reports on the musical input infants are exposed to in their first months of life using daylong recordings showed variability in nature, frequency, and source across families (Hippe et al., 2024; Mendoza & Fausey, 2021) (see Section 2.2.2).

In summary, infants are exposed to various auditory inputs daily, with a great predominance of speech from caregivers who instinctively adapt their prosody, rhythm, and vocabulary. Next, I will see how daylong recording methods contributed to better quantify and describe these auditory inputs in more ecological settings.

2.2. Daylong recordings in infancy

2.2.1. A technological advance

The first studies on the auditory environment in infants were conducted in semi-controlled environments in laboratory settings or through parental questionnaires (Hart & Risley, 1992a, 1995; Harwood et al., 1999; Keller et al., 2004a). For example, (Hart & Risley, 1992b) asked parents to come to the lab and interact with their child naturally as they would do at home; in a second phase, an experimenter went to the families' homes to record auditory data and take notes about the children's and parents' behaviors. These methods, although ingenious, inevitably induced an observer's paradox that could bias the results. However, the idea of recording the infants' daily auditory environment was already there. As technology improved, new tools that allowed more

ecological data collection started to draw researchers' interest. Specifically, **daylong recordings** emerged in developmental studies in the early 2000's. This data collection method involves providing families with a recording device to use at home without any intervention from an experimenter, thereby bypassing the observer's bias. In addition to recording data in infants' natural environment, this method has other advantages. First, it allows the collection of longer recordings, yielding more substantial datasets. Second, it saves time and energy for the experimenters since they are no longer directly involved in the data collection process. Third, the presence of an experimenter at home can be considered very intrusive for the families and cause many refusals to participate: with autonomous daylong recordings, this effect is dimmed. Last, because parents actively participate in data collection, they often feel more implicated in the study, which is particularly important in longitudinal studies that usually suffer from high attrition rates.
In parallel with the emergence of **daylong recordings** devices, automatic analysis tools were also developed. Indeed, auditory data used to be manually annotated, which was highly time-consuming and might have limited the datasets' size. Automatic analysis softwares detect specific sound properties in the raw recording to provide relatively reliable annotations depending on the quality of the recording, the language spoken, and the goal of the analysis. Manual annotation has not yet been completely set aside because it can be used as a control for validation studies (Busch et al., 2018; Canault et al., 2016a; Cristia et al., 2021a; Pae et al., 2016b; Xu et al., 2009a), or constitute precise transcriptions for specific research questions (Oller et al., 2010a; Orena et al., 2019b).

The Language ENvironment Analysis (LENA) System (Foundation, 2015) is a daylong recording tool that is particularly adapted for developmental research in ecological conditions. The system includes both a recording device and automated analysis software. The hardware part consists of a small recorder and a t-shirt with a front pocket where the recorder can be inserted. This way, infants can wear the recording device for a full day without any difficulty or inconvenience. The software provides estimations of the number of words and conversational turns produced by the infant and by the adults of the surrounding environment, as well as quantitative information about synthesized sounds like music or television. Overall, this technological revolution has been a major breakthrough for developmental research (Bergelson et al., 2023; H. Ganek & Eriks-Brophy, 2018a) and the study of daily auditory environments in infancy. The increase of the daylong recording community in the last decade has also led to the emergence of new open tools for processing, storing, and sharing these data. In

particular, Gautheron et al. (2023)'s ChildProject package constitutes a standardized tool for data management, including performance evaluation measures. Next, I propose a short literature review of daylong recording studies to introduce some major studies that exploited the power of these tools.

2.2.2. Daylong recording studies

Daily auditory environments have been investigated in a variety of populations, as young as 2 months of age and in both typically and atypically developing children (Aragon and Yoshinaga-Itano (2012), Caskey et al. (2014a), Christakis et al. (2009), and Warren et al. (2010), see H. Ganek and Eriks-Brophy (2018b) for a review). Research questions ranged from characterizing the auditory environment in specific populations to assessing links between daily auditory inputs and various factors (this second aspect will be developed in section 2.3. Works on typically developing populations revealed that the proportion of adult word counts remained stable while children's vocalizations and child-caregiver interactions increased with age (Gilkerson & Richards, 2008a; Pae et al., 2016a). This was first evidenced in the "Natural Language Study (NLS)", a normative study conducted by the LENA Research Foundation (Gilkerson & Richards, 2008b) that included 329 American families. Daylong LENA recordings were done monthly for a full day in monolingual English-learning infants aged 2 to 48 months. All infants were longitudinally followed for at least six months, and a subset of participants continued for six additional months, yielding a total of 2,682 recording sessions. The authors quantified the linguistic outputs using the three main LENA metrics: Adult Word Count (AWC), Child Vocalization Count (CVC), and Conversational Turn Count (CTC). Regression tests showed no significant correlation between age and AWC, but significant positive correlations between age and both CVC and CTC. These results are consistent with the following studies conducted in other linguistic environments in the same age ranges (Cristia et al., 2021b; Gilkerson & Richards, 2009a; Pae et al., 2016b). Adult-infant interactions were more extensively examined in Warlaumont et al. (2014a)'s study, which looked at the temporal contingencies between adults and key child's speech-like vocalizations. Specifically, they showed that (1) when an adult's speech follows a key child's vocalization, this child's vocalization is more likely to be speech-related; and (2) when a child's speech-like vocalization receives an adult speech feedback, the subsequent child's vocalization is more likely to be speech-like. Overall, these findings suggest the idea of a "social feedback loop" between infants and caregivers that facilitates language development. However, although observed in children with autism spectrum disorder (ASD) as well, this social feedback loop was reduced because of the lower amount of

speech-like vocalizations produced by ASD infants and less contingent feedbacks from the adults. Daylong recording studies in clinical populations were mainly interested in ASD children (Warlaumont et al., 2014a; Warren et al., 2010), preterms (Caskey et al., 2011, 2014b) and infants with hearing loss (Aragon & Yoshinaga-Itano, 2012), but more thorough investigations are needed to better characterize the daily auditory environment in atypically developing infants and provide clinical recommendations.
Other fine-grained descriptions of typically developing infants' soundscapes focused on musical inputs (Hippe et al., 2024; Mendoza & Fausey, 2021) or temporal dynamics (Warlaumont et al., 2022). For example, Mendoza and Fausey (2021) described the musical input received by 6-to-12-month-old American infants at home. Because automatic music annotation is not possible, musical segments were manually annotated and coded based on their features, voices, and tunes. Mendoza and Fausey (2021) found that infants were exposed to "a mix of live, vocal, recorded, and instrumental music". Live music accounted for 30% of the musical input, vocal music 50%, recorded music 75%, and instrumental music 75%. Hippe et al. (2024) used the same methods in a longitudinal follow-up from 6 to 24 months of age and found that on average, infants were more exposed to speech than music and that these musical inputs originated predominantly from electronic devices. Moreover, they showed that while the percentage of speech addressed to infants increased with age, the percentage of music intended for them did not vary. These works are some examples of studies that require manual coding of raw audio data. Another example of manually annotated data lies in works on multilingualism (Oller et al., 2010b; Orena et al., 2019a; Ramírez & Hippe, 2024), as automatic recognition of the spoken language is not possible. For instance, Oller et al. (2010b)'s case study on a trilingual infant involved automatic annotation to extract speech segments, which were then manually annotated to quantify each language in the environment. Finally, daylong recordings have been used in intervention-based contexts, like in Pae et al. (2016a)'s study that showed a positive effect of language feedback to parents on AWC and CTC variables in Korean families. Importantly, this kind of study opens new remediation perspectives for at-risk children.

This paragraph on daylong recording studies exposed some of the research questions that have been addressed using the LENA system. This technology offers many possibilities but also has some limitations. In the next section, I will briefly review the studies that explored the links between daily auditory input and language development.

2.3. The daily auditory environment as a predictor for language skills

The daily auditory inputs described in the previous section play a crucial role in language development. Indeed, a range of studies have shown links between the daily auditory environment in infancy and concurrent or later language abilities.

In section 2.1.1 we saw that speech inputs from the caregivers constitute a great part of infants' auditory environment. Accordingly, researchers wondered to what extent these speech inputs were related to the development of linguistic abilities. Some studies that used in-lab child-caregiver interactions found that quantitative measures of caregiver's speech (number of utterances, number of words, mean length of utterance) were positively correlated with the number of words produced by the child between 14 and 24 months of age (Hoff, 2003; Hurtado et al., 2008; Huttenlocher et al., 1991). The same question was addressed using the LENA system and showed that higher (AWC) was associated with more words produced by the child (Gilkerson & Richards, 2009b). However, it seems that the number of CTC appears to be a better predictor of infants' productions and vocabulary size. Indeed, a higher proportion of adult-child interactions has been associated with larger vocabulary size Gilkerson et al. (2018a). In line with these results, manually annotated data that allowed to differentiate between speech addressed to the child versus overheard speech showed that a higher proportion of child-direct speech was correlated with the amount of child's vocalizations (Oller et al., 2010b), and the expressive vocabulary size at 24 months of age (Weisleder & Fernald, 2013b). More generally, both the quantity and quality of child-caregiver interactions have been associated with linguistic and cognitive developmental outcomes (Gilkerson & Richards, 2009a; Rowe, 2008, 2012; Warlaumont et al., 2014b; Zimmerman et al., 2009). These results indicate that what matters is not only the average number of adult vocalizations, but also the number of child-caregiver interactions.

Individual variations in caregivers' auditory inputs have been explored. For example, in a fine-grained analysis of the subphonemic properties of caregivers' inputs, Cristia (2011) showed that the variability in the acoustic realization of the /s/ fricative consonant predicted infant's ability to discriminate this sound from another very similar fricative consonant (/ʃ/). This suggests that caregivers' inputs play a role at all levels of language development, including phonological representations. On a broader scale, variations in the quantity and quality of the auditory inputs linked with SES have been showed to impact linguistic skills, cognitive abilities, and academic achievement

(Bulgarelli & Bergelson, 2023; Gilkerson et al., 2018a; Hart & Risley, 1995; Hirsh-Pasek et al., 2015; Romeo et al., 2018; Topping et al., 2011). The impact of SES on auditory environment, however, has been recently questioned (Bergelson et al., 2023).

The association between infants' linguistic outcomes and non-linguistic inputs such as TV has been explored as well. Christakis et al. (2009) showed that increased exposure to TV was associated with decreased number and duration of child vocalizations. Greater exposure to TV was also correlated with reductions in the number of adult vocalizations and conversational turns, suggesting a negative impact on child-adult interactions. However, Zimmerman et al. (2009) emphasized that the exposure to TV alone was not responsible for the decrease in infants' vocalizations, but was a consequence of the reduced number of opportunities for adult-child interactions induced by TV watching. Gilkerson and Richards (2009a) found similar negative correlations between average daily hours of TV exposure and the main LENA metrics: AWC, CVC, and CTC. In addition, they found a negative correlation between TV exposure and scores of expressive language.

2. Auditory environment and language development: Summary

The present section provided a description of the auditory inputs infants are exposed to in their first months of life. Caregivers' speech productions and IDS constitute a great part of this input and play a crucial role in infants' development of linguistic and cognitive abilities. The increasing use of daylong recording methods, especially since the release of the LENA system, has allowed the collection of more substantial datasets and more extensive investigation of infants' daily auditory inputs in ecological settings. Finally, the literature showed that quantitative and qualitative measures of early auditory environment can be used as predictors for language outcomes.

The last two sections focused on internal and external components of language development: speech perception abilities, and auditory environment. In the following section, I will introduce the literature on vocabulary acquisition, showing that both internal and external factors play a crucial role in first word learning.

3. Vocabulary acquisition in early infancy

How do infants learn their first words? How do they transition from passive listeners to fluent speakers in such a short period of time? Which mechanisms, strategies, and cues do they rely on? How to study vocabulary acquisition in infants? These questions have drawn researchers' interest for several decades and continue to be explored. Indeed, vocabulary acquisition is a crucial step in infants' journey of language development, as words constitute one of the main linguistic units in all languages. In this section, I first describe the major challenges infants face when acquiring their first words, then I introduce substantial theories about the internal and external factors involved in word learning, and finally, I present my research focus and the experimental procedures that I used in this PhD work.

3.1. Word learning in infancy: challenges

By 24 months of age, typically developing infants' productive vocabulary ranges from 0 to more than 600 words (http://wordbank.stanford.edu/data/?name=vocab_norms). This high individual variability might be explained by the challenges infants face when learning their first words. This first subsection provides a brief overview of why acquiring words as a young infant is such a demanding process.

3.1.1. Speech segmentation

The first challenge for infants is to segment the continuous speech stream. It might seem obvious to fluent speakers that speech is a combination of words ordered in sequences. However, as nicely exposed by Kuhl (2004), "[u]nlike written language, spoken language has no reliable markers to indicate word boundaries. Acoustic analysis of speech shows that there are no obvious breaks between syllables or words [...]."[5] Indeed, acquiring words of the native language(s) as a baby is like arriving in a foreign country as an adult and trying to comprehend an unfamiliar language. All that could be perceived would be a continuous stream of combined sounds. However, unlike adults, young infants overcome this difficulty with no explicit learning and within a few months only. Although spoken language does have pauses, there is not systematically a pause at each word boundary as illustrated in Figure 1.7. Thus, infants cannot rely on brief periods of silence to segment words.

5. Note that some languages do not mark word boundaries in their written forms, especially in ideographic systems where symbols represent ideas and not words, like in Mandarin Chinese for example.

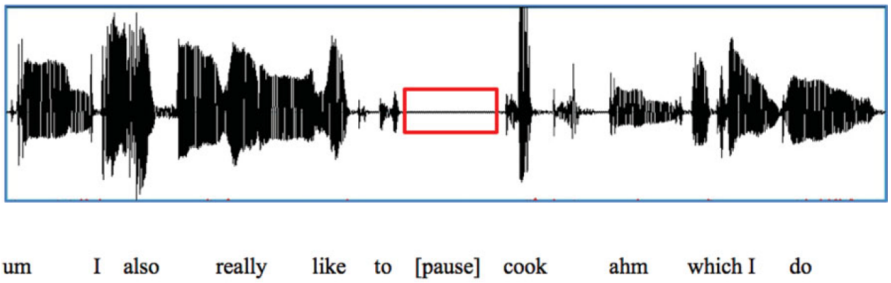


Figure 1.7. – Pauses in speech do not mark word boundaries: example from Kahng (2018). The sound wave of the sentence "um I also really like to cook, ahm which I do" is represented, illustrating a distinctive pause between two segments of the auditory sequence. Notably, there is not a pause at every word boundary.

Some words might occasionally occur in isolation, but this happens too rarely to allow word learning (Aslin et al., 1996; Woodward & Aslin, 1990). Additionally, coarticulation effects, word contractions, and speech rate variations might increase the difficulty of the task. Across and within languages, there is also great variability in word length and number of words in sentences: infants cannot rely on duration cues either. Word segmentation thus constitutes a first tremendous task for young infants.

3.1.2. Word-to-world mapping

A second challenge for word learners is to map the identified sequences of speech sounds (words) with their referents. In a famous philosophical thesis, Willard Quine (1960) exposed this problem through a striking example (Quine & Van, 1960). Let's imagine a linguist wandering in a foreign land, and suddenly seeing a rabbit passing by. The linguist notices a native speaker pointing to the rabbit and saying "*gavagai*" at the exact moment the rabbit runs by. Here comes the dilemma: how does the linguist know what *gavagai* refers to? In this situation, it might refer to the rabbit as a whole, but also to any rabbit's part like its ears or tail; it could mean something more general like "animal" or "mammal"; it might refer to what the rabbit is doing (running, jumping, rushing), or even to the marks that the rabbit leaves on the ground. How to know? Quine emphasizes the infinite possibilities that exist between a word and its mapping to the world. Moreover, it is important to note that a great majority of words have an arbitrary relationship with their referents (De Saussure, 1989; Hockett & Hockett, 1960). This means that the word itself does not contain any information that could help find its referent. This is particularly straightforward when considering the different forms that languages use for the same referent: for example, the word *dog* is *chien* in French, *perro*

in Spanish, *Hund* in German, or *koira* in Finnish. There is no reason why a *dog* is more likely to be called a *chien* than a *Hund*... This is what has been called the "*arbitrariness of the sign*" by Ferdinand de Saussure (De Saussure, 1989). Onomatopoeia and some rare examples of iconic words are exceptions that do not occupy a significant portion of everyday vocabulary. This task of associating speech segments and their referent in the real world relates to the referential nature of words. Inevitably, infants need to understand this property of words to overcome the word-to-world mapping problem (see Section 3.2.1). Evidence for understanding the referential nature of words has been found as young as 3 months of age (Quinn & Eimas, 1996; Quinn et al., 1993).

3.1.3. Generalization and categorization

In connection with the previous challenge, generalization and categorization constitute another demanding set of processes for infants. Going back to Quine's example, mapping a word form to its referent in the real world is not sufficient to learn a new word. Indeed, human minds perceive the world into organized categories that are reflected in their languages [6]. Thus, imagine that the rabbit encountered by the linguist was in fact domesticated and belonged to the native speaker. In this case, *gavagai* could be the rabbit's name and refer only to this particular rabbit; then, another rabbit would not be called *gavagai*. If *gavagai* meant "rabbit", then the linguist should **generalize** this word to all the rabbits. In other words, all rabbits would be included in the same **category** termed *gavagai*. But if *gavagai* meant "small furred animal", then the **generalization** would be larger and expanded to all small furred animals, not only rabbits. In this case, a larger **category** including all small furred animals would be stored under the term *gavagai*. This example illustrates the strong interdependence between generalization and categorization. Indeed, each process can be included in the definition of the other:

- "**Generalization** involves extracting the statistical regularity between individual items or in the sequential order of items. In relation to word learning, generalization involves extending the association of a word and its referent to similar items from the same **category**." (Axelsson et al., 2016, p. 2)
- "Categories are equivalence classes of different (i.e., discriminable) entities and **categorization** is the ability to form such categories and treat discriminable entities as members of an equivalence class. [...] This ability to overlook differences for the sake of generality is important [...]. There are bi-directional relationships between categorization and knowledge: (a) knowing that an animal shares prop-

6. See Koerner (1992) for a review on the Sapir-Whorf hypothesis.

> erties with members of the category 'cats' enables us to categorize this animal as a cat; and (b) knowing that an animal is a cat enables us to predict that the animal has properties of cats. The first ability could be defined as categorization and the second as inductive projection. I will refer to both abilities as inductive **generalization**." (Sloutsky, 2003, p. 246)

These definitions are suited for semantic aspects, but generalization and categorization are also important at the acoustic level. Indeed, different instances of the same word will rarely be produced exactly the same way. Variability can appear in the prosody, in the coarticulation with surrounding words, in the speaker (different speakers), in the speaker's emotional or general state, etc. Thus, children also face generalization/categorization challenges at the acoustic level. They must detect the different instances of the same word, and associate them into the same lexical entry.

3.2. First word acquisition: overcoming the challenges

How do infants overcome the challenges of word learning? Previous research has found that infants rely on a set of external cues and internal cognitive processes. Several theories have attempted to explain how these external and internal factors interact and contribute to word learning. Here, I present some comprehensive models that consider multiple factors and how they support word learning at different stages of development.

3.2.1. Internal factors

To deal with the many challenges associated with word learning, infants are equipped with internal biases, strategies, and cognitive processes that emerge gradually throughout development. As seen previously in this chapter, infants are born with perceptual biases that naturally orient their attention towards speech (see Section 1.2.1).

The constraints/principles theory. It has been proposed that these perceptual biases are bolstered by cognitive constraints described by Golinkoff et al. (1994). The term "principles" was preferred to "constraints", because the latter may indicate some *limitations* to the learner, and reflect the impossibility to reach certain information or knowledge. In the sense of *principles*, constraints "potentiate learning by limiting hypotheses that the learner considers" (Golinkoff et al., 1994, p. 127). Accordingly, *principles* are defined as "intelligent strategies that have the effect of restricting the search space by heightening certain hypotheses over others" (Golinkoff et al., 1994, p. 127). The six principles are organized in two tiers as shown in Figure 1.8. The

first tier (top part) corresponds to principles that are present at the very beginning of word learning. No assumption was made about the innate or acquired nature of these principles, but the first tier is argued to come first in the developmental trajectory and help in first words learning. The second tier (bottom part) can be seen as an "upgrade" or "evolution" of the first one and coincides with the vocabulary spurt. Its principles develop as vocabulary size increases and as new word-learning strategies emerge.

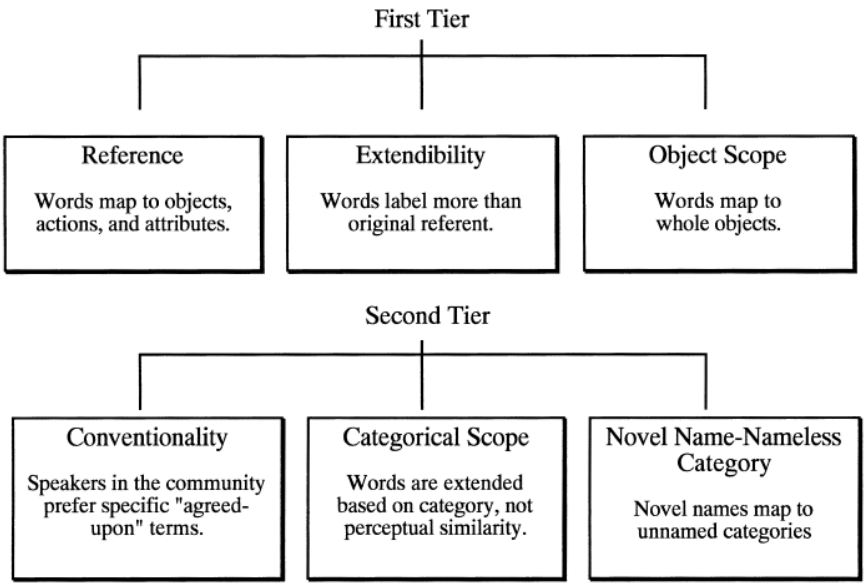


Figure 1.8. – The constraint/principles of word learning proposed by Golinkoff et al. (1994). This figure is a reformatted version from Hollich et al. (2000a).

The six principles can be described as follows:

1) **Reference** relates to the most fundamental aspect of language: it is the fact that "words symbolize, or stand for, objects, actions, or events" (Hollich et al., 2000a, p. 4). Importantly, words have a "stand for" rather than "go with" relationship with their referents, in that they can be used in any context, even when the referents are not present or visible, and still refer to the same entity.
2) **Extendibility** involves that words refer to categories as opposed to proper names that refer to a unique person or entity.
3) **Object Scope** stands for the way objects are preferentially mapped to words: a word will be preferentially associated with an object rather than an action, event, or abstract idea, and to a whole object rather than an object's part or property.
4) **Conventionality** is related to the principle of reference. Words refer to entities of the surrounding world (reference), but because the purpose of language is to

communicate, the phonological forms used to refer to entities must match the phonological forms used and understood by the other speakers. It involves an ability to distinguish between phonological forms used in the familiar context only, and those that will be understood by anyone outside the family circle.

5) **Categorical scope** complements the extendibility principle by relying on taxonomic categories. This principle refines the ability to categorize objects by going beyond perceptual similarity alone (for example, objects' functions).

6) **Novel Name-Nameless Category (N3C)** states that infants will look for nameless objects in their environment when presented with a new word. This principle helps infants map new words to objects for which they do not have associated a phonological form yet.

The WOLVES model. In more recent work, Samuelson (2021) proposed a theory of how words are learned, from hearing a word and identifying its referent in a scene to building robust lexical representations. The Word Object Learning via Visual Exploration in Space (WOLVES) model focuses on the internal processes involved in single concrete noun learning, including low-level perceptual and higher-level cognitive processes. It considers both visual and auditory modalities, constituting the two building blocks of this model: Visual Exploration in Space, and Word-Object Learning (see Figure 1.9).

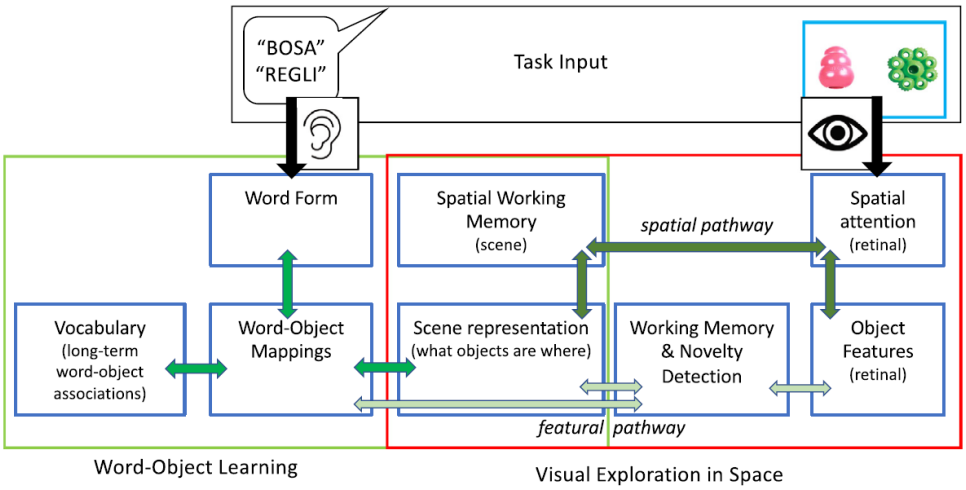


Figure 1.9. – The WOLVES model, from Samuelson (2021).

According to WOLVES, infants process the auditory and visual inputs in parallel when presented with a new word and its referent. Within the Visual Exploration in Space (red box in Figure 1.9), two distinct neural pathways operate to (1) identify object fea-

tures and detect novelty (*featural pathway*) and (2) identify object location and create a memory trace of the visual scene (*spatial pathway*). Here, cognitive processes are involved at various levels, from early visual fields to attentional processes and working memory. Spatial working memory and scene representation bind visual processes with auditory processes of the Word-Object Learning block (in light green in Figure 1.9). Visual scene representation (in which novelty has been detected) connects with the auditory word form representation through word-object mapping processes. Long-term memory traces of the word and referent develop as the number of word-object mappings' repetitions increase, creating robust vocabulary entries. In brief, the WOLVES model "provides insight into how processes of visual exploration, object recognition, working memory, novelty detection, and association learning jointly influence early word-object learning and vocabulary development" (Samuelson, 2021, p. 120).
Importantly, WOLVES model has been argued to account for the inter-individual variability in developmental trajectories of word learning. Samuelson (2021) suggests that this variability might be explained by differences in the development and strength of each process involved, at any level. This implies that infants might use different strategies to compensate for their weakness at a given time in development. For instance, some infants might compensate for weak spatial attention with strong working memory to create robust word-object mapping. Interestingly, this model opens new opportunities for characterizing individual developmental trajectories and predicting word-learning abilities in infancy.

3.2.2. Capitalizing on external cues: the Emergentist Coalition Model

In the previous subsection, I emphasized the internal processes infants rely on to learn words. However, these cognitive processes are not self-sufficient and require external inputs (Kartushina & Mayor, 2019; McMurray et al., 2012; Mollica & Piantadosi, 2017; Polka & Werker, 1994). Here, I will focus on the role of external cues in word learning and how infants use them to improve their word-learning strategies. I present a theory proposed by Hollich et al. (2000b): the Emergentist Coalition Model. In this model, three main types of cues are distinguished and described from more domain-general to more domain-specific:

- **Attentional cues** include domain-general attentional factors like perceptual salience, temporal contiguity, and novelty. Indeed, evidence from previous work showed that young infants are preferably attracted to salient objects from their environment (Cohen, 1972; Tummeltshammer et al., 2014), multimodal synchronization (Bahrick & Lickliter, 2000; Gogate & Bahrick, 1998), and novelty (Saayman et al.,

1964; Tyrrell et al., 1991).

- **Social cues** comprises eye gaze, pointing, and speakers' intention. With experience, infants start to understand that following adults' eye gaze direction, pointing, and attending to the current social context are efficient cues for mapping words to their referents.
- **Linguistic cues** are domain-specific features in the language input such as prosody, syllabic sequences (distributional cues), and grammar.

The emergentist coalition model posits that all these cues are available to infants from birth, but are not used equally at the same stages of word learning. Rather, infants attribute different weights to each of them at different times in development. This weighting changes as infants' experience grows, allowing them to update their strategies and capitalize on the most relevant cues. Indeed, the predominant use of attentional cues at the early stages of word learning tends to orient infants toward the most salient object or the new object in their environment for word-referent mapping. As they encounter more situations where adults name objects that are not the most interesting, salient or the new ones, infants start to understand that the attentional cues are no longer sufficient. Using probabilistic inferences, they change the weighting from attentional to social cues, focusing on the adult's gaze for example. As illustrated in Figure 1.10, attentional cues might be useful at first for understanding the referential nature of language. Then, infants might shift to social and linguistic cues to segment the speech stream into words and map them with their referents.

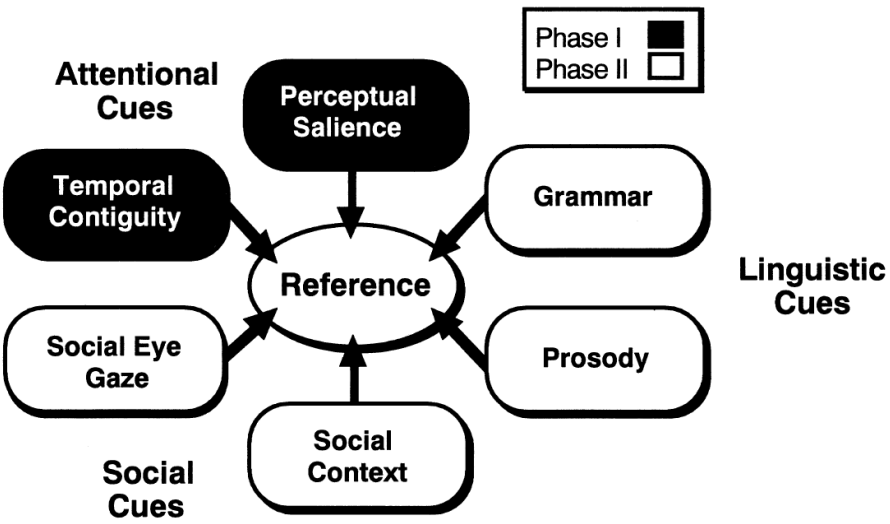


Figure 1.10. – The emergentist coallition model. Figure from *Hollich et al. (2000b)*.

This shift from domain-general (attentional) to domain-specific cues (social, linguistic) is associated with a maturation of the word-learning strategies, hence the "emergentist" aspect of the model. Indeed, the term refers to the idea that the change in cues' weighting reflects "emergent principles of word learning that move from immature to mature, from domain-general to domain-specific, principles" (Hollich et al., 2000b, p. 18). Like in Samuelson (2021)'s WOLVES model, there is this idea of maturation and evolution through learning, both in the cognitive processes and strategies involved and in the cues that are used to map words and objects.
A similar change in weighting occurs for linguistic cues, which are argued to be exploited at all stages of word learning. In the first months of life, infants show a preference for the prosodic patterns of speech and especially for their native language (Höhle et al., 2009; Santolin et al., 2019). Prosodic cues have been argued to play a crucial role in the speech segmentation challenge, but also in word-to-referent mapping processes (Teixidó et al., 2018). The computation of transitional probabilities between syllables has been observed as young as 8 months (Saffran et al., 1996) and allows infants to better understand the characteristics of their language to identify words in the speech stream. Then, morpho-syntactic information helps infants to map words to their referents: these grammatical cues contribute later in development to word learning, between 12 and 18 months of age (Gelman & Taylor, 1984; Katz et al., 1974; Waxman & Markow, 1995).
Meylan and Bergelson (2022) proposed that word learning can be described in three main phases: "brute-force early associations (before ~10 months), a comprehension boost (12–14 months), and leveraging efficiency (after 17 months)" (Meylan & Bergelson, 2022, p. 84). Interestingly, the same changes in the dominant cues as the emergentist coalition model are mentioned: in the first phase, word-referent mappings are bolstered by attentional associative cues; in the second phase, social cues and transitional probabilities are exploited; and in the third phase, all types of cues are processed with increased efficiency, allowing better linguistic prediction and faster word learning.

In sum, the emergentist coalition model depicts an integrative framework that grasps the complexity of word learning. It emphasizes the essential interaction between multiple cues and cognitive processes that evolve through development and experience. Although this kind of model emerged from experimental data, it still needs to be explored. Next, I will focus on methods that allowed the study of vocabulary acquisition in infancy, and present my research interest.

3.3. How to study vocabulary acquisition in infancy?

Previously, I introduced theoretical frameworks that highlighted the crucial role of both internal processes and external cues in word learning. The models presented above showed that word learning is a complex process that involves many factors. Thus, how do we study word learning in infants? In the next subsections, I present my research interest and the main methodological tools that I used in the present work.

3.3.1. Focusing on associative word learning

One approach to addressing the complexity of a process is to decompose it, thereby reducing the scope of research. In this work, I focused on associative word learning, a crucial aspect of vocabulary acquisition. The following paragraphs present the main findings on associative word learning and its developmental trajectory.

"Associative word learning refers to the mechanism for linking a word to a referent" (Tsui et al., 2019, p. 5). In section 3.1, we saw that mapping words to their referents is one of the main challenges infants have to overcome to grow their vocabulary, and that different cognitive strategies are involved at different stages of development. Associative word learning is highly related to other skills such as object categorization, speech segmentation, and word form recognition (Golinkoff & Hirsh-Pasek, 2006; Hollich et al., 2000a). From a developmental perspective, the ability to map words to their referents is observed relatively early, but first in an immature state in that infants attribute a simple "go with" relationship to words and their referents (Nazzi & Bertoncini, 2003). Thanks to repeated co-occurrences of the phonological forms and their associated objects and to developing statistical mechanisms, infants understand the "stand for" relationship and start to create robust lexical representations. Moreover, variability and richness in the auditory input have been shown to be beneficial for new word learning, especially for nouns (Gleitman et al., 2005; Snedeker & Gleitman, 2004). Next, I will review some of the findings from previous work that looked at associative word learning abilities and word comprehension in infancy.

Infants show linguistic associative abilities as young as 6 months of age. Indeed, Tincoff and Jusczyk (1999) were the first authors to show that 6-month-olds could understand some familiar words and successfully associate a phonological form to its referent. In this study, infants were presented with videos showing one of their parents (for example, their mother) associated with either the correct phonological form

("mommy" in that case) or the incorrect form ("daddy" in that same example) in isolation. Results showed that infants looked longer at the screen when the correct word was associated with the video. Later, Bergelson and Swingley (2012) brought additional behavioral evidence of word comprehension between 6 and 9 months of age. In this study, infants were presented with pictures of fruits and body parts, two categories that are very frequent in young infants' daily environment. Results showed that on average, infants looked longer at the picture that matched the auditory stimuli compared to the other picture. Additionally, electrophysiological evidence showed that 6-month-olds were able to learn new pseudo-objects/pseudo-words associations; however, when tested again the next day, results showed that they did not consolidate these new associations in long-term memory, as opposed to 14-month-olds (Friedrich & Friederici, 2011). These results support the idea that although 6-month-olds have associative learning abilities, it is not mature enough to build robust lexical representations. At 7 months of age, however, Gogate and Bahrick (2001) showed that they could associate moving images and isolated vowels when presented synchronously, and remember these associations four days after. Another work from Jusczyk and Hohne (1997) showed that from 8 months of age, infants were more interested in listening to words they were presented with two weeks before than unfamiliar words. These results show that before one year of age infants display simple associative learning abilities, and can recognize words they were familiarized with a few days before. However, associative word learning *per se* develops during the second year of life.

Evidence for associative word learning has been found at 14 months of age. Indeed, using a switch paradigm (see Section 3.3.3), Werker et al. (1998) showed that 14-month-olds, but not 8-to-12-month-olds, were able to form word-object associations with short exposure. These results were obtained in highly controlled conditions, using isolated words produced in IDS that were presented during both the habituation and the test phases. Interestingly, infants recovered the associations only when the objects were moving, suggesting that the object's movement helps infants infer the referential link between the phonological form and the object. These results were later reproduced in the study of Höhle et al. (2020), who additionally found that phonetic variability but not visual variability helped infants to succeed in the associative word learning task. Phonetic variability was introduced by presenting words produced by multiple speakers, resulting in spectral features. In this study, infants had to learn two associations between two pseudo-words and two pseudo-objects, in conditions that were very similar to Werker et al. (1998)'s study. However, in Höhle et al. (2020), the audi-

tory stimuli constituted a minimal pair, differing in the voicing of the first consonant. Several associative word learning studies have used auditory minimal pairs to test the lexical-phonological feature of word learning, which I develop in 3.3.2.

Between 18 and 24 months of age, infants typically shift to more rapid growth in productive and expressive vocabulary size: this is the vocabulary spurt (L. Bloom, 1973; Goldfield & Reznick, 1990a; Nelson, 1973). During this period, infants show increased word recognition (Fernald et al., 1998, 2006) and associative word learning abilities (Gershkoff-Stowe & Hahn, 2007; Werker et al., 2002). Also known as **fast mapping**, the ability to learn new word-object associations with only short exposure has been argued to bolster the vocabulary spurt (Goldfield & Reznick, 1990b; Reznick & Goldfield, 1992). In addition, poor associative word learning abilities have been associated with language delays and Specific Language Impairment (Alt & Plante, 2006; Ellis et al., 2015). More precisely, these studies showed that impaired phonological processing abilities impacted infants' performance in the associative word learning task. For example, in Ellis et al. (2015)'s study infants had to learn two pseudo-word/pseudo-object associations involving a minimal pair. Results showed that both typically developing infants and Late Talkers were able to learn the new associations. However, moment-by-moment analysis of looking time-courses suggested more robust lexical representations and more control of attentional processes in the typical group than the Late talkers. This was observed through differences in looking patterns, indicating different learning trajectories and strategies between the two groups. Late Talkers and infants from the control group matched in their cognitive scores measured with the Bayley Scale of Infant Development (see section 3.3.3), but differed in the Language score and in the number of words understood and produced. This additionally suggests that minimal pair associative word learning abilities predict language outcomes such as receptive and expressive vocabulary size (see also Kemp et al. (2017) and Werker et al. (2002)). In the next subsection, I will develop the phonological-lexical aspect of word learning.

3.3.2. Lexical-phonological aspect of word learning

The lexical-phonological aspect of word learning refers to the role of phonological processes in the emergence of new lexical representations. When learning new words, infants must be able to distinguish between similar-sounding words, such as 'bat' and 'cat,' which have completely different referents. These two words constitute a minimal pair, in that they differ only by one phonological unit: here, the initial consonant. Because all languages of the world comprise multiple minimal pairs, infants must develop

the ability to process them. In particular, this involves a close relationship between refined phonological processing and associative word-learning abilities. As introduced in the previous subsection, this interplay has been studied using minimal pair associative word learning tasks.

In the late 1990s, Stager and Werker (1997) tested 14-month-olds ability to learn new similar-sounding pseudo-words "bih" and "dih". Because the contrast /b/-/d/ is phonemic in English, they expected English-learning infants to easily discriminate between the two pseudo-words and successfully learn the new pseudo-word/pseudo-object associations. Unexpectedly, results did not indicate that infants learned the associations. In a subsequent experiment, Stager and Werker (1997) showed that 14-month-old infants could learn such associations when the pseudowords were dissimilar (e.g., “lif” vs “neem”); these results were further confirmed the next year using the same task (Werker et al., 1998). To check whether the results with the minimal pair were due to impaired phonological processing, Stager and Werker (1997) ran an additional experiment in which infants were tested on their phonetic discrimination abilities. Results showed that they indeed distinguished "bih" from "dih". Pursuing the investigations, Werker et al. (2002)s’ study suggested that the ability to learn new words involving a minimal pair emerged gradually between 17 and 20 months of age. Moreover, they found that 17-month-olds’ performance in the minimal pair word learning task correlated with receptive vocabulary size.

Nevertheless, more recent studies explored the factors influencing successful minimal pair associative word learning and showed more nuanced results. Using the same task as Stager and Werker (1997), Rost and McMurray (2009) showed that **acoustic variability** facilitated infants’ minimal pair associative word learning. In this study, two conditions were presented: in the first one, one speaker produced all auditory stimuli, and in the second one, objects’ labels were produced by multiple speakers. Results indicated that 14-month-olds learned the new associations only in the multiple-speaker condition. Höhle et al. (2020) obtained the same results when testing speaker variability in 14-month-olds, but showed that visual variability did not help learning.

Other factors were shown to have an impact on similar-sounding word learning, such as the **salience** and the **position** of the contrast. Indeed, Curtin et al. (2009) found that 15-month-olds habituated to word-object pairs using the [dit] - [dɪt] contrast successfully learned the associations, whereas infants habituated to the contrast [dit] - [dut]

did not. The authors argued that this was due to the vowels' formant structure: the [dit] - [dɪt] contrast is based on F1 difference (vowel height), whereas the [dit] - [dut] contrast is based on F2 difference (backness). Thus, it has been suggested that because F1 contains more acoustic energy, it is more salient to infants. Based on the same idea of formant difference salience, Archer and Curtin (2018) showed that 14-month-old English learners successfully learned new associations involving the contrast /r/-/l/. Indeed, the authors argued that while the F2/F3 formant transition from bilabials /b/ and /d/ to the vowel /ɪ/ involves minimal change, the transition from the liquids /r/ and /l/ to the vowel /ɪ/ is more salient. Because they showed successful associative word learning using this minimal pair contrast in 14-month-olds, Archer and Curtin (2018) claimed that the unsuccessful learning from Stager and Werker (1997)' infants was due to contrast salience. These results were consistent with Narayan et al. (2010)'s auditory discrimination study showing that acoustic salience can directly impact speech perception. In addition, salience has been shown to interact with the position of the contrast: at 10 months of age, infants successfully discriminated consonants' voicing and place of articulation contrasts, but only when placed in word-initial position; at 16 months of age, infants could discriminate word-final contrast in place of articulation but still not in voicing (Zamuner, 2006). Archer et al. (2016) further found that 12-month-olds successfully discriminated place of articulation contrasts in word-final position for voiced stops but not voiceless stops; yet, they could not discriminate either contrast in medial position. At 20 months of age, however, infants discriminated place of articulation contrasts for both voiced and voiceless stops, and in both word positions. This indicates that voiced stops may be more acoustically salient than voiceless and that salience interacts with position. In Archer et al. (2014)'s study, this interaction between position and salience has been found to impact 14-month-olds' learning of trisyllabic words involving a /b/-/d/ contrast. Indeed, infants succeeded when the contrast was contained in the stressed medial syllable, but not in the initial stressed syllable. These results were explained by the larger difference in F2 transition when the contrast is placed in the medial position, again showing the interplay between position and salience of the contrast. These findings were further strengthened by Höhle et al. (2021) who found that including vowel variability in a /bVk/-/dVk/ place of articulation contrast (where "V" stands for any vowel, thus introducing formant transition variability) facilitated word learning at 14 months of age.

Finally, task-dependent factors have been shown to impact infants' minimal pair word learning abilities. For example, C. T. Fennell and Waxman (2010) explored the effect of

the **referential status** of the labels during the habituation phase. More precisely, they compared three conditions in which the habituation phase comprised: 1) the novel object labeled within a sentence; 2) the novel object and familiar words accompanied with the corresponding label produced in isolation; 3) the same objects as condition 2 but familiar objects's labels were replaced by familiar exclamations. Results showed that infants successfully learned the new object's label in conditions 1 and 2 but not in condition 3. In another work, Thiessen (2007) showed that the **distributional information** of the phonological units involved in the contrast impact infants' minimal pair word learning. This other factor has been studied by presenting infants with the /d/-/t/ contrast, either in the same contexts (ex: "daw"-"taw", or "dawgoo"-"tawgoo") or in different contexts (ex: "dawboo"-"tawgoo"). Results showed that infants' performances were better in the multiple-context condition than in the same-context condition.

More generally, task demands have been argued to play a role in infants' learning of new similar-sounding words (Höhle et al., 2021; Rost & McMurray, 2010). Indeed, most studies introduced above used the Switch task to test infants' minimal pair associative word learning abilities. This task relies on infants' ability to detect a mismatch between what they were habituated to in the first phase and what they are presented with in the test phase. Some authors suggested that this might be very demanding for infants, and showed successful word learning using positive identification tasks (Ballem & Plunkett, 2005; Ellis et al., 2015; Yoshida et al., 2009). In the next subsection, I develop this task-related issue and introduce the paradigms, tasks, and standardized tools that have been used in the context of this PhD work.

3.3.3. Paradigms and standardized scales

Because of infants' limited attentional resources and expressive abilities, indirect measures of word learning abilities and vocabulary size are needed in developmental research. Vocabulary acquisition in infancy has been investigated at both electrophysiological and behavioral levels (Koenig & Cole, 2013). In the present work, I chose to rely on indirect behavioral measures, using eye movement recordings to assess word learning abilities. In addition, I measured expressive and receptive vocabulary size using a standardized scale of infant development. These methods are further described in the next paragraphs.

Looking-while-listening paradigms In contrast to adults, infants cannot be tested on their perceptual or receptive abilities with button presses or verbal answers. Thus,

paradigms involving looking behaviors have been developed in developmental research, such as **preferential looking** procedures. Specifically, the Intermodal Preferential Looking Paradigm (IPLP) involves the simultaneous presentation of auditory and visual stimuli and relies on infants' preference for a given auditory-visual association. During IPLP, two pictures are presented side-by-side and accompanied by an auditory stimulus that matches only one of the two pictures. In each trial, the named object is the target, and the other is the distractor. For instance, in Golinkoff et al. (1987)'s study on word comprehension, 16-month-olds were presented with two simultaneous video clips on two television monitors, and a central speaker produced a speech sound that matched one of the two videos. If a shoe and a boat were displayed, the infant could hear "Where's the boat? Find the boat!". An experimenter sat behind the infants to code their gaze orientation in real time. Latency to look at the expected screen and total fixation time toward the match and non-match screens were analyzed. The authors showed that infants tended to look longer at a familiar named object than at an unnamed object, providing a measure of word recognition (Fernald et al., 1998).

Thanks to increasing technologies and in an attempt to capture real-time language processing time course, a modified version of the IPLP emerged: the Looking-While-Listening (LWL) procedure. LWL has been described as "an offshoot of the IPLP [that] measures the time course of young children's gaze patterns to two images in response to speech, allowing researchers to investigate on line, moment-to-moment speech processing" (Golinkoff et al., 2013, p. 333). Developed by Anne Fernald in the 1990s, this method is now extensively used in language development studies. The procedure is very similar to IPLP but differs in how visual fixation data are coded and analyzed. Indeed, in LWL, infants' gaze is monitored online and recorded using a video camera or an eye-tracker (Swingley, 2011). First tested in Fernald et al. (1998)'s study, the setup was similar to Golinkoff et al. (1987). However, a video camera hidden between the two screens replaced the experimenter/coder of Golinkoff's study and recorded the infants' eye movements. The video was synchronized with stimuli presentation through a superimposed time code, allowing offline frame-by-frame coding of infants' eye movements for the first time. Fernald's new coding method allowed greater temporal resolution for analyzing response latencies from stimulus onset: notably, they showed that 24-month-olds were 300ms faster than 15-month-olds in shifting their attention towards the target picture (Fernald et al., 1998). Manual coding of infants' looking times via hidden camera monitoring is still used today (Archer & Curtin, 2018; Archer et al., 2016; C. T. Fennell & Waxman, 2010; Höhle et al., 2021; Kemp et al., 2017), although some

researchers have attempted to implement offline automatic coding (Archer et al., 2016; Meints & Woodford, 2008).

The LWL procedure can be adapted into several tasks. Among them, the Switch task has been extensively used in the context of early associative word learning research. It consists of two phases: a learning phase and a test phase. During the learning phase, infants are presented with two alternating word-object associations: object A is presented with label A, and object B is presented with label B. During the test phase, two types of trials are presented: "Same" (control) trials and "Switch" trials. In "Same" trials, one of the two newly learned associations is presented (object A with label A); in the "Switch" trials, an incorrect association is presented (object A with label B). If infants successfully associated the objects with their labels during the learning phase, the "Switch" trials that violate this pairing should surprise them and induce greater looking times. However, if the object-label associations were not successfully learned, both types of trials should be equally familiar and yield comparable looking times (Swingley, 2011; Werker et al., 1998).

As mentioned in the previous subsection, some authors argued that the Switch task's demands were too high for infants. Thus, some authors preferred to use the **visual choice paradigm** (sometimes called *preferential looking*). Like in the Switch task, infants are first presented with two new word-object associations, either in a familiarization or habituation phase (see Box 3.3.1). Then, in the test phase, infants are presented with the two new objects displayed side-by-side accompanied by one of the two labels. If infants learned the word-object associations during the first phase, they are expected to look preferentially at the picture corresponding to the label. If not, looking time toward both objects should be equal. Usually, control trials involving familiar words (objects that infants already know) are presented (Ballem & Plunkett, 2005), but not systematically (Mani & Plunkett, 2008). Importantly, Yoshida et al. (2009) reproduced Stager and Werker (1997)'s study on 14-month-olds' ability to learn the word-objects associations involving the "bih"-"dih" contrast, but with a visual choice paradigm instead of the Switch task. Whereas infants did not learn the new word-object associations in Stager and Werker (1997), they successfully learned them in Yoshida et al. (2009). These results suggest that the Switch task's demands might have hindered infants' associative learning abilities.

The LWL paradigm has also been used with eye-tracking, a pupil detection technology that allows to automatically "track with great precision saccades, fixations, gaze durations, regressions, blinks, pupillometry, among many other variables" (Ellis et al.,

More about: Familiarization and habituation.

During the learning phase of visual choice paradigms, the number of trials can be fixed so that all infants are presented with the same number of trials: in this case, this is a familiarization. Another option is to monitor infants' looking time online and determine a "learning" criterion that conditions the end of the learning phase: this is an habituation. The looking time value of each trial is compared to the mean looking time of the previous trials in real time. When the relative looking time drops below the habituation threshold (between 50% and 65% (C. T. Fennell, 2011)),the learning phase ends. The main advantage of habituation is that it adapts to infants' learning pace. However, the maximum number of habituation trials and the habituation criterion should be carefully chosen to avoid high attrition rates or infants' boredom. Although infants might become bored or fail to learn the associations during familiarization, the fixed number of learning trials can be used to determine potential learning differences between two experimental groups (Keller et al., 2004a).

Box 3.3.1 – Familiarization and habituation

2015, p. 26). The development of head-mounted and remote eye-tracking systems has made this technology very convenient for developmental research. Figure 1.11 shows an example of a setup with a remote eye-tracker, barely visible when attached to the screen. Although eye-tracking systems come with their own set of challenges (Oakes, 2012), they override the inconveniences of manual and/or video-based coding (possible errors, experimenter-related costs) and offer more detailed and precise measurements of gaze behaviors. Thus, eye-tracking allows researchers to ask questions that cannot be answered using classical looking-time measures. Indeed, although total looking time can inform about infants' associative word learning abilities (Fernald et al., 1998; Höhle et al., 2021; Kemp et al., 2017), eye-tracking provides more detailed insight into the time courses of the processes involved, specific looking patterns and learning strategies. This has been evidenced in Ellis et al. (2015), who did not find significant differences in mean looking time towards target between typical infants and late talkers. However, they found different looking patterns: late talkers showed fewer fixations, a more narrow looking area, and spent less time fixating on details of the objects compared to typical infants. Late talkers also displayed a bias towards the right side of the screen whereas typical infants looked at both sides of the screen with more global exploration gaze. Finally, analyses of the average time course throughout trials showed that a *target divergent time*[7] could be identified in typical infants but not in late talkers. These divergent

7. "the time at which fixation to the "target" object diverges from fixations to the "distractor"[...]."

looking patterns have been interpreted as differences in control of attentional processes, familiarity/novelty preference, and robustness of the lexical representations. Outside the scope of early language acquisition, other eye-tracking studies have shown that this method offered new insights into infants' processes and how they change with age (Gredebäck et al., 2009).



Figure 1.11. – Example of a setup with the Tobii X3-T120 eye-tracker. The infant is seated on a caregiver's lap, facing the screen where stimuli are presented. The eye-tracker (circled in red) is a screen-based system, clipped directly on the bottom part of the screen and barely visible. *Credits: Estelle Hervé.*

Standardized scales of vocabulary development Considerable individual variability exists in language acquisition. At 24 months of age, some infants can produce hundreds of words, while others do not say a word. To account for this variability and establish normative scales of vocabulary development, several standardized tools have been developed. Here I introduce two of them:

1. The Mac-Arthur Bates Communicative Development Inventories (MCDI) is a set of standardized questionnaires based on parental reports. Created and standardized by (Fenson et al., 1993), the MCDI comprises two versions: the Words and Gestures version for infants from 8 to 16 months, and the Words and Sentences version for infants from 16 to 26 months. For the present work, I used the Words and Sentence version that contains a list of 683 vocabulary items. For each of them, parents are asked to indicate if their child "understands" or "understands and produces"

(Ellis et al., 2015, p. 52)

them. The vocabulary items were chosen to represent the most common words that infants know at this age. MCDI's validity has been reinforced through good agreement scores with other standardized measures of language and behavioral data (Fenson et al., 1994; Fernald et al., 2006; Gilkerson & Richards, 2009a; Stone & Yoder, 2001).

2. The Bayley Scales of Infant Development (BSID) assess infants' development from 16 days to 42 months of age via five subscales: Cognition, Language, Motor, Socio-Emotional, and Adaptive Behavior. The Language subscale is divided into two parts: Receptive Vocabulary and Expressive Vocabulary. Each subscale contains several items that must be administered by a trained experimenter, and in the presence of a caregiver. Indeed, the BSID include the administration of tests that explore acquired skills, learning strategies, and problem solving, as well as direct observation of the child and active participation from the caregiver. The BSID have undergone several revisions, with the most recent being the BSID IV. The BSID can be used for clinical purposes to detect developmental disorders, or in developmental research. Cognitive and Language scores assessed with the BSID have been correlated with infants' auditory input (Caskey et al., 2014b) and language outcomes (Fellman et al., 2004; Heilmann et al., 2005).

If I needed to choose between these two tools, I should consider that the MCDI provides more detailed information about which and how many words infants produce and understand, whereas the BSID yields only standardized developmental scores. However, the MCDI might be less reliable as it is filled by the parents, as opposed to the BSID which is administered objectively by a trained experimenter. Finally, both tools offer different but complementary information about infants' vocabulary acquisition and have been extensively used in developmental studies.

3. Vocabulary acquisition in early infancy: Summary

In this last section, I reviewed the main challenges infants face when learning their first words: segmenting the speech stream, mapping words to their referents, and organizing words into categories. Then I introduced the main theories that attempted to account for infants' ability to transition from passive listeners to speakers, showing that both external and internal cues play a role in this process. Lastly, I reduced the scope of the research to minimal pair associative word learning abilities, focusing on the interaction between word-referent mapping skills and phonological processes. Adapted paradigms and tasks were described, thereby introducing the experimental methods used in the present work.

4. Chapter's summary

This first chapter provided an overview of the primary theoretical and methodological frameworks that underpinned this PhD work. While not exhaustive, the chapter aimed to review the key stages of language acquisition, from infants' first auditory experiences in the womb to first words learning and vocabulary expansion during the second year of life. From a developmental perspective, the chapter emphasized three main aspects of this research:

1. First, I explored infants' speech perception abilities across two crucial brain structures of the processing pathway, introducing key developmental stages of neural maturation. I reviewed pivotal studies that demonstrated the shift from phonetic to phonological speech perception, and the main theories that attempted to explain this developmental change. Finally, I described the cortical MMR and the subcortical FFR, two brain responses that provide complementary insights into the development of speech encoding and constituting potential biomarkers of language outcomes.
2. Second, I focused on the external inputs that participate in the development of linguistic abilities in the first two years of life. After describing the main constituents of this auditory input, I introduced the LENA system and reviewed some key studies that used this technology to describe infants' daily auditory environment. Then, I presented evidence of the interplay between daily auditory inputs and linguistic outcomes.
3. Third, I addressed the question of vocabulary acquisition in the first two years of life. After describing the main challenges associated with word learning, I reviewed substantial theories highlighting the crucial interaction between internal processes and external cues in word learning. Finally, I focused on associative word learning, a key ability that infants develop during their first two years of life. Particularly, minimal pair associative word learning tasks have been shown to efficiently explore the interplay between phonological representations and the emergence of new lexical entries. I end this third section with a description of the paradigms and methodological tools used in the presented work.

Central to this literature review are the roles of external and internal factors in vocabulary acquisition and the interaction between them throughout the developmental trajectory. In the following chapter, I introduce the objectives of this work and the main hypotheses.

Chapter 2. General aims and hypotheses

Contents

1 Objectives of the research 91
2 Main hypotheses 93
2.1 Cortical and subcortical brain responses to speech sounds in early infancy 93
2.2 Daily auditory environment 94
2.3 Minimal pair associative word learning 94

1. Objectives of the research

Chapter 1 highlighted that language development is a complex process involving rapid changes across multiple levels, including anatomical, electrophysiological, environmental, behavioral, motor, cognitive, and social. The last decade of developmental research has shown that solving the great mystery of how infants acquire language is a long and collective inquiry that requires considering all these levels. Consequently, the field has recently witnessed a growing interest in the analysis of multimodal data gathered in longitudinal prospective designs (Dale et al., 2023). The main objective of my PhD work was to **contribute to this scientific effort by longitudinally collecting neurophysiological responses of phonetic encoding, environmental and linguistic factors, and word learning abilities, associated with vocabulary acquisition**.

Longitudinal studies are particularly relevant in developmental research because they capture individual trajectories. As presciently argued by Bates et al. (1995), "The Average Child is a fiction, a descriptive convenience [...]. Theories of language development can no longer rely on this mythical being. Any theory worth the name will have to account for the variations that are reliably observed in early language learning." (Bates et al., 1995, p. 26). This view might be considered too harsh because cross-sectional studies are necessary for describing average developmental trajectories, targeting general devel-

opmental milestones (Kuhl, 2004), and determining critical periods (Werker & Hensch, 2015). However, grasping individual variation with longitudinal data offers a unique opportunity to decipher the complex interplay between the multiple factors involved in language development. Importantly, tracking changes over time provides robust knowledge for designing more tailored clinical interventions that respond to the unique needs of each child.

The first two years of human life represent an ideal window for studying early language development. This period of high brain plasticity provides infants with a strong sensitivity to linguistic inputs, inducing important changes at the neural and behavioral levels. In addition to encompassing critical periods for the emergence of phonological representations, the first two years of life constitute the building blocks of first word acquisition, both at the receptive and productive levels. This is also a period of increasing interactions between linguistic, cognitive, and social skills. Infants start to engage more actively with their caregivers, progressively relying on social cues for language development. Thus, the impact of the daily auditory environment and of the interactions with peers is also growing during this period. Finally, previous research has highlighted the role of early language development in predicting later cognitive, academic, and social outcomes (Duff et al., 2015; Kuhl et al., 2005; Lee, 2011; Schaadt et al., 2023). Thus, delays or difficulties in language acquisition during this period may signal developmental challenges that could impact future learning, making early study crucial for identifying potential issues.

In this context, I created a new multimodal dataset collected longitudinally in a sample of infants aged 3 to 24 months. Crucially, this longitudinal language development corpus allowed us to explore the interactions between external factors and internal processes supporting the building of vocabulary. More specifically, I explored:

1. The emergence of phonological representations at the cortical level as well as spectro-temporal encoding of speech sounds at the subcortical level (Chapter 4). This was done by developing an innovative experimental protocol allowing the recording of infants' cortical and subcortical responses to speech sounds simultaneously in one session.
2. The role of specific linguistic features present in the environment by collecting daylong auditory recordings longitudinally in a weekly manner. Using Natural Language Processing methods, I examined the proportion of speech input infants receive in their daily life, the amount of infant-caregiver interactions, and their

relation with age (Chapter 5).

3. The interaction between phonological and lexical processing during and after the vocabulary spurt period and the relation between associative word learning abilities and vocabulary size (Chapter 6).

This research has been elaborated so that each level of interest constituted an independent study while providing converging information for a potential integrated neurodevelopmental model of language acquisition. The originality of my work comes from the age span of the targeted population, the diversity of the methodologies, the longitudinal aspect, and the nature of the multimodal nature of the measures, which coexist in a single research. Next, I expose the main hypothesis of each study.

2. Main hypotheses

2.1. Cortical and subcortical brain responses to speech sounds in early infancy

In Chapter 4, investigated the developmental trajectory of cortical and subcortical responses to speech sounds in infants. Based on the literature and previous findings, I formulated the following hypotheses:

- **Cortical development hypothesis**: I hypothesized that the cortical contribution to speech sound encoding, reflected by the mismatch response (MMR), will show a developmental trend marked by a decrease in MMR amplitude with increasing age. This would indicate an age-related refinement of cortical processes involved in speech perception, as older infants develop more efficient auditory processing abilities.
- **Subcortical development hypothesis**: At the subcortical level, I expected the frequency-following response (FFR) to exhibit a developmental trajectory characterized by either a U-shaped or logarithmic growth curve. Specifically, I anticipated that the encoding of spectro-temporal features of the speech stimulus would mature with age, as reflected by improvements in various indices of subcortical auditory encoding.
- **Interaction hypothesis**: Finally, I hypothesized that there would be a gradual shift in the relative contributions of cortical and subcortical structures to speech sound encoding as infants develop. While subcortical response might indicate a phonetic encoding of the speech sound features, I expected cortical activity

to become more prominent over time, indicating an increasing contribution of higher-level auditory processing as phonological representations emerge.

2.2. Daily auditory environment

In Chapter 5, I aimed to investigate the auditory environment of 3-to-24-month-old infants in French-speaking households using daylong audio recordings. Based on prior research, I propose the following hypotheses:

- **Validation of LENA system hypothesis**: I hypothesized that the automatic annotations generated by the LENA system would show strong correlations with the annotations provided by other automatic tools with different underlying algorithms. Key metrics, such as adult word count (AWC), child vocalizations (CV), and conversational turns (CT) would be considered to assess the reliability of the LENA system in European French-speaking households.
- **Social loop hypothesis**: I expected an increase in child vocalizations and conversational turns with age. Moreover, based on Warlaumont et al. (2014b)'s results, I expected a social loop revealed by an interaction between infants' and adults' proportion of subsequent speech vocalizations.

2.3. Minimal pair associative word learning

Chapter 6 aimed at investigating the interaction between phonological processing skills and associative word learning abilities in infants at 18 and 24 months of age. Using a minimal pair word learning task based on a voicing contrast, I also explored individual differences by exploring the relation between looking behaviors and vocabulary size. Based on previous findings and the specific goals of this study, I propose the following hypotheses:

- **Developmental hypothesis**: I hypothesized that 24-month-olds would exhibit better minimal pair associative word learning abilities compared to 18-month-olds. Specifically, 24-month-olds should demonstrate higher accuracy, as indicated by a greater proportion of looking time toward the target object in the eye-tracking task. Moreover, older infants were expected to learn faster, requiring fewer familiarization trials than younger infants.
- **Looking patterns hypothesis**: I predicted different time-course trajectories of gaze behavior between the two age groups. Specifically, 24-month-olds were expected to look earlier at the target object during the test phase and show more consistent

gaze patterns across trials and conditions (word learning versus word recognition) compared to 18-month-olds. This would reflect more efficient processing of word-object associations and stronger phonological-lexical representations in the older infants.

- **Interactions hypothesis**: I hypothesized that individual differences in vocabulary size (both productive and receptive) and neurodevelopmental scores would modulate infants' looking behaviors on the associative word learning task. Infants with larger vocabularies and higher cognitive development scores were expected to show better associative word learning abilities, including higher accuracy and faster learning.

Before delving into the exploration of these hypotheses, I review the methodological considerations and challenges associated with research in young infants, in the next chapter. I also provide details about the recruitment procedure and logistical aspects that apply to all subsequent chapters.

Chapter 3. Working with infants, a challenge

Research with young infants constitutes a great challenge at many levels. When conducting experiments in such a vulnerable population, one must ensure the well-being of the participants while meeting ethical, theoretical, and methodological requirements. This chapter addresses some of the logistical and methodological considerations that every developmental scientist might encounter, as well as the strategies I implemented for my Ph.D. work.

Contents

1 Recruitment and logistical aspects 97
1.1 Collaboration with the Babilou group 97
1.2 Le Labo des Minots: a new Babylab in Marseille 100
2 Electrophysiological recordings in infants: challenges and perspectives 102

1. Recruitment and logistical aspects

The recruitment issue was regularly brought out in discussions with peers during scientific events or informal conversations. All developmental scientists are confronted with this question: **where and how will I recruit the babies?** This question gets even more challenging when the research involves longitudinal follow-ups: **what are the best conditions for families to stay involved in the study until the end?** This section shows how I answered those questions and exposes the recruitment and follow-up strategies that I used for the present work.

1.1. Collaboration with the Babilou group

1.1.1. The beginning of the collaboration

When elaborating the recruitment strategy, I had to consider several factors. First, I needed to include many infants between 3 and 24 months of age that would be followed

for several months. Second, I knew that infants would be seen six times along this longitudinal follow-up and that I could accept only a two-week delay around each targeted time point. Thus, I needed to find a way to schedule the recording sessions so that it would not depend on the caregivers' agenda (or find very available parents, which was not the best option). Moreover, although possible, I didn't want to ask parents to bring their children to my lab for data collection, as I didn't have a Babylab at the beginning of recruitment. Because all of my experimental devices were mobile, the ideal solution would have been to visit infants in their daily environments rather than asking families to come to us. Finally, I needed to find a convenient organization to collect the LENA data weekly (see Chapter 5), as parents would not get access to the LENA repository for transferring the data.
With all these constraints, the best option was to rely on daycare centers. By chance, I heard about the Babilou group from the Southern-East region, France, whose team already took part in other research projects and was very interested in contributing to developmental research. I could easily contact the South-Est director of the group, who quickly agreed to start the collaboration. Three Babilou daycare centers from Aix-Marseille metropole were involved: Schuman (run by Christine Leduc), la Duranne (run by Jennifer Jorand), and République Dames (run by Claire Ligneau). Clément François managed all the administrative procedures to obtain ethical approval from the National Committee, and an official collaboration convention was established between the lab and the three daycare centers.

1.1.2. Description of the daycare centers

Schuman. Located within the university campus for literature, arts, languages, and humanities in Aix-en-Provence, Babilou Schuman was the closest daycare center to my lab. It was also the one where caregivers were most familiar with scientific research, as many of them worked at the university or were researchers themselves. However, given the multicultural nature of the academic world, many families were composed of non-native French speakers and were multilingual. The center could accommodate between 40 and 50 infants and had a majority of infants above 24 months old (around 60 %) [1]. Among 38 targeted infants, 37 could be contacted, and a final sample of 20 participated in the study.

La Duranne. Babilou la Duranne was also in Aix-en-Provence but on the periphery of the town. La Duranne is a recent district that hosts many housing but also many companies. Most of the families thus originated from a rather high socio-economic

level. The center could accommodate between 40 and 50 infants and half of them were older than 24 months old[1]. Among 53 targeted infants, 47 could be contacted, and a final sample of 21 participated in the study.

République Dames. Situated in downtown Marseille, this daycare center was the farthest from my lab, typically a 30-minute drive by car, which could extend to an hour during rush hour. Babilou République Dames was a small daycare in a busy area of the city. A greater variability in the socio-economic level was observed among these families. Multilingual families were also frequent in this area. The center could accommodate between 30 and 35 infants, and half of them were older than 24 months old[1]. Among 38 targeted infants, 37 could be contacted, and a final sample of 20 participated in the study.

1.1.3. Recruitment procedure

Families were first met during the annual welcome meeting organized by the managers of the daycare centers. In each center, the project was presented to them as a group, but many parents were absent. Moreover, the "group effect" did not encourage them to volunteer or ask questions. Thus, a second round was necessary, during which parents were met individually when they came to pick up their child at the end of the day. This second round was more time-consuming but also more efficient. Indeed, I targeted only the families that met my inclusion criteria, thanks to the information provided by the daycare centers' managers. During these face-to-face interactions, there was a great variability in parents' interest in the project. Some immediately agreed to participate, others were completely against it, and others needed to take some time to think about it. For parents who said yes, those who wanted to take some time, and the caregivers who could not decide (because they were the grandparents or other relatives for example), documents were systematically provided. They comprised the information sheet with all the details about the study, two versions of the consent form, and a flyer with a QR code to access a short explanatory video (see Annexe A). In the video, the entire project was explained, with methodologies described and illustrated. The documents also included my contact information, so that families could contact me to have more details on the study or update me about their participation choice.
When provided with the list of infants from the daycare centers and their birth dates, I realized that only a few infants started daycare before or at 3 months of age. Thus, I

1. Older infants would leave the center each year and new young infants would arrive, yielding more recruitment opportunities.

decided to include all infants between 3 and 18 months of age. In the end, the collaboration with Babilou allowed 117 families to be targeted according to my inclusion criteria. I was able to contact 107 of those families, and a final sample of 49 agreed to participate in the study. Finally, the success rate of recruitment in the nurseries reached 42% (40%, 53%, and 31% for la Duranne, Schuman, and République Dames respectively). Among the 49 participants, 16 could be seen at least three times, 16 were seen only two times, and the 17 remaining were seen only once, contributing to cross-sectional analyses only (see Annexe B).

1.1.4. Benefits and difficulties

Benefits. As already mentioned, working with daycare centers was very convenient because it allowed to both recruit and collect data directly within the centers: everything happened at the same places. Recording sessions were scheduled directly with the daycare working team, preventing timing issues due to busy parents' agendas. Thanks to this collaboration, I was able to contact families that fitted perfectly with my inclusion criteria, and importantly I could build trusting relationships with both the caregivers and the infants, which was crucial for the longitudinal follow-up. Recording data in infants' daily environment was also a great advantage for the quality of the data: infants were usually calm and cooperative.

Difficulties. The first difficulty I encountered was related to the participants' age. Few infants were younger than 6 or 8 months, so I was unable to include the expected number of 3- and 6-month-olds. Additionally, although I tried to anticipate which families would meet my criteria, not all parents agreed to participate. Parents of very young children were especially inclined to refuse. A second challenge arose during data collection: a childminder needed to be present with the child during recordings. This required careful coordination between my team and the Babilou staff. Specifically, one Babilou worker needed to be available for the entire recording session, which had to be planned in advance. While this arrangement generally worked, there were times when daycare centers were understaffed, preventing the childminder from joining us. As a result, I had to reschedule or cancel some recordings.

1.2. Le Labo des Minots: a new Babylab in Marseille

At the beginning of the third year of the project, a new Babylab, named 'Le Labo des Minots,' was created in Saint-Charles University campus in Marseille. This new facility

became a key hub for recruitment and data collection. Created by Isabelle Dautriche and Clément François, this new Babylab gathers researchers in language development, social cognition, and sensorimotor skills. The 'Labo des Minots' welcomes infants from 3 months to 12 years of age and is equipped with EEG, fNIRS, and eye-tracking systems, a motricity circuit, a cabin for on-screen experiments, and two Faraday cages. This environment was perfectly adapted to my project. Thus, I started a collaboration shortly after its opening.

Families were contacted by a list of participants who were already involved in other studies run in the Babylab and had agreed to be contacted for other studies. Although this list was constrained by the inclusion criteria of those studies, it generally aligned with the requirements of my project. With this arrangement, I was able to recruit and record 21 children in the end. However, I could not find a way to record LENA data weekly with families from the Babylab. Therefore, I decided not to ask these families to participate in the LENA recordings. Moreover, the longitudinal follow-up was harder to maintain with these families. Most of these infants were seen only once or twice and mainly contributed to cross-sectional EEG data. Thanks to my collaboration with the Babylab, I was able to increase the number of young participants (6, 8, and 10 months old) who were difficult to recruit in the daycare centers.

Compared to daycare centers, it was faster and easier to recruit young infants in the Babylab (between 6 and 10 months of age), because (1) the recruitment was carried out by the lab manager, so it was less time-consuming, (2) parents could register on a participants list, so I was sure to contact only families interested by scientific research, and (3) because infants usually started daycare "late" (after 8 or 10 months of age), I recruited few infants below 8 months of age; there was not such constraint at the Babylab. However, the Babylab was not adapted for LENA data recordings, and follow-ups were more difficult to maintain, possibly because the relationships with the families were not as strong as those with daycare participants.

In sum, both recording locations had their advantages and disadvantages. Nevertheless, some methodological challenges inherent to my recording devices and targeted population were faced, independently of the site. Next, I address this topic by focusing on EEG in infants, which constituted an important part of this work.

2. Electrophysiological recordings in infants: challenges and perspectives

Decades of developmental research have led to adapted tools and methods specifically tailored to the unique capabilities and limitations of infants. However, experimental procedures in infants remain challenging because they require extensive care and thoroughness. This is particularly true for EEG, an electrophysiological device that I used to assess early speech perception abilities in infants (see Chapter 4).
During my PhD, I learned how to use EEG and to apply it to research involving young infants. Throughout this process, I encountered various challenges and discovered the existing procedures and pipelines from the literature. My documentation and learning journey is summarized in Hervé et al. (2022), an overview of the challenges and perspectives associated with developmental cognitive studies involving electrophysiological recordings. This paper also includes a set of guidelines for systematic electrophysiological data collection and analysis in this specific population. The full journal version of the paper is included in Annex C.

Hervé, E., Mento, G., Desnous, B., & François, C. (2022). Challenges and new perspectives of developmental cognitive EEG studies [Publisher: Elsevier]. NeuroImage, 260, 119508. https://doi.org/10.1016/j.neuroimage.2022.119508 (cit. on pp. 4, 40, 41, 88).

The following chapter is a striking example of how can EEG be used to investigate speech perception abilities in infants at the cortical and subcortical levels.

Chapter 4. Development of cortical and subcortical responses to non-native speech sounds during infancy

Contents

1 Introduction 104
2 Methods 106
2.1 Participants 106
2.2 Stimuli 106
2.3 Procedure 107
2.4 EEG data processing 109
2.5 Statistical analyses 113
3 Results 116
3.1 Subcortical FFRs 116
3.2 Cortical ERPs 123
4 Discussion 133
4.1 Recording simultaneous cortical and subcortical responses in infants: feasibility 133
4.2 Subcortical encoding of speech sounds 134
4.3 Cortical responses to speech sounds 138
4.4 Limitations and future directions 142
5 Conclusion 142

1. Introduction

The emergence of robust speech encoding abilities is critical for language development in human infants. Independently of the native language, the developmental trajectory of language functions systematically involves a shift from universal to language-specific speech perception (Kuhl, 2004; Werker & Tees, 1984b). Importantly, consonant perception becomes language-specific between ten and twelve months (Best & McRoberts, 2003; Tsushima et al., 1994; Weikum et al., 2012; Werker & Tees, 1984b). The opposite developmental trajectories for native versus non-native phonetic contrasts discrimination known as perceptual narrowing (Cheour, Ceponiene, et al., 1998; Polka & Werker, 1994; Rivera-Gaxiola, Csibra, et al., 2000; Werker & Tees, 1984b) may result from an interplay between an initial sensitivity to the acoustic-phonetic features of speech sounds, a constantly increasing exposure to the native language, and neural maturational changes (Gervain & Werker, 2013; Maurer & Werker, 2014; Vouloumanos & Werker, 2007).

The first two years of life constitute a critical period for early speech perception wherein important neuroplasticity mechanisms such as myelination occur from the brainstem to the auditory cortex (Holland et al., 2014; Moore et al., 1995; Su et al., 2008; Werker & Hensch, 2015). At the same time, the neural system's high sensitivity to the environment (Hensch, 2005) also plays a critical role in shaping early sensory systems, as revealed by studies in premature babies or sensory-deprived animals (Duffy & Mitchell, 2013; François et al., 2021; Pasman et al., 1992). Early phonological development is thus a perfect model for deciphering the interplay between environmental and genetic factors during human development (Werker & Hensch, 2015).

The maturation of the auditory pathway has been extensively studied during early development using cortical ERP (see Hervé et al. (2022)). Specifically, the mismatch response (MMR) is elicited by deviant stimuli inserted in a series of repeated standard stimuli (Näätänen et al., 2007). The MMR provides a unique measure of auditory discrimination, constituting a window to the neural representation of the standard stimulus and the short-term memory trace it creates (Näätänen et al., 2007). Several studies have reported positive MMRs in young infants (Garcia-Sierra et al., 2011; L. Liu et al., 2023; Virtala et al., 2022), interpreted as an immature brain response (He et al., 2007, 2009; L. Trainor et al., 2003; L. J. Trainor, 2012). Interestingly, the MMR has been successfully used to study the encoding of various speech features (Cheour, Alho, et al., 1998; Cheour, Ceponiene, et al., 1998; Dehaene-Lambertz, 2000b; Dehaene-Lambertz

& Dehaene, 1994; François et al., 2021; Pang et al., 1998; Rivera-Gaxiola, Csibra, et al., 2000). Changes in VOT or formants can give insights into the early developmental changes associated with perceptual narrowing for consonants and vowels (Cheour, Ceponiene, et al., 1998; L. Liu et al., 2023; Rivera-Gaxiola, Csibra, et al., 2000). For instance, Cheour, Ceponiene, et al. (1998) showed in a cross-linguistic longitudinal design that exposure to native language shapes early speech perception by inducing a switch from acoustic to phonological processing.

Subcortical responses to syllables provide crucial information on the development of infants' encoding of relevant speech features. Recently, the frequency-following response (FFR), a neural correlate of subcortical encoding of syllables or musical tones, has gained interest in human auditory physiology Kraus and Chandrasekaran (2010). The FFR is generated mainly by the brainstem (G. Bidelman & Powers, 2018; Coffey et al., 2016, 2019, 2021) and corresponds to a sustained, phase-locked activity of subcortical neurons that are fine-tuned to precise spectro-temporal features of sounds (Kraus et al., 2017). Recent work in infants and newborns suggests that the FFR is an ideal brain index of the early development of the auditory brainstem response (S. Anderson et al., 2015; Escera, 2023; Gorina-Careta et al., 2022; Ribas-Prats et al., 2019). Neonatal FFRs to syllables suggest that the encoding of speech sounds' periodicity is already mature at birth (Arenillas-Alcón et al., 2021; Ribas-Prats et al., 2019; Ribas-Prats et al., 2022), although the tracking of temporal features may continue to improve until 1 month of age (Ribas-Prats et al., 2023). The fine-grained encoding of the fundamental frequency (F0) still matures until three months of age (Jeng et al., 2010, 2016), whereas the encoding of speech sounds' fine structure and harmonics continues to mature up to adulthood (S. Anderson et al., 2015; K. L. Johnson et al., 2008; Skoe et al., 2015). A gap still exists in our understanding of the development of subcortical responses to complex sounds between 6 and 24 months of age. Further, only a few adult studies have collected and analyzed both ERPs and FFRs in adults (G. M. Bidelman, 2015; Elmer et al., 2017; Galbraith & Kane, 1993), but no study has systematically analyzed cortical and subcortical responses to speech sounds in infants.

Here, I investigated the early developmental changes of MMRs and FFRs to non-native artificial speech sounds in a within-subject semi-longitudinal design. I developed an innovative multi-feature oddball paradigm in which infants were passively exposed to highly controlled synthetic syllables that differed only by the trajectory of the second formant (F2; see Figure 4.11a). The stimuli were presented at two stimulation rates

to gather cortical ERPs and subcortical FFRs optimally in one experimental session. Crucially, the brain responses collected during my paradigm in a semi-longitudinal design allowed me to study the early developmental trajectory of two processing steps of the auditory hierarchy known to reflect the fidelity of speech sound encoding. Such an experimental paradigm was chosen to provide new insights into how cortical and subcortical structures may contribute to the early development of speech perception. The first objective was to assess the feasibility of this paradigm in young infants, by obtaining cortical and subcortical responses comparable to those found in previous studies that used separate recording sessions. Then, based on previous work, I expected that the cortical contribution in the encoding of speech sound should increase with age, compared to the contribution of subcortical structures. Specifically, I expected a decrease in MMR amplitude with age. At the subcortical level, I expected either a U-shaped or a logarithmic growth curve trajectory in the encoding of temporal and spectral features of the syllable with age, reflected by various indices of the FFR.

2. Methods

2.1. Participants

A total of sixty-seven healthy full-term babies were included in this experiment at 6, 8, 10, 18, or 24 months (respectively T_6, T_8, T_{10}, T_{18} and T_{24}). They had no history of hearing impairment, ear infection, or neurological disorder and were raised in a dominant French-speaking environment. Three additional babies were recruited but excluded from the analysis because of technical issues (n=1) or fussiness (n=2). My final sample included both cross-sectional and longitudinal data, for a total of 126 recordings. Infants were recruited and tested in three daycare centers in southeast France (Aix-Marseille metropole). The study has been approved by SUD-EST VI Protection of People Committee (2022-A02281-42), and informed consent was obtained from the caregivers before the recordings. Infants received a "baby researcher" diploma and a t-shirt for compensation.

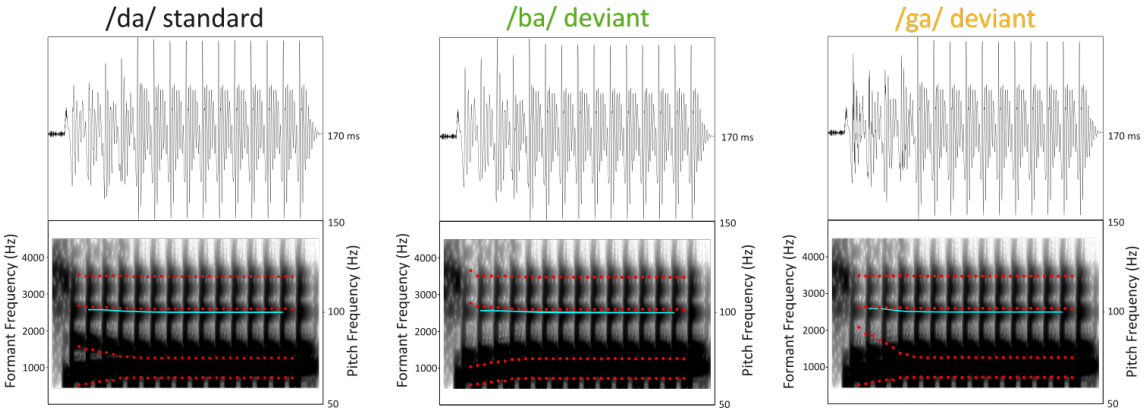
2.2. Stimuli

Stimuli were complex auditory sounds that consisted of three synthesized syllables: /da/, /ba/, and /ga/ (see Figure 4.11a). These syllables were created with a Klatt-based synthesizer (Klatt, 1980) by Skoe and Kraus (2010). The three syllables' fundamental frequency and total duration were identical, with values of 100.3 Hz and 170 ms, respec-

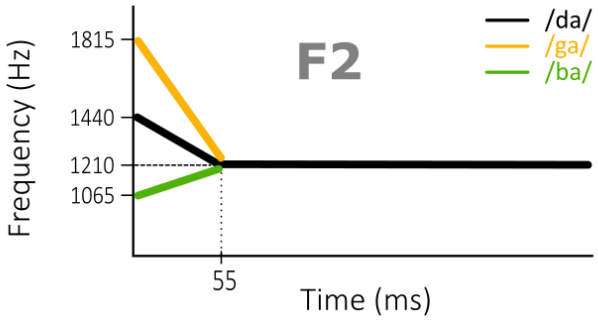
tively (10 ms onset period, 45 ms consonant transition, and 115 ms steady-state vowel). Importantly, the three stimuli differed only by the trajectory of the second formant in the consonant-to-vowel transition part (/da/: 1440-1210 Hz; /ba/ 1065-1210 Hz; /ga/ 1815-1210 Hz; see Figure 4.1b). In French, voiced stops also differ in the VOT: here, the syllables' acoustic features correspond to English phonemes. All other formants had the same ranges of values: the first formant (F1) from 550 to 700 Hz, the third formant (F3) from 2500 to 2420 Hz, and the fourth formant (F4) from 3245 to 3020 Hz. In the vowel part, all the formants were stable in all syllables (F1: 700 Hz; F2: 1210 Hz; F3: 2420 Hz; F4: 3020 Hz; see Figure 4.11a).
Stimuli were delivered to the right ear (as in Hornickel et al. (2009) and Jeng et al. (2010)) at 67-dB volume through an ear-inserted shielded earphone (Etymotics ER 3; Etymotic Research Inc., Elk Grove Village, IL, USA) using custom MATLAB® scripts. Volume was calibrated with a Bruel and Kjaer artificial ear (model 4152) and a 2250-Light-G4 sound level meter beforehand and controlled during data recording through a Fireface UCX sound card (RME, Haimhausen, Germany). To reach the high temporal precision required for FFR recordings, triggers were included in the second channel of a stereo audio track, thus containing the auditory stimulus in the first channel and the trigger in the second channel. A splitter was used to send the stimulus to the inserted earphone on one side and the trigger to the BIOSEMI receiver on the other. The inserted earphone and the stimulation cable were shielded to minimize electric leakage from the stimulation equipment to the scalp electrodes.

2.3. Procedure

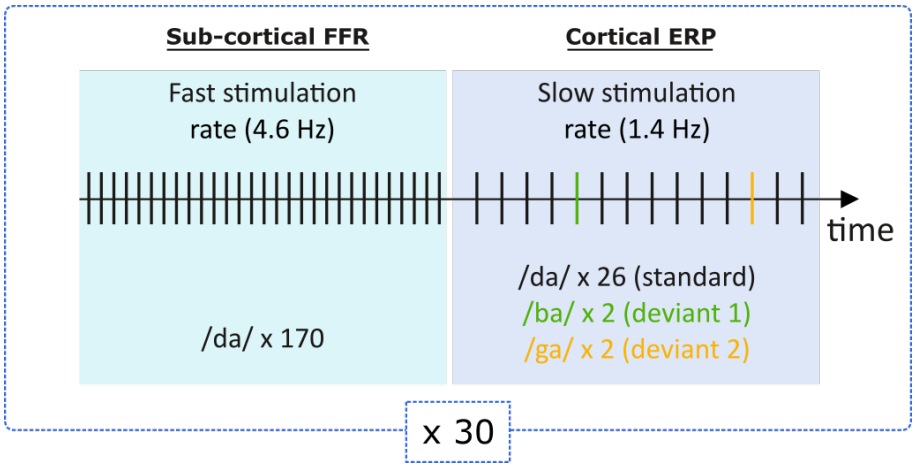
The entire recording session took place in dimly lit and quiet rooms of the daycare centers in the presence of a professional daycare worker. Younger infants were tested while seated on a caregiver's lap or a play mat. Older infants (18 and 24-month-olds) could also be seated on a child-friendly foam sofa. After a period of 10 to 20 minutes of preparation, infants were presented with the "clustered stimulus presentation" paradigm adapted from G. M. Bidelman (2015) for about 30 minutes. This procedure allows for recording both the subcortical and cortical brain activities, which require different numbers of trials to be robustly analyzed.



(a) The three CV syllables used during the interleaved FFR/ERP stimulation paradigm: /da/, /ba/, and /ga/. The sound wave and spectrogram are represented for each stimulus. The red dots depict the four first formants, and the blue trace depicts the fundamental frequency. The three syllables lasted 170 ms and had a fundamental frequency of 100.3 Hz. The /da/ syllable was used in both fast and slow stimulation rate sequences.



(b) Visualization of the F2 trajectory of each stimuli.



(c) Illustration of one stimulation bloc, divided into Fast Stimulation Rate (FSR) and Slow Stimulation Rate (SSR) sequences. Each block started with an FSR sequence that contained 170 repetitions of the /da/ syllable presented at a 4.6 Hz rate. SSR sequences consisted of 30 syllables: 26 standards /da/, 2 deviants /ba/, and 2 deviants /ga/, presented in a semi-random order at a 1.4 Hz rate.

Figure 4.1. – Stimuli and experimental design.

Stimuli were delivered in alternating blocks of fast (4.6 Hz) and slow (1.4 Hz) stimulation rates (see Figure 4.1c), with a Stimulus Onset Asyncrony (SOA) of 220 ms and 710 ms, respectively. In the Fast Stimulation Rate (FSR) sequences, the syllable /da/ was presented 170 times in alternating polarities (Ribas-Prats et al., 2019) to isolate the neural response by minimizing stimulus artifact and cochlear microphonic (Aiken & Picton, 2008; Gorga et al., 1985). In the Slow Stimulation Rate (SSR) sequences, a series of 30 stimuli was presented with the standard /da/ and two types of deviants, /ba/ and /ga/. The SSR sequences were created following the multiple deviant oddball paradigms (François et al., 2021; Partanen et al., 2013) with 26 standards and four deviants (2 for each of the two deviant types; 6.67% probability for each deviant). A total of 30 blocks with alternating FSR and SSR sequences were presented.
The EEG signal was recorded at a sampling rate of 16 384 Hz using a Biosemi amplifier system (Biosemi ActiveTwo, Amsterdam University) from 16 active Ag-Cl scalp electrodes mounted on an infant-sized elastic cap at standard positions (International 10/20 system sites: Fp1, Fp2, F3, F4, T7, C3, C4, T8, P3, P4, O1, O2, Fz, Cz, Pz, and Oz). Additionally, two electrodes were placed on the right and left mastoids for offline re-referencing and one electrode was placed at the Fpz location for FFR recordings.

2.4. EEG data processing

Data processing was performed using custom scripts based on MATLAB® version R2022b (The MathWorks Inc, 2022), EEGLAB® toolbox (Delorme & Makeig, 2004), and adapted functions of the Brainstem Toolbox (Skoe & Kraus, 2010). Before applying cortical and subcortical processing steps separately, continuous data was visually inspected. From this observation, bad channels were identified: if one or two channels were detected, they were interpolated using EEGLAB *pop_interp* function; if more than two channels were bad, the whole recording was excluded from analyses.

2.4.1. Subcortical FFRs

The FFR activity was extracted from the EEG data gathered during the FSR sequences only. The signal was computed using the two mastoids and the Fpz electrode as required by the Biosemi device ($(Left) + (Right)/-2 = Fpz - (LA + RA)/2$). Next, data were band-pass filtered between 80 and 1500 Hz (Ribas-Prats et al., 2023), then segmented into -40 to 200 ms epochs from stimulus onset. Epochs with activity exceeding ± 30 μV were removed from further analyses as recommended in Ribas-Prats et al. (2019), Ribas-Prats et al. (2022), and Ribas-Prats et al. (2023). The mean number of trials

rejected for FSR epochs was 438.8 (sd = 384.5; 8.6% ± 7.5%). The neural responses to positive and negative stimulus polarities were averaged by adding them ([Rarefaction + Condensation] /2). This was done to eliminate the cochlear microphonic, an electrical potential generated by the cochlear hair cells that resembles the FFR in many aspects (Chandrasekaran & Kraus, 2010). Finally, the delay induced by the Etymotic Earphone tube was added to the mean FFR response. Recordings with less than 2500 remaining trials were excluded from analyses (n = 5). A visual inspection of the temporal FFR was performed and allowed to detect 7 additional recordings to be rejected. A final sample of 114 recordings were kept for analysis (n(T_6) = 7, n(T_8) = 17, n(T_{10}) = 32, n(T_{18}) = 31 and n(T_{24}) = 27). I extracted measures of temporal (time domain) and spectral (frequency domain) encoding from these 114 individual FFRs. These measures are listed and detailed below.

a. Time domain

Time-domain measures were computed to obtain indicators of neural phase-locking to the stimulus periodicity and magnitude of the response.

Neural lag. The neural lag was considered to estimate the FFR latency, reflecting the transmission delay induced by the auditory system's neural conduction velocity (Arenillas-Alcón et al., 2021). It was obtained by computing a cross-correlation between the stimulus and the neural response. First, an r-value was calculated to determine the extent to which the two signals were correlated as a function of the time shift between them. The neural lag corresponded to the time shift between 3 and 10 ms (Arenillas-Alcón et al., 2021; Ribas-Prats et al., 2019; Ribas-Prats et al., 2022; Ribas-Prats et al., 2023) with the highest absolute r-value. To perform the cross-correlation, the stimulus was first band-pass filtered between 80 and 1500 Hz and resampled to 16384 Hz to match the frequency content and the sampling rate of the neural response, as recommended by Krizman and Kraus (2019) and Skoe and Kraus (2010).

Signal-to-noise ratio (FFR). The signal-to-noise ratio of the response (SNR_{FFR}) was taken as an index of the magnitude of the FFR, as well as subcortical response quality. Specifically, I computed the Root Mean Square (RMS) in the prestimulus period (-40 to 0 ms from response onset) and in three different time windows corresponding to three regions of interest in the stimulus: the whole response (0 to 170 ms from response onset), the consonant-to-vowel transition part (10 to 55 ms from response onset), and

the vowel part (55 to 170 ms from response onset). Individual variability in response-to-stimulus delays was taken into account by adding individual neural lag values to each of the selected regions (except the pre-stimulus region). The RMS was obtained by computing the square root of the mean squared values at each timepoint within the given time window (Skoe & Kraus, 2010). The Signal-to-Noise Ratio (SNR) corresponded to the ratio between the RMS of the prestimulus period and the RMS of each of the three time windows, as follows: $SNR_{FFR} = RMS_{FFR} \ / \ RMS_{prestim}$.

b. Frequency domain

Frequency-domain measures were obtained to assess participants' encoding of the pitch and formant structure of the speech sound /da/. Specifically, analyses focused on the encoding of the F0 (100.3 Hz) and the first formant F1 (550-700 Hz in the CV transition, 700 Hz in the vowel part). Because the values of the second formant F2 (1210-1440) were close to the upper limit of the filter applied to the data and the FFR phase-locking limit (1500 Hz; Krizman and Kraus (2019)), I chose not to explore F2 encoding.

Spectral amplitude at frequency peak. The Fast-Fourrier Transforms (Fast-Fourrier Transform (FFT)) of the neural response within the CV transition (10-55 ms) and the vowel part (55-170 ms) were obtained using the *fft()* MATLAB® function. Then, I obtained the spectral amplitudes at F0 peak (100.3 Hz) and F1 peak (700 Hz) by computing the maximum peak amplitude between 99.3 and 101.3 Hz (100.3 ± 1 Hz) and between 699 and 701 Hz (700 ± 1 Hz) for each frequency peak respectively. Spectral amplitude was considered as an indicator of the magnitude of neural phase-locking at the two frequencies of interest.

Signal-to-noise ratio (F0 and F1). The signal-to-noise ratios around the F0 peak (SNR_{F0}) and the F1 (SNR_{F0}) peak were used to estimate the accuracy of the response's spectral content encoding. For the (SNR_{F0}), maximum *signal spectral power* was expected between 95.4 Hz and 105.4 Hz: this frequency window was chosen to match a ± 5 Hz frequency window around F0 peak (100.3 Hz), as recommended in previous studies (Arenillas-Alcón et al., 2021; Ribas-Prats et al., 2019; Ribas-Prats et al., 2022). The *noise spectral power* frequency window was determined using the same studies' criteria: the lower band window of the noise was computed as the difference between the lower filter limit (80 Hz) and the lower edge of the signal's window (105.4 Hz), which led to a delta

of 15.4 Hz. This delta was reported at the upper band of the noise window, starting at the signal's higher limit (105.4 Hz). In the end, the *noise* windows were [80-95.4] and [105.4-120.8] Hz. The mean spectral power was calculated and then normalized using the following formula (Ribas-Prats et al., 2022):
SNR_{F0} = 10 * log_{10} (Signal spectral power/Noise spectral power)
The same procedure was applied to the thirteen harmonics of the F0 (HH_{1-13}), using the same frequency deltas for signal and noise. To obtain the (SNR_{F1}), the mean SNRs of the fourth (501.5 Hz), fifth (601.8 Hz), and sixth harmonics (702.1 Hz) were averaged (HH_{4-6}) to match the frequency values of F1 in the CV transition (550-700 Hz), as recommended by Krizman and Kraus (2019). The SNR of the sixth harmonic (HH_6) was taken for the SNR_{F1}) of the vowel part (700 Hz).

Pitch error and pitch strength. Pitch-tracking accuracy of the F0 contour was assessed by computing a cross-correlation between the F0 contour of the stimulus and the F0 contour of the response. This computation was performed within the time window corresponding to the steady part of the stimulus (55-170 ms); for the response, the neural lag was added to the onset of the time window. The pitch error corresponded to the mean absolute Euclidian distance between these two F0 contours, where lower values indicated higher pitch-tracking accuracy. While pitch error measured the accuracy of pitch encoding, pitch strength was used as a measure of the magnitude of neural phase-locking (periodicity) to the stimulus F0.

2.4.2. Cortical ERPs

Only the SSR sequences were included for cortical ERP analysis. The EEG signals were re-referenced offline to the averaged activity of the mastoids and band-pass filtered between 1 and 30 Hz using a butterworth filter (order 2, *pop_basicfilter* EEGLAB function). These data were split into -100 to 500 ms epochs from stimulus onset and were baseline corrected. Epochs with activity exceeding ± 150 μV were rejected from further analyses as done in previous studies (François et al., 2017, 2021). The first three trials of each sequence (always standards) were systematically removed from the analyses to override novelty effects and provide a sufficient number of standard trials to create a short-term memory trace (Näätänen et al., 2007). Finally, an additional epoch rejection was performed manually by visually browsing the remaining trials. After all these preprocessing steps were performed, 1 recording with less than 150 standard (/da/) trials and 10 of each deviant (/ba/ or /ga/) was excluded from analyses. Individual averages were calculated for the standard (/da/) and for each deviant (/ba/ and /ga/), and grand

averages per group were computed. The difference wave was obtained for each deviant by subtracting the averaged response to the standard to the averaged response to the deviant. Further visual inspection of the individual averaged responses to each stimulus allowed for rejecting 17 additional recordings. For the 108 remaining recordings, the mean SSR epochs rejection rate was approximately 34% (sd = 14% ; /da/ standard: 39%, sd = 12%; /ba/ deviant: 31%, sd = 14%; /ga/ deviant: 32%, sd = 14%).

Measures of cortical ERPs included latency and amplitude of the maximum peak, and mean amplitudes within different time windows of the averaged responses to standards and the difference waves. First, the automatic auditory response was assessed by analyzing the obligatory P150 component of the responses to standards, a reliable indicator of auditory processing in early infancy and precursor of the adult P1 component (Čeponien, Rinne, & Näätänen, 2002; Coch & Gullick, 2012; Kushnerenko, Ceponiene, Balan, Fellman, Huotilainen, & Näätänen, 2002). The peak amplitude and mean amplitude between 50 and 250 ms were extracted, as recommended by Kushnerenko, Ceponiene, Balan, Fellman, Huotilainen, and Näätänen (2002). Then, responses to deviants were explored by extracting peak amplitude and mean amplitudes of the difference waves, within two different time windows. The choice for the time windows was data-driven, and determined by computing cluster-based permutations as recommended by Werwach et al. (2022). The procedure for time-windows extraction is detailed in the next section on statistical analyses.

2.5. Statistical analyses

Our final dataset consisted of both cross-sectional and longitudinal data. Some participants were recorded only once at one of the five targeted ages (n=31); infants seen at least two times or more were included in the *longitudinal* sample (n=97 of the 126 recordings, involving 36 different infants). Therefore, data were explored using three approaches: "group comparison", "longitudinal sample", and "age-based regressions". For the **group comparison** analyses, data were split into two groups depending on the recording age: the "young" included data collected at T_6, T_8, and T_{10}, and the "old" included recordings at T_{18} and T_{24}. The group separation between 10 and 18 months was chosen because of the critical developmental shift in phonological representations of consonants during this period (Eimas et al., 1971; Werker & Tees, 1984b; Werker et al., 1981). Statistical analyses consisted of comparisons between young and old.
The **longitudinal sample** included only participants that were seen at least two times, and data were split into five age groups corresponding to the five assessment points

ages (T_6, T_8, T_{10}, T_{18} and T_{24}). Statistical analyses consisted of multiple comparisons between assessment point groups.
Age-based regressions consisted of linear models with the age (in days) of participants as the fixed factor.

Because rejection criteria differed between cortical and subcortical data processing, the number of recordings included in each analysis differed as well. Thus, subcortical and cortical response analyses are described separately.

2.5.1. Subcortical FFRs

After applying the rejection criteria (see section 2.4.1), 114 recordings were kept for subcortical response analyses (n(T_6) = 7, n(T_8) = 17, n(T_{10}) = 32, n(T_{18}) = 31 and n(T_{24}) = 27). Normality was assessed for all variables using Shapiro-Wilk tests. Depending on the distribution of each variable, parametric or non-parametric tests were performed. When necessary, data were log-transformed to obtain a normal distribution and analyzed again with parametric tests to strengthen statistical power.

Group comparison. The *young* group included 56 recordings (mean age = 279 days, SD = 45.1 days, range: 176-322 days; mean birth weight = 3346 g; SD = 404 g, range: 2700-4500 g; mean gestational age = 40.1 weeks; SD = 1.32 week, range: 37-42.1 weeks) and the *old* group had 58 recordings (mean age = 639 days, SD = 91.9 days, range: 540-754 days; mean birth weight = 3420 g; SD = 368, range: 2700-4300 g; mean gestational age = 40.1 weeks; SD = 1.34 week, range: 37.3-42.1 weeks). For non-normally distributed variables, non-parametric Wilcoxon signed-rank tests were performed and p-values were reported. T-tests were applied to normally-distributed data, with p-values reports.

Longitudinal data. From the 114 subcortical recordings, 80 were included in the longitudinal sample (n(T_6) = 6, n(T_8) = 11, n(T_{10}) = 19, n(T_{18}) = 22 and n(T_{24}) = 22), for a total of 58 different infants. After testing the normality of the data distributions with the Shapiro-Wilk test, either one-way ANOVAs or Kurskal-Wallis tests (an equivalent for non-parametric ANOVA) with assessment point (T_6, T_8, T_{10}, T_{18}, T_{24}) as the independent variable were applied to each measure. When p-values indicated significant group effect, post-hoc tests were conducted using Tukey's Honest Significant Difference and paired Wilcoxon signed-rank tests for parametric and non-parametric variables respectively.

Age-based regressions. All recordings (n=114) were included in these analyses. Regression with age was assessed for each measure using Pearson-product correlations and linear models with the participant's age (in days) as predictor.

2.5.2. Cortical ERPs

After data rejection (see section 2.4.1), 108 recordings were kept for cortical response analyses (n(T_6) = 7, n(T_8) = 19, n(T_{10}) = 28, n(T_{18}) = 27 and n(T_{24}) = 27). For difference wave analysis, non-parametric cluster-based permutation tests (Maris and Oostenveld (2007), $p < .05$, $\alpha = 0.05$, 1000 permutations, 2 channels minimum cluster size) were performed using Brainstorm and MATLAB® version R2022b (The MathWorks Inc, 2022), to identify relevant time windows and electrodes for which the amplitude of standard and deviant ERPs significantly differed. The tests were applied independently for each deviant and assessment point to frontal, central, and parietal electrodes (F3, Fz, F4, C3, Cz, C4, P3, Pz, P4), at all time points between 100 and 500 ms after stimulus onset, as recommended in Werwach et al. (2022). All time windows (including the 50-250 time window applied to standard ERPs) were then split into three smaller same-sized time windows to account for potential variability in the components' shape across age (Werwach et al., 2022). Normality was assessed for all variables using Shapiro-Wilk tests.

Group comparison. The *young* group included 54 recordings (mean age = 274 days, SD = 45.4 days, range: 176-322 days; mean birth weight = 3335 g; SD = 406 g, range: 2700-4500 g; mean gestational age = 40.1 weeks; SD = 1.32 week, range: 37-42.1 weeks) and the *old* group had also 54 recordings (mean age = 646 days, SD = 92.1 days, range: 540-754 days; mean birth weight = 3455 g; SD = 362, range: 2700-4300 g; mean gestational age = 40.1 weeks; SD = 1.3 week, range: 37.3-42.1 weeks). Statistical analyses consisted of three-way ANOVAs with age group (6-10 months, 18-24 months), laterality (left, center, right), and anteroposterior location (frontal, central, parietal) as factors, and main measures (peak latency, peak amplitude and mean amplitudes) from the determined time windows as the dependent variables. When p-values indicated significant effects, post-hoc tests were conducted using Tukey's Honest Significant Difference tests.

Longitudinal data. From the 108 cortical recordings, 77 were included in the longitudinal sample (n(T_6) = 6, n(T_8) = 13, n(T_{10}) =16, n(T_{18}) = 21 and n(T_{24}) = 21). Three-way ANOVAs with assessment point (T_6, T_8, T_{10}, T_{18}, T_{24}), laterality (left, center, right), and anteroposterior location (frontal, central, parietal) as factors, and main measures (peak latency, peak amplitude and mean amplitudes) as the independent variables were com-

puted. When p-values indicated significant effects, post-hoc tests were conducted using Tukey's Honest Significant Difference tests.

3. Results

3.1. Subcortical FFRs

3.1.1. Group comparison

Time domain. Temporal representations of the averaged FFRs of each group are depicted in Figure 4.2 A. All variables of the time domain followed a non-normal distribution. Corresponding group comparison statistics (see Table 4.1) indicated no significant difference in the neural lag (p = .80). However, significant increases in the SNR_{FFR} of the whole response (p = .039), the CV transition (p = .013), and the vowel (p = .044) were found between the younger and the older group. Nonetheless, as shown in the violin plots reported in Figure 4.3, there seem to be many outliers in the dataset, although the greater number of high values in the older group might indicate a substantial tendency for better response quality after 18 months. Nevertheless, further exploration of the data without outliers (values higher than 1.5 interquartile range (IQR) below first quartile (Q1) or higher than 1.5 IQR above Q1, as recommended in Ribas-Prats et al. (2022)) was conducted. The three SNR_{FFR} variables could be log-transformed to obtain normal distributions and compute parametric statistics. T-tests on these subsets of data indicated only a significant difference for the SNR_{FFR} of the CV transition (p = .021).

Frequency domain. Spectral contents of the CV transition and vowel segments of the averaged FFRs of each group are represented in Figure 4.2 B. All variables of the frequency domain followed a non-normal distribution, except the pitch strength and the spectral amplitude at F1 frequency (700 Hz) in the CV transition part. Corresponding group comparison statistics (see Table 4.1) indicated significant differences between the two groups for the SNR_{F0} of the vowel (p = .045), the SNR_{F1} of the CV transition (p = .049) and the vowel (p = .036). Pitch error significantly decreased between the younger and the older (p = .002), and pitch strength significantly increased (p = .003). After outlier removal, all effects were maintained. However, the spectral amplitudes at F0 (100.3 Hz) and F1 (700 Hz) in the CV transition and vowel parts of the response yielded different results before and after outlier removal. First, no significant effect was found when considering the whole dataset. After outliers removal (using the same IQR method as for the time domain variables), the four variables of the spectral amplitude could be

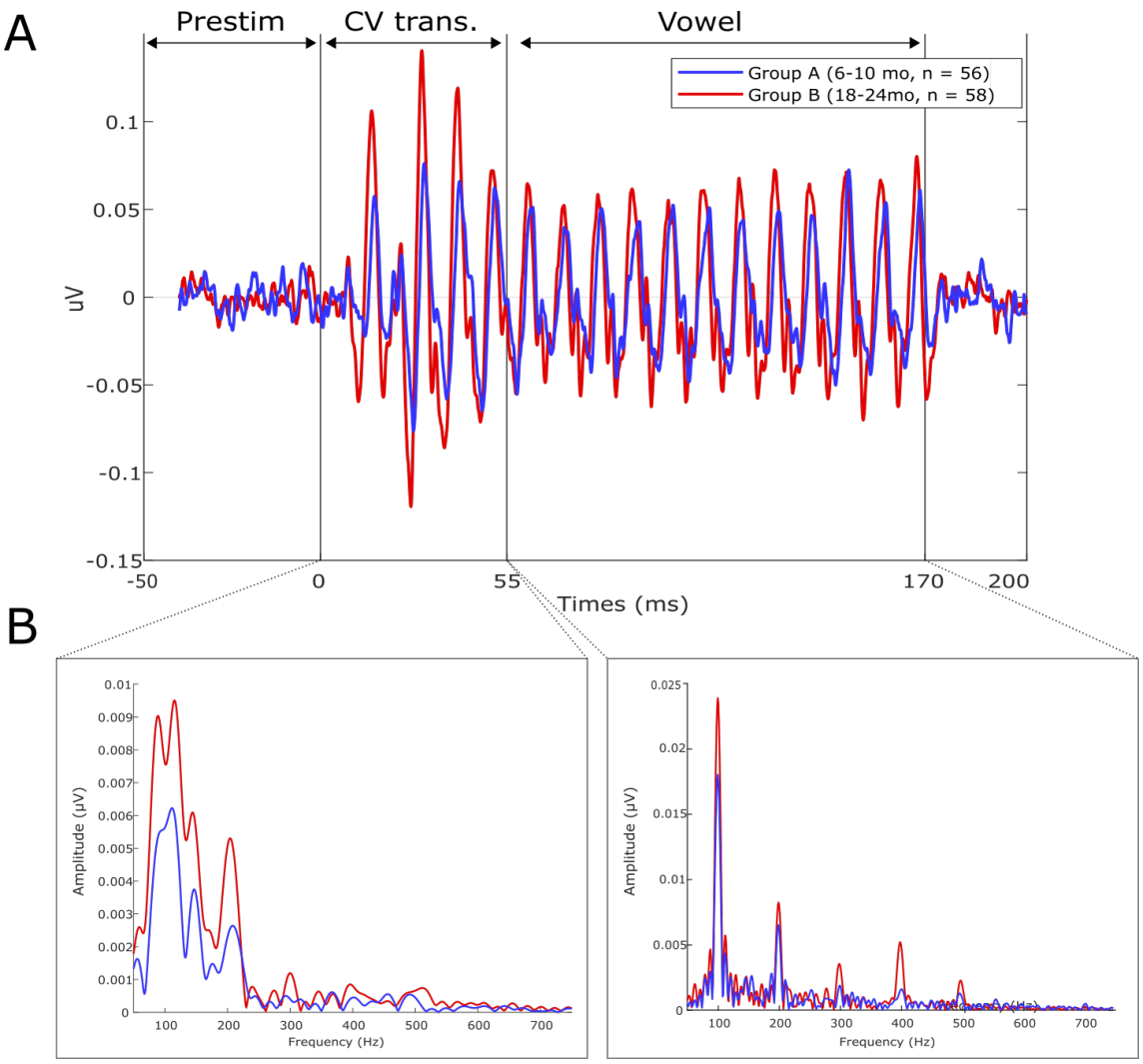


Figure 4.2. – Temporal (A) and spectral (B) representations of the grand averaged FFRs by group. B: The Left panel indicates The spectral content of the CV transition (10-55ms), and of the vowel part (55-170ms) is shown in the left and right panel, respectively.

log-transformed. T-tests indicated a significant difference in the spectral amplitude at F1, only in the vowel part of the response (p = .05). No statistical differences were found for the three other measures of spectral amplitude (at F0 in the CV transition, at F0 in the vowel transition, and at F1 in the CV transition).

3.1.2. Longitudinal sample

Time domain. Temporal representations of the averaged FFRs at each assessment point are depicted in Figure 4.4. All variables of the time domain followed a non-normal

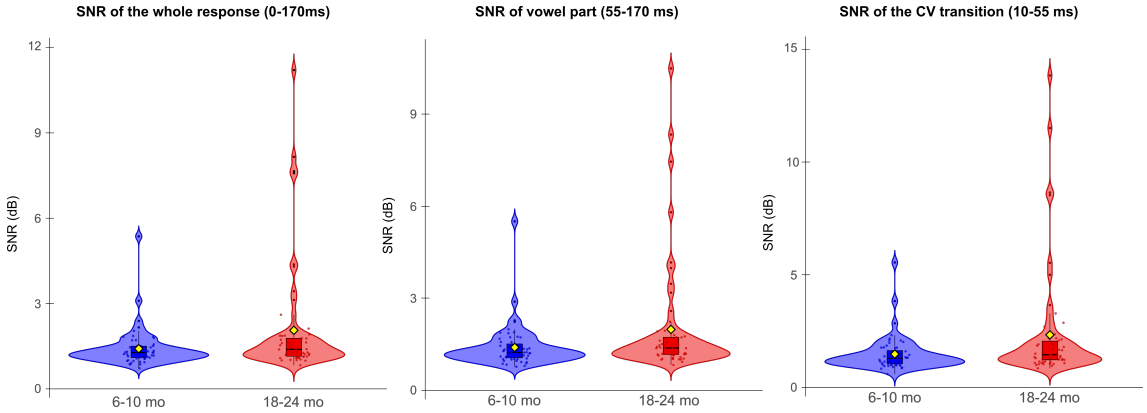


Figure 4.3. – Violin plots of the SNR_{FFR} for the three time windows of interest: whole response, vowel, and CV transition.

distribution. Like in the group comparison analysis, multivariate analysis of covariance with assessment point (T_6, T_8, T_{10}, T_{18}, T_{24}) as the independent variable did not show any significant change in neural lag (p = .54). However, a significant effect of assessment point was found for the SNR_{FFR} computed on the whole response (0-170ms, p = .037) and the CV transition part (10-55 ms, p = .004) after outliers rejection. A marginally significant effect was found for the SNR_{FFR} of the vowel part (55-170ms, p = .093).

Frequency domain. When computed on the whole longitudinal sample, about half of the subcortical measures in the frequency domain had a normal distribution (SNR_{F0} in the vowel part, SNR_{F1} in CV transition and vowel part, pitch strength, spectral amplitude at f1 peak for CV transition and vowel), and the others were non-normally distributed (SNR_{F0} in CV transition, pitch error, amplitude at F0 peak for the CV transition and vowel part). After outliers removal, all measures were normally distributed, for which ANOVA results are reported here. A strong assessment point effect was found for the SNR_{F0} in the vowel part (p < .001), but not in the CV transition. Similarly, a significant effect was also found for the SNR_{F1} in the vowel part (p = .007), but not in the CV transition. ANOVAs on both measures of pitch tracking (pitch error and pitch strength) indicated significant assessment point effect (p = .002 and p = .04 respectively). Finally, all four measures of the spectral amplitude (at F0 and F1 peaks, in the CV transition and vowel part) were significantly modulated by the assessment point (see Table 4.1 for p-values). For three spectral measures that yielded significant assessment point effect (SNR, amplitude at frequency peak in the CV transition, and amplitude at frequency peak in the vowel part), mean values were computed and developmental patterns are illustrated in Figure 4.5.

	Group comparison		Longitudinal sample	
Measure	**All data**	**Without outliers**	**All data**	**Without outliers**
Time domain				
SNR_{FFR} of the response (0-170 ms)	**.039** * [b]	.057 [a][c]	**.02** * [e]	**.0375** * [d]
SNR_{FFR} of the CV transition (10-55 ms)	**.013** * [b]	**.0215** * [a][c]	**.006** ** [e]	**.004** ** [d]
SNR_{FFR} of the vowel (55-170 ms)	**.044** * [b]	.14 [a][c]	.055 . [e]	.093 . [d]
Neural lag	.80 [b]	.80 [b]	.54 [e]	.54 [d]
Frequency domain				
SNR_{F0} of the CV transition (10-55 ms)	.84 [b]	.71 [b]	.48 [e]	.24 [d]
SNR_{F0} of the vowel (55-170 ms)	**.045** * [b]	**.045** * [b]	**.034** * [d]	**<.001** *** [d]
SNR_{F1} of the CV transition (10-55 ms)	**.049** * [b]	**.016** * [e]	.57 [d]	.20 [d]
SNR_{F1} of the vowel (55-170 ms)	**.036** * [b]	**.016** * [e]	.16 [d]	**.007** ** [d]
Pitch error	**.002** ** [b]	**.001** ** [b]	.058 . [e]	**.002** ** [d]
Pitch strength	**.004** ** [a]	**.004** ** [a]	**.04** * [d]	**.04** * [d]
Spectral amplitude at F0 (CV transition)	.35 [b]	.48 [b]	**.04** * [e]	**.003** ** [d]
Spectral amplitude at F0 (vowel)	.32 [b]	.88 [a][c]	**.04** * [e]	**.006** ** [d]
Spectral amplitude at F1 (CV transition)	.71 [a]	.55 [a][c]	**.001** ** [d]	**<.0001** *** [d]
Spectral amplitude at F1 (vowel)	.11 [b]	**.05** * [a][c]	.14 [d]	**.02** * [d]

Table 4.1. – Results of the comparative statistics computed in the group comparison and longitudinal analyses. P-values are reported.
[a]T-test; [b]Wilcoxon signed-rank test; [c]Log-transformed; [d]One-way ANOVA; [e]Kruskal-Wallis test.

3.1.3. Age-based regressions

Interactions between all FFR metrics and age were explored. Figure 4.6 reports the correlation matrix that was obtained from these tests, with *r* values reported. Colors indicate the *p*-value. The correlation matrix revealed that age predicted the values of the SNR_{FFR} (at all time windows), the SNR_{F0} in the vowel part only, the pitch error,

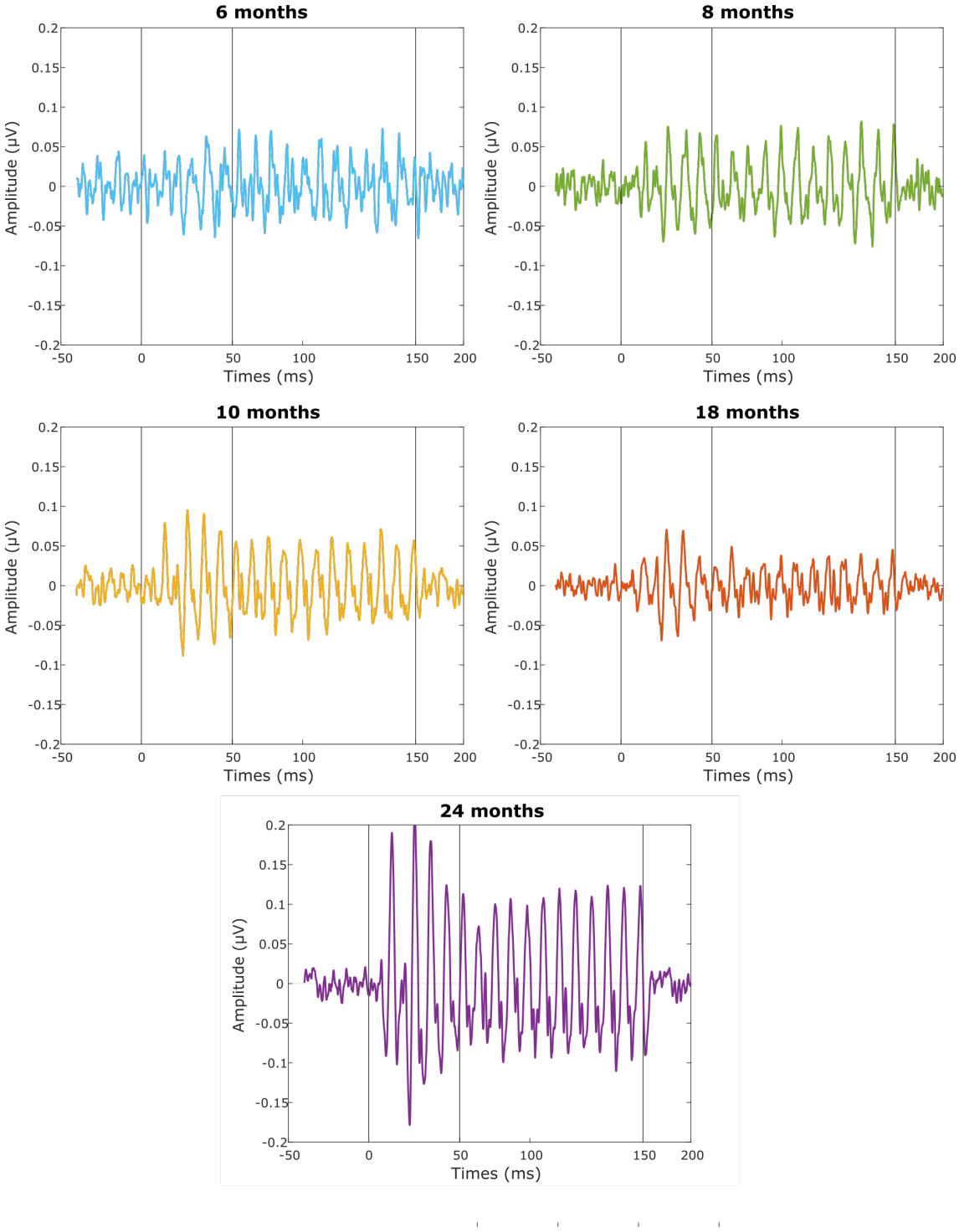


Figure 4.4. – Temporal representations of the grand averaged FFRs by age group.

the pitch strength, and the amplitude of F1 in both time windows (CV transition and vowel region). As expected, most metrics of the temporal domain were correlated to

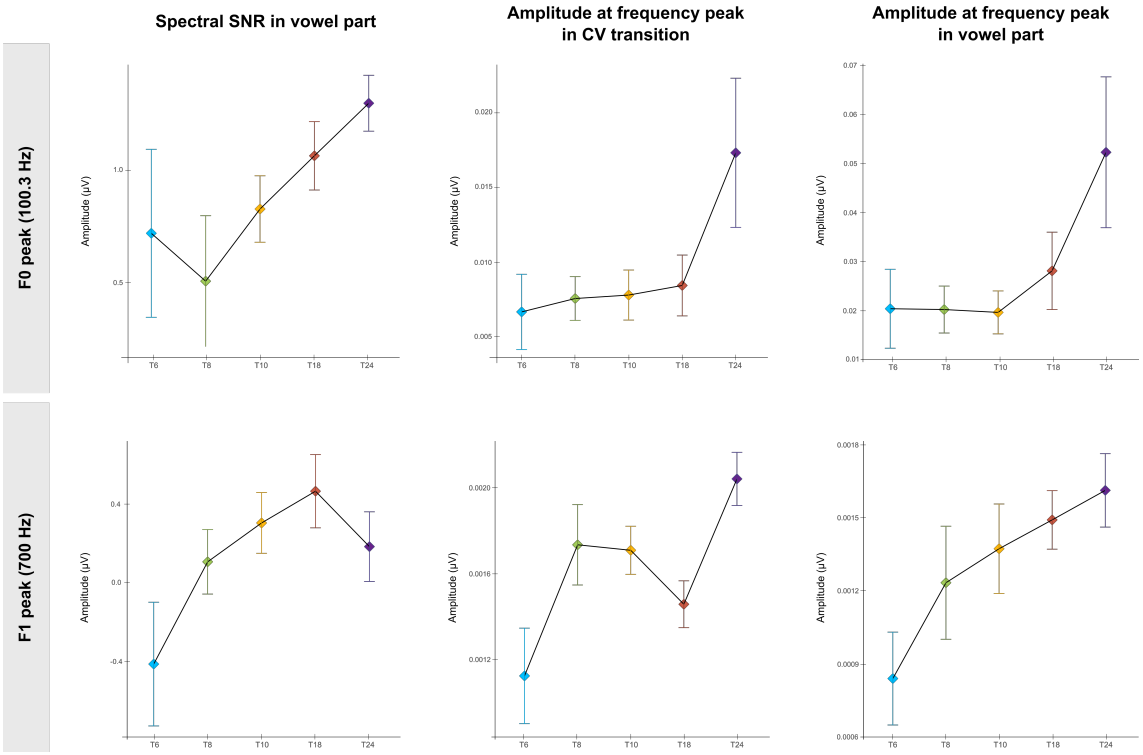


Figure 4.5. – Illustration of the mean values for FFR measures in the frequency domain where significant effects of group were found (spectral SNR, spectral amplitude at frequency peak in the CV transition and in the vowel part at both F0 and F1 peaks) across the five assessment points.

one another, except the neural lag, and most metrics of the frequency domain were correlated to one another. Interestingly, some metrics of the temporal domain were also correlated with metrics of the frequency domain, like the pitch tracking metrics (pitch error and pitch strength), which were correlated with all other metrics except the neural lag.

The highest correlation scores with age were found for pitch error and pitch strength, which seem to be reliable indicators of speech encoding maturation. As seen in figure 4.7, pitch tracking error was negatively correlated with age (r = -0.34, p < .001) whereas pitch strength was positively correlated with age (r = 0.32, p < .001). This was not surprising, as pitch strength has a strong negative correlation with pitch error (r = -0.75).

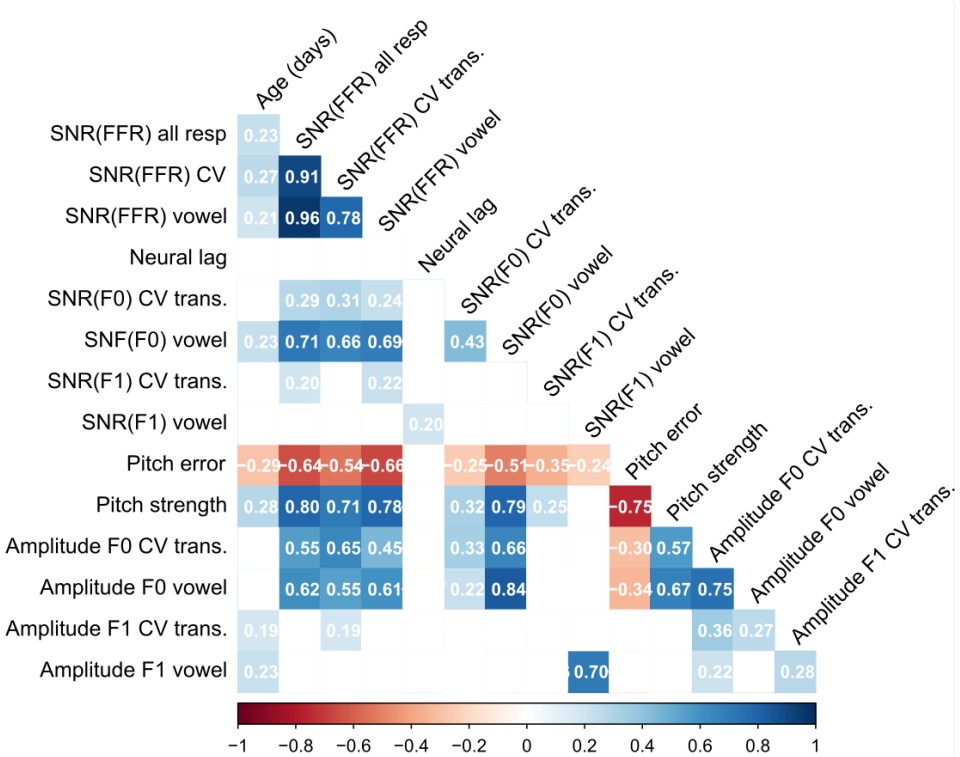
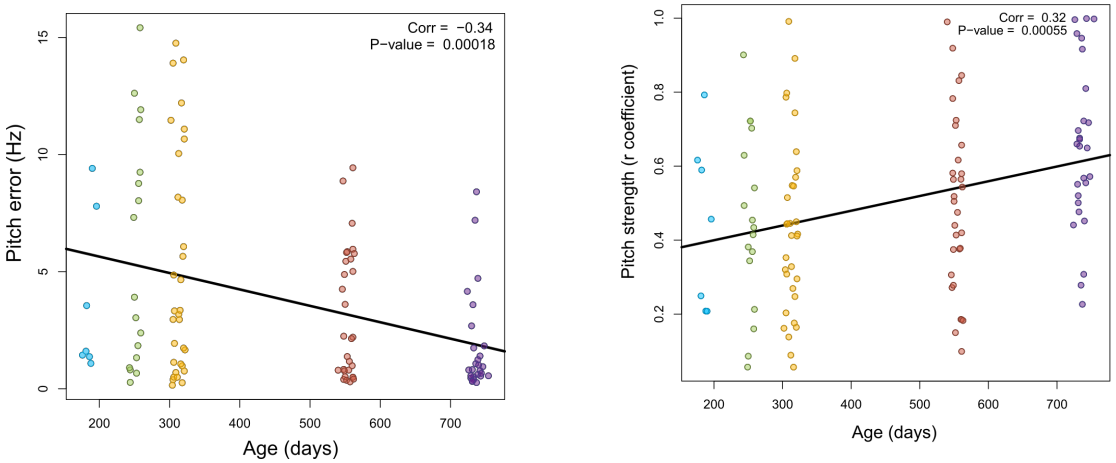


Figure 4.6. – Correlation matrix between age (in days) and all FFR measures. The colors indicate the *r* value: the darker the color, the higher the *r* value. Red and blue squares indicate negative and positive *r* values respectively. White numbers indicate the *p*-value.



(a) Pitch error as a function of age

(b) Pitch strength as a function of age

Figure 4.7. – Correlations between pitch tracking metrics (pitch error and pitch strength) and age. Colors indicate age: see Figure 4.4 for color code.

3.2. Cortical ERPs

3.2.1. Cluster-based permutations.

To determine the time windows in which the response to standards significantly differed from the response to deviants, I applied cluster-based permutation analyses as described above (see Section 2.5.2). After computing all significant clusters, I extracted the longest cluster for each deviant from all assessment points and electrode groups (frontal, central, parietal). The longest cluster for the deviant /ba/ was found at 8 months of age at parietal electrodes, between 300 and 480 ms. The longest cluster for the deviant /ga/ was found at 24 months at frontal electrodes, between 180 and 300 ms. The two chosen time windows for difference wave analyses were thus **180-300 ms** and **300-480 ms**.

3.2.2. Group comparison

Response to standards. For the cortical ERPs, I first assessed the quality of the responses to the standard /da/. Figure 4.8 shows the mean response to the standard for 6-18 months vs. 18-24 months groups at C3 electrode. As described by Čeponien, Rinne, and Näätänen (2002), the auditory ERP is dominated by the P1 component (or P150, a positive component that peaks at around 150 ms after stimulus onset). Thus, Figure 4.8 indicate the obligatory ERP P150, as well as the N250 and P350, typically observed in cortical responses to auditory stimuli in infants (Kushnerenko, Ceponiene, Balan, Fellman, Huotilainen, & Näätänen, 2002).

Results from the three-way ANOVAs computed on the P150 component with each measure of interest as the dependent variable are reported in Table 4.2. A group effect was found for the latency and amplitude of the maximum peak within the 50-250 time window, with a significant decrease of latency ($F(1) = 35.1$, $p < .001$) and a significant increase of amplitude ($F(1) = 89.7$, $p < .001$). The mean amplitude between 50 and 250 ms also increased significantly with age ($F(1) = 97.9$, $p < .001$). The smaller time windows analysis revealed significant group effects in the first (tw1: 50-116 ms, $F(1) = 291$, $p < .001$) and second (tw2: 117-183 ms, $F(1) = 85.8$, $p < .001$) time windows, but not in the third one (tw3: 184-200 ms, $p = .22$). Effects of laterality (left, center, right) were significant for the mean amplitudes in tw1 ($F(2) = 5.7$, $p < .01$) and tw2 ($F(2) = 3.9$, $p < .05$) only. Post hoc statistics revealed significant differences between left and central electrodes ($p < .01$) and between left and right electrodes ($p < .05$) in tw1. A significant difference was reported between left and central electrodes ($p < .05$) in tw2. Effects

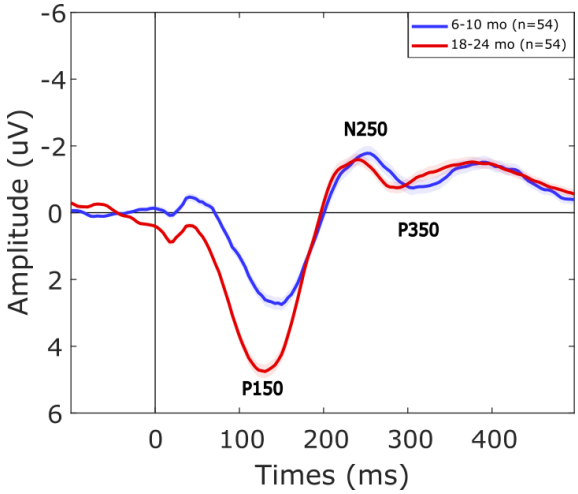


Figure 4.8. – Response to standard /da/ at C3 electrode. Shaded areas indicate the standard deviation. The P150, N250, and P350 peaks are indicated, following Kushnerenko, Ceponiene, Balan, Fellman, Huotilainen, and Näätänen (2002).

	Basic statistics		ANOVA results		
Measure	**6-10 months**	**18-24 months**	**Group**	**Laterality**	**Antero-posterior**
Peak latency (μV)	166 (49.7)	149 (38.9)	F(1) = 35.1, p < .001	p = .59	p = .24
Peak amplitude (μV)	1.10 (3.65)	3.11 (3.43)	F(1) = 89.7, p < .001	p = .18	F(2) = 74.4, p < .001
Mean amplitude (μV)	0.52 (2.01)	1.27 (2.21)	F(1) = 97.9, p < .001	p = .054	F(2) = 98.7, p < .001
tw1 (50-116 ms)	0.36 (1.43)	1.74 (1.25)	F(1) = 291, p < .001	F(2) = 5.7, p < .01	F(2) = 91.1, p < .001
tw2 (117-183 ms)	1.72 (1.93)	2.73 (1.98)	F(1) = 85.8, p < .001	F(2) = 3.9, p < .05	F(2) = 151.6, p < .001
tw3 (118-250 ms	-0.51 (1.96)	-0.66 (1.77)	p = .22	p = .91	F(2) = 9.9, p < .001

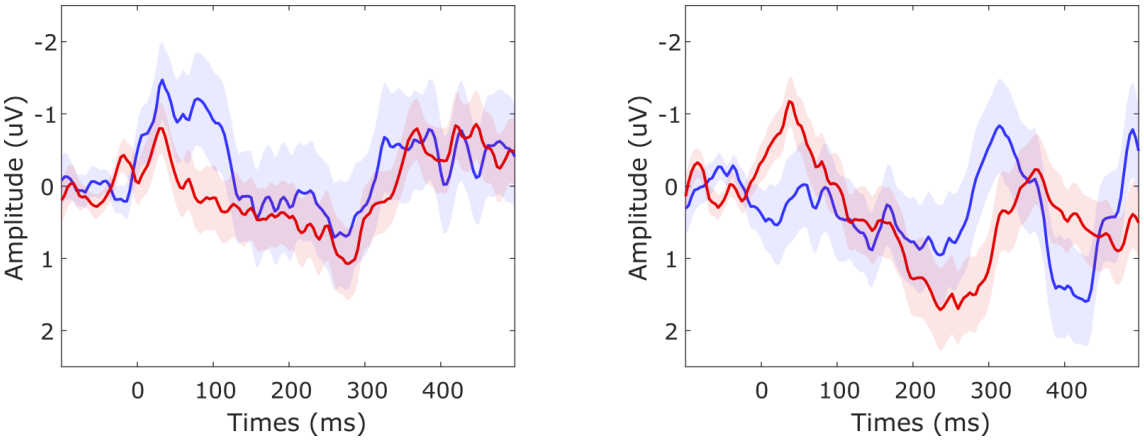
Table 4.2. – Basic statistics and results of the ANOVAs computed on the standard response in the 50-250 ms time-window. In the basic statistics columns, the mean value is indicated and standard deviations are reported in brackets.

of the anteroposterior location were found for all measures except the peak latency ($p = .24$). For the peak amplitude ($F(2) = 74.4$, $p < .001$), the mean amplitude of the whole time window ($F(2) = 98.7$, $p < .001$), and the mean amplitude in the smaller tw2 ($F(2) = 3.9$, $p < .05$), post hoc statistics indicated significant differences between all regions (frontal-central: $p < .001$; parietal-central: $p < .001$; parietal-frontal: $p < .001$). In the other smaller time windows, significant differences were found for parietal-central

($p < .001$) and parietal-frontal ($p < .001$) regions in tw1, and for frontal-central ($p < .001$) and parietal-frontal ($p < .001$) regions in tw2.

Difference waves analysis. The difference waves between standard and deviants were computed for each group (6-10 months and 18-24 months). Figure 4.9 shows the temporal representation of the difference waves for /ba/ and /ga/ deviants at F3 electrode.

Three-ways ANOVAs with age group (6-10 months, 18-24 months), laterality (left, center, right), and anteroposterior location (frontal, central, parietal) indicated group effects for both deviants and in both time windows (180-300 ms and 300-480 ms). Table 4.3 reports the basic statistics and results of the ANOVAs.



(a) Difference wave to /ba/ deviant, F3 electrode.

(b) Difference wave to /ga/ deviant, F3 electrode.

Figure 4.9. – Difference waves to deviants /ba/ and /ga/ at F3 electrode, group comparison. Blue: 6-10 mo; red: 18-24 mo.

For the difference wave between the standard /da/ and the deviant /ba/, a group effect was found in the whole 180-300 ms time-window, indicating larger amplitude in the old compared to the young group ($F(1) = 6.87$, $p < .01$). An effect of antero-posterior location was also found in this time-window, indicating a significantly higher amplitude at frontal than parietal electrodes ($p < .05$). The Group x Anteroposterior analysis was not significant ($p = .39$). ANOVAs applied to the smaller time-windows revealed a significant group effect in the first time window (tw1: 180-220 ms, $F(1) = 4.71$, $p < .05$) but no effect of antero-posterior location ($p = .85$). In the 300-480 ms time window, the group effect was marginally significant ($p = .08$) in the whole time window.

	Basic statistics		**ANOVA results**		
	6-10 months	**18-24 months**	**Group**	**Laterality**	**Antero-posterior**
	MMR / ba/ - 180-300 ms				
Peak latency	241 (33.8)	244 (34.4)	p = .22	p = .86	p = .78
Peak amplitude	0.17 (6.73)	0.44 (5.23)	p = .49	p = .84	p = .30
Mean amplitude	-0.01 (4.45)	0.37 (3.25)	F(1) = 6.87, p < .01	p = .19	F(2) = 3.22, p < .05
tw1 (180-220 ms)	-0.19 (4.50)	0.35 (3.19)	F(1) = 4.71, p < .05	p = .39	p = .85
tw2(220-260 ms)	0.08 (4.36)	0.41 (3.26)	p = .19	p = .59	p = .39
tw3(260-300 ms)	0.08 (4.48)	0.34 (3.30)	p = .30	p = .64	F(2) = 3.09, p < .05
	MMR / ba/ - 300-480 ms				
Peak latency	393 (52.8)	395 (52.7)	p = .49	p = .40	p = .069
Peak amplitude	-1.85 (7.88)	-1.42 (5.50)	p = .34	p = .32	p = .81
Mean amplitude	-1.03 (4.88)	-0.78 (3.06)	p = .08	p = .14	F(2) = 6.84, p < .01
tw1 (300-360 ms)	-0.62 (4.86)	-0.53 (2.91)	p = .73	p = .74	p = .20
tw2 (360-420 ms)	-1.15 (4.95)	-1.12 (3.11)	p = .89	p = .44	F(2) = 3.70, p < .05
tw3 (420-480 ms)	-1.31 (4.81)	-0.67 (3.14)	F(1) = 6.02, p < .05	p = .35	p = .14
	MMR / ga/ - 180-300 ms				
Peak latency	241 (33.6)	247 (33.2)	F(1) = 5.97, p < .05	p = .84	p = .38
Peak amplitude	0.36 (7.02)	1.41 (5.21)	F(1) = 7.07, p < 0.01	p = .83	F(2) = 3.39, p < .05
Mean amplitude	0.22 (4.63)	0.79 (3.43)	F(1) = 13.98, p < .001	p = .82	F(2) = 5.70, p < .01
tw1 (180-220 ms)	0.46 (4.44)	0.86 (3.35)	p = .11	p = .90	p = .60
tw2(220-260 ms)	0.42 (4.83)	0.85 (3.46)	p = .11	p = .91	F(1) = 4.33, p < .05
tw3(260-300 ms)	-0.20 (4.60)	0.64 (3.47)	F(1) = 10.43, p < .01	p = .88	p = .19
	MMR / ga/ - 300-480 ms				
Peak latency	398 (51.4)	390 (51.3)	F(1) = 5.42, p < .05	p = .88	p = .73
Peak amplitude	-0.40 (7.70)	-0.24 (5.87)	p = .71	p = .86	p = .30
Mean amplitude	-0.15 (4.52)	-0.17 (3.26)	p = .90	p = .85	F(2) = 3.86, p < .05
tw1 (300-360 ms)	-0.64 (4.41)	-0.37 (3.43)	p = .29	p = .77	p = .67
tw2 (360-420 ms)	0.16 (4.66)	-0.37 (3.19)	F(1) = 4.16, p < .05	p = .93	F(2) = 3.33, p < .05
tw3 (420-480 ms)	0.03 (4.44)	0.23 (3.11)	p = .41	p = .97	p = .21

Table 4.3. – Difference wave statistics, group comparison.

The mean amplitude was significantly more negative in the young group than in the old group (F(1) = 6.02, $p < .05$), but no effect of antero-posterior location was found. For the deviant /ba/, the difference waves were not significantly different between left, center, and right regions, in all time windows. ANOVAs with maximum peak latency as the dependent variable yielded no significant effect overall.

For the difference wave between the standard /da/ and the deviant /ga/, group effects were found in the whole 180-300 ms time window, indicating a greater mean amplitude value (F(1) = 13.98, $p < .001$) and a greater peak amplitude (F(1) = 7.07, $p < 0.01$) in the old group compared to the young group. An effect of anteroposterior location revealed a significant difference between parietal and frontal electrodes ($p < .05$), with greater amplitude at frontal electrodes. Additionally, the Group x Anteroposterior analysis was significant (F(2) = 8.75, $p < .001$). The mean amplitude was significantly larger in the old group than in the young group in the last smaller time window (tw3: 260-300 ms, F(1) = 10.43, $p < .01$), with a shift from negative to positive amplitude. The Group x Anteroposterior analysis in tw3 was also significant (F(2) = 3.03, $p < .05$). In the 300-480 ms time window, a significant group difference was found for the peak latency (F(1) = 5.42, $p < .05$). No effect was found for the peak amplitude and mean amplitude over the whole time window. However, group (F(1) = 4.16, $p < .05$) and antero-posterior location (F(2) = 3.33, $p < .05$) effects were found in the smaller tw2 (360-420 ms). These results indicated a shift from positive to negative amplitude between young and old infants, and significant difference between parietal and frontal electrodes.

3.2.3. Longitudinal sample

Response to standards. Cortical brain responses to the standard /da/ at each assessment point are illustrated in Figure 4.10. Visual inspection of the data suggest a progressive increase of the P150 amplitude and decrease of the latency. Three-ways ANOVAs confirmed these observations (see Table 4.4), as a group effect was found for all measures of interest: maximum peak latency (F(4) = 12.78, $p < .001$), maximum peak amplitude (F(4) = 38.27, $p < .001$), mean amplitude between 50 and 250 ms (F(4) = 43.32, $p < .001$), and mean amplitude in all smaller time windows (tw1: 50-116 ms, F(4) = 87.30, $p < .001$; tw2: 117-183 ms, F(4) = 58.67, $p < .001$; tw3: 184-220 ms, F(4) = 17.09, $p < .001$). A laterality effect was found in tw2 (117-183 ms: F(2) = 4.67, $p < .01$), and post hoc tests indicated significant differences between all regions of interest (frontal-central: $p < .05$; parietal-central: $p < .001$, parietal-frontal: $p < .001$). The effects of antero-posterior location indicated significant differences between all regions of interest for

the mean amplitude (frontal-central: $p < .05$; parietal-central: $p < .001$, parietal-frontal: $p < .001$), and only at parietal-central and parietal-frontal regions for the maximum peak amplitude and mean amplitudes of the smaller time windows.

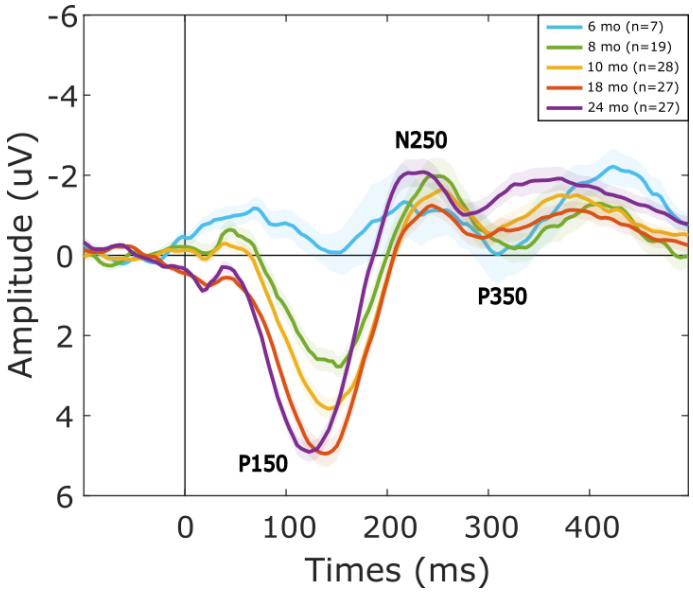


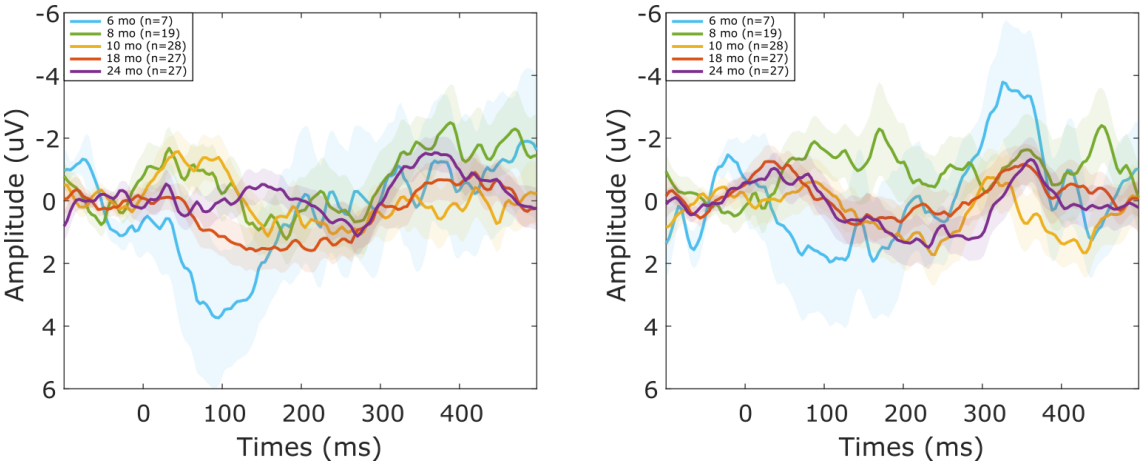
Figure 4.10. – Response to standard /da/ at C3 electrode for all assessment points.

	Basic statistics					ANOVA results		
Measure	**6 mo**	**8 mo**	**10 mo**	**18 mo**	**24 mo**	**Group**	**Laterality**	**Antero-posterior**
Peak latency	168 (59.1)	180 (53.3)	163 (44.4)	152 (30.5)	147 (42.3)	F(4) = 12.78, p < .001	p = .54	p = .56
Peak amplitude	-0.20 (2.66)	0.12 (3.63)	1.96 (3.50)	3.98 (2.90)	2.87 (3.73)	F(4) = 38.27, p < .001	p = .50	F(2) = 62.21, p < .001
Mean amplitude	-0.03 (1.52)	0.23 (1.85)	0.70 (2.02)	1.71 (2.11)	1.06 (2.33)	F(4) = 43.32, p < .001	p = .21	F(2) = 73.07, p < .001
tw1 (50-116 ms)	-0.61 (1.37)	0.05 (1.38)	0.55 (1.23)	1.61 (1.17)	1.90 (1.31)	F(4) = 87.30, p < .001	F(2) = 4.67, p < .01	F(2) = 44.56, p < .001
tw2 (117-183 ms)	0.29 (1.41)	1.42 (1.56)	2.11 (1.79)	3.42 (2.04)	2.52 (1.83)	F(4) = 58.67, p < .001	p = .08	F(2) = 134.13, p < .001
tw3 (118-250 ms	0.23 (1.62)	-0.79 (1.87)	-0.57 (1.98)	0.10 (1.52)	-1.25 (1.75)	F(4) = 17.09, p < .001	p = .59	F(2) = 7.19, p < .001

Table 4.4. – Basic statistics and results of the ANOVAs computed on the standard response in the 50-250 ms time-window. In the basic statistics columns, the mean value is indicated and standard deviations are reported in brackets.

Difference waves analysis. As in the group comparison analysis, difference waves obtained for each deviant are shown in Figure 4.11. Results of the three-way ANOVAs

with maximum peak latency, maximum peak amplitude and mean amplitudes at all time windows of interest as dependent variables and group, laterality, and anteroposterior location as factors are reported in Table 4.5. From these tests, group effects were revealed in both 180-300 ms and 300-480 ms time windows, as well as in most smaller time windows, for both deviants.



(a) Difference wave to /ba/ deviant, C3 electrode.

(b) Difference waves to /ga/ deviant, C3 electrode.

Figure 4.11. – Difference waves to deviants /ba/ and /ga/ at C3 electrode, all assessment points.

For the difference wave between the standard and the /ba/ deviant, results of the three-way ANOVAs indicated a significant effect of group in the 180-300 ms time window for the peak latency ($F(4) = 4.84$, $p < .001$), peak amplitude ($F(4) = 3.51$, $p < .01$) and mean amplitude ($F(4) = 9.39$, $p < .001$). Significant effects of groups were also found in the mean amplitude of the smaller time-windows tw1 (180-220 ms: $F(4) = 6.04$, $p < .001$) and tw2 (220-260 ms: $F(4) = 4.55$, $p < .01$), but not in tw3 (260-300 ms: $p = .08$). Importantly, post-hoc tests revealed that the mean amplitude within the 180-300ms time window was significantly higher at 8 than 10 months ($p < .001$), 8 than 18 months ($p < .01$), and at 24 than 10 months ($p < .001$). In the 300-480 ms time window, no significant effect of group was found for the maximum peak latency ($p = .51$). However, significant effects of group were found for the maximum peak amplitude ($F(4) = 4.33$, $p < .01$) and the mean amplitude ($F(4) = 9.68$, $p < .001$). This was confirmed by a significant effect of group for the mean amplitude in the smaller time windows tw1 (300-360 ms: $F(4) = 6.98$, $p < .001$) and tw2 (360-420 ms: $F(4) = 4.41$, $p < .01$), but not in tw3 (420-480 ms: $p = .17$).

	Basic statistics					ANOVA results		
	6 months	**8 months**	**10 months**	**18 months**	**24 months**	**Group**	**Laterality**	**Antero-posterior**
	MMR / ba/ - 180-300 ms							
Peak latency	386 (46.7)	400 (45.2)	406 (55.3)	384 (50.7)	401 (52.7)	F(4) = 4.84, p < .001	p = .92	p = .36
Peak amplitude	-3.27 (7.25)	-1.43 (8.38)	0.87 (6.69)	-0.27 (5.66)	0.54 (6.06)	F(4) = 3.514, p < .01	p = .74	p = .08
Mean amplitude	-1.96 (4.54)	-0.85 (5.01)	0.46 (3.61)	-0.11 (3.06)	0.26 (3.22)	F(4) = 9.39, p < .001	F(2) = 5.78, p < .01	p = .35
tw1 (180-220 ms)	-3.28 (4.44)	-0.94 (4.06)	0.18 (3.38)	-0.46 (3.39)	0.28 (3.27)	F(4) = 6.04, p < .001	p = .49	p = .29
tw2(220-260 ms)	-1.77 (4.36)	-0.61 (5.63)	0.56 (3.49)	-0.12 (2.95)	-0.13 (3.12)	F(4) = 4.55, p < .01	p = .81	p = .22
tw3(260-300 ms)	-0.84 (4.56)	-1.00 (5.25)	0.65 (3.94)	0.25 (2.78)	0.64 (3.25)	p = .08	p = .70	F(2) = 3.20, p < .05
	MMR / ba/ - 300-480 ms							
Peak latency	391 (48.5)	397 (57.2)	386 (53.7)	392 (52.0)	395 (53.6)	p = .51	p = .90	p = .16
Peak amplitude	-0.291 (8.35)	-2.36 (7.58)	0.565 (8.28)	-1.02 (5.24)	-2.14 (5.64)	F(4) = 4.33, p < .01	p = .95	p = .15
Mean amplitude	-0.51 (5.24)	-1.14 (4.50)	0.25 (5.25)	-0.53 (2.80)	-1.20 (3.07)	F(4) = 9.68, p < .001	p .12	F(2) = 8.07, p < .001
tw1 (300-360 ms)	0.3 (4.67)	-0.80 (4.71)	1.02 (5.12)	-0.23 (2.60)	-1.15 (2.98)	F(4) = 6.98, p < .001	p = .90	p = .17
tw2 (360-420 ms)	-0.29 (5.48)	-1.49 (4.10)	0.04 (5.20)	-1.01 (2.88)	-1.64 (3.12)	F(4) = 4.41, p < .01	p = .41	F(2) = 2.97, p = .05
tw3 (420-480 ms)	-1.56 (5.45)	-1.14 (4.68)	-0.31 (5.37)	-0.36 (2.85)	-0.82 (3.07)	p = .17	p = .24	F(2) = 3.57, p < .05
	MMR / ga/ - 180-300 ms							
Peak latency	232 (28.5)	236 (34.3)	246 (30.6)	250 (32.7)	242 (34.8)	F(4) = 5.12, p < .001	p = .99	p = .55
Peak amplitude	0.38 (4.82)	-1.05 (7.62)	2.25 (7.19)	0.69 (5.31)	2.51 (4.82)	F(4) = 7.77, p < .001	p = .80	p = .13
Mean amplitude	-0.33 (3.39)	-0.38 (5.04)	1.30 (4.79)	0.41 (3.57)	1.50 (3.07)	F(4) = 17.88, p < .001	p = .41	F(2) = 3.56, p < .05
tw1 (180-220 ms)	-0.10 (3.31)	0.10 (5.08)	1.20 (4.47)	0.74 (3.21)	1.50 (3.36)	F(4) = 3.37, p < .01	p = .95	p = .83
tw2(220-260 ms)	0.172 3.02	-0.28 (5.36)	1.69 (5.24)	0.49 (3.56)	1.49 (3.01)	F(4) = 5.33, p < .001	p = .87	F(2) = 3.02, p < .05
tw3(260-300 ms)	-1.07 (3.74)	-0.96 (4.65)	1.01 (4.64)	0.002 (3.88)	1.50 (2.86)	F(4) = 7.20, p < .01	p = .35	p = .24
	MMR / ga/ - 300-480 ms							
Peak latency	386 (46.7)	400 (45.2)	406 (55.3)	384 (50.7)	401 (52.7)	F(4) = 4.96, p < .001	p = .80	p = .33
Peak amplitude	-3.27 (7.25)	-1.43 (8.38)	0.87 (6.69)	-0.27 (5.66)	0.54 (6.06)	F(4) = 5.29, p < .001	p = .63	p = .24
Mean amplitude	-1.96 (4.54)	-0.85 (5.01)	0.46 (3.61)	-0.11 (3.06)	0.26 3.22)	F(4) = 17.35, p < .001	p = .15	F(2) = 3.44, p < .05
tw1 (300-360 ms)	-3.28 (4.44)	-0.94 (4.06)	0.18 (3.38)	-0.46 (3.39)	0.28 (3.27)	F(4) = 11.74, p < .001	p = .29	p = .85
tw2 (360-420 ms)	-1.77 (4.36)	-0.61 (5.63)	0.56 (3.49)	-0.12 (2.95)	-0.13 (3.12)	F(4) = 4.01, p < .01	p = .62	F(2) = 3.66, p < .05
tw3 (420-480 ms)	-0.84 (4.56)	-1.00 (5.25)	0.65 (3.94)	0.25 (2.78)	0.64 (3.25)	F(4) = 4.86, p < .001	p = .71	p = .35

Table 4.5. – Difference wave statistics,all assessment points.

Post-hoc tests indicated that mean amplitudes between 300 and 480 ms were significantly higher at 10 than 8 months ($p < .001$), at 10 than 18 months ($p < .05$), at 10 than 24 months ($p < .001$), and at 18 than 24 months ($p < .05$). Effects of anteroposterior location were also found for the mean amplitudes between 300 and 480 ms ($F(2) = 8.07$, $p < .001$) and between 360 and 420 ms ($F(2) = 2.97$, $p = .05$). Post hoc tests indicated greater amplitudes in frontal than parietal (300-480: $p < .001$; 360-420: $p < .05$) and central regions (300-480 ms: $p < .05$).

For the difference wave between the standard and the /ga/ deviant, results of the three-way ANOVAs indicated a significant effect of group in the 180-300 ms time window for the peak latency ($F(4) = 5.12$, $p < .001$), peak amplitude ($F(4) = 7.77$, $p < .001$) and mean amplitude ($F(4) = 17.88$, $p < .001$). Effects of groups were also found in the mean amplitude of all smaller time-windows tw1(180-220 ms: $F(4) = 3.37$, $p < .01$), tw2 ($F(4) = 5.33$, $p < .001$), and tw3 ($F(4) = 7.20$, $p < .01$). Post-hoc tests revealed that the mean amplitude within the 180-300ms time window was significantly different between all age groups, except between 6 and 8 months ($p = .99$), between 6 and 18 months ($p = .22$), and between 10 and 24 months ($p = .94$). Effects of anteroposterior location were also found for the mean amplitudes in the 180-300 ms time window ($F(2) = 3.56$, $p < .05$) and in the small 220-260 ms time window ($F(2) = 3.02$, $p < .05$). In the 300-480 ms time window, group effects were found for the maximum peak latency ($F(4) = 4.96$, $p < .001$), the maximum peak amplitude ($F(4) = 5.29$, $p < .001$) and the mean amplitude ($F(4) = 17.35$, $p < .001$). Group effects were also found for the mean amplitude in the smaller time windows tw1 (300-360 ms: $F(4) = 11.74$, $p < .001$), tw2 (360-420 ms: $F(4) = 4.01$, $p < .01$), and tw3 (420-480 ms: $F(4) = 4.86$, $p < .001$). Post hoc tests indicated that mean amplitudes between 300 and 480 ms were significantly different between all age groups except between 10 and 18 months ($p = .11$), between 10 and 24 months ($p = .91$), and between 18 and 24 months ($p = .44$). Effects of anteroposterior location were found for the mean amplitudes between 300 and 480 ms ($F(2) = 3.44$, $p < .05$) and between 360 and 420 ms ($F(2) = 3.66$, $p < .05$). Post hoc tests indicated significant differences between parietal and frontal regions in both time windows (300-480: $p < .05$; 360-420: $p < .05$).

In order to better understand the developmental trajectory of the difference waves, the mean amplitudes for each deviant and each time window of interest as a function of age group are represented in Figure 4.12. Only time windows for which significant group effects were reported in the ANOVAs are included in the figure. The thick solid lines

indicate the pattern of the mean amplitude in the larger time windows (180-300 ms or 300-480 ms). Visual observation of these patterns indicated a positive peak at 10 months of age at all time windows and deviants. We can also observe differences in the patterns associated with the /ba/ deviant and the one associated to the /ga/ deviant: whereas the mean amplitude increases up to 10 months of age and then linearly decreases between 8 and 24 months of age for the /ba/ deviant, the mean amplitude decreases between 10 and 18 months then increases again between 18 and 24 months for the /ga/ deviant.

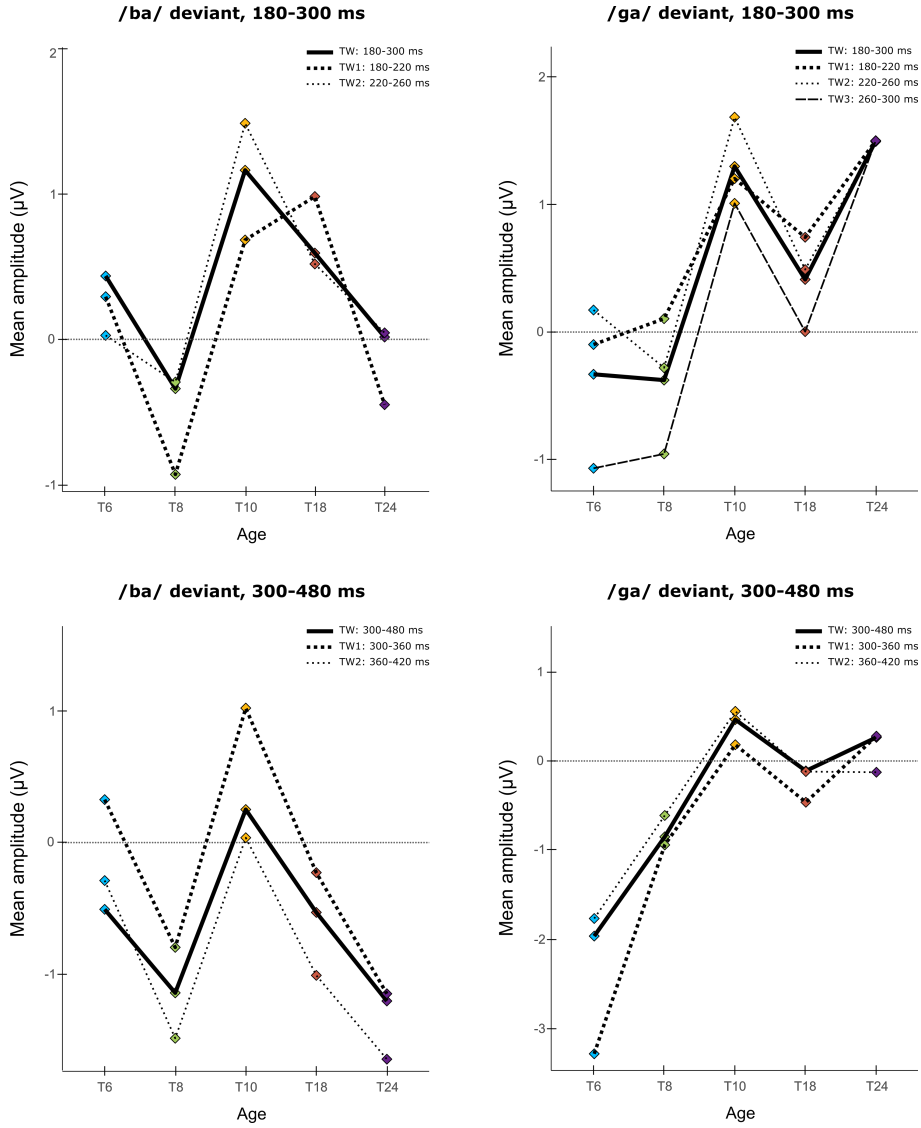


Figure 4.12. – Developmental patterns of the mean amplitudes of the difference wave as a function of assessment point, for all time windows and deviants.

4. Discussion

In this study, I assessed the feasibility of an innovative paradigm that allowed for recording cortical and subcortical responses to speech sound in infants, in a single session. Results indicated that such a procedure is feasible in developmental populations and yielded consistent results with previous work. I first discuss the feasibility and possible improvements of this method. Then, I further discuss the results obtained at the cortical and subcortical levels.

4.1. Recording simultaneous cortical and subcortical responses in infants: feasibility

Early cortical and subcortical responses to speech sounds were recorded multiple times in previous work, but mostly separately and for different research questions. Following G. M. Bidelman (2015)' work, I designed a novel protocol to record early brain responses to speech sounds at both levels in a single session (to my knowledge, this has not been tested previously with this procedure). The procedure in G. M. Bidelman (2015) was called "clustered stimulus presentation" and relied on the alternating presentation of auditory stimuli at variable Inter-Stimulus Intervals (ISIs). In this study, analyses for data quality assessment indicated reliable brain responses at both cortical and subcortical levels, with minimal differences in amplitude and latencies between traditional and clustered paradigms. Additionally, the benefits of studying both responses in a single session were highlighted by showing a 3-fold increased recording efficiency. Indeed, the clustered ISIs allowed a reduction in testing time while accommodating the stimulation rates to each type of response. However, results revealed subtle differences in the latencies compared to classical procedures: cortical ERP latencies (P1, N1) were increased under the clustered paradigm, likely due to adaptation effects, though the amplitudes remained consistent. Overall, G. M. Bidelman (2015) provided critical evidence for successful recording of cortical and subcortical responses simultaneously. In more recent work, T.-H. Z. Cheng and Zhao (2024) also validated the clustered procedure for simultaneous cortical and subcortical recordings in adults. A few other studies gathered cortical and subcortical brain responses in adults, showing that the auditory pathway is not merely a hierarchical system but involves bidirectional communication where both early and higher auditory regions are involved in speech sound processing (Font-Alaminos et al., 2021; Shiga et al., 2015). However, I report only one study that collected cortical and subcortical brain responses in infants. In Novitskiy et al. (2022), lexical tones were presented in the syllable /ga/ with a fixed ISI,

thus limiting the number of repetitions to a thousand. As Richard et al. (2020) argued in their review on neonatal FFR studies, a higher number of trials is needed in infants compared to adults, to increase the SNR of the response. Thus, the "clustered stimulus presentation" paradigm might offer a good practical solution for obtaining reliable and comprehensive data on auditory processing in infants at multiple levels of the auditory pathway, without requiring multiple, time-consuming experimental sessions.

One major goal of the present study was to assess the feasibility of this paradigm in very young infants, between 3 and 24 months. Knowing the challenges associated with developmental EEG studies (Hervé et al., 2022), particular care was taken with the implementation. Indeed, cortical and subcortical brain responses required different recording parameters that needed to be gathered in one experimental setup. Specifically, the collection of subcortical brain responses requires a very high sampling rate (Chandrasekaran & Kraus, 2010; Krizman & Kraus, 2019; Skoe & Kraus, 2010) and a vertical (or horizontal, see Billings et al. (2019) and Galbraith et al. (2000) for discussions) montage. Moreover, because the brainstem responds to millisecond delays, high temporal resolution was needed regarding the synchronization of the response with stimuli onsets. The best solution we v found with the engineers of the lab was to code the trigger information directly in the auditory input with MATLAB®, to override jitter issues. The input consisted of a stereo file: the sound wave of the stimulus was presented on one channel, and the other included the trigger information. A splitter was designed and built by our lab engineer to send the auditory signal to the ear on one side, and the trigger information to the Biosemi software on the other side. A BabyFace soundcard was used to monitor the intensity of the sound and thus obtain perfectly controlled stimuli, both in the temporal precision and the intensity. Regarding data collection in infants, it was no more challenging than other classical EEG studies. At the practical level, this procedure in infants was thus largely possible. Regarding data quality and the resulting brain responses, I was also convinced by the protocol. I develop the results in the following sections.

4.2. Subcortical encoding of speech sounds

Overall, the metrics extracted from the FFRs yielded comforting results, consistent with previous work. As shown by the visual exploration of the FFRs in the time and frequency domains, I obtained reliable measures, comparable with other FFR in the same age ranges (S. Anderson et al., 2015; Jeng et al., 2010; Ribas-Prats et al., 2023). Like in previous FFR studies, known technical challenges were faced, particularly in terms

of maintaining sufficient SNR due to recording data in infants. This was reflected in the results obtained for the SNR of the response in the vowel region of the stimulus: greater SNR values were observed in the CV transition than in the vowel region, and no significant group effect was found in the vowel region whereas the SNR in the CV transition yielded significant age group effects. Considering the larger duration and increased periodicity of the vowel region compared to the CV transition region, I expected a group effect and higher SNR in the vowel than in the CV transition region, but this was not the case. As noted by Jeng et al. (2010), individual recordings in infants often have lower SNR compared to grand-averaged waveforms, making it difficult to isolate reliable frequency-following responses (FFR) at the individual level. Future studies might focus on enhancing SNR through improved signal processing techniques or recording protocols, to increase the reliability of FFR as a biomarker of early speech perception. As a result, the neural lag, which was used as an index of the encoding of the stimulus periodicity and of the delay of the brain response, was not significantly different between age groups, and neither correlated with age. This was surprising but not inconsistent with previous work. Indeed, studies that found significant differences in the neural lag focused on very young infants, before 6 months of age (Arenillas-Alcón et al., 2021; Ribas-Prats et al., 2019; Ribas-Prats et al., 2022). Thus, I might have revealed that by 6 months of age, the neural lag is no longer modulated by infants' speech sound encoding and cannot be used as a biomarker of auditory processing. These results are further supported by Ribas-Prats et al. (2022)'s results in infants born after fetal growth restriction, where no significant difference with typically developing infants was found for the neural lag.

Crucially, my results contribute to increasing our understanding of how subcortical encoding of speech sounds develops in the first two years of life. In the temporal domain, the SNR of the FFR at the CV transition region of the stimulus was modulated by infants' age. In addition, an increase in the overall temporal SNR was revealed. As mentioned above, this increase was not found in the SNR of the vowel region, suggesting that the overall SNR increase was mostly due to the encoding of the CV transition region. These results reflect an increase in the magnitude of neural activity over time, associated with higher time locking of subcortical neurons to the CV transition region of the stimulus (Arenillas-Alcón et al., 2021).
In the frequency domain, previous work in neonates and infants suggested that the encoding of F0 matures rapidly. Importantly, S. Anderson et al. (2015) and Jeng et al. (2011) have shown a robust F0 encoding as young as 3 months of age. Moreover, longitudinal

assessments between 1 and 10 months of age in Jeng et al. (2010) and between 0 and 6 months in Ribas-Prats et al. (2023) showed an increase in the encoding of the F0, as indicated by measures of pitch tracking and spectral amplitude. Cross-sectional studies in younger (2-7 months) vs. older (7-12) infants (Van Dyke et al., 2017) and in infants vs. adults also indicated more robust encoding of the F0 across time. Here, I found consistent results indicated by measures of pitch tracking (pitch strength and pitch error): a significant group effect indicated more robust encoding of the F0 over time. Moreover, the developmental trajectory of the frequency amplitude at F0 peak also revealed an increasing pattern (Figure 4.5). Interestingly, this pattern indicates a slow increase between 6 and 18 months of age, followed by a strong significant increase of the amplitude at 24 months. This enhanced F0 encoding might be the result of synaptic pruning in the subcortical auditory pathways at the end of the second year of life (Moore & Linthicum, 2007; Petanjek et al., 2011), which reinforces the connectivity of neural networks within the auditory system. This change in subcortical encoding is consistent with age-related increases in the encoding of spectral features, which allow young children to distinguish subtle variations in pitch, tone, and other spectral components across a wide range of auditory stimuli (Novitskiy et al., 2022; Werner & Kitzes, 2011). Such enhancements in F0 encoding at this age might not be specific to language but could instead indicate a general readiness of the auditory system to process complex acoustic information, a fundamental skill necessary for language and music perception (Eggermont & Moore, 2012). Moreover, this enhanced spectral encoding may reflect the brain's continued plasticity at age 2, allowing the auditory system to remain highly adaptable to new and diverse sounds, including nonnative syllables. This period of high sensitivity to environmental sounds might also indicate an effect of experience with auditory stimuli (Kral & Sharma, 2012; Kuhl, 2004).

The encoding of the higher frequencies revealed a different developmental pattern. Consistent with S. Anderson et al. (2015), Arenillas-Alcón et al. (2021), and Ribas-Prats et al. (2023), I found a progressive increase in the encoding of F1 in the vowel part, revealed by developmental patterns extracted from measures of the spectral SNR and amplitudes at the F1 peak frequency. In contrast with the encoding of the fundamental frequency, the developmental changes for the amplitude at F1 peak in the CV transition appear more irregular. An unexpected decrease seemed to be observed at 18 months, however statistical analyses indicated that it was not significant. This nonlinear pattern of the developmental trajectory of the amplitude at F1 peak in the CV transition region could be due to the less stable spectral content and shorter duration of the consonant burst regions, compared to vowels. Nevertheless, this enhanced encoding of the fine

spectral content of speech sounds might reflect early cross-modal integration between subcortical auditory structures and cortical regions associated with language processing. During this period, auditory information, especially speech-related spectral cues, becomes more precisely relayed to and synchronized with higher cortical language areas, supporting the emergence of top-down processes recruited for speech sound processing (Friederici, 2005; Poeppel et al., 2008). The improved encoding of F1 at the subcortical level not only facilitates phonetic discrimination but also aids in building the neural circuitry that will support more complex language functions as children continue to develop (Werker & Tees, 2005). In addition, a *consonant bias* observed as young as 8 months of age (Nazzi et al., 2016; Nishibayashi & Nazzi, 2016; Von Holzen & Nazzi, 2020) could reinforce these top-down processes between speech perception and phonological categorization over time, and increase the neural encoding of consonant-related spectral features during critical periods of phonological-lexical skills development.

Finally, I found that the measures of pitch tracking such as pitch error and pitch strength could be reliable indicators of speech encoding abilities over time. Although correlations with age were found for several measures, in both temporal (SNR of the response) and spectral levels (spectral amplitude at F1 frequency), the pitch error and pitch strength showed the highest correlation scores. Moreover, these two measures of pitch tracking were strongly correlated with all the other measures except the neural lag. These pitch-tracking measures are particularly interesting because, while typically considered spectral measures, they capture information about both spectral and temporal processing. This dual aspect is critical in understanding speech encoding because the auditory system uses both frequency and timing information to faithfully represent and distinguish sounds. As Cabrera and Gervain (2020) highlighted, temporal features of a sound are encoded not just in the temporal domain but also reflected in the spectral domain. This overlap means that auditory encoding is not strictly separated between frequency and timing information: both are intertwined in the auditory representation. Furthermore, pitch-tracking measures may reflect dual peripheral coding processes that capture the temporal characteristics of a sound in complementary ways. According to Cabrera and Gervain (2020)'s findings, these processes involve both phase-locking mechanisms (neurons firing at intervals that match the sound frequency, encoding temporal information) and tonotopic mapping (neurons tuned to specific frequencies, encoding spectral information). Together, these mechanisms enable the auditory system to represent fine spectro-temporal features of speech sounds. The strong correlation between pitch tracking measures and other measures further underscores their

comprehensive role. Since they correlate well with temporal and spectral encoding metrics, pitch tracking appears to capture a broad aspect of auditory processing. It likely reflects the synchronization of neural responses over time, a critical factor in accurate speech perception and language development (Friederici, 2005; Hornickel et al., 2009; Kraus & Nicol, 2005; White-Schwoch et al., 2015).

4.3. Cortical responses to speech sounds

At the cortical level, I also obtained reliable auditory brain responses across all age groups examined. The multi-featured oddball paradigm involved three synthetic syllables (/ba/, /da/, /ga/)adapted for English speakers (K. L. Johnson et al., 2008), thus considered non-native for the French infants. I hypothesized that infants would show a decrease in MMR amplitude, as found in previous work using nonnative contrasts (Cheour, Ceponiene, et al., 1998; Werker & Tees, 1984b). However, the patterns of results were not this clear-cut. Before summarizing and discussing these results, I briefly report the cortical responses obtained to the standard /da/.

The quality of the response was first explored by looking at the ERP to the standard syllable /da/. I focused on the P150 component, characterized by its positive peak around 150 ms following stimulus onset, shown to reflect automatic response to auditory stimuli in infancy and precursor of the adult P1 component (Čeponien, Rinne, & Näätänen, 2002; Coch & Gullick, 2012; Kushnerenko, Ceponiene, Balan, Fellman, Huotilainen, & Näätänen, 2002). The component was successfully observed in all groups, and measures of amplitude and latency were extracted between 50 and 250 ms (Kushnerenko, Ceponiene, Balan, Fellman, Huotilainen, & Näätänen, 2002). Statistical analysis of these measures revealed significant developmental changes. Notably, both group comparison (young vs. old) and longitudinal analyses (all five assessment points: 6, 8, 10, 18, and 24 months) indicated an increase in the amplitude and a decrease in the latency of the P150 response. These results reflect a clear trajectory of maturation of the cortical auditory response and emphasize the importance of the P150 as a reliable neural marker of speech encoding and auditory perception in early childhood.

The MMR to the deviants /ba/ and /ga/ were explored by extracting latency and amplitude measures in specific time windows. To determine these time windows, cluster-based permutation analyses were performed on each deviant and age group, and for all electrodes of interest (F3, Fz, F4, C3, Cz, C4, P3, Pz, P4) as recommended in Werwach et al. (2022). From these analyses, two main time-window where the

response to standard and deviant significantly differed were extracted: 180-300 ms (found for deviant /ga/), and 300-480 ms (found for deviant /ba/). These time windows are consistent with the latency ranges typically observed in MMR studies in infants (Kailaheimo-Lönnqvist et al., 2020; H.-M. Liu et al., 2003; Rivera-Gaxiola et al., 2005). Interestingly, a clear difference in latency was observed, with an earlier response to /ga/ compared to /ba/. This difference in latency might be attributed to acoustic properties and perceptual salience. Indeed, Hornickel et al. (2009) investigated the subcortical responses to the same /ba/, /da/, and /ga/ syllables and found that the neural encoding of these stop consonants occurred in a hierarchy of salience. Importantly, /ga/ elicited the strongest response, followed by /da/, and /ba/ producing the weakest response. This hierarchy suggests that /ga/ has more salient acoustic features, which facilitates earlier and more robust neural processing. To better understand this phenomenon, I analyzed the F2 values, which is the only difference between the three stimuli. I observed that while the spectral deviation of both deviants' F2 from the standard's F2 at the onset of the syllables were similar (±375 Hz [1]), the slope of F2 between 0 and 55 ms differed between the three syllables. Indeed, the frequency differences between 0 and 55 ms were 145 Hz, -230 Hz, and -605 Hz for /ba/, /da/, and /ga/ respectively. Thus, the more abrupt consonant-vowel transition in /ga/ leads to greater salience and earlier neural encoding (Hornickel et al., 2009). The fact that I found earlier cortical time windows for the /da/-/ga/ contrast than for the /da/-/ba/ contrast suggests an acoustic processing mechanism, based on the salience of the consonants themselves rather than a phonological processing. Considering that the stimuli were nonnative, it is likely that infants relied on the acoustic properties of the speech sounds rather than their phonological processing skills. However, this might be true only for older infants (10-month-olds and older), as the phonological categories emerge typically around 10 months of age for consonants (Werker & Tees, 1984b). To further investigate this assumption, I will now discuss the results found for the difference waves.

Exploration of the difference waves in the two time windows (180-300 ms and 300-480 ms) revealed different developmental patterns for the two deviants (see Figure 4.12). The mean amplitudes were taken as the main measure of change detection, as done in previous work (François et al., 2017; Friederici, 2005; Rivera-Gaxiola et al., 2005). Variation of the mean amplitudes across age showed that for the deviant /ba/, the amplitude shift from more positive to more negative values between 6 and 8 months of age, then increases to more positive values at 10 months of age, creating a positive peak at 10

1. For deviant /ba/: 1210-1440 = -375 Hz; for deviant /ga/: 1815-1440 = 375 Hz.

months of age. Then the mean amplitude decreases again linearly up to 24 months of age. For the /ga/ deviant, the mean amplitude progressively increases from more negative to more positive values between 6 and 10 months, then decreases between 10 and 18 months, and increases again between 18 and 24 months. For both deviants, the patterns are similar in the earliest (180-300 ms) and latest (300-480 ms) time window. However, the slopes are steeper in the earliest time window, but the difference between the range of values is larger in the latest time window. It has been argued that positive MMRs (p-MMRs) reflect high attentional demands during auditory discrimination, while negative MMRs (n-MMRs) indicate more automatic processing. Both types of MMR are found in young infants, with p-MMRs being associated with immature MMR, and a shift from p- to n-MMR has been argued to reflect a maturation of the automatic detection change brain response (Friederici et al., 2002; He et al., 2007; Leppänen et al., 1997; L. Trainor et al., 2003; Werwach et al., 2022).

Considering these elements, I interpret the patterns of the mean amplitude as follows. Before 10 months, the decrease toward more negative values indicates the maturation of the MMR, which transits towards a more mature response as the infants' brains develop. At this stage, the categorical perception is not established (Kuhl et al., 1992, 2008; Werker & Tees, 1984b). However, between 8 and 10 months of age, the increased exposure to the native language induces the emergence of phonological representations of the native repertoire. Thus, at 10 months of age, infants start to change their strategy and rely on top-down processes, recruiting their native phonological representations. Because infants were presented with synthetic nonnative syllables, the attentional demands might have been increased as they tried to process the stimuli as native phonemes. Even if the stimuli were processed as nonnative speech sounds, infants were still able to discriminate the deviants from the standard, by relying on the phonetic properties. This would explain the positive peak in the mean amplitude of the difference waves at 10 months, reflecting these high attentional demands. This interpretation might be nuanced by the results obtained from a short speech perception task conducted in adults. Because the phonemes /b/, /d/, and /g/ exist in French with different spectro-temporal features than in English, I wanted to know if these stimuli would have been perceived as native or non-native by French adult speakers. The perceptual experiment conducted with 9 French speakers revealed that most of the time, the synthesized consonants were perceived as their corresponding voiceless, i.e., /pa/, /ta/, and /ka/. Thus, it is possible that the syllables were perceived as native /pa/, /ta/, /ka/. In the developmental pattern associated with the /ba/ deviant, the linear decrease of the mean amplitude toward

more negative values between 10 and 24 months suggests that the /ba/ might have been perceived as a native /pa/ and yielded more mature negative MMR as infants got older. However, a different pattern is observed for the /ga/ deviant. The mean amplitude stays positive at 10, 18, and 24 months, with no significant difference between 10 and 24 months. A possible explanation could be that the greater salience of the spectral features of the /ga/ may have prevented infants from processing the syllable as a native /ka/, as opposed to more experienced adults. Indeed, as experience with the native language grows, speakers encounter a higher number of instances of the same phoneme, thus expanding their phonological representations (Kuhl et al., 1992, 2008).

The interpretation suggested here is supported by models of phonological refinement like the Perception Assimilation Model PAM (Best, 1993, 1994, 1995a, 1995b) or the Native Language Magnet theory NLM from (Kuhl, 2000; Kuhl et al., 1992). These models propose that as individuals are increasingly exposed to the speech sounds of their native language, their ability to perceive and categorize non-native phonetic distinctions diminishes. From the PAM perspective, the syllables /ba/, /da/, and /ga/ might have been perceived as "weak" or "non-prototypical" forms of the French phonemes /pa/, /ta/, and /ka/. Considering the "magnet effect" of the NLM, the distance from the prototypical form of the native phonemes might have varied across syllables, yielding different patterns in the mean amplitudes of the differences waves. Indeed, the /ba/ deviant might have been closer to a French /pa/ than the /ga/ deviant was to a French /ka/. Overall, these models emphasize a shift from phonetic to phonological processing, highlighting the role of experience in shaping auditory perception. This supports the idea that exposure to specific phonetic categories during critical periods of language development fosters robust neural representations. My findings not only reinforce the theoretical underpinnings of PAM and NLM but also highlight the interplay between experience and perception in the development of speech perception abilities.

Time constraints did not allow for analyzing the correlations between brain responses and later language development. Nevertheless, MMR amplitudes to speech sounds have been previously associated with later language outcomes in both typical (Rivera-Gaxiola et al., 2005; Schaadt et al., 2023; Von Holzen et al., 2018) and clinical populations (Bonte et al., 2007; Hämäläinen et al., 2018; Linnavalli et al., 2017; van der Leij et al., 2013), as well as infants raised in multilingual environments (Garcia-Sierra et al., 2011). More nuanced results were also found, as in the work of Kailaheimo-Lönnqvist et al. (2020) in 6-month-olds. They used a multi-feature oddball paradigm involving pseudo-words

rather than syllables. Three different deviants were presented: duration, frequency, and vowel change. Language outcomes were assessed at 6 and 12 months of age via the Finnish version of the standardized parental questionnaire Infant–Toddler Checklist (ITC). They found a correlation between the amplitude of the P1 component in the averaged response to standards and the difference between the linguistic scores at 6 and 12 months of age. Moreover, the MMR to frequency deviant at 6 months old was positively correlated with the score of prelinguistic development at 6 months of age.

4.4. Limitations and future directions

One of the main limitations of this work, which is also a strength, is the synthetic nature of the stimuli. Indeed, using artificial syllables allowed for controlling the spectro-temporal features with high precision, which was crucial to assessing the subcortical encoding of these features. However, the syllables sounded 'robotic' and it was not clear if they were perceived as native or nonnative by the infants, as exposed in the discussion. To go even further, one might wonder to what extent these stimuli were perceived as speech sounds. Thus, future work should replicate this study and include a control condition with native stimuli produced by a real native speaker.
Moreover, I did not have the time to analyze the correlations between cortical and subcortical indices of speech sound encoding and later language outcomes, although I collected neurodevelopmental data. Future exploration of the dataset should include these analyses. This way, I could decipher which measures of early cortical and subcortical speech sound encoding constitute the best biomarkers of later language outcomes (Garcia-Sierra et al., 2016; Kuhl et al., 2005; Liebig et al., 2020; White-Schwoch et al., 2015).
Finally, I would like to explore the relations between cortical and subcortical measures of speech sound encoding, to further investigate the interactions between the two structures of the auditory pathway in critical periods of development.

5. Conclusion

To conclude, this study showed that recording cortical and subcortical brain responses to speech sounds in young infants using clustered ISIs is possible. Visual inspection of the data combined with the computation of analyses similar to previous studies yielded consistent results with the existing literature. In addition, I contributed to a better understanding of how subcortical encoding of spectro-temporal features of speech

sounds develops over the first two years of life. Cortical brain responses to synthetic nonnative syllables opened a new discussion in the scope of perceptual narrowing theories.

The first two years of life is a period of high plasticity that involves multiple critical periods for speech perception, modulating later linguistic skills. Thus, longitudinal studies are essential to track these rapid changes and provide fine-grained descriptions of these developmental trajectories.

Chapter 5. Investigation daily auditory environment in infants with LENA

Contents

1 Introduction 146
2 Related Work 147
2.1 Experimental studies 147
2.2 Validation studies 148
3 Rationale 149
3.1 Participants 149
3.2 Procedure 150
4 Tools and Methods 150
4.1 LENA 151
4.2 VTC and VCM 154
5 Dataset 154
6 Investigating the dataset 155
6.1 Testing age 155
6.2 Interactional metrics 157
7 Discussion 159
7.1 Dataset investigation: linguistic input and contingencies 159
8 Limitations 161
9 Conclusion 162

The present chapter is dedicated to the study of infant's daily auditory input through daylong recording data collected with the LENA system. Because this dataset has been presented in a conference paper, most sections of this chapter correspond to Hervé et al. (2024) (see Annex D for the published version).

1. Introduction

Infants rely on their daily auditory environment to develop language and other cognitive skills. Pioneering studies interested in these early auditory inputs used observatory experiments in laboratory settings or short recordings that were manually annotated (Hart & Risley, 1992b; Keller et al., 2004b). In the last decades, technological advances brought new tools that allowed the collection and analysis of more considerable and ecological datasets. Daylong recordings are now increasingly used in developmental studies (Bergelson et al., 2023; H. Ganek & Eriks-Brophy, 2018a), especially since the release of the Language Environment Analysis (LENA) system in 2004.

Daily auditory environments have been described in a variety of populations (Aragon & Yoshinaga-Itano, 2012; Caskey et al., 2014a; Christakis et al., 2009; Warren et al., 2010), highlighting the positive effects of early caregiver-infant interactions on language development (Bergelson & Aslin, 2017; Gilkerson & Richards, 2008a; Warlaumont et al., 2014a). Nonetheless, only two datasets were collected in French-speaking households (Canault et al., 2016b; Orena et al., 2019a). Here, we expand the literature by describing an original dataset of infants' daylong audio recordings gathered in twenty French-speaking families.

I focused on 3-to-24-month-old babies for several reasons: (1) it allows a direct comparison with Canault et al. (2016b)'s and Warlaumont et al. (2014a)'s results ; (2) it includes crucial steps for language development, including the emergence of phonemic categories between 6 and 10 months (Cheour, Ceponiene, et al., 1998; Werker & Tees, 1984a), and the vocabulary spurt between 18 and 24 months (Benedict, 1979; Goldfield & Reznick, 1990b; Nazzi & Bertoncini, 2003) ; (3) the two first years of life constitute a critical period where caregiver-infant interactions are early precursors of later language outcomes and cognitive skills (Bergelson & Aslin, 2017; Gilkerson & Richards, 2009b; Warlaumont et al., 2022; Weisleder & Fernald, 2013b).

LENA's output correlation with human annotations has been assessed in several languages, suggesting good reliability (Busch et al., 2018; H. V. Ganek & Eriks-Brophy, 2017; Gilkerson et al., 2015; Pae et al., 2016a; Weisleder & Fernald, 2013b; Xu, Yapanel, Gray, Gilkerson, et al., 2008). However, only one study provided evidence for LENA system reliability in European French, yielding relatively good results (Canault et al., 2016b). Moreover, LENA validation studies implied listening to the continuous raw audio recordings. In addition to being highly time-consuming, this approach raises critical ethical issues associated with data privacy (Casillas & Cristia, 2019; Cychosz et al., 2020). Here, we override these difficulties by comparing the LENA metrics outputs with other annotation systems.

The paper has the following contributions: (i) describe a new French corpus of auditory LENA-recorded data, (ii) compare different automatic annotation tools, (iii) provide a picture of the daily auditory environment in French-speaking families in 3-to-24-months infants, (iv) show the potential of the data set by replicating daylong recordings-based results on developmental trajectories.

2. Related Work

This section is a brief overview of the existing literature regarding daylong recording studies in developmental populations. We identified two main types of studies: (*i*) experimental studies that used daylong recordings as a tool to answer a specific research question and (*ii*) validation studies that focused on assessing the reliability of the recording and analysis tools themselves.

2.1. Experimental studies

Daylong recording studies in infants often involved the LENA system. Four years after its release, the first LENA normative study, the “Natural Language Study (NLS)” was conducted by the LENA Research Foundation (Gilkerson & Richards, 2008a). This report relied on the three main LENA metrics (Adult Word Count: AWC; Child Vocalization Count: CVC; Conversational Turn Counts: CTC) to describe daily auditory environments in 329 English-speaking infants aged 2 to 48 months. Then, more experimental works involving daylong recordings in infants began to emerge (see H. Ganek and Eriks-Brophy (2018a) for a review).

Studies that focused on the characteristics of the daily auditory environment in typically developing infants revealed that children’s vocalizations and child-caregiver interactions increased with age within the first two years of life (Gilkerson & Richards, 2008a; Pae et al., 2016a). Warlaumont et al. (2014a) proposed a "social feedback loop" in which contingencies between adult-child and child-adult speech-like vocalizations contribute to increasing interactions between infants and adults through age. Additionally, a higher proportion of adult-child interactions has been associated with larger vocabulary size (Weisleder & Fernald, 2013b). The impact of various factors like multilingualism (Oller et al., 2010b; Orena et al., 2019a; Ramírez & Hippe, 2024), socio-economic status (Bergelson et al., 2023), exposure to TV Christakis et al. (2009) and Zimmerman et al. (2009), musical inputs (Mendoza & Fausey, 2021), activity during the day (Soderstrom & Wittebolle, 2013) and temporal dynamics of the surrounding sounds (Warlaumont et al.,

2022) have been investigated as well. Daylong recording studies in clinical populations showed the importance of understanding infants' daily soundscape for early language intervention (Aragon & Yoshinaga-Itano, 2012; Caskey et al., 2011; Warlaumont et al., 2014a; Warren et al., 2010).

Overall, age ranges, sample sizes, and recording spans varied across studies. Some infants were included as early as 2 months of age (Aragon & Yoshinaga-Itano, 2012; Bergelson & Aslin, 2017; Zimmerman et al., 2009), while others started after 12 months of age (Oller et al., 2010b; Warren et al., 2010; Weisleder & Fernald, 2013b). Children could be followed longitudinally within various periods (Gilkerson et al., 2018b; Sy et al., 2023) but not systematically (Bergelson & Aslin, 2017; Weisleder & Fernald, 2013b).

Although most studies relied on the LENA system, methodological choices regarding data collection and analysis were various. For example, some authors chose to rely on preexisting datasets that already fitted their research questions (Aragon & Yoshinaga-Itano, 2012; Christakis et al., 2009; Warren et al., 2010). For data analysis, the LENA metrics were mostly used although some preferred to develop their own tools (Al Futaisi et al., 2019; Lavechin et al., 2020; MacWhinney, 2000; Räsänen et al., 2021).

2.2. Validation studies

The first LENA validation study was led in 2008 on American English, as part of the NLS (Xu, Yapanel, Gray, Gilkerson, et al., 2008). Human annotations were compared with automatic outputs provided by the LENA software to determine agreement scores, measured with Pearson's correlations. LENA's AWC and CVC reached $r = 0.82$ and $r = 0.76$ respectively, indicating reliable LENA annotations for subsequent English-speaking environment studies (Christakis et al., 2009; Gilkerson et al., 2017a; Warren et al., 2010; Xu, Yapanel, Gray, & Baer, 2008a; Zimmerman et al., 2009).

Later, the same validation procedure was applied to other languages, focusing on the three main LENA metrics (AWC, CVC, and CTC). Overall, the AWC metric was the most reliable, although several authors reported that, on average, the LENA's estimations were lower than the human counts (Canault et al., 2016b; Xu et al., 2009b). Agreement scores for AWC were reported in Spanish ($r = 0.80$, Weisleder and Fernald (2013b)), Mandarin ($r = 0.72$, Gilkerson et al. (2015)), Korean ($r = 0.72$, Pae et al. (2016a)), and Dutch ($r = 0.87$, Busch et al. (2018)). CVC and CTC's reliability were not systematically assessed and yielded variable results, ranging from $r = 0.52$ (Busch et al., 2018) to $r = 0.84$ (Gilkerson & Richards, 2008a) (see Appendix A in Annex D). In French, we found Canault et al. (2016b)'s report as the only existing validation study so far. They manually annotated and transcribed 324 ten-minute samples recorded in 3-to-48-

month-olds: Pearson's correlation scores were $r = 0.64$ for AWC and $r = 0.71$ for CVC. These results suggest good reliability for LENA metrics in French, although slightly below the abovementioned languages.
Cristia et al. (2021b)'s comprehensive validation study in three different linguistic and socio-cultural environments calls for more validation studies with more detailed and systematic methods. However, the concurrent emergence of annotation tools (Al Futaisi et al., 2019; Lavechin et al., 2020; MacWhinney, 2000; Räsänen et al., 2021) tends to increase methodological variability. To converge toward a standardized pipeline for daylong data management, Gautheron et al. (2023) developed the ChildProject package. It is compatible with many existing annotation formats and allows annotation systems comparisons. Here, we relied on these tools to compare LENA's metrics with measures extracted from the `Voice Type Classifier (VTC)` from Lavechin et al. (2020) and the `VoCalisation Maturity analysis (VCM)` from Al Futaisi et al. (2019).

3. Rationale

3.1. Participants

Infants were recruited between 3 and 18 months of age in three daycare centers from a French major city. An official collaboration between our team and the daycare centers was established to facilitate both participants' recruitment and data collection. we met parents in person to communicate the project and obtain their informed consent. The French Ethics Committee Review Board approved the study, which was conducted according to the guidelines of the Declaration of Helsinki (World Health Organisation, 2008). Parents filled out a questionnaire to ensure that infants did not have any hearing, cognitive, or developmental disorders and that they were raised in a dominant French-speaking environment. Other metadata were gathered through this questionnaire: number of caregivers, musical practice of the caregiver(s), linguistic environment (which language(s) spoken around the child), and socio-economic status (SES) assessed via profession.
Independently of their age at the inclusion date, we followed infants until 24 months of age when possible, or as long as possible otherwise. Twenty infants were involved, with a mean age at inclusion of 12 months ($m = 360$ days, $sd = 132.5$). Six additional babies were recruited but excluded from the analysis because parents did not provide enough recordings (<5).

3.2. Procedure

As mentioned above, we collaborated with three daycare centers that became a hub for data collection. Once parents had given their informed consent, they were provided with the LENA materials: a recorder and a t-shirt with a frontal pocket. Each infant had one unique recorder that they kept until the end of data collection. Clothing size was adapted to infants and changed throughout the months when needed. To help parents get used to the LENA system, we gave them some oral instructions when possible so they could ask questions and we could make sure they understood everything. Additionally, all families were given an instruction sheet that was taken from LENA's support materials and adapted to our study. Instructions comprised information about when and how often to record, how to use the device, various recommendations, and the procedure for device deposit and pickup at daycare. Parents also had our contact information and could reach us whenever they needed.
Parents were asked to have their child wear the recorder once a week for a full day, preferentially at home or during the weekends. To limit attrition, we accepted recordings at daycare occasionally, when they could not record another day or if they forgot. The frequency of the recordings was hard to maintain for some families, so we had to send them kind reminders sometimes. But overall, all families were very involved and consistent. We recommended that during a recording day, they never turn the recorder off, limit noisy environments, and let the recorder nearby while showering and during bedtimes. Once the recording was completed, they were asked to bring the LENA recorders back to the daycare center once a week. Then, the investigator could transfer the data to the database the same day, so parents could get the recorder back and start over for a new week. At the end of data collection (when the child reached 24 months or when families decided to stop), infants were given a "baby researcher" diploma and a customized t-shirt as a reward.

4. Tools and Methods

Daylong recordings present a set of challenges in terms of processing. The first constraint is the data set size: we gathered 10^4 hours of highly heterogeneous recordings (both across and within recordings) that need to be sampled. Existing literature has used the notion of *hot spots* (areas in the recordings with a high density of speech events) as well as a method consisting of human labeling of extremely short sound events (Semenzin et al., 2021). Due to the required infrastructure, the latter approach was not

considered for our work. Instead, we applied state-of-the-art computational techniques and packages to perform step-by-step reliability tests and calculate agreement scores between them. The automatic tools we used for our analyses are the LENA suite (Gilkerson & Richards, 2008a) and a set of tools developed or adapted whithin the framework of ChildProject (Lavechin et al., 2020).

4.1. LENA

We used the LENA system (Foundation, 2015) for both data collection and analysis. [1] This tool released in 2004 by the Research Foundation combines a wearable audio recorder with an automated analysis software.

The recording hardware consists of a small digital language processor (DLP) (dimensions: 3-3/8" x 2–3/16" x 1/2" [2]; weight: 2-3 ounces [3]) that disposes of a display screen, two buttons for powering and recording, and a USB port for data transfer and charging. The LENA system also provides specific clothing designed for children (size starting at 3 months) with a front pocket adapted to the DLP's size (see Figure 5.1). The recorder is inserted in the pocket by the parents and secured with press studs. Once activated, the DLP can save up to 16 hours of raw audio data that can then be uploaded in WAV format. Experimenters usually recommend that parents leave the recorder on until it turns off, thus reaching the maximum number of recording hours and completely draining the battery. Once recharged, the same or a new participant can use the DLP again without any constraint on the number of recordings.

The LENA software includes a repository that stores data by participant and an automatic analysis program. The repository is optimized for longitudinal follow-up of children because data from repeated recordings of a given participant are grouped in the same folder and can be compared over time. Reports on quantitative measures extracted from the automatic annotation software are provided and retrievable at any time. The annotation software automatically segments the raw recordings based on acoustic features such as intensity or pitch (Xu, Yapanel, Gray, & Baer, 2008b). Its goal is not to access the semantic content of the recordings (Gilkerson & Richards, 2020), thus it does not provide transcriptions but only annotations. The program first identifies speech segments against other types of sounds (noise, media, silence) and classifies all segments into one of the eight predetermined categories. The segments are next

1. This section has been modified from the original paper version to provide more details about the LENA system.
2. 8.6 x 5.6 x 1.3 centimeters
3. between 60 and 80 grams

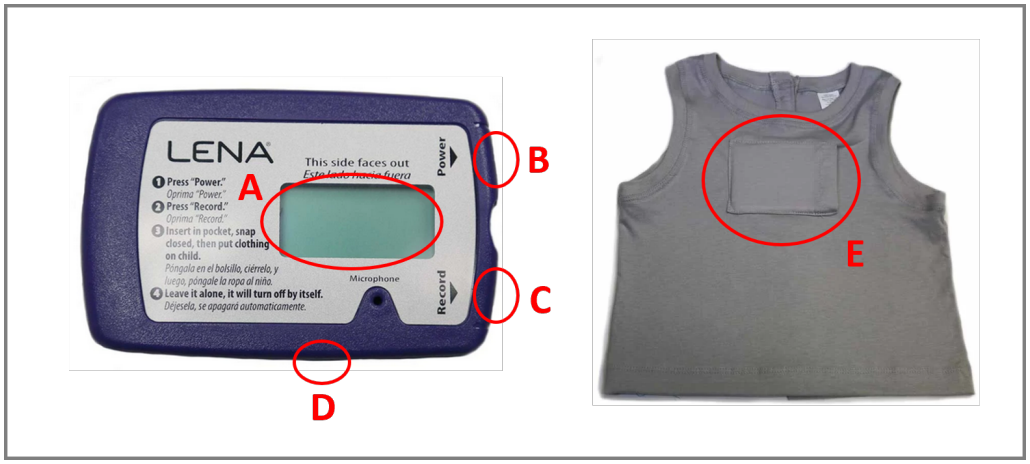


Figure 5.1. – LENA hardware. Left: DLP with display screen (A), powering (B) and recording (C) buttons, and USB port (D). Right: LENA clothing with front pocket (E).

compared to a human-trained model and passed through a silence likelihood ratio test. Based on that, the segments marked close to silence are labeled 'unclear' and excluded from further analysis. The eight categories include four "human vocal activity" categories and four "others" categories as follows:

1. **Human vocal activity categories:**
   - **Key child**: referred to as *CHild Near (CHN)*, segments where the child wearing the device (*key child* or *target child*) is producing sounds. These segments are further categorized as 'vegetative sounds' (sneezing or burping), 'fixed signal sounds' (crying or laughing), or 'vocalizations' (speech-like sounds defined by qualities like duration, rhythm, and pitch control (Gilkerson & Richards, 2008b; Oller et al., 2010a). The 'vocalizations' segments are used to create the CVC, an estimation of the number of words produced by the key child.
   - **Adult male**: referred to as *Male Adult Near (MAN)*, segments where a male voice has been identified near the child. It includes speech and non-speech (laugh, screams) vocalizations.
   - **Adult female**: referred to as *Female Adult Near (FAN)*, it corresponds to the same adult category as Adult Male but for female voices. The MAN and FAN segments are further analyzed by a speech recognition algorithm that provides an estimation of the Adult Word Count (AWC).
   - Other child: referred to as *Other Child Near (CXN)*, segments identified as child vocalizations but do not originate from the target child.
2. **Other categories:**

- **TV/electronic sounds**: referred to as *TV Near (TVN)*, segments identified as TV, media (radio) or any other electronic sound; gives an estimation of how much time these auditory inputs were present in the child's environment.
- **Noise**: referred to as *NOise Near (NON)*, noisy segments where any of the other categories could be identified. Another close category called Fuzzy (FUZ) includes low-amplitude and low-likelihood segments. Both are considered noisy segments.
- **Silence**: referred to as *SILence (SIL)*, indicates the duration of silence segments.
- **Overlapping sounds**: referred to as *OverLap Near (OLN)*, indicates the duration of overlapping segments, i.e. segments with a high likelihood of speech but with overlaps.

An estimation of the number of conversational turns (Conversational Turn Count, CTC) per recording is computed using the AWC and CVC (Child Vocalization Count) variables. A CTC corresponds to the subsequent occurrence of child and adult vocalization within five seconds, in any order but without interruption from another human vocal activity segment (Gilkerson & Richards, 2020). The software provides accessible reports on the repository that include all these metrics. In particular, reports generate graphs to inform about the proportion and amount of noise, silence meaningful (CHN, MAN, FAN, CXN) and distant (OLN, FUZ) speech. CVC, AWC, and CTC values are also displayed on graphs and are additionally compared to the LENA Research Foundation's normative study results (Foundation, 2015; Gilkerson et al., 2017b). A key strength of this software is that all these variables and reports are available the same day data are uploaded. Additional information can be extracted using the Advanced Data Extractor (ADEX), a free software created and shared by the LENA Research Foundation (Ford et al., 2008; Foundation, 2015). ADEX gives more detailed measures and estimations, like separating word counts for male and female adults, or obtaining the duration of the other child vocalizations, overlapping speech, TV, noise, and silence. It gives more control over the data: for example, researchers can pre-filter the data to focus on a specific type of speech or interaction and can choose the time resolution and intervals that they want to analyze.

In summary, the LENA technology is very convenient for developmental research as it meets all expectations regarding dataset size, ecological validity, and processing time for optimized assessment of daily auditory environments in infants.

4.2. VTC and VCM

The ChildProject suite starts processing recordings with the *voice-type-classifier (VTC)* (Lavechin et al., 2020), which relies on the state-of-the-art speech diarization tool, *pyannote* (Bredin et al., 2020). *VTC* identifies sound activity segments that can be mapped to some of LENA's categories: target key child (KCHI), other children (CHI), female (FEM), and male (MAL). Another tool, *VoCalisation Maturity analysis (VCM)* (Al Futaisi et al., 2019), refines the output of *VTC*. *VCM* is grounded on the state-of-the-art signal processing and emotion recognition tool, *SMILE* (Eyben et al., 2010), and more precisely on the Geneva Minimal Acoustic Parameter Set (GeMAPS) (Eyben et al., 2015). It adds information to the labeled categories (e.g., speech from the target child) by determining whether the targeted speech is Canonical (CNS), Non-canonical (NCS), cries (CRY), or other sounds (noise, laughter). Such classification has been used in Casillas et al. (2017), for example.

5. Dataset

The corpus currently consists of 8286 hours of LENA daylong recordings. Table 5.1 indicates the mean, minimum, and maximum values for the recording period (age span), number of recording sessions, and length of the recordings.

	Avg	Min	Max	Sum
Age span (months)	9.85	3	18	-
# sessions	27.0	6	66	540
Duration (hours)	414	87	1022	8286

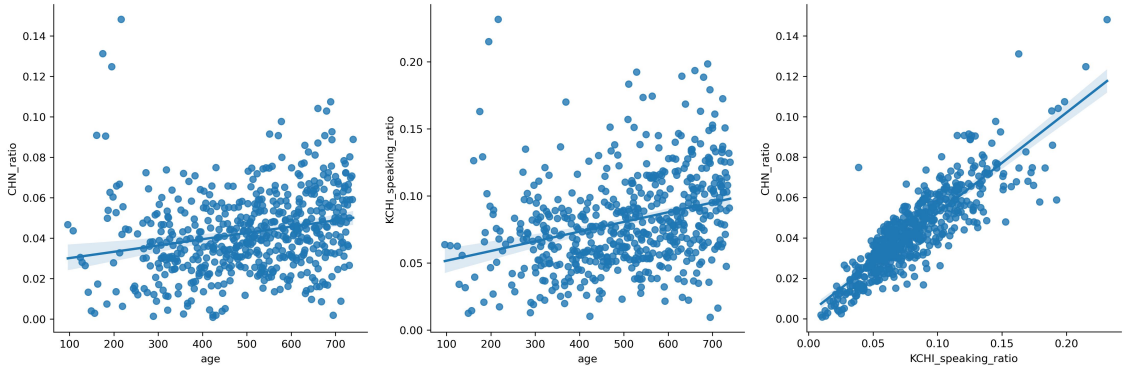
Table 5.1. – The data set. N = 20 children. Age span: number of months between the first and the last recording.

Table 5.1 reflects a high variability in parents' use of the LENA device. As mentioned earlier, we asked them to turn the DLP on in the morning and leave it until it automatically turns off after 16 hours of recording. However, some families turned the device on and off multiple times during the day or stopped the recording before reaching 16 hours. Thus, we observed variability across participants in recording length and number. Additionally, there was variability in the recording span: not all children started the recordings at the same age, and not all were followed until 24 months of age. Given these observations, we selected a sample of children that 1) had at least 9 months of recording span and 2) provided at least 10 recordings. These thresholds allowed us to

focus on more representative datasets while maintaining a sufficient number of data points to observe developmental trajectories. A sample of 10 children met these two criteria and were selected for complementary analyses (see Table 5.2).

	Avg	Min	Max	Sum
Age span (months)	14.2	10	18	-
# sessions	40.9	12	66	409
Duration (hours)	637	192	1022	6366

Table 5.2. – Selected children for individual longitudinal metrics and plots. N = 10 children (age recording span ≥ 9 months; number of recordings ≥ 10).



(a) LENA **(b)** VTC **(c)** Correlation ($r = 0.862$)

Figure 5.2. – Target Child Speaking Time Ratio for (a) LENA (controlled for child id, $\beta = 0.159$; SE = 0.027; p < 0.001); (b) VTC ($\beta = 0.242$; SE = 0.029; p < 0.001) and (c) correlation plot between LENA and VTC. All children included (n=20). Age in days.

6. Investigating the dataset

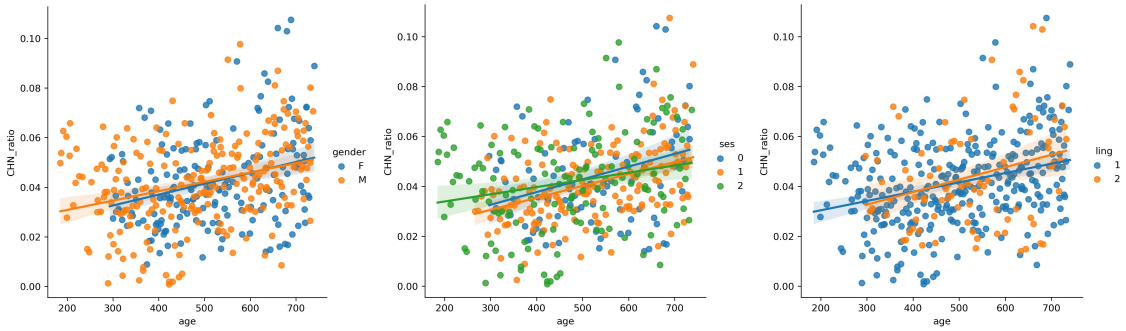
6.1. Testing age

Our first goal was to test the reliability of the metrics extracted with the three targeted tools (LENA, VTC, VCM) on our data. A crucial check for our dataset consisted of testing whether children's production evolved with age. We expected an increase in children's speech-related metrics (such as speaking time and ratio, vocalization counts, etc.), while adults' metrics would remain stable. Finally, the voices of other children present in the recordings were expected to increase as the siblings of the target child followed their

own development. We began by examining the production time ratio (the percentage of the recording time occupied by a given category), for example, for the target child as shown in Figure 5.2 (see Appendix D in Annex D for other categories).

More precisely, we examined the production ratio calculated from LENA (sum of the duration of intervals labeled with *CHN* as the speaker, divided by recording duration), VTC (using the same approach with the label *KCHI*), as well as the correlation between the two measures. Importantly, the ratios obtained for the different voices from these two tools were highly correlated. More generally, all the metrics extracted with both approaches were highly correlated for all comparable categories.

We tested whether age remained a dominant factor when controlling for the available metadata. In Figure 5.3, we plot the production ratio alongside gender, socio-economic status (high vs. low), and linguistic context (monolingual vs. plurilingual).



(a) Gender **(b)** Socio Economic Status **(c)** Linguistic Context

Figure 5.3. – Target child Production Time (from LENA) differentiating for (a) gender ; (b) socio-economic status ; (c) linguistic background. Only age is significant. The three other variables are not (see Appendix D in Annex D for details. Selected children. Age in days.

From these figures, a general observation is that children's speaking ratio increased with age. This was tested by conducting a linear mixed model analysis using pymer4 (Jolly, 2018). We treated *'target child speaking ratio'* as the dependent variable and *'age'* as the fixed effect, with *'child ID'* as the random effect. Only *'age'* had a significant effect on the target child production ratio ($\beta = 0.261$, $SE = 0.028$, and $p < 0.001$), controlling for *'gender'*, *'linguistic environment'*, and *'socio-economic status'* (all not significant).

VCM metrics allowed us to refine our evaluation of child production with age. We applied it to our set of selected children and found (Figure 5.4) that the increase in speaking ratio was due to an increase in real child speech (both canonical and non-canonical) rather than to a variation in the proportion of cries, laughter, or noise. In addition, we replicated findings from Warlaumont et al. (2014a), who tested the evolution of

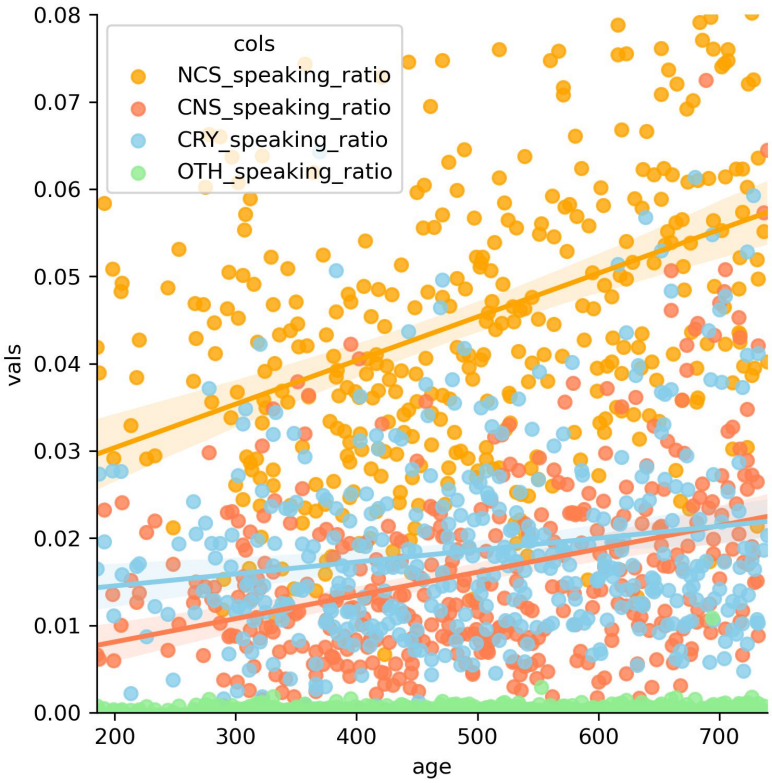
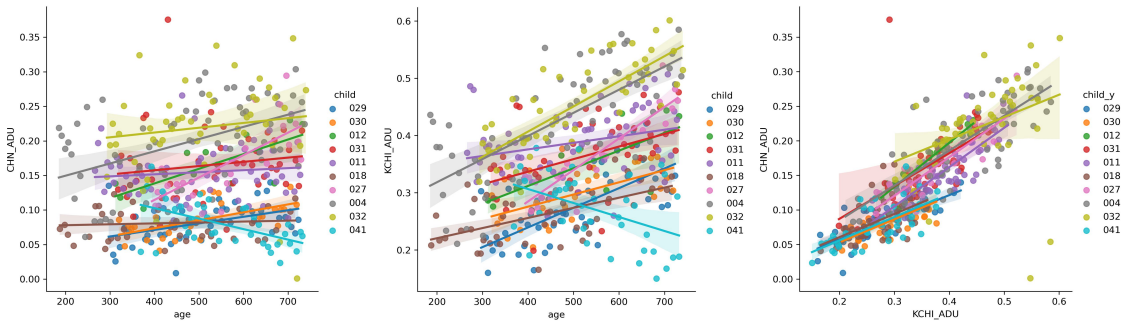


Figure 5.4. – Speech vs. Non-speech VCM extraction for the selected children.

speech-related vocalization ratio (versus non-speech-related) produced by the target child. Using our own pipeline and analysis, we also found a significant increase in this ratio with age while controlling for all metadata ($\beta = 0.219$, $SE = 0.022$, $p < 0.001$) (see Appendix C.1 in Annex D).

6.2. Interactional metrics

Conversations arguably constitute the most important aspect of the linguistic environment. First, we approached this by examining temporal contingencies (K. Bloom et al., 1987; Warlaumont et al., 2014a) between the child and the other voices in the recordings. Specifically, we analyzed instances of target child productions followed by another participant, as well as target child productions preceded by other participants. Working with manually transcribed data, Nikolaus et al. (2022) explored the effect of time-contingent responses on children's intelligibility. These studies found that (1) caregivers provided more time-contingent responses to intelligible utterances from the child and (2) children produced more intelligible utterances if their caregivers were responsive. Then, we investigated the "social feedback loop" as proposed by Warlaumont



(a) LENA **(b)** VTC **(c)** Correlation ($r = 0.831$

Figure 5.5. – ADULT>CHILD>ADULT contingencies for (a) LENA ($\beta = 0.131$; SE = 0.024; p < 0.001), (b) VTC ($\beta = 0.130$, $SE = 0.024$, $p < 0.001$) and (c) correlation plot. Selected children. Age in days.

et al. (2014a).

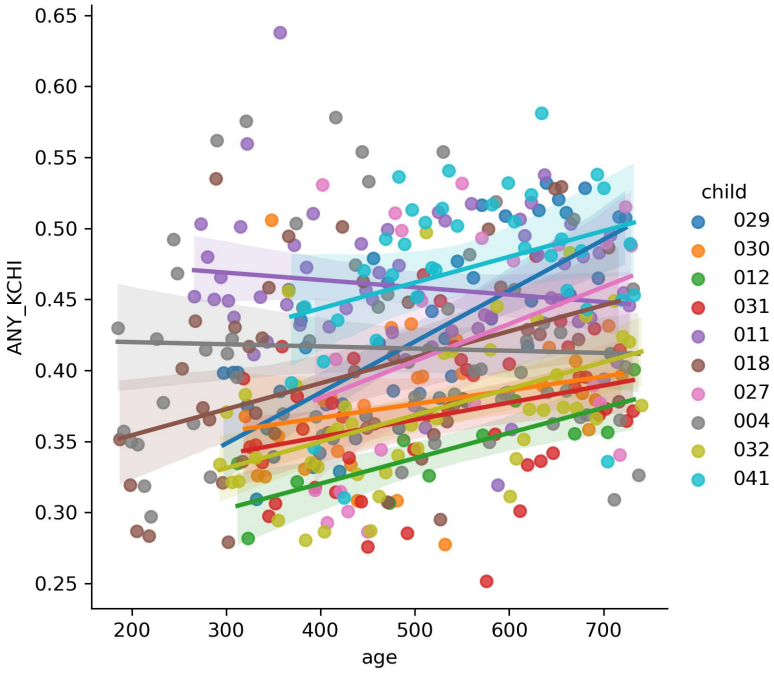


Figure 5.6. – ANY>CHILD>ANY contingencies from VTC speech extraction. ($\beta = 0.158$, $SE = 0.030$, $p < 0.001$)

I employed a similar approach for both tools at our disposal. For *CHILD>ADULT* contingencies, we selected all target child productions and checked whether there was a production from an adult (MAL + FEM) participant **1 second after**. While LENA metrics extract similar information, we aimed to use the same method for both tools to enable a direct comparison. Figure 5.5 depicts the comparison for *CHILD>ADULT* contingencies.

We also looked at ANY>CHILD contingencies by considering any activity occurring **2 seconds before** [4] before a child production. Figure 5.6 provides the VTC extraction for *ANY>CHILD* contingencies (additional combinations are included in Appendix E in Annex D). We tested, for VTC data, the relationship between *age* and temporal contingencies for *CHILD>ADULT* ($\beta = 0.326$, $SE = 0.026$, $p < 0.001$) and *ANY>CHILD* ($\beta = 0.158$, $SE = 0.030$, $p < 0.001$), while controlling for gender, linguistic environment, and socio-economic status. This result was further refined by replicating the second finding from Warlaumont et al. (2014a) on our dataset (and with LENA metrics this time). Specifically, we confirmed that children's speech-related productions tend to elicit more feedback from adults (see Appendix C.2 in Annex D).

Finally, we also replicated Warlaumont's result about the social loop on our data. We tested whether initial speech-related CHILD productions that were followed by an ADULT production 1 second after [5] were more likely to be followed by a speech-related CHILD production than a non-speech CHILD production within 3.5 seconds, compared to initial speech-related CHILD productions that did not receive an adult response. [6] This is illustrated in Figure 5.7, which shows the difference in speech-related / non-speech related ratio in child's production after an adult's response versus no response. This difference is positive, indicating that adult responses to children's speech-related productions tend to increase the proportion of child speech-related follow-ups.

7. Discussion

7.1. Dataset investigation: linguistic input and contingencies

The collection and investigation of our original corpus in French households facilitated the comparison of different analysis tools and the replication of previous results within the same age ranges.

First, we identified a robust relationship between target child production metrics and age. As expected, we did not observe a similar significant change for adult voices.

4. I considered allowing for a longer gap for children's follow-up to be appropriate. To count "turns", LENA metrics use a 5-second threshold.

5. Following Warlaumont et al. (2014a) and Nikolaus et al. (2022), who followed Oller et al. (2010b), which used a 1-second window to extract relevant vocal activity to investigate the "social feedback loop" for children between 8 to 48 months.

6. This 3.5 second is between the end of initial CHILD production and the start of the following one. The delay is to allow for a potential ADULT response to occur in between. It is not possible to use the ADULT production for proposing a simpler time threshold for the following utterance since there is not always an ADULT production in-between.

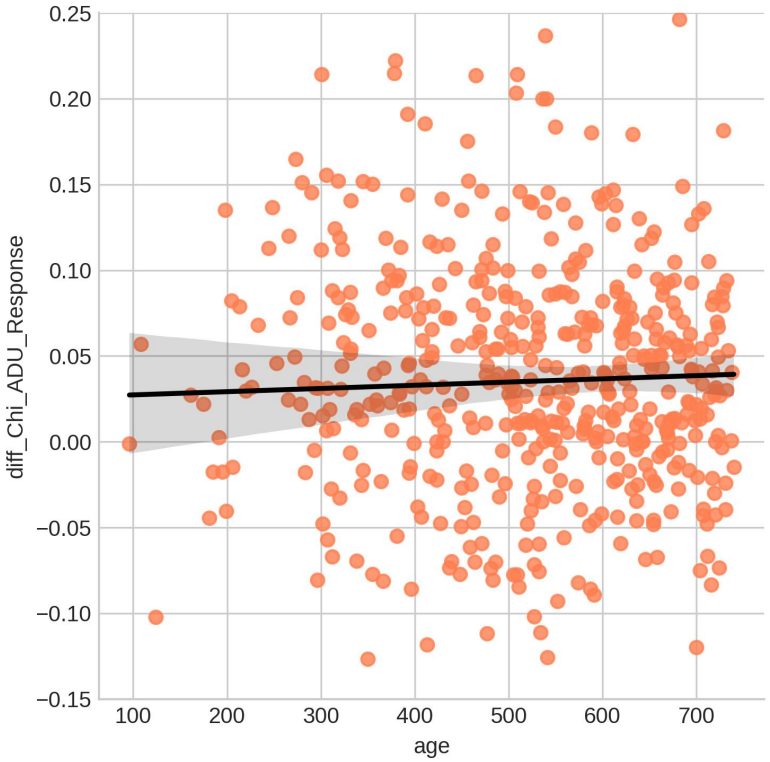


Figure 5.7. – Child ratio difference between Child speech and non-speech productions depending on whether or not an initial child speech-related utterance was responded by an adult or not (positive mean : indicating a tendency for more speech-related responses), replication of Warlaumont et al. (2014a).

However, the evolution was also positive and significant for other children's voices ($\beta = 0.179$, $SE = 0.027$, $p < 0.001$). This can be attributed to the behavioral path of siblings as well as other children in kindergartens. All these results were obtained using both LENA and VTC pipelines. Finally, we found coherent results in line with existing literature and across the tools we used. These results strengthen our confidence in both our recording protocol and in the metrics extraction and analysis.

These comments hold for interactional metrics as well: the increases in temporal contingencies involving the children were consistent with findings from previous studies (Nikolaus et al., 2022; Warlaumont et al., 2014a). We calculated temporal contingencies in a way that ensured these increases were not influenced by the overall amount of child productions.

Furthermore, our more detailed analysis (depicted in Figure 5.4) and the replication of Warlaumont's first results, showed that the increase in child's productions with age was due to speech-related productions and not to vegetative sounds or noise. In summary,

our data show that children do produce more speech while growing up in their first two years of life. These increased productions are temporally contingent on other speakers, regardless of the initiator of the interaction (target child or other speaker).

This is further elaborated by the replication of Warlaumont's third result about the social loop that showed the benefit of follow-up productions of adult feedback on children's speech. Contrary to Warlaumont et al. (2014a) (but in line with Bergelson et al. (2023) we did not find any effect of parental SES, gender, or linguistic environment on children's productions. This was the same for temporal contingencies used as a proxy for measuring linguistic interaction with the child.

Other studies have attempted to dive further into linguistic metrics from large child-caregiver datasets. Some refined the speech contingencies to distinguish corrective feedback or to approach the grammaticality evaluation of the productions in these datasets (Nikolaus et al., 2023). However, those datasets have been manually transcribed, while ours did not (and will not) receive such a transcription. A second major difference is the age range of the children. Most of the existing studies involved children from 12-24 months up to later ages. This inevitably changes the nature and characteristics of the metrics that can be extracted. Another trend of work focuses on phonetic learning based on daylong recordings, such as Lavechin et al. (2024) on perceptual attunement. Finally, in a more cultural and typological direction, Bergelson et al. (2023) studied a global sample of 1,001 child-centered audio capturing 2- to 48-months-olds from many countries and various cultural backgrounds.

8. Limitations

One major limitation of this work is the absence of manual annotation. For legal and ethical reasons, we are not in position to perform extensive manual annotations of raw audio data, as well as sharing raw audio. We needed to find other ways to check the reliability of our dataset. By replicating previous results from the literature and comparing different computational tools, we reinforced our trust in our dataset and overcame this constraint. All metrics derived from the corpus related to this paper as well as for future work will be made available in the `LLDC` public repository on Ortolang institutional platform https://www.ortolang.fr/. Moreover the code for producing the analyses presented in this paper is available at : https://github.com/prevotlaurent/LENA_CMCL.

9. Conclusion

This work first constitutes a replication of earlier results on daylong recordings, on a completely new and independent data set, using two different tools. This contributes to answer Cruz Blandón et al. (2023)'s call for more and better meta-studies on long recordings. Indeed, despite their creation cost, daylong recordings' significance is growing in cognitive science. Showing that these data sets, despite their noisy nature, do present enough reliability to gain insights about the children's language and communicative skills development is crucial.

The second contribution consists of the characterization of our data set itself. It is a large data set (>8000 hours of recordings) that is still growing at the time of writing. It is unique by being truly longitudinal with some children's environments being recorded over a two-year span. Finally, other experimental data were collected longitudinally in the same sample of infants. These are beyond the scope of this paper but open up the possibilities of cross-analyses between the characterization of the linguistic environment and other experimental results regarding language development. The present study, therefore, constitutes a crucial first step in this direction through the thorough exploration of the data set, and by considering individual variability.

From here, we now plan to refine the analyses by entering into more linguistic characterizations of these productions in terms of richness. We will consider tools that allow for more phonological measures such as Räsänen et al. (2021) and more content-based metrics (Nikolaus et al., 2023, 2024) that have been used so far only on transcript-based corpora from CHILDES (MacWhinney, 2000) and that can be now explored on daylong recordings thanks to the improvement of automatic speech recognition engines.

Chapter 6. Minimal pair associative word learning in 18- and 24-month-olds

Contents

1 Introduction 163
2 Methods 167
2.1 Participants 167
2.2 Associative word learning task 167
2.3 Vocabulary size and language and cognitive development assessment 171
2.4 Data analysis 172
3 Results 174
4 Discussion 184
4.1 Dataset investigation 184
4.2 Focus on the consonant contrast 187
4.3 The vocabulary spurt: a critical period for word-learning mechanisms? 189
4.4 Links with language outcomes 191
4.5 Limitations and future directions 192

1. Introduction

It is now largely established that children produce their first words at around 12 months of age on average. Nevertheless, evidence of familiar word comprehension at 6 months of age (Bergelson & Swingley, 2012; Tincoff & Jusczyk, 1999) suggests that infants learn their first words before they can produce them (Bates et al., 1979). This rapid acquisition is remarkable given the complexity of the word-learning process and the abilities required to process multiple multimodal information simultaneously (Hollich et al., 2000a, 2000b; Samuelson, 2021). Accordingly, interest in word comprehension

and word learning abilities has grown in the developmental community. The Switch task has been extensively used in infants to assess their ability to form new word-object associations after minimal exposure (see Tsui et al. (2019) for a meta-analysis). In this task, infants are first presented with two new objects and their associated labels in an habituation phase. In the subsequent test phase, two types of trials are presented: 'Same' trials, where objects are presented with the corresponding correct label, and 'Switch' trials, where objects are presented with the incorrect label (usually, the other object's label). Infants are expected to look longer at the pictures during 'Switch' trials due to surprise induced by the violation of the word-object pairing. Using this task, Werker et al. (1998) showed that 14-month-olds (but not 8-to-12-month-olds) successfully learned pseudo-objects/pseudo-words associations involving two very different CVC pseudowords (/nim/ and /lif/).

The comprehension of the first words coincides with a substantial change in speech perception abilities between 6 and 10 months of age. Indeed, speech sound discrimination transits from universal to language-specific, with an increasing sensitivity to native contrasts and a decreasing sensitivity to nonnative ones (Cheour, Ceponiene, et al., 1998; Kuhl et al., 2008; Werker & Tees, 1984a). The contingency between the emergence of the native phonemic repertoire and the receptive vocabulary might reflect an underlying connection between phonological processes and associative word-learning abilities. This assumption has been explored in a range of studies that evaluated infants' abilities to learn similar-sounding pseudo-words. Stager and Werker (1997)'s landmark work tested 14-month-olds' capacity to learn the pseudo-words /bih/ and /dih/ in a Switch task. The two pseudo-words constituted a minimal pair in that they differed only by one phoneme: here, the distinction was based on the place of articulation of the initial consonant. Although infants successfully discriminated these two auditory sequences and could learn two other dissimilar pseudo-words, they failed at the minimal pair associative word learning task. Following these results, further investigations were conducted. In subsequent work, Werker et al. (2002) suggested that minimal pair associative word learning abilities emerged gradually between 17 and 20 months of age. Höhle et al. (2020) and Rost and McMurray (2009), however, argued that 14-month-olds could successfully learn new minimal pair words when introducing speaker variability in the auditory stimuli. Further explorations of the factors influencing minimal pair word learning showed that the salience (Archer & Curtin, 2018; Curtin et al., 2009; Narayan et al., 2010), position (Archer et al., 2014; Nazzi & Bertoncini, 2009), referential status (C. T. Fennell & Waxman, 2010), or phonemic context (Höhle et al., 2021; Thiessen, 2007)

of the contrast modulated infants' performance on the task.

Importantly, some authors argued that the task demands of the Switch task might also impede infants' performance on minimal pair associative word learning (Höhle et al., 2021; Rost & McMurray, 2010). Indeed, this task implies that infants create sufficiently strong phonological-lexical representations to detect a mismatch between a word form they were shortly exposed to and another similar-sounding word form. Additionally, even if they successfully mapped each label to its referent, their confidence level might not be high enough to reject a mismatch association during "Switch" trials because of the similarity between the *matching* and the *mismatching* labels. Inspired by Ballem and Plunkett (2005)'s methodological procedure, Yoshida et al. (2009) tested 14-month-olds' performance on a minimal pair associative word learning task involving a /b/-/d/ contrast in word-initial position, like in Stager and Werker (1997)'s experiment. However, they did not use a Switch task but a two-alternative visual choice procedure called *Looking-While-Listening* (LWL) (Fernald et al., 1998, 2006; Swingley, 2011) or *preferential looking* (Golinkoff et al., 1987). This procedure starts with a habituation phase similar to the Switch task, but the test phase is different. Instead of one object, both new objects are displayed side-by-side, accompanied by one label. Infants are expected to look longer at the object associated with the label (the target) than at the other object (the distractor). Thus, the task involves choosing the best match for the auditory stimuli rather than rejecting a mismatch association. Using the LWL procedure, Yoshida et al. (2009) found that infants were able to learn two similar-sounding words, thereby questioning the findings of Stager and Werker (1997) obtained with the Switch task.

Most studies investigating minimal pair associative word learning in infants focused on 14-month-olds. Nevertheless, 18 months constitutes a pivotal age where the so-called "vocabulary spurt" characterizes the increasing speed at which toddlers' productive and receptive vocabulary size expands daily (L. Bloom, 1973; Goldfield & Reznick, 1990a; Nelson, 1973). Studies on word recognition suggest that phonological representations are actively involved in lexical processing at this age: for example, Swingley and Aslin (2000) tested infants' word recognition abilities of correctly pronounced and mispronounced words. The mispronunciations were created by changing one phoneme (usually the initial consonant), which yielded minimal pairs such as "baby"-"vaby" or "dog"-"tog". Using a visual choice paradigm and eye-tracking measures, the authors showed that infants poorly recognized words when mispronounced, indicating detailed encoding

of words' phonological units and strong interaction between phonological and lexical representations. François et al. (2024), however, provided evidence for minimal pair associative word learning in 18-month-olds. Specifically, they collected visual fixation measures in a minimal pair word-learning task in 18 healthy full-terms and 18 late-to-moderate preterms (mean gestational age at birth 33.7 weeks) at 18 months of corrected age. While familiar word recognition was preserved, they observed a negative impact of late-to-moderate premature birth on the ability to learn minimal pairs differing in VOT. Crucially, full-term toddlers could learn to map the pseudowords forming a minimal pair with novel referents, while preterms could not. To my knowledge, no study tested this phonological-lexical interaction during associative word learning processes at 24 months of age.
Moreover, most studies that involved minimal pairs in associative word learning tasks have used minimal pairs differing in place of articulation. Nevertheless, the salience and position of the contrast within the word have been shown to modulate infants' performance on discrimination and associative word learning tasks (Archer et al., 2014; Höhle et al., 2021; Narayan et al., 2010). Specifically, Archer et al. (2016) found that 20-month-olds but not younger infants successfully discriminated place of articulation contrasts for both voiced and voiceless stop consonants. Twelve-month-old infants could only discriminate the same contrasts for voiced stop consonants, suggesting that voiced are more salient than voiceless stops. Therefore, if voiceless stops are perceived as less salient before 20 months of age, 24-month-olds should be better than 18-month-olds at learning a minimal pair based on a voicing contrast. However, to my knowledge, no study has tested this hypothesis.

Here, I explored the interaction between phonological processing and associative word learning abilities during a developmental window pivotal for vocabulary acquisition. More specifically, I assessed associative word learning abilities in 18- and 24-month-olds using a minimal pair based on a voicing contrast. I used eye-tracking to obtain fine-grained measures of looking behavior previously shown to reflect different learning mechanisms, as in Ellis et al. (2015). Additionally, I considered individual differences by exploring the link between word learning performance and measures of vocabulary size and cognitive development.
We expected that 24-month-olds should exhibit better minimal pair word learning performances, as indicated by higher proportions of looking time toward target objects, than 18-month-olds. In addition, I expected different dynamics of infants' gaze across trials and conditions (word learning vs. word recognition) between the two age groups:

older infants should look earlier at the target object and should display less variability in their looking patterns than younger infants. Overall, 24-month-olds should learn better (higher accuracy) and faster than 18-month-olds (fewer familiarization trials). These patterns of looking behavior should be modulated by infants' productive and receptive vocabulary size at the individual level as well as their language and cognitive development scores (Werker et al., 2002). Therefore, I predicted that greater vocabulary size and higher neurodevelopmental scores should be associated with better associative word learning performance.

2. Methods

2.1. Participants

Infants were recruited at 18 or 24 months in the Southeast of France via three daycare centers of the Babilou group and a BabyLab (le Labot des Minots, Marseille). All participants were raised in a dominant French linguistic environment and had no known hearing or developmental disorder. The present study was conducted according to guidelines laid down in the Declaration of Helsinki, with written informed consent obtained from a parent or guardian for each child before any assessment or data collection. All procedures involving human subjects in this study were approved by a national Bioethical Committee (2022-A02281-42).

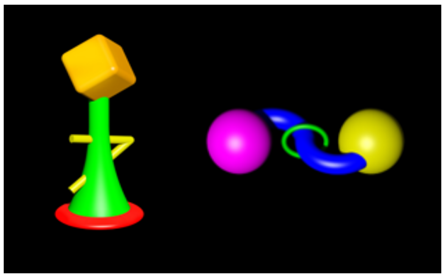
A total of 40 infants participated in the study. Among them, 26 were recruited at 18 months of age and 14 infants at 24 months. Among those 26 infants enrolled at 18 months, 20 could be re-evaluated at 24 months: these infants were considered a longitudinal group. This yielded a final total of 60 recording sessions: 26 recordings in the 18-month group (20 longitudinal + 6 cross-sectional) and 34 recordings in the 24-month group (20 longitudinal + 14 cross-sectional). From the 60 recordings, 6 were removed because of bad calibration quality (n = 3), high data loss (n = 2), or both (n = 1) in the test phase . Finally, analyses were performed on a sample of 54 recordings, including 21 recordings at 18 months of age and 33 at 24 months of age. In the 24-month group, there were 16 remaining infants from the longitudinal group after bad data rejection.

2.2. Associative word learning task

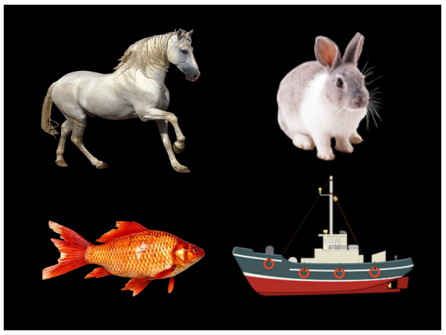
Stimuli. I created four pairs of visual-auditory stimuli from two pseudo-objects and two pseudo-words. The auditory stimuli were natural CVCV pseudo-words "BALI" / "PALI" constituting a minimal pair based on a voicing distinction in the initial stop consonant.

These stimuli were recorded in an anechoic room by a French female speaker. To enhance the attractiveness of the stimuli and the salience of the initial syllable (that holds the phonemic contrast), prosodic variability was induced by using an IDS register (Ma et al., 2011) and different stress patterns (Floccia et al., 2011). The two bisyllabic pseudo-words were first recorded in isolation with four different prosodic contours (flat, large span falling, rising, and small span falling, see Figure 6.2a) for the familiarization phase (see **Procedure** below). The two pseudo-words and four familiar words (rabbit, fish, horse, boat) were also produced within three different carrier sentences starting with one occurrence of the target word in isolation, followed by a child-directed sentence including the target word at the end. For example, one sentence was "*BALI? Oh, regarde le BALI!*" [1] This sentence structure was repeated four times to create 20-second auditory sequences that were used in the test phase (see Procedure). The mean VOT values (in ms) for the onset consonants (/b/ and /p/) were -80 ms (SD = 34) and 13 ms (SD = 4), respectively. All auditory stimuli were edited with Audacity (v 3.1.2) to match for intensity.

The visual stimuli comprised two colorful toys-like pseudo-objects, that had already been used in a previous study (François et al., 2024). The two pseudo-objects were different in both shape and color; I referred to them as "object 1" and "object 2" (see Figure 6.1a). Pseudo-words and pseudo-objects were paired to create two associations: “bali” associated with object 1 and “pali” associated with object 2 (familiarization type 1); “bali” associated with object 2 and “pali” associated with object 1 (familiarization type 2). In the test phase, four familiar objects were selected to create two pairs: horse-fish and rabbit-boat (see Figure 6.1b).



(a) Pseudo-objects: object 1 on the left, object 2 on the right.



(b) Familiar objects used in the test phase.

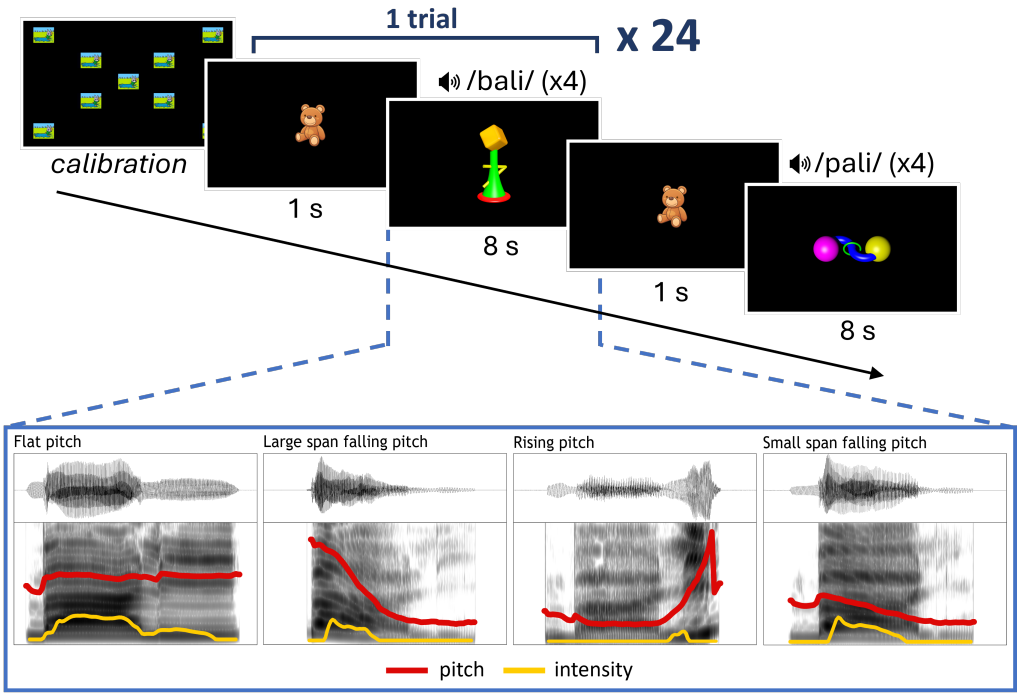
Figure 6.1. – Visual stimuli: pseudo-objects and familiar objects.

1. Translation: "BALI? Oh, look at the BALI!".

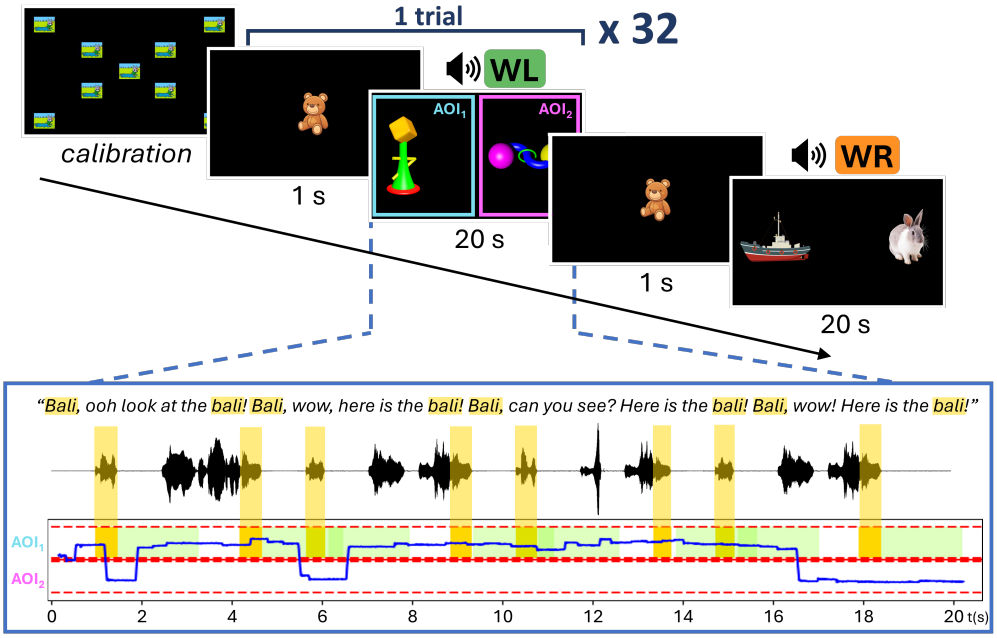
Procedure. Infants were seated alone on a small chair or a caregiver's lap in a small and dimly lit room of the daycare center (see Figure 1.11 on page 86). Participants' position was adjusted to have their eyes 70 centimeters away from the screen. Stimuli were presented with Tobii Pro Lab software (v.1.181.37603, Danderyd, Sweden) on a Dell LCD monitor with a resolution of 1680 x 1050 pixels, and eye gaze was recorded using a Tobii X3-T120 eye-tracker at a sampling rate of 120 Hz. The experimenter was seated next to the child to monitor the gaze online via the Tobii Pro Lab software and check if he/she looked at the screen. The experimental procedure was adapted from the task initially developed by Yoshida et al. (2009), which deploys the IPLP suitable to test minimal pair word learning in young toddlers (Fernald et al. (1998); see Figure 6.2 for an illustration of the procedure). Instead of a habituation phase, however, the test phase was preceded by a familiarization phase in which infants had to learn two associations between the two pseudowords and the two new objects (see Box 3.3.1 p.85 for the difference between habituation and familiarization). Each phase started with a calibration using a short cartoon video that was shrinking at nine calibration points under the experimenter's control.

Familiarization types 1 ("bali" associated with object 1) and 2 ("bali" associated with object 2) were randomly assigned to each participant. For the longitudinal group, I made sure that infants saw the same familiarization type at 18 and 24 months of age. During a familiarization trial, an attention-grabber first appeared at the center of the screen for 1 second to fix the infant's gaze. Then one object appeared at the center of the screen for 8 seconds while being labeled 4 times in isolation in each of the four prosodies (Figure 6.2a). Prosodies were always presented in the same order within a trial: flat, large span falling, rising, and small span falling. To maintain infants' attentional engagement with the experimental stimuli, the object slowly moved according to one of the six possible following movements: lateral starting left, lateral starting right, vertical starting upwards, vertical starting downwards, zooming in, and zooming out. A total of 24 familiarization trials were presented, alternating between the two objects (12 trials for each object). For a given familiarization type, the same trials were presented in the same order for all infants.

The test phase consisted of 32 trials: 12 trials in which the two new objects appeared simultaneously on the screen (Word Learning (WL) condition) and 20 trials that displayed pairs of familiar objects (Word Recognition (WR) condition). The WR trials served as the control condition to ensure that infants were doing the task as expected (e.g., looking longer at the labeled object).



(a) Top part: time course of two familiarization trials. Bottom part: the four prosodies corresponding to the four occurrences of the pseudo-words within one trial.



(b) Top part: time course of two test trials: one word-learning (WL) followed by one word-recognition trial (WR). The blue and pink frames represent the area of interests (AOIs) for analyses. In this example, the blue AOI (AOI_1) corresponds to the target side, and the pink AOI (AOI_2) to the distractor side. Bottom part: time windows where target pseudo-words are heard (in yellow) in the sentence, and the time windows for analyses (in green). The dark blue line represents the horizontal displacement of the eye on the screen as a function of time.

Figure 6.2. – Time course of familiarization and test trials.

The familiar objects were chosen according to a standardized list of words that infants understand at 18 and 24 months of age (Fenson et al., 1993): horse, rabbit, fish, boat. As in the familiarization phase, a test trial started with a 1-second display of a centered attention-getter. This was followed by a 20-second sequence wherein the two objects appeared side by side and moved vertically in synchrony. During this sequence, the target object was labeled 8 times, alternating between the label in isolation and four sentences containing the label, all in IDS (see Figure 6.2b).

The four sentences were always the same and were always presented in the same order across trials ("BALI? Oh, regarde le BALI. BALI? Waouh, il y a le BALI! BALI? Tu vois, il y le BALI. BALI? Waouh, il y a le BALI!" [2].). Test trials were presented in blocks of 8 trials, interspersed with short attractive videos to maintain infants' attention towards the screen. The first 3 blocks contained 4 WL trials and 4 WR trials; the last block contained 8 WR trials. The order of trial presentation and the side of the target object were semi-randomized and were the same for all participants in a given familiarization type.

2.3. Vocabulary size and language and cognitive development assessment

Receptive and expressive vocabulary size was obtained at 24 months of age using an adapted version of the MacArthur-Bates communicative development inventories (MCDI), a standardized and reliable questionnaire filled out by the parents. The survey comprises a list of 683 words, for which parents have to indicate if their child understands and/or produces them. Language and cognitive development were assessed at 24 months of age using the Bayley Scale of Infant Development (BSID) IV, a standardized scale of development for infants from 0 to 42 months old. This scale comprises seven subtests: Cognitive, Receptive Communication, Expressive Communication, Fine Motor, Gross Motor, Socio-Emotional, and Adaptive Behavior. For this study, infants were only assessed on Cognitive, Receptive Communication, and Expressive Communication. Among the 33 infants kept for analyses at 24 months of age, 25 completed the MCDI and the BSID; 6 additional infants completed only the BSID because the MCDI questionnaire was never returned by the parents.

2. Translation: "Bali, ooh look at the bali! Bali, wow, here is the bali! Bali, can you see? Here is the bali! Bali, wow! Here is the bali!"

2.4. Data analysis

General processing parameters. For the analysis of proportion of fixations on the stimuli, I used only individual gaze samples that were classified as belonging to fixations. Preprocessing of the raw eye-tracking data and identification of fixations were performed with the Tobii I-VT fixation filter implemented in the Tobii Pro Lab Software (v.1.181.37603, Danderyd, Sweden). Gaze sample positions were averaged from the left and the right eye, missing data were interpolated linearly using the Gap fill-in function with a maximum gap length of 75 ms, and noise reduction was performed with the symmetric Moving Median filter with a window size of 3 ms. Fixations were then identified by using the I-VT Gaze Filter with a velocity threshold of 30°/s with a window length of 20 ms. Finally, multiple adjacent fixations were merged into one fixation using a maximum interval of 75 ms and a maximum visual angle of the eyes between separate fixations of 0.5 degrees. For the familiarization phase, the area of interest (AOI) was defined as the visible area on the screen (1680 pixels width x 1050 pixels height). For the test phase, two rectangular AOIs encompassing 790 x 1050 pixels in width and height each — one on the left side and one on the right side of the screen — were defined, leaving a space of 100 pixels between the AOIs at the center of the screen. Depending on the trial and familiarization type, the left AOI was either coded as "target" or "distractor", and the right AOI was coded with the other option (see Figure 6.2b). Mean looking time analyses consisted of computing the mean proportion of looking time toward each AOI. At the temporal level, eight time windows of interest were considered from the onset of each word, starting at 500 ms and ending at 2300 ms after word onset as recommended by Ellis et al. (2015). The time windows were either averaged (**Mean looking times**) or analyzed independently (**time windows**).

***Post-hoc* habituation criterion.** To test whether infants' performance in the test phase was related to infants' looking behavior during the familiarization phase, I calculated a *post-hoc* habituation criterion for each participant. Based on previous studies that used habituation in the learning phase (Archer & Curtin, 2018; Curtin et al., 2009; C. T. Fennell & Waxman, 2010; Yoshida et al., 2009), infants were considered habituated when they showed a significant decrease in looking time during familiarization, indicating a decrease in interest. The habituation criterion was defined as a decrease in average looking time to below 60% of the average looking time of the first three trials, which served as a baseline (Archer et al., 2016; François et al., 2024). To assess habituation, a moving average of looking time across three consecutive trials was calculated and compared to the baseline. When this moving average met the habituation criterion,

the infant was deemed to have habituated. For participants who met the habituation criterion, the number of trials reached for habituation is reported; for those who never reached the criterion, the habituation trial is set to zero.

Mean looking times. Mean looking times were first computed for the familiarization trials. To control for object preference, the mean proportion of looking time toward each object during familiarization was computed, and paired t-tests were run. The same tests were applied to each familiarization type separately to control for word-object association preference. For test trials, the overall proportion of looking at the screen during each condition (WR and WL) was computed, and differences in looking time were assessed with t-tests. The overall performance of each group was measured in each condition by computing the mean proportion of fixation on the target over the eight time-windows defined from each word onset. Paired t-tests were run to assess the effect of age and condition. Performances and comparison test results within the longitudinal group were also reported. To control for object preference in the WL condition, the mean proportions of looking at object 1 and object 2 were computed separately for targets and distractors, and for each age group. T-tests were run to assess differences in looking time when each object is a target versus when it is a distractor and between object 1 and object 2. Comparisons between groups were performed using two-sample t-tests assuming unequal variance (Welch approximation to the degrees of freedom).

Time windows. Looking behaviors of each group by condition were further explored by splitting data into time windows. For each condition, the mean proportion of fixation on the target was retrieved for each time window and compared to chance level using t-tests. Group comparisons were also performed using t-tests. As illustrated in Figure 6.2b, there were overlaps between the time window of the last word of a sentence and the time window of the following isolated word. Thus, additional analyses were run by word "type" (in isolation versus at sentence end) to eliminate this temporal overlap. Moreover, this separation allowed testing the effect of word occurrence context on the infants' performance.

Time courses. More fine-grained temporal dynamics of looking patterns were explored by computing the time courses of the fixations across trials. Grand averages were calculated for each word to obtain six different looking patterns through time and AOI. Comparisons between age groups (18 vs. 24 months) and object status (target vs. distractor) were computed. For each group, a measure was computed, as an indicator

of the time at which fixations to the target significantly differ from the fixations to the distractor (Borovsky et al., 2012, 2013). Inspired by the work of Ellis et al. (2015), analyses consisted of point-by-point t-tests at each 50-ms bin between the mean proportion of looking time to the target and the mean proportion of looking time to the distractor. The sampling rate of the eye tracker (one sample each 8.3 ms) prevented us from using smaller time bins which may have allowed for a more fine-grained temporal analysis. I report target divergence time as the first time point within the 500 to 2300 ms post-word-onset-window in which the difference in looking time on the target versus distractor AOI was significant for at least 250 ms (five significant consecutive t-tests, textitp < .05). All other time points after this *target divergence time* that met these two criteria were also reported.

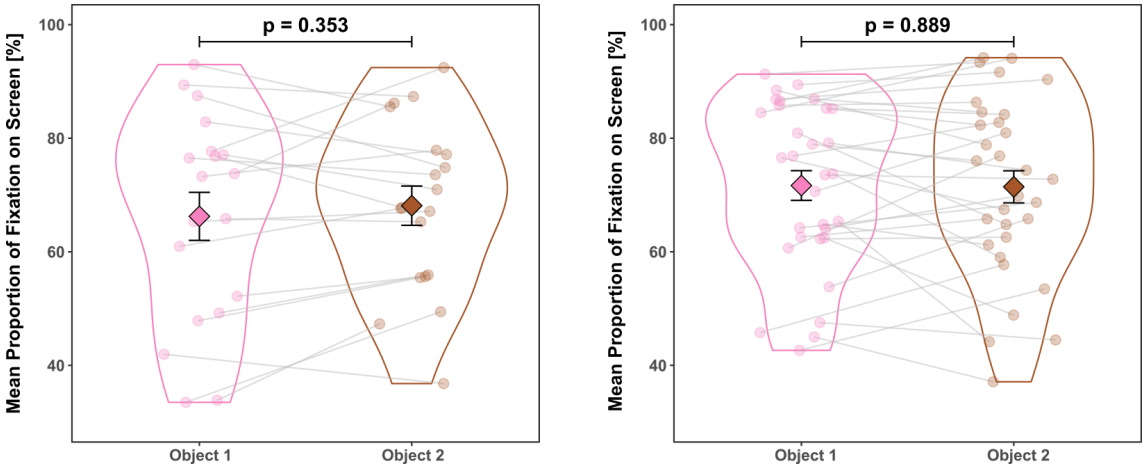
Correlations. Correlations were assessed between measures from the associative word learning task and neurodevelopmental measures (language and cognitive scores; vocabulary size) for infants that completed the BSID and/or the MCDI. Measures from the associative word learning task included the habituated trial, the mean proportion of looking toward the target in the WL condition, and the mean proportion of looking toward the target in the WR condition. These measures were considered at 18 and 24 months of age and compared to the neurodevelopmental measures collected at 24 months. Neurodevelopmental measures included the BSID scores of global language [3], receptive language, expressive language, and cognitive development, as well as expressive and receptive vocabulary size extracted from the MCDI. Correlation matrices were computed, and significant correlations were reported.

3. Results

Familiarization trials. To assess the equivalence of visual attention allocation between the two objects during the familiarization phase, I computed the mean looking time at the screen for each object presentation. Figure 6.3 shows the distribution of the mean looking time toward object 1 (pink) and object 2 (brown) for each object in 18-month-olds (6.3a) and 24-month-olds (6.3b). T-tests between objects 1 and 2 in each group revealed no significant difference in the mean looking time to screen ($t(18)$ = -0.953, p = .353 in 18-month-olds; $t(30)$ = 0.141, p = .889 in 24-month-olds). When split by familiarization type, the comparison between mean looking time to object 1 and

3. a global score of language development that is a combination of the receptive and expressive subscales scores

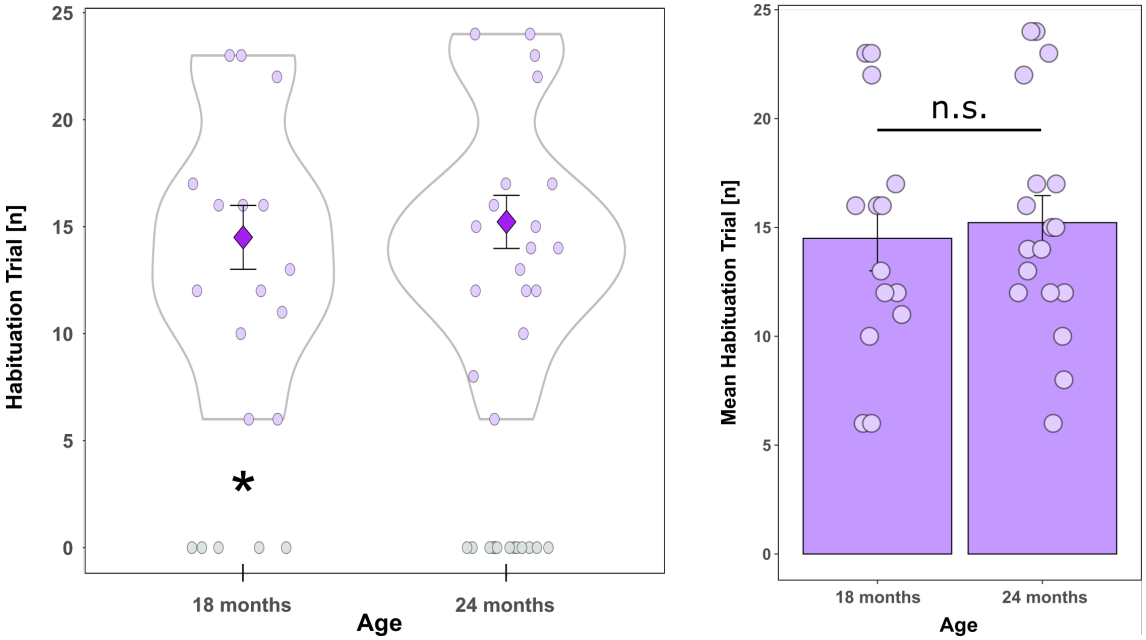
mean looking time to object 2 indicated no significant differences (familiarization type 1: $t(8)$ = -0.608,p = .56 at 18 mo, $t(13)$ = 0.762, p<0.0001.459 at 24 mo; familiarization type 2: $t(9)$ = -0.700, p = .502 at 18 mo, $t(16)$ = -0.684, p = .504 at 24 mo). These results indicate that both objects were equally attended to during both familiarization types and that infants did not prefer one object over the other which might have biased learning of the object-word associations.



(a) 18 months **(b)** 24 months

Figure 6.3. – Mean proportion of fixation at screen during familiarization phase for objects 1 (pink) and 2 (brown), in each age group.

***Post-hoc* habituation criterion.** According to the habituation criterion detailed above, a habituation trial was determined for each infant. Figure 6.5a illustrates the number of habituated infants (in purple) compared to non-habituated infants (in gray). At 18 months, 73.68% of the infants reached the habituation criterion (n = 14), and 58.06% at 24 months (n = 18). The mean habituation trial was calculated on habituated infants only at 18 months (m = 14.5) and 24 months (m = 15.22), as illustrated in Figure 6.4b. One-sided tests of proportions in each group showed that participants were overall habituated in the 18 months group ($\chi^2(1)$ = 4.263, p = .019) but not in the 24 months group ($\chi^2(1)$ = 0.806, p = .185). No significant difference was found between the two age groups for the habituated trial ($t(27.231)$ = -0.372, p = .713).

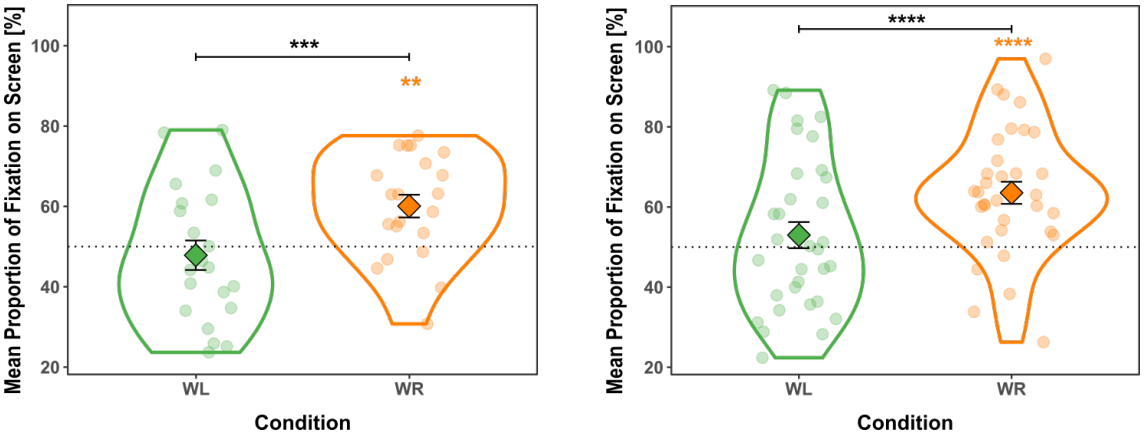


(a) Mean habituation trial for each age group. Purple: habituated infants; gray: non-habituated infants (mean habituated trial = 0).

(b) Mean habituation trial: comparison between 18 and 24 months for habituated infants only.

Figure 6.4. – Mean habituation trial by age group.

Mean looking times. In the test phase, comparison between mean looking time to screen in the WL and WR conditions indicated that infants looked significantly longer at the screen during WR trials than during WL trials, at both 18 ($t(20) = -4.186$, $p < 0.0001$) and 24 months ($t(32) = -5.253$, $p < 0.0001$). Moreover, t-tests on the average proportion of fixation on target against chance (50%) computed for each group and each condition revealed that infants looked significantly above chance at the screen during WR trials at both 18 ($t(20) = 3.568$, $p = .002$) and 24 months ($t(32) = 4.905$, $p < 0.0001$), but not during WL trials ($t(20) = -0.584$, $p = .565$ at 18 mo; $t(32) = 0.912$, $p = .368$ at 24 mo). These results are illustrated in Figure 6.5.



(a) 18 months **(b)** 24 months

Figure 6.5. – Distribution of the mean proportion of fixation to the screen during the test phase in each condition, split by age group. The dotted line represents the chance (50%).

Infants' performance in both conditions were assessed at 18 and 24 months of age by calculating the mean proportion of fixation to the target. Figure 6.6 represents the distribution of the mean proportions of fixations to target for all infants (left panel), habituated infants (middle panel), and infants from the longitudinal sample (right panel). T-tests against chance for each group and each condition revealed that infants' preferred looking time to the target AOI was not significantly above chance in the WL condition, neither at 18 nor 24 months of age (see Table 6.1). No significant difference was found between age groups. In the WR condition, results for all infants and infants from the longitudinal sample were identical: 24-month-olds looked significantly above chance at the target ($t(32) = 9.419$, $p < .0001$; $t(15) = 7.490$, $p < .0001$), but 18-month-olds were at chance ($t(20) = 1.958$, $p = .064$; $t(15) = 1.831$, $p = .087$). There was also a significant difference between the two age groups (Welch t-test, $t(39.363) = -3.899$, $p < .0001$; $t(15) = -2.942$, $p = .01$). Different results were found for habituated infants: preferred looking time to the target was significantly above chance at both 18 ($t(13) = 2.174$, $p = .049$) and 24 months ($t(17) = 6.806$, $p < .0001$). Moreover, group comparisons revealed that 24-month-olds looked significantly more at the target than 18-month-olds (Welch t-test, $t(27.947) = -2.838$, $p = .008$).

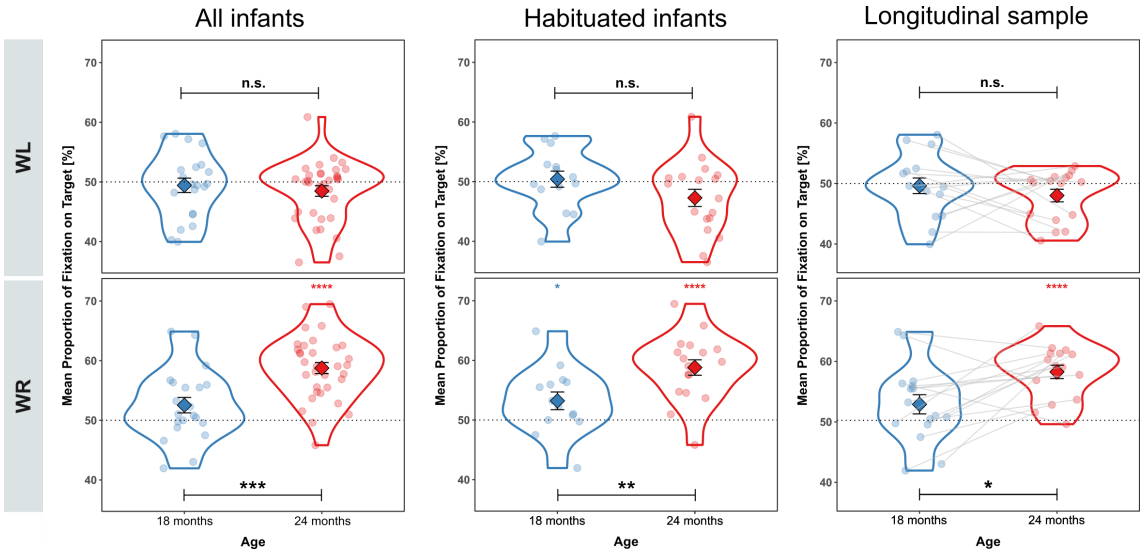
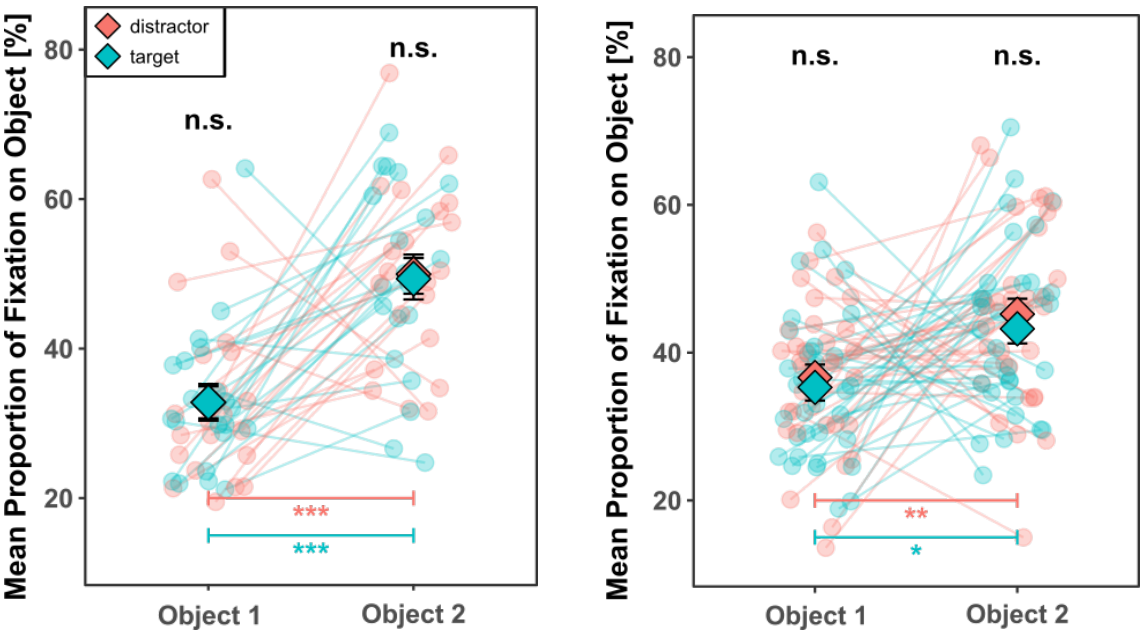


Figure 6.6. – Mean proportion of fixation to target during test phase for 18-month-olds (blue) and 24-month-olds (red), split by condition. Left panel: all recordings (n = 54); middle panel: habituated infants only (n = 32); right panel: longitudinal sample (n = 16). The dotted line represents the chance level (50%). Results of t-tests against chance for each word type and condition (colored stars) and group differences (black lines and stars) are reported (*$p < .05$; **$p < .01$; ***$p < .001$).

The absence of effect in the WL condition was further explored by looking at the mean proportion of looking at object 1 and object 2 depending on whether it was a target or a distractor (see Figure 6.7. Results revealed no significant differences in looking proportions to targets and distractors for object 1 (18 months: $t(20) = 0.001$, $p = .999$; 24 months: $t(32) = 0.594$, $p = .557$) and object 2 (18 months: $t(20) = 0.26$, $p = .798$; 24 months: $t(32) = 1.512$, $p = .14$). Overall, infants looked more at object 2, whether it was a target (18 months: $t(20) = -4.577$, $p < .0001$; 24 months: $t(32) = -2.427$, $p = .021$) or a distractor (18 months: $t(20) = -4.269$, $p < .0001$; 24 months: $t(32) = -2.734$, $p = .01$).

Condition	Sample	Group	mean	t	df	p
WL	All participants	18 months	49.429	0.473	20	.641
		24 months	48.476	-1.720	32	.095
		comparison	0.953	0.637	40.09	.528
	Habituated	18 months	50.417	0.309	13	.762
		24 months	47.280	-1.886	17	.077
		comparison	3.137	1.589	29.872	.123
	Longitudinal	18 months	49.617	-0.297	15	.770
		24 months	48.007	-1.910	15	.075
		comparison	1.610	0.964	15	.350
WR	All participants	18 months	52.539	1.958	20	.064
		24 months	58.762	9.419	32	**>.0001**
		comparison	-6.223	-3.899	39.363	**>.0001**
	Habituated	18 months	53.222	2.174	13	**.049**
		24 months	58.804	6.806	17	**>.0001**
		comparison	-5.582	-2.838	27.947	**.008**
	Longitudinal	18 months	52.905	1.831	15	.087
		24 months	58.263	7.490	15	**>.0001**
		comparison	-5.358	-2.942	15	**.010**

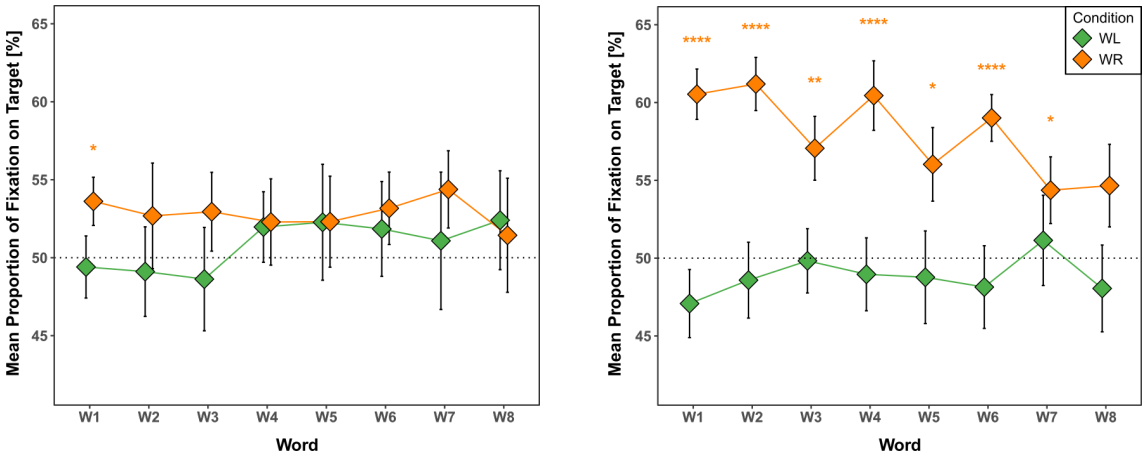
Table 6.1. – Word learning: summary statistics of mean looking times. Results for the 18 and 24 months row correspond to t-tests against chance within a group, and comparison rows indicate the group comparisons tests (the mean difference is reported in the "mean" column).



(a) 18 months **(b)** 24 months

Figure 6.7. – Distribution of the mean proportion of fixation to object 1 and object 2 depending on whether it was a target or a distractor. Results of t-tests between object 1 and object 2 for target and distractor separately (colored lines and stars) and between target and distractor for each object (black notation) are reported (*$p < .05$; **$p < .01$; ***$p < .001$).

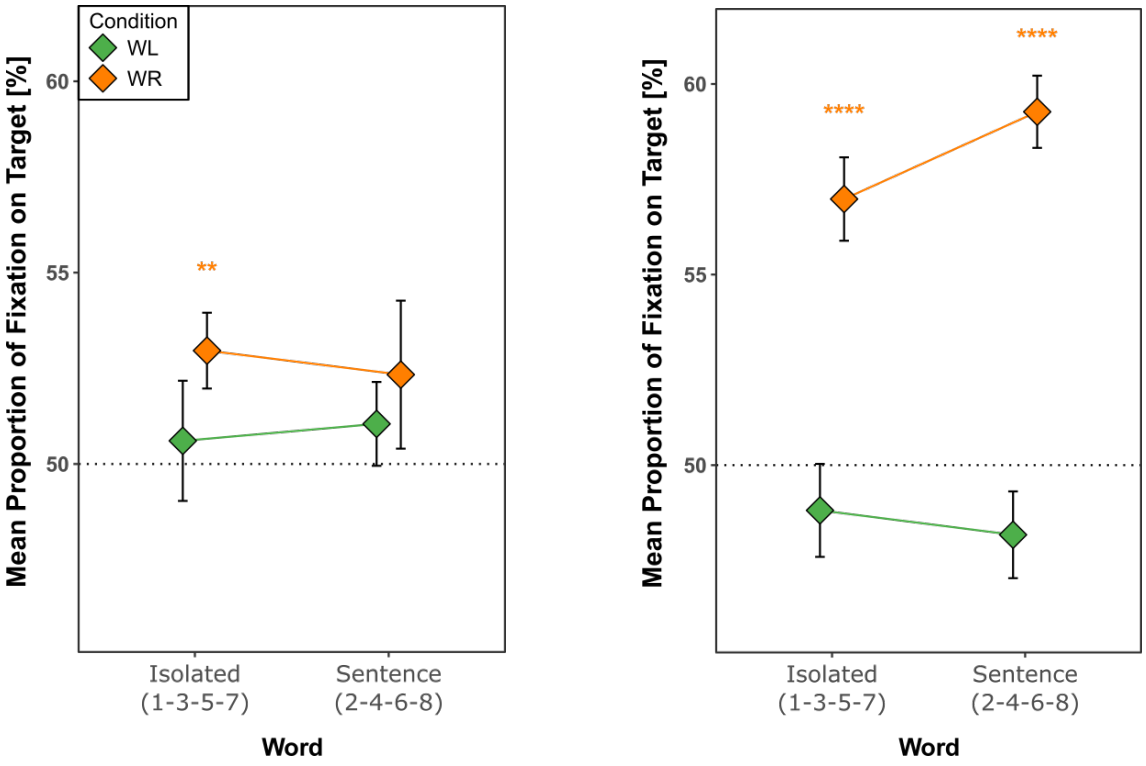
Time windows. To achieve a more detailed understanding of preferential looking behavior towards the target stimulus, I computed the mean proportion of fixations directed at the target for each discrete time window corresponding to individual word occurrences (Figure 6.8). T-tests against chance for each time window indicated no significant difference in the WL condition in both groups. Interestingly, significant differences were found in the first time window (W1) at 18 months (W1: $t(20) = 2.338$, $p = .03$) and at all time windows except the last one (W1-7) at 24 months (W1: $t(32) = 6.506$, $p < .0001$, W2: $t(32) = 6.555$, $p < .0001$, W3: $t(32) = 3.443$, $p = .002$, W4: $t(32) = 4.679$, $p < .0001$, W5: $t(32) = 2.546$, $p = .016$, W6: $t(32) = 6.001$, $p < .0001$, W7: $t(32) = 2.037$, $p = .05$). Significant differences between age groups were found only in the WR condition at W1 ($t(50.195) = -3.091$, $p = .003$), W2 ($t(30.245) = -2.244$, $p = .032$), W4 ($t(43.112) = -2.294$, $p = .027$), and W6 ($t(36.278) = -2.114$, $p = .042$). In the subgroup of habituated infants, I observed slight pattern differences at both ages. At 24 months, the results were similar to those in the full sample of infants in the WR condition. However, they were significantly below chance during W1 ($t(17) = -2.266$, $p = .037$) and at chance during all other time windows. At 18 months, results were similar to those in the full sample of infants in the WL condition. However, in the WR condition the mean proportion of fixation to target was significantly above chance at W6 ($t(13) = 2.160$, $p = .05$), but not at W1 ($t(13) = 1.448$, $p = .17$). Group comparisons in the habituated sampled yielded only one significant difference in the WL condition, at W4 ($t(29.983) = 2.459$, $p = .02$).



(a) 18 months **(b)** 24 months

Figure 6.8. – Mean proportion of fixation on target in each time window (500 – 2300 ms after word onset) corresponding to the 8 occurrences of the target word within a test trial. The dotted line represents the chance level (50%). Results of t-tests against chance for each time window and condition are reported (*$p < .05$; **$p < .01$; ***$p < .001$).

Data from each time window were averaged by word type (in isolation or at sentence end), and the mean proportion of fixation to target was tested against chance for each group and condition (see Figure 6.9b). Again, no difference was found in the WL condition in both groups and word types. However, infants looked significantly above chance to target for isolated words at 18 months ($t(20) = 2.991$, $p = .007$) but not for sentence words ($p = .24$), and for both word types at 24 months (isolated: $t(32) = 6.383$, $p < .0001$; sentence: $t(32) = 9.791$, $p < .0001$). In the habituated sample, results were similar at 24 months, but no significance was reached at 18 months.



(a) 18 months **(b)** 24 months

Figure 6.9. – Mean proportion of fixation to target by word type for each condition and age group: words in isolation ("isolated", word occurrences number 1, 3, 5 and 7 of the test trial) and words at the end of a sentence ("sentence", word occurrences number 2, 4, 6, and 8 of the test trial). Results of t-tests against chance for each word type and condition are reported (*$p < .05$; **$p < .01$; ***$p < .001$).

Time courses. Time courses of the proportion of looking to target and distractor for each pseudo-word and each age group are represented in Figure 6.10. Note that looking patterns for target and distractor are not symmetrical because the data in each time course do include the fixations outside the AOI during test trials. The target divergence

time is marked by a red arrow for each pseudo-word and age group. No target divergence time was found for 18-month-olds: significant t-tests between target and distractor are reported but never reached the criterion of five consecutive points. Target divergence times in 24-month-olds were found at 2655 ms after word 1 onset for "bali" and 551 ms after word 2 onset for "pali". However, in both cases, infants were looking to the distractor. For "bali" trials, two other time points indicating a sustained look towards the target were reported at 48 and 26 ms after word onsets but were outside the 500-2300 ms time window from word onset. For "pali" trials, three other time points were reported within the targeted time window from word onset: two of them were found after a word in isolation (W5 and W7), and one was found after a word in sentence (W6). In these three time points, infants were looking at the distractor.

Correlations. I explored the relations between measures of associative word learning and language outcomes using Pearson correlations. A correlation matrix was obtained for each age group with the following variables: age (in days), weight at birth, gestational age at birth, cognitive and language outcomes at 24 months (expressive and receptive vocabulary size from the MCDI, scores of the cognitive and language scales of the BSID, the expressive and receptive languages subscales of the BSID), and the mean proportion of looking to target in WR and WL conditions. For participants from the habituated sample, the habituation trial (number) was added to the correlation matrix. At 18 months, a significant positive correlation was found between the habituation trial and the global language score from the BSID (see Figure 6.11). No significant correlations were found between the mean proportion of looking to target in the WL condition and any other variable, but positive correlations were found between the mean proportion of looking to target in the WR condition and the cognitive and the global language scores from the BSID. At 24 months, no correlations were found between the habituation trial and cognitive or language outcomes. However, a significant negative correlation was found between the proportion of looking to target in the WL condition and the expressive vocabulary from the MCDI. No correlation was found between WR performances and language outcomes at 24 months.

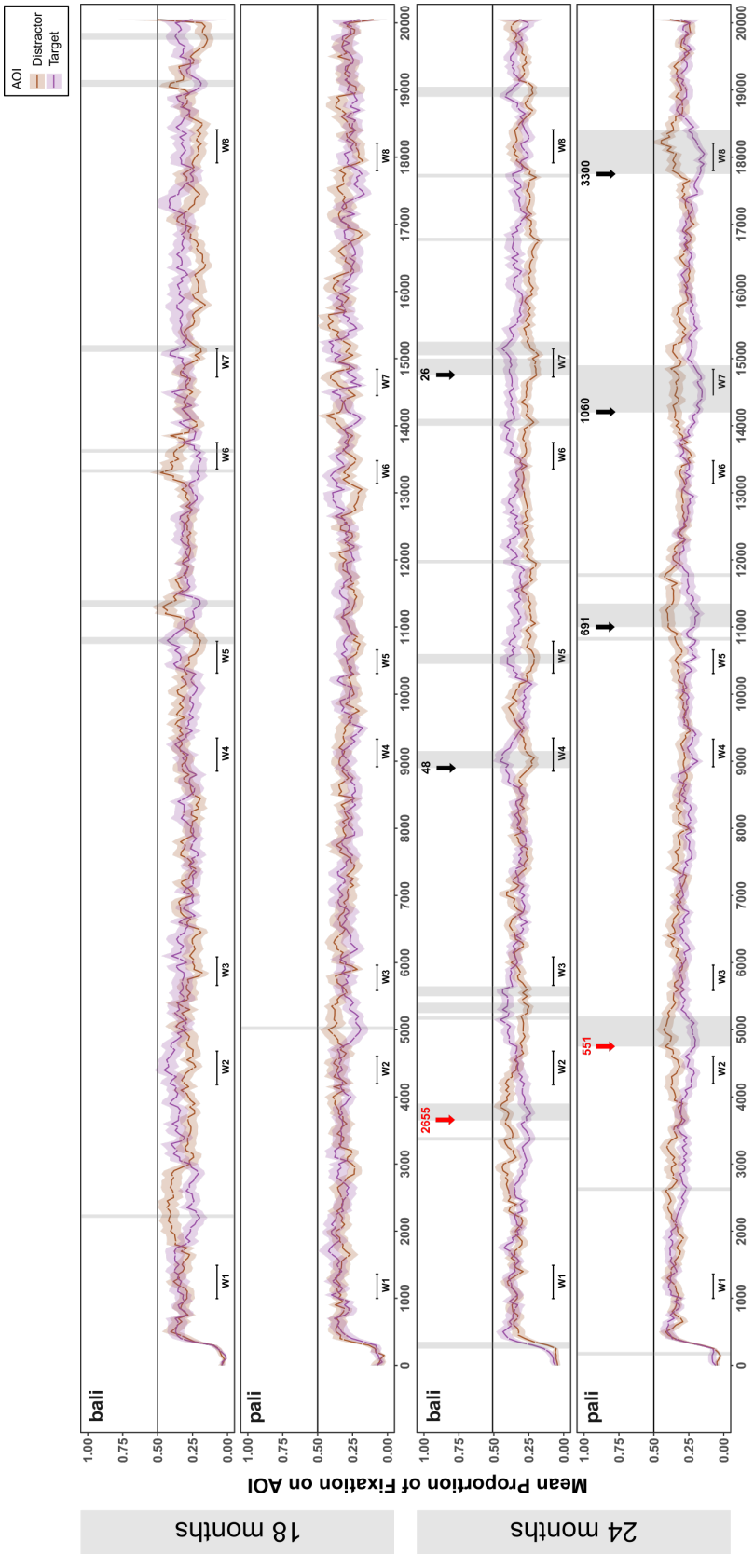


Figure 6.10. – Time courses of looking to target and distractor during WL trials. Red arrow: target divergence time; black arrow: subsequent time points where five consecutive significant t-tests between target and distractor have been reported. The numbers above the arrows indicate the positive latency from the closer word onset.

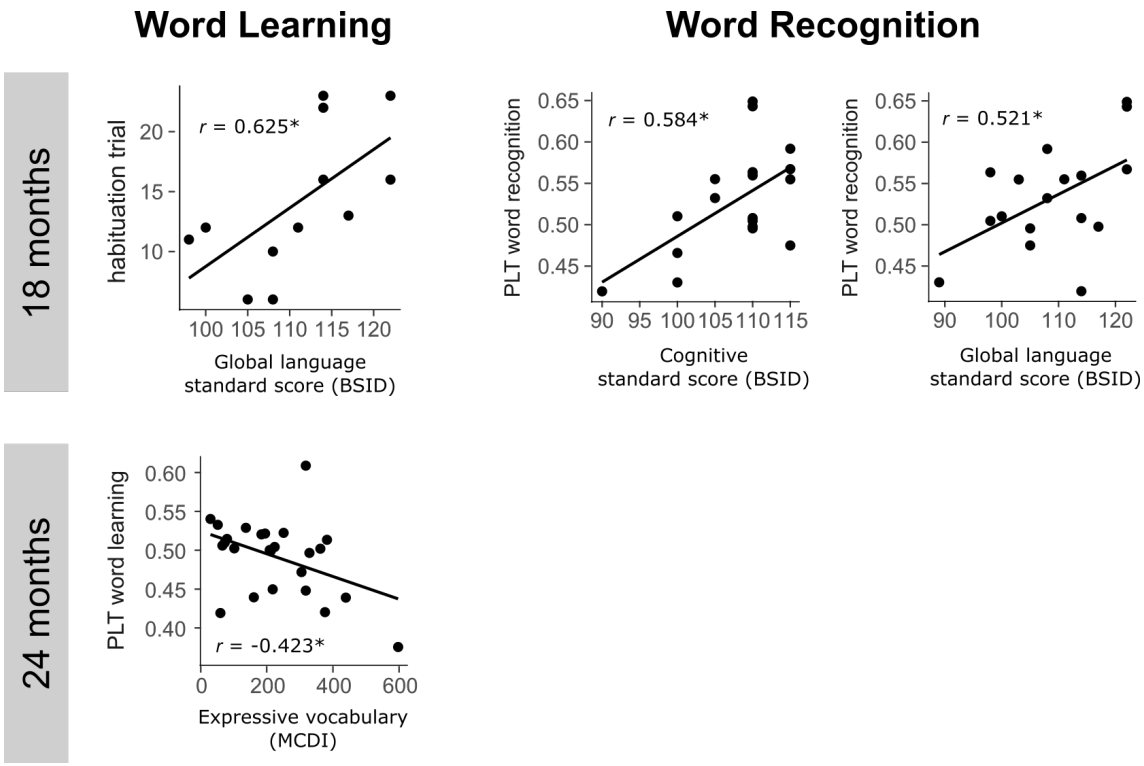


Figure 6.11. – Significant correlations between performances in the WR and WL trials and neurodevelopmental measures from the MCDI and the BSID. Significance levels of the correlations are reported (*$p < .05$; **$p < .01$; ***$p < .001$).

4. Discussion

This study investigated 18- and 24-month-old infants' minimal pair associative word learning abilities using a LWL paradigm adapted from Yoshida et al. (2009). Importantly, I used eye-tracking to obtain mean proportions of looking times as well as the fine-grained time courses of looking patterns in each group. Overall, the results indicate that infants from both age groups recognized familiar words but did not learn the new pseudo-words that constituted a minimal pair. In this section, I further discuss these results.

4.1. Dataset investigation

In word learning trials, classical measures of the mean proportion of looks to targets revealed that infants performed at chance level, suggesting they did not learn the new associations. Additionally, no significant difference between the 18- and 24-month-old

groups were found. To better understand these findings, I examined the mean looking times for familiar word recognition trials. The data indicated that 24-month-olds performed significantly above chance in WR trials, but 18-month-olds did not. While the pattern of results observed in 24-month-olds might be explained by decreased associative word learning abilities or difficulties in processing the initial consonant difference of the minimal pair, 18-month-olds performance might relate to attentional resources, lexical processing mechanisms, or task-related challenges. Before delving into interpretations of these unexpected results, I further explored the dataset.

Unlike most LWL studies focusing on associative word learning in infants (Curtin et al., 2009; Graf Estes & Hurley, 2013; Höhle et al., 2021; Werker et al., 1998), my task involved familiarization rather than habituation in the learning phase. In addition to ensuring that infants were exposed to the same learning input, this choice was made to increase the chances that infants learned the novel pseudo-words while still allowing for computing a habituation criteria *post-hoc*. Results showed that a majority of infants reached the habituation criterion at 18 months (73.68%) but not at 24 months (58,06%). These results might be explained in two ways: (1) 18-month-olds but not 24-month-olds learned the associations, or (2) 18-month-olds lost interest faster than 24-month-olds due to lower attentional resources. This second hypothesis was considered because of the results from the mean looking times to target, suggesting that both groups did not perform above chance on the WL trials. Indeed, if infants were overall habituated at 18 months, the mean looking times to targets indicate that they did not learn the new associations. For infants who met the habituation criterion (habituated sample), age-related differences were found the number of trials required to reach habituation. This suggests that habituated infants at 18 and 24 months may have similar attentional resources or interest in the stimuli. To determine whether infants learned the new associations and if the habituation trial modulated their looking behaviors, I examined the mean proportion of looking to target for the habituated sample separately. Habituated infants were at chance in the WL condition, indicating that the habituation criterion did not impact infants' performance in the word learning task. However, 18-month-old habituated infants looked significantly longer at the target during WR trials. These findings support the idea that the ability to shift attention away from new auditory stimuli, as reflected in reaching the habituation criterion, may be associated with enhanced lexical processing.

Using a familiarization also allowed for controlling looking behaviors during the learning phase. Analysis of the mean looking times at the screen for each new object

revealed that infants looked equally at both objects at 18 and 24 months. Thus, results in the test phase could not be attributed to an object preference in the familiarization phase, or to one association being easier to learn compared to the other. However, analyses of the mean proportions of fixation during the test phase showed that overall infants looked more at the screen during WR than WL trials. Thus, the difference in the proportion of looking to target between WR and WL trials mentioned above might be influenced by the reduced amount of time spent looking at the screen during WL trials. I propose three alternative explanations for this effect of condition on mean looking time to screen: (1) infants might have displayed a preference for *familiarity* over *novelty* (Horst et al., 2011; Houston-Price & Nakai, 2004; Mather & Plunkett, 2011): infants rather spent time looking at objects they already knew than unknown objects. (2) This *familiarity* vs *novelty* preference might have also played a role in the opposite direction: because all infants were exposed to 24 trials during the familiarization phase, they might have lost interest in the new objects that became the familiar ones; in contrast, the objects from WR trials became more interesting because they were "new" in the task. (3) Another possibility is that the task demands might have been too high, which could explain infants' faster disinterest during WL trials. Interestingly, the first two hypotheses maintain the possibility that infants learned the new associations.

To further investigate infants' success in learning the novel pseudo-words, I conducted analyses on distinct time windows and on trial time courses. Results in the WL condition were consistent with mean proportions of fixations to target and did not indicate successful word learning in both age groups. However, results in the WR conditions revealed that in contrast to the "mean looking time" analyses, 18-month-olds recognized the familiar words. Moreover, different patterns of word recognition were found between 18 and 24 months of age: while older infants looked significantly above chance at the target in almost all time windows (except the last one), younger infants looked longer at the target only after the occurrence of the first word. This result can be interpreted as a difference in attentional resources or cognitive control. In the first case, younger infants lost interest faster than older infants, whereas in the second case, they failed to inhibit the drive to look at the other object, although they recognized the target word. Additional analyses by word "type" showed that the words' context of occurrence in the test phase had an impact on the younger infants' performances. Indeed, 18-month-olds were looking at the target preferentially after an isolated word, whereas 24-month-olds benefited from both contexts of occurrence. If I set aside the attentional explanation, these findings might be related to stimulus salience: isolated

words might be more salient and easier to process than words embedded in a sentence. It could also reflect a primacy effect (the first word is more salient), as evidenced by the time-windows analyses.
Finally, time-course analyses involving the "target divergence time" computation in the WL condition were also informative. Indeed, while no target divergence time was found in 18-month-olds, a target divergence time was found for each pseudo-word in 24-month-olds, along with several additional time windows in which they looked significantly longer at one object. However, infants were looking at the distractor: this further confirms that infants did not learn the associations in both age groups. Nevertheless, the fact that significant clusters were found during which infants looked longer at the distractor at 24 months suggests that: 1) the presentation of new words did have an effect on the preferred looking behavior: infants did understand that there was a new word associated with an object they did not know before; 2) they were able to extract this new word from the sentences; and 3) they did understand the task, i.e. that they were supposed to look at one of the two visual alternatives. Although these demonstrate that 24-month-olds did not successfully learn the associations during familiarization, they suggest differing exploration strategies during the test phase between 18 and 24 months. Indeed, 24-month-olds' patterns of sustained looking at the distractor may reflect a more intentional decision-making process, as they seem to spend more time considering each object as the target. In contrast, the looking patterns of 18-month-olds appeared more chaotic, as they seemed to switch rapidly between objects without sustained attention, which would explain the absence of clusters in the target divergence time analysis. However, this interpretation should be further assessed with analyses of the number and duration of the switches from one object to the other.

4.2. Focus on the consonant contrast

A major goal of this study was to investigate the interaction between early phonological processing and associative word learning abilities. Accordingly, a minimal pair involving a voicing contrast on the initial consonant was used to create the two pseudowords "bali" and "pali", following previous studies' methodologies (Archer & Curtin, 2018; Höhle et al., 2020; Pater et al., 2004; Rost & McMurray, 2009; Yoshida et al., 2009). The investigation of my data set indicated that, overall, infants failed to learn the new associations at 18 and 24 months, although they successfully recognized familiar words. Analyses of the time courses in the WL condition revealed different looking patterns between age groups and indicated a possible emergence of lexical representation in 24-month-olds. These results seem to be in line with my initial expectations, i.e., that

24-month-olds should be better at processing phonological and lexical information simultaneously in the context of word learning. However, my data did not fully support this claim. Therefore, I will instead discuss the possible challenges that may have hindered infants from learning the two similar-sounding words.

First, the nature of the consonant contrast might be questioned. I chose the /p/-/b/ contrast based on voicing because previous work on minimal pair associative word learning has mostly focused on infants younger than 18 months and often involved place or manner (Archer & Curtin, 2018; Archer et al., 2014; Fais et al., 2012; C. T. Fennell & Waxman, 2006; Werker et al., 2002) of articulation. It is important to highlight that voicing contrasts, such as the difference between /b/ and /p/, require fine phonetic discrimination. Previous research has shown that the ability to process these contrasts can develop later than other contrasts like place of articulation (Archer et al., 2016; Zamuner, 2006). The discrimination of /p/ and /b/ is thus particularly complex due to the relatively subtle acoustic differences between voiced and voiceless consonants and might involve fine-tuned phonological processing skills that are still maturing in infants around 18 to 24 months. Indeed, several studies have demonstrated that infants tend to have more difficulty learning words that differ by voicing, compared to contrasts involving place or manner (Pater et al., 2004; Rost & McMurray, 2010; Werker et al., 2002; Yoshida et al., 2009; Zamuner, 2006). This could be due to the temporal properties of voicing contrasts, as they often rely on small differences in VOT, which can be harder for infants to detect and process. Specifically, Archer et al. (2016) showed that the perception of voicing contrast improves with age, and may not be fully developed for word learning in 18-month-olds. They further highlighted that voicing contrasts might be more perceptually salient after 20 months: therefore, I expected better performance from 24-month-olds than 18-month-olds in my task involving a minimal pair based on voicing. However, my findings show that 24-month-olds may still struggle with learning new words based on this contrast. One possible explanation is that minimal pair associative word learning based on voicing may require a more advanced level of phonological processing than previously assumed. Despite infants' perceptual abilities at this age, the cognitive load associated with this particular task may still be too high.

Second, the prosodic variability introduced in the familiarization phase might also have impacted the results. Previous work in infants has highlighted the crucial role of prosodic cues in various aspects of language development (Christophe et al., 2003; Jusczyk et al., 1999; Mehler et al., 1988; Nazzi et al., 2000, 2006; Teixidó et al., 2018). In

particular, the exaggerated prosodic patterns of IDS have been shown to facilitate word learning (Gervain & Werker, 2013; Graf Estes & Hurley, 2013; Han et al., 2022; Thorson, 2018), especially in the younger infants (Ma et al., 2011; Shukla et al., 2011). Accordingly, most early word learning studies used auditory stimuli produced in IDS. In the present study, I followed these recommendations by presenting isolated words produced in IDS. To increase the prosodic variability and enhance the salience of the /b/-/p/ contrast, I also introduced stress variability. Because the contrast was in word-initial position, I used more trochaic patterns (stress on the first syllable) than iambic patterns (stress on the last syllable), as shown in Figure 6.2a. This choice was made following Floccia et al. (2011)'s work in 20- to 24-month-old English-learning infants, showing that word learning was successful only when the contrast was in a stressed syllable. These results highlighted the strong interaction between stress and phonemic contrast position in minimal pair associative word learning that overrode English infants' sensitivity to native trochaic stress patterns at this age (Graf Estes & Bowen, 2013). Accordingly, I expected this effect to be stronger in French-learning infants, as the French language does not use lexical stress, resulting in a "deafness" to lexical stress in young French learners (Dupoux et al., 1997, 2001). Considering my results, it is possible that the stress patterns I used in the familiarization phase did not facilitate learning. Infants may have processed the trochaic patterns as non-native, as French's dominant stress pattern is iambic (stress on the last syllable), resulting in decreased word processing efficiency. In addition, some general theories of early word learning argue that infants rely on prosodic cues in the earlier stages of word learning and then switch their focus to other social and linguistic cues when they start to expand their vocabulary (Hollich et al., 2000b). Because infants in my study were 18 months old or older, they might already have transited into this later phase of word learning and rely less on prosodic cues. Thus, the enhanced prosodic features of the stimuli did not facilitate learning. However, even if the prosodic feature of the stimuli impacted infants' word learning, other considerations related to general word-learning mechanisms will be discussed in the next paragraph.

4.3. The vocabulary spurt: a critical period for word-learning mechanisms?

The vocabulary spurt is a substantial period in development during which infants transition from slow to fast word learners, accelerating the growth of both their expressive and receptive vocabularies (Goldfield & Reznick, 1990a). Starting between

18 and 20 months, this period also marks changes in the predominant word learning mechanisms and attended cues. My results further supported this claim by showing that 18- and 24-month-olds failed to learn two new pseudo-words when constrained to learn through simple word-object association. The lack of social context variability, the absence of human interactions, and the fact that words were presented in isolation during the learning phase might have contributed to these results.

Indeed, Hollich et al. (2000b)' Emergentist Coalition Model proposes that when learning their first words, infants rather rely on domain-general mechanisms and attentional cues such as perceptual salience, temporal contiguity, or novelty. Thus, associative word learning might be preferred during this period, as it relies on simple associationist mechanisms and statistical regularities of word-referent co-occurrences. This is also a period where infants are particularly sensitive to the exaggerated prosodic contours of the IDS (Brandone et al., 2007; Golinkoff & Hirsh-Pasek, 2006; Nurmsoo & Bloom, 2008; Teixidó et al., 2018). This model is consistent with previous work on associative word learning abilities in 14-month-olds, who successfully learned new word-object associations with simple temporal co-occurrences of objects and their associated labels (C. Fennell & Byers-Heinlein, 2014; Werker et al., 1998). The Emergentist Coalition Model also predicts that older infants with larger vocabulary sizes shift their focus to social and linguistic cues, and develop learning mechanisms based on the social context, other's gaze orientation and intentions, and syntactic cues. This second phase of word learning that occurs after the vocabulary spurt onset is supported by evidence from *cross-situational* word learning studies (Smith & Yu, 2008; Vlach & Sandhofer, 2011; Vouloumanos & Werker, 2009; Yu & Smith, 2011). *Cross-situational* learning implies that "each time a new word is uttered, children entertain a whole set of situationally plausible meanings, and learning entails pruning the potential referential candidates as new instances of the word. The word-meaning mapping thus starts as a one-to-many association." (Dautriche & Chemla, 2014). This change in learning mechanisms is progressive, and infants may still rely on associative mechanisms after the vocabulary spurt, but to a lesser extent (Vlach & Sandhofer, 2011; Vouloumanos & Werker, 2009). Moreover, inter-individual variability implies that infants differ in their strengths and weaknesses at each stage of their developmental trajectory, resulting in different compensatory strategies. Thus, different mechanisms might coexist at a given age, across and within infants (Samuelson, 2021; Yu et al., 2012). My study addressed the question of individual variability in early word learning by investigating the interactions between infants' performance on the associative word learning task and measures of cognitive

and language development.

4.4. Links with language outcomes

Results from the correlation analyses indicated the emergence of an inverted pattern between 18 and 24 months of age: overall, correlations between measures of WL/WR and neurodevelopmental scores were positive at 18 months and negative at 24 months. As illustrated in Figure 6.11, I considered two main measures of infants' lexical processing: the habituation trial and the mean proportion of looking time in both WL and WR conditions. First, although infants did not show evidence of successful minimal pair word earning, I found significant correlations between the habituation trial at 18 months (for habituated participants only) and later language score of the BSID. In contrast, no correlation between the habituation trial and language outcomes were found at 24 months. These results further support the claim that the habituation criterion might be a reliable indicator of lexical processing at 18 months. Furthermore, the absence of correlation at 24 months reinforces the idea that more experienced infants may not prefer associative word learning mechanisms. Indeed, if infants with larger vocabulary shifted to more complex word learning mechanisms, the habituation criteria may not be related to their linguistic skills at this period of development.

Second, mean proportion of looking time to target during the WR trials at 18 months was positively correlated with later cognitive and language scores from the BSID. In line with previous work, word recognition abilities at 18 months were correlated with later language outcomes (Fernald et al., 2013; Legacy et al., 2018; Marchman & Fernald, 2008). However, this was not the case at 24 months. Although my findings that 24-month-old infants displayed better word recognition performance than 18-month-olds are consistent with previous work (Fernald et al., 1998), the absence of correlation with language outcomes was not (Fernald et al., 2006). The reasons for these unexpected results are not clear but may be related to task-related limitations (see next section).

Only one correlation between lexical processing and language outcomes was found at 24 months: the mean proportion of looking to target during WL trials was negatively correlated with the expressive vocabulary (MCDI). Poorer associative word learning abilities at 24 months were associated with larger vocabulary, interpreted as a shift in word learning strategy. The fact that more correlations with lexical skills were found at 18 than 24 months might be due to a developmental difference, as argued by Werker et al. (2002). In this study, they found that vocabulary size predicted word-learning

performance at 14 months but not at 20 months. The authors proposed that there might be a vocabulary size threshold from which infants become more experimented learners and become more efficient at learning new words. I further suggest that the decrease in individual variability might induce an effect on the correlation with vocabulary size, as infants' associative word-learning skills have reached a critical threshold. The present findings at 18 months seem to support this hypothesis. Whereas Werker et al. (2002) argue in favor of a single mechanism influenced by a reorganization of the mental representations (continuity hypothesis), my data seem to indicate a change in learning mechanisms between 18 and 24 months (discontinuity hypothesis). My logic here is that if infants have reached a vocabulary size threshold at 18 or 20 months that no longer depends on their learning mechanisms, I should not have found a correlation at 24 months. Thus, I propose that the inverted pattern of correlation observed between 18 and 24 months reflects a shift in the developmental trajectory of word learning, involving a change in learning mechanisms. However, I need to be very careful with this interpretation because my task has some limitations that should not be overlooked.

4.5. Limitations and future directions

The present work presents some limitations that must be acknowledged. First, some methodological choices might be questioned, such as the . As already discussed, I wanted to expose infants to the same learning input and compare their performances on this basis. However, it is very likely that the infants who reached an early habituation trial found the learning phase too long and lost interest. Moreover, an increase in data loss through trials suggests that, overall, infants lost interest in the last trials, indicating that the task was a bit too long. These issues related to infants' attentional demands might have impacted the results. Stimuli-related limitations were also identified: I chose only four familiar objects to limit the cognitive load of the task, and I tried to find words that (1) were frequent and known at 18 months, (2) were bisyllabic, and (3) had a "masculine" gender[4]. However, I ended up with three semantically related objects (animals) and one word from a different semantic category, which may have induced object preference in infants. In the test phase, I found that infants seemed to have a preference for Object 2 (Figure 6.7: this might also have impacted the results.

To further explore the factors that may have prevented infants from learning the new

4. In French, common nouns have a gender.

associations, I could examine the effect of the prosodic cues in the familiarization phase. Notably, a more thorough investigation of the effect of the prosodic cues should include a condition with native stress patterns. Furthermore, future studies should include an additional control condition in the test phase where the familiar words also constitute a minimal pair. This condition could help disentangle the effect of the associative word learning mechanism from the effect of the phonological contrast and, more generally, study the impact of the interaction between lexical representations and phonological distinction during word learning. Finally, future research should include a longitudinal assessment of word learning abilities spanning a greater developmental period (from 12 to 24 months, for example).

Finally, my findings highlight the importance of further investigation into the developmental trajectory of voicing contrast processing. While 24-month-olds may show improved phonetic discrimination in phonemic discrimination tasks, the complexity of mapping minimal pairs based on voicing contrasts to new objects may require more advanced phonological-lexical interactions, which may continue to develop beyond the second year of life. Future studies could explore whether different learning contexts (e.g., cross-situational learning) could enhance infants' ability to learn voicing-based word distinctions after the vocabulary spurt period.

General discussion

Contents

1 Focus on the main results 195
2 From speech perception to vocabulary acquisition 197
3 Limitations and future directions 199

1. Focus on the main results

This Ph.D. work tackled the fascinating question of how infants acquire their first words. After decades of research involving multiple disciplines, such as developmental psychology, neurosciences, biology, audiology, linguistics, and computational modeling, the field is still progressing. Indeed, vocabulary acquisition is a complex process that involves rapid changes at many levels: perceptual, motor, social, and cognitive. Thus, current research integrates diverse scientific approaches to link multimodal data, intending to decipher the mechanisms, cues, and cognitive resources involved in this early process. With the present work, I aimed to contribute to this scientific effort by providing a new multimodal longitudinal dataset. Specifically, I focused on early neural correlates of speech perception, environmental factors, and associative word learning abilities that might predict vocabulary size at 24 months.

Before discussing the contributions of this work, I present a brief reminder of our main findings. Since each associated chapter already includes a discussion of the results, I will not develop further.

Chapter 4 was dedicated to the exploration of early speech perception abilities in the first two years of life. We developed an innovative paradigm inspired by G. M. Bidelman (2015) that allowed for recording early subcortical and cortical brain responses in a single session. We first confirmed the reliability of this paradigm in a young population, spanning several periods of infants' development (6 to 24 months) with their associated challenges. Visual inspection and analyses of metrics extracted from the

cortical ERPs and subcortical FFRs indicated consistent results with previous studies using similar methods in separate sessions and in the same age ranges (S. Anderson et al., 2015; Cheour, Ceponiene, et al., 1998; Jeng et al., 2010; Ribas-Prats et al., 2023; Rivera-Gaxiola, Csibra, et al., 2000; Werwach et al., 2022). Additional exploratory analyses of the FFR revealed specific developmental patterns for temporal and spectral encoding of speech sounds between 6 and 24 months of age. Specifically, different developmental trajectories were found for the encoding of the fundamental frequency and the higher harmonics, supported by maturational changes in the brain (Gorina-Careta et al., 2021; Kraus et al., 2017; Moore & Linthicum, 2007). At the cortical level, we used a multi-featured oddball paradigm with consonant deviants based on a F2 difference. MMRs were elicited in at all age groups, reflecting an automatic change detection in the auditory cortex. Group comparisons revealed differences in the encoding of these phonetic differences, consistent with the perceptual narrowing theory of phonological representations.

Chapter 5 explored the characteristics of infants' auditory environment using the LENA system. I provided a new LENA corpus in French households, thus expanding the resources and literature on "non-English" datasets. I assessed the reliability of LENA's automatic annotation software on French data, replicating (Canault et al., 2016a)'s findings. Because I could not manually annotate the data, I compared the outputs of different automatic annotation software, revealing consistent measures across systems. Quantitative analyses of the corpus indicated an increase in adults' and infants' speech over time, associated with a significant increase in adult-infant interactions. Results from Warlaumont et al. (2014b) on the emergence of a *social loop* with age were replicated and highlighted the importance of the social factors for infants' language production.

Chapter 6 focused on the interactions between phonological processing and associative word learning abilities during a crucial period of increasing vocabulary size (vocabulary spurt). Using a Looking-While-Listening paradigm and eye-tracking measures, I assessed infants' looking behaviors through classical measures of mean looking times and more fine-grained looking patterns during the task. Unexpectedly, results indicated that infants did not learn the two new words at both ages. The implications for the developmental trajectory of word learning mechanisms were discussed, suggesting an association between the vocabulary spurt and a shift in word learning strategies.

Next, I discuss some general implications of the present work and perspectives for

future directions.

2. From speech perception to vocabulary acquisition

A central theme of this research was the interaction between speech perception abilities and vocabulary acquisition. This could have been explored through the first study only (Chapter 4), by adding correlational analyses between neural correlates of speech perception and the scores of the MCDI and the BSID. Based on previous work, I expect that I would have found correlations between early speech perception abilities and language outcomes (Duff et al., 2015; Kuhl, 2004; Schaadt et al., 2023). However, the scope of this project expanded to unraveling the interactions between multiple factors, considering both internal and external parameters. Notably, the same sample of infants was involved in the different studies and data collected, constituting a single dataset. Thus, this work also tackles the nature vs. nurture debate from an original perspective.

Even if these data were collected in the same infants, independent analyses were performed. Swingley and Aslin (2000) highlighted that "To understand the complex relationship between receptive and productive lexical representations, they must be assessed independently, a point made by many investigators". This statement suggests that to better understand how a complex system works, we must first understand each element of this system independently. This can apply to the relationship between speech perception and vocabulary acquisition in that they form a complex system but might imply different developmental trajectories and recruit different mechanisms, although connected. To disentangle the divergences and commonalities of these processes, it is necessary to first study them separately. This is also a way of eliminating potential confounding factors, such as attentional resources or memory load.

This work additionally addresses the question of what it means to learn in early development (a word or a phonological category in this context). At the phonological level, speech perception might be considered as a low-level process that consists of encoding the features of the auditory inputs. However, research on the emergence of phonological representations suggests the implication of higher-level processes of categorization. As discussed in Chapter 4, infants shift from phonetic encoding to phonological categorization. But in that context, can we use the term "learn"? Researchers usually mention the "emergence" of phonological representations, which implies that infants do not "learn" the phonemic categories. Rather, the categories emerge progressively with exposure

to a consistent native input. In parallel, this reinforces the idea that the environment plays a critical role in this process. *Acquiring* and *learning* are two notions that are often used alternatively in the context of vocabulary development but bear strong different meanings.
Do "word learning" and "vocabulary acquisition" refer to the same concept? I would argue that the first one refers to the active process of linking a new word to a new referent, as in the task presented to infants in Chapter 6. Infants actively create new permanent lexical entries and grow their vocabulary. The second term rather relates to the progressive increase in vocabulary size with age. From this perspective, vocabulary acquisition would be a consequence of word learning. But what does "word learning" imply in terms of mental representations? In the task used in Chapter 6, infants were not tested on their ability to form new robust lexical representations. Rather, they were tested on their ability to associate a word with its referent and recall this association within a short period of time. This associative ability is necessary for word learning, but in this specific task, can we really talk about "word learning"? I argued in the discussion of the Chapter 6 that infants may rely on other word learning mechanisms such as cross-situational word learning. If this is the case, word learning becomes very similar to the progressive emergence of the phonological categories through repeated exposure: new lexical entries emerge with the multiple instances received from different contexts. These remaining open questions highlights the complexity of language development and suggests that further work is needed in developmental populations to "crack the speech code" (Kuhl, 2004).

Finally, to better understand the complexity of language development and decipher the cues, mechanisms, and processes involved, research should extend its scope to more diverse populations. Indeed, the present research includes a well-known bias, especially regarding the targeted population. Healthy full-term infants were recruited from similar or very close SES, growing in France and going to the daycare center every day or mostly every day. This bias for indo-European developed societies in language development studies has been previously discussed and called "WEIRD" for Western Educated Industrialized Rich Democratic (Draper et al., 2023; Freitas et al., 2022). This bias raises concerns about the findings from WEIRD samples because they may not generalize to the broader global population. For example, parenting practices, education systems, and exposure to technology can vary dramatically across cultures, influencing cognitive and linguistic development in ways that are underrepresented in WEIRD-dominant research (Golinkoff et al., 2015). Focusing on these specific samples

might result in extracting norms, values, and behaviors specific to these groups rather than universal human patterns. Thus, as researchers we must consider the question of the reliability of the models and theories that are mostly drawn on these WEIRD studies. These models are based on ethnocentric assumptions that influence our conceptions of "normal", "typical", or "universal" development.

3. Limitations and future directions

The longitudinal design of this study coupled with the multimodality of the data collected made this project particularly challenging. Indeed, the final sample did not include real longitudinal follow-ups, as many infants were recruited relatively at a relatively old age (at 10 months or after), and some of them were included at a young age but did not complete all the recording sessions because of the inherent deadlines of the PhD. Thus, the lack of data at many assessment times for most infants is a central limitation that prevented us from assessing intra-individual changes within the developmental trajectories of each data type.

Although the dataset collected for this thesis is rich and offers numerous avenues for further exploration, the scope of this work was necessarily constrained by time limitations. As such, not all the analyses I had initially envisioned were possible within the timeframe of this PhD. Nonetheless, additional analyses could yield deeper insights into the interactions between the neural, environmental, and cognitive factors assessed here, and provide a strong foundation for future research. Notably, the next step would be to assess the interactions between each separate data gathered in the three main studies. For instance, looking at the links between early correlates of speech perception and quantitative measures of infants' interactions with their caregivers. Importantly, the present dataset would be ideal for computing an integrative neurodevelopmental model of vocabulary acquisition, using computational engineering methods. Such a model could inform about the relative contribution of each factor to language acquisition over time, thus contributing to better understand the balance between external and internal influences in this period of intense development. Beyond the impact on fundamental research, this kind of neurodevelopmental model would be particularly relevant for clinical intervention. Indeed, comprehensive knowledge of the parameters that could enhance or impede language development is a valuable asset for providing infants with earlier and more tailored solutions to language or cognitive impairments. Finally, the measures gathered in this research might also constitute a valuable dataset

for later language and cognitive outcomes: notably, phonological processing and daily environments have been associated with later reading skills (Navarrete-Arroyo et al., 2024; Puolakanaho et al., 2007; Snowling et al., 2003) and might be used as preventive factors of dyslexia (Lyytinen et al., 2008; Snowling & Melby-Lervåg, 2016).

From a broader perspective, future work should continue in this direction and include more longitudinal designs with multimodal datasets. Although challenging, this approach will provide a deeper, more nuanced understanding of language development, considering the richness and inter-individual variability of this process. Moreover, such interdisciplinary work constitutes valuable opportunities for progress across all scientific communities, allowing researchers to bridge gaps between neuroscience, developmental psychology, linguistics, and computational modeling while contributing to both theoretical research and concrete clinical applications.

Bibliography

Abrams, D. A., Nicol, T., Zecker, S., & Kraus, N. (2009). Abnormal cortical processing of the syllable rate of speech in poor readers. *Journal of Neuroscience, 29*(24), 7686–7693 (cit. on p. 37).

Abramson, A. S., & Lisker, L. (1973). Voice-timing perception in spanish word-initial stops. *Journal of Phonetics, 1*(1), 1–8 (cit. on p. 41).

Acke, F. R., De Vriese, C., Van Hoecke, H., & De Leenheer, E. M. (2022). Twelve years of neonatal hearing screening: audiological and etiological results. *European Archives of Oto-Rhino-Laryngology, 279*(7), 3371–3378 (cit. on p. 39).

Agheyisi, R., & Fishman, J. A. (1970). Language attitude studies: a brief survey of methodological approaches. *Anthropological linguistics*, 137–157 (cit. on p. 29).

Aiken, S. J., & Picton, T. W. (2008). Human cortical responses to the speech envelope. *Ear and hearing, 29*(2), 139–157 (cit. on p. 109).

Akhtar, N. (2005). The robustness of learning through overhearing. *Developmental Science, 8*(2), 199–209 (cit. on p. 60).

Al Futaisi, N., Zhang, Z., Cristia, A., Warlaumont, A., & Schuller, B. (2019). Vcmnet: weakly supervised learning for automatic infant vocalisation maturity analysis. *2019 International Conference on Multimodal Interaction*, 205–209 (cit. on pp. 148, 149, 154).

Alho, K., Sainio, K., Sajaniemi, N., Reinikainen, K., & Näätänen, R. (1990). Event-related brain potential of human newborns to pitch change of an acoustic stimulus. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section, 77*(2), 151–155 (cit. on p. 54).

Alho, K., Sajaniemi, N. K., Niittyvuopio, T., Sainio, K. O., & Näätänen, R. (1990). Erps to an auditory stimulus change in preterm and fullterm infants. In *Psychophysiological brain research* (pp. 139–142). Tilburg University Press. (Cit. on p. 54).

Alt, M., & Plante, E. (2006). Factors that influence lexical and semantic fast mapping of young children with specific language impairment. *Journal of Speech, Language, and Hearing, 49*(5) (cit. on p. 79).

Anderson, D. R., Huston, A. C., Schmitt, K. L., Linebarger, D. L., Wright, J. C., & Larson, R. (2001). Early childhood television viewing and adolescent behavior: the recontact study. *Monographs of the society for Research in Child Development*, i–154 (cit. on p. 61).

Anderson, D. R., & Pempek, T. A. (2005). Television and very young children. *American behavioral scientist*, *48*(5), 505–522 (cit. on p. 61).

Anderson, S., Parbery-Clark, A., White-Schwoch, T., & Kraus, N. (2015). Development of subcortical speech representation in human infants. *The Journal of the Acoustical Society of America*, *137*(6), 3346–3355. https://doi.org/10.1121/1.4921032 (cit. on pp. 57, 105, 134–136, 196).

Andruski, J. E., & Kuhl, P. K. (1996). The acoustic structure of vowels in mothers' speech to infants and adults. *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP'96*, *3*, 1545–1548 (cit. on p. 60).

Aragon, M., & Yoshinaga-Itano, C. (2012). Using language environment analysis to improve outcomes for children who are deaf or hard of hearing. *Seminars in Speech and Language*, *33*(04), 340–353 (cit. on pp. 64, 65, 146, 148).

Archer, S. L., & Curtin, S. (2018). Fourteen-month-olds' sensitivity to acoustic salience in minimal pair word learning. *Journal of Child Language*, *45*(5), 1198–1211 (cit. on pp. 81, 83, 164, 172, 187, 188).

Archer, S. L., Ference, J., & Curtin, S. (2014). Now you hear it: fourteen-month-olds succeed at learning minimal pairs in stressed syllables. *Journal of Cognition and Development*, *15*(1), 110–122 (cit. on pp. 81, 164, 166, 188).

Archer, S. L., Zamuner, T., Engel, K., Fais, L., & Curtin, S. (2016). Infants' discrimination of consonants: interplay between word position and acoustic saliency. *Language Learning and Development*, *12*(1), 60–78 (cit. on pp. 81, 83, 84, 166, 172, 188).

Arenillas-Alcón, S., Costa-Faidella, J., Ribas-Prats, T., Gómez-Roig, M. D., & Escera, C. (2021). Neural encoding of voice pitch and formant structure at birth as revealed by frequency-following responses. *Scientific Reports*, *11*(1), 6660. https://doi.org/10.1038/s41598-021-85799-x (cit. on pp. 57, 105, 110, 111, 135, 136).

Arenillas-Alcón, S., Ribas-Prats, T., Puertollano, M., Mondéjar-Segovia, A., Gómez-Roig, M. D., Costa-Faidella, J., & Escera, C. (2023). Prenatal daily musical exposure is associated with enhanced neural representation of speech fundamental frequency: evidence from neonatal frequency-following responses. *Developmental Science*, *26*(5), e13362 (cit. on p. 57).

Aslin, R. N. (1980). Some developmental processes in speech perception. *Child phonology*, *2*, 67–96 (cit. on pp. 44, 45).

Aslin, R. N., Woodward, J. Z., LaMendola, N. P., Bever, T. G., et al. (1996). Models of word segmentation in fluent maternal speech to infants. *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*, 117–134 (cit. on p. 69).

Au, T. K.-f., Knightly, L. M., Jun, S.-A., & Oh, J. S. (2002). Overhearing a language during childhood. *Psychological science*, *13*(3), 238–243 (cit. on p. 60).

Axelsson, E. L., Williams, S. E., & Horst, J. S. (2016). The effect of sleep on children's word retention and generalization. *Frontiers in psychology*, *7*, 1192 (cit. on p. 70).

Bahrick, L. E., & Lickliter, R. (2000). Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Developmental psychology*, *36*(2), 190 (cit. on p. 74).

Ballem, K. D., & Plunkett, K. (2005). Phonological specificity in children at 1; 2. *Journal of child language*, *32*(1), 159–173 (cit. on pp. 82, 84, 165).

Bates, E. (1999). Language and the infant brain. *Journal of communication disorders*, *32*(4), 195–205 (cit. on p. 30).

Bates, E., Benigni, L., Bretherton, I., Camaioni, L., & Volterra, V. (1979). *The emergence of symbols: cognition and communication in infancy*. New York: Academic Press. (Cit. on p. 163).

Bates, E., Dale, P. S., & Thal, D. (1995). Individual differences and their implications for theories of language development. *The Handbook of Child Language*, 95–151 (cit. on pp. 27, 91).

Bates, E., Marchman, V., Thal, D., Fenson, L., Dale, P., Reznick, J. S., Reilly, J., & Hartung, J. (1994). Developmental and stylistic variation in the composition of early vocabulary. *Journal of child language*, *21*(1), 85–123 (cit. on p. 27).

Bateson, M. C. (1979). The epigenesis of conversational interaction: a personal account of research development. In M. Bullowa (Ed.), *Before speech: the beginning of human communication* (pp. 63–77). Cambridge University Press. (Cit. on p. 59).

Bateson, M. C. (1975). Mother-infant exchanges: the epigenesis of conversational interaction. *Annals of the New York Academy of sciences*, *263*(1), 101–113 (cit. on p. 59).

Benedict, H. (1979). Early lexical development: comprehension and production. *Journal of child language*, *6*(2), 183–200 (cit. on p. 146).

Bergelson, E., & Aslin, R. N. (2017). Nature and origins of the lexicon in 6-mo-olds. *Proceedings of the National Academy of Sciences*, *114*(49), 12916–12921 (cit. on pp. 146, 148).

Bergelson, E., Casillas, M., Soderstrom, M., Seidl, A., Warlaumont, A. S., & Amatuni, A. (2019). What do north american babies hear? a large-scale cross-corpus analysis. *Developmental science*, *22*(1), e12724 (cit. on p. 60).

Bergelson, E., Soderstrom, M., Schwarz, I.-C., Rowland, C. F., Ramírez-Esparza, N., R. Hamrick, L., Marklund, E., Kalashnikova, M., Guez, A., Casillas, M., et al. (2023).

Everyday language input and production in 1,001 children from six continents. *Proceedings of the National Academy of Sciences, 120*(52), e2300671120 (cit. on pp. 60, 63, 67, 146, 147, 161).

Bergelson, E., & Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences, 109*(9), 3253–3258 (cit. on pp. 78, 163).

Best, C. T. (1993). Emergence of language-specific constraints in perception of non-native speech: a window on early phonological development. In *Developmental neurocognition: speech and face processing in the first year of life* (pp. 289–304). Springer. (Cit. on pp. 48, 141).

Best, C. T. (1994). The emergence of native-language phonological influences in infants: a perceptual assimilation model. *The development of speech perception: The transition from speech sounds to spoken words, 167*(224), 233–277 (cit. on pp. 48, 49, 141).

Best, C. T. (1995a). A direct realist view of cross-language speech perception. *Speech perception and linguistic experience, 171* (cit. on pp. 48, 141).

Best, C. T. (1995b). Learning to perceive the sound pattern of english. *Advances in infancy research, 9*, 217–217 (cit. on pp. 48, 141).

Best, C. T., & McRoberts, G. W. (2003). Infant perception of non-native consonant contrasts that adults assimilate in different ways. *Language and speech, 46*(2-3), 183–216 (cit. on pp. 43, 104).

Bidelman, G., & Powers, L. (2018). Response properties of the human frequency-following response (ffr) to speech and non-speech sounds: level dependence, adaptation and phase-locking limits. *International journal of audiology, 57*(9), 665–672 (cit. on pp. 56, 105).

Bidelman, G. M. (2015). Towards an optimal paradigm for simultaneously recording cortical and brainstem auditory evoked potentials. *Journal of Neuroscience Methods, 241*, 94–100 (cit. on pp. 105, 107, 133, 195).

Billings, C. J., Bologna, W. J., Muralimanohar, R. K., Madsen, B. M., & Molis, M. R. (2019). Frequency following responses to tone glides: effects of frequency extent, direction, and electrode montage. *Hearing Research, 375*, 25–33 (cit. on p. 134).

Birnholz, J. C., & Benacerraf, B. R. (1983). The development of human fetal hearing. *Science, 222*(4623), 516–518 (cit. on p. 39).

Bisiacchi, P. S., Mento, G., & Suppiej, A. (2009). Cortical auditory processing in preterm newborns: an erp study. *Biological psychology, 82*(2), 176–185 (cit. on p. 54).

Bloom, K., Russell, A., & Wassenberg, K. (1987). Turn taking affects the quality of infant vocalizations. *Journal of child language*, *14*(2), 211–227 (cit. on p. 157).

Bloom, L. (1973). *One word at a time: the use of single word utterances before syntax*. The Haugue: Mouton. (Cit. on pp. 79, 165).

Bonte, M. L., Poelmans, H., & Blomert, L. (2007). Deviant neurophysiological responses to phonological regularities in speech in dyslexic children. *Neuropsychologia*, *45*(7), 1427–1437 (cit. on pp. 54, 141).

Borovsky, A., Burns, E., Elman, J. L., & Evans, J. L. (2013). Lexical activation during sentence comprehension in adolescents with history of specific language impairment. *Journal of communication disorders*, *46*(5-6), 413–427 (cit. on p. 174).

Borovsky, A., Elman, J. L., & Fernald, A. (2012). Knowing a lot for one's age: vocabulary skill and not age is associated with anticipatory incremental sentence interpretation in children and adults. *Journal of experimental child psychology*, *112*(4), 417–436 (cit. on p. 174).

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. (1999). Training japanese listeners to identify english/r/and/l: long-term retention of learning in perception and production. *Perception & psychophysics*, *61*, 977–985 (cit. on p. 45).

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training japanese listeners to identify english/r/and/l: iv. some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America*, *101*(4), 2299–2310 (cit. on p. 45).

Brandone, A. C., Pence, K. L., Golinkoff, R. M., & Hirsh-Pasek, K. (2007). Action speaks louder than words: young children differentially weight perceptual, social, and linguistic cues to learn verbs. *Child development*, *78*(4), 1322–1342 (cit. on p. 190).

Brandt, A., Gebrian, M., & Slevc, L. R. (2012). Music and early language acquisition. *Frontiers in psychology*, *3*, 327 (cit. on p. 62).

Bredin, H., Yin, R., Coria, J. M., Gelly, G., Korshunov, P., Lavechin, M., Fustes, D., Titeux, H., Bouaziz, W., & Gill, M.-P. (2020). Pyannote. audio: neural building blocks for speaker diarization. *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 7124–7128 (cit. on p. 154).

Brown, R. (1973). *A first language: the early stages*. Harvard University Press. (Cit. on p. 29).

Bruner, J. (1974). From communication to language—a psychological perspective. *Cognition*, *3*(3), 255–287. https://doi.org/https://doi.org/10.1016/0010-0277(74)90012-2 (cit. on p. 59).

Bryant, G. A. (2022). Vocal communication across cultures: theoretical and methodological issues. *Philosophical Transactions of the Royal Society B*, *377*(1841), 20200387 (cit. on p. 61).

Bulgarelli, F., & Bergelson, E. (2023). Linking acoustic variability in the infants' input to their early word production. *Developmental Science*, e13545 (cit. on p. 67).

Burnham, D., Kitamura, C., & Vollmer-Conna, U. (2002). What's new, pussycat? on talking to babies and animals. *Science*, *296*(5572), 1435–1435 (cit. on p. 60).

Busch, T., Sangen, A., Vanpoucke, F., & van Wieringen, A. (2018). Correlation and agreement between language environment analysis (lena™) and manual transcription for dutch natural language recordings. *Behavior research methods*, *50*, 1921–1932 (cit. on pp. 63, 146, 148).

Butterfield, E., & Siperstein, G. (1970). Influence of contingent auditory stimulation upon non-nutritional suckle. *Third symposium on oral sensation and perception: The mouth of the infant*, 313–334 (cit. on p. 40).

Cabrera, L., & Gervain, J. (2020). Speech perception at birth: the brain encodes fast and slow temporal information. *Science advances*, *6*(30), eaba7830 (cit. on pp. 40, 137).

Canault, M., Le Normand, M.-T., Foudil, S., Loundon, N., & Thai-Van, H. (2016a). Reliability of the language environment analysis system (LENA™) in European French [Publisher: Springer]. *Behavior research methods*, *48*, 1109–1124 (cit. on pp. 63, 196).

Canault, M., Le Normand, M.-T., Foudil, S., Loundon, N., & Thai-Van, H. (2016b). Reliability of the language environment analysis system (lena™) in european french. *Behavior research methods*, *48*, 1109–1124 (cit. on pp. 146, 148).

Cartmill, E. A. (2023). Overcoming bias in the comparison of human language and animal communication. *Proceedings of the National Academy of Sciences*, *120*(47), e2218799120 (cit. on p. 39).

Casillas, M., Brown, P., & Levinson, S. C. (2020). Early language experience in a tseltal mayan village. *Child Development*, *91*(5), 1819–1835 (cit. on p. 60).

Casillas, M., Bunce, J., Soderstrom, M., Rosemberg, C., Migdalek, M., Alam, F., Stein, A., & Garrison, H. (2017). Introduction: the aclew das template. (Cit. on p. 154).

Casillas, M., & Cristia, A. (2019). A step-by-step guide to collecting and analyzing long-format speech environment (lfse) recordings. *Collabra: Psychology*, *5*(1), 24 (cit. on p. 146).

Caskey, M., Stephens, B., Tucker, R., & Vohr, B. (2011). Importance of parent talk on the development of preterm infant vocalizations. *Pediatrics*, *128*(5), 910–916 (cit. on pp. 65, 148).

Caskey, M., Stephens, B., Tucker, R., & Vohr, B. (2014a). Adult talk in the nicu with preterm infants and developmental outcomes. *Pediatrics*, *133*(3), e578–e584 (cit. on pp. 64, 146).

Caskey, M., Stephens, B., Tucker, R., & Vohr, B. (2014b). Adult talk in the NICU with preterm infants and developmental outcomes [Publisher: American Academy of Pediatrics Elk Grove Village, IL, USA]. *Pediatrics*, *133*(3), e578–e584 (cit. on pp. 65, 87).

Čeponien, 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 (cit. on p. 54).

Čeponien, R., Rinne, T., & Näätänen, R. (2002). Maturation of cortical sound processing as indexed by event-related potentials. *Clinical Neurophysiology*, *113*(6), 870–882 (cit. on pp. 113, 123, 138).

Chandrasekaran, B., & Kraus, N. (2010). The scalp-recorded brainstem response to speech: Neural origins and plasticity. *Psychophysiology*, *47*(2), 236–246. https://doi.org/10.1111/j.1469-8986.2009.00928.x (cit. on pp. 110, 134).

Chandrasekaran, B., Sampath, P. D., & Wong, P. (2010). Individual variability in cue-weighting and lexical tone learning. *The Journal of the Acoustical Society of America*, *128*(1), 456–465 (cit. on p. 45).

Cheng, T.-H. Z., & Zhao, T. C. (2024). Validating a novel paradigm for simultaneously assessing mismatch response and frequency-following response to speech sounds. *Journal of Neuroscience Methods*, *412*, 110277 (cit. on p. 133).

Cheng, Y.-Y., Wu, H.-C., Tzeng, Y.-L., Yang, M.-T., Zhao, L.-L., & Lee, C.-Y. (2015). Feature-specific transition from positive mismatch response to mismatch negativity in early infancy: mismatch responses to vowels and initial consonants. *International Journal of Psychophysiology*, *96*(2), 84–94 (cit. on p. 53).

Cheour, M., Alho, K., Čeponiené, R., Reinikainen, K., Sainio, K., Pohjavuori, M., Aaltonen, O., & Näätänen, R. (1998). Maturation of mismatch negativity in infants. *International journal of psychophysiology*, *29*(2), 217–226 (cit. on pp. 44, 54, 104).

Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., & Näätänen, R. (1998). Development of language-specific phoneme representations in the infant brain. *Nature Neuroscience*, *1*(5), 351–353. https://doi.org/10.1038/1561 (cit. on pp. 42, 44, 45, 54, 104, 105, 138, 146, 164, 196).

Christakis, D. A., Gilkerson, J., Richards, J. A., Zimmerman, F. J., Garrison, M. M., Xu, D., Gray, S., & Yapanel, U. (2009). Audible television and decreased adult words, infant vocalizations, and conversational turns: a population-based study. *Archives of pediatrics & adolescent medicine*, *163*(6), 554–558 (cit. on pp. 64, 67, 146–148).

Christakis, D. A., Zimmerman, F. J., DiGiuseppe, D. L., & McCarty, C. A. (2004). Early television exposure and subsequent attentional problems in children. *Pediatrics*, *113*(4), 708–713 (cit. on p. 61).

Christophe, A., Nespor, M., Teresa Guasti, M., & Van Ooyen, B. (2003). Prosodic structure and syntactic acquisition: the case of the head-direction parameter. *Developmental Science*, *6*(2), 211–220 (cit. on p. 188).

Clarkson, M. G., Clifton, R. K., Swain, I. U., & Perris, E. E. (1989). Stimulus duration and repetition rate influence newborns' head orientation toward sound. *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology*, *22*(7), 683–705 (cit. on p. 54).

Coch, D., & Gullick, M. M. (2012). Event-related potentials and development. *The Oxford handbook of event-related potential components*, 475–511 (cit. on pp. 53, 113, 138).

Coffey, E. B. J., Herholz, S. C., Chepesiuk, A. M. P., Baillet, S., & Zatorre, R. J. (2016). Cortical contributions to the auditory frequency-following response revealed by MEG. *Nature Communications*, *7*(1), 11070. https://doi.org/10.1038/ncomms11070 (cit. on pp. 56, 105).

Coffey, E. B. J., Nicol, T., White-Schwoch, T., Chandrasekaran, B., Krizman, J., Skoe, E., Zatorre, R. J., & Kraus, N. (2019). Evolving perspectives on the sources of the frequency-following response. *Nature Communications*, *10*(1), 5036. https://doi.org/10.1038/s41467-019-13003-w (cit. on pp. 56, 105).

Coffey, E. B., Arseneau-Bruneau, I., Zhang, X., Baillet, S., & Zatorre, R. J. (2021). Oscillatory entrainment of the frequency-following response in auditory cortical and subcortical structures. *Journal of Neuroscience*, *41*(18), 4073–4087 (cit. on pp. 56, 105).

Cohen, L. B. (1972). Attention-getting and attention-holding processes of infant visual preferences. *Child development*, 869–879 (cit. on p. 74).

Crawley, A. M., Anderson, D. R., Wilder, A., Williams, M., & Santomero, A. (1999). Effects of repeated exposures to a single episode of the television program blue's clues on the viewing behaviors and comprehension of preschool children. *Journal of educational psychology*, *91*(4), 630 (cit. on p. 61).

Cristia, A. (2011). Fine-grained variation in caregivers'/s/predicts their infants'/s/category. *The Journal of the Acoustical Society of America*, *129*(5), 3271–3280 (cit. on p. 66).

Cristia, A., Dupoux, E., Gurven, M., & Stieglitz, J. (2019). Child-directed speech is infrequent in a forager-farmer population: a time allocation study. *Child development*, *90*(3), 759–773 (cit. on p. 60).

Cristia, A., Lavechin, M., Scaff, C., Soderstrom, M., Rowland, C., Räsänen, O., Bunce, J., & Bergelson, E. (2021a). A thorough evaluation of the Language Environment Analysis (LENA) system [Publisher: Springer]. *Behavior research methods*, *53*, 467–486 (cit. on pp. 32, 63).

Cristia, A., Lavechin, M., Scaff, C., Soderstrom, M., Rowland, C., Räsänen, O., Bunce, J., & Bergelson, E. (2021b). A thorough evaluation of the language environment analysis (lena) system. *Behavior research methods*, *53*, 467–486 (cit. on pp. 64, 149).

Cristia, A., & Seidl, A. (2014). The hyperarticulation hypothesis of infant-directed speech. *Journal of child language*, *41*(4), 913–934 (cit. on p. 61).

Cruz Blandón, M. A., Cristia, A., & Räsänen, O. (2023). Introducing meta-analysis in the evaluation of computational models of infant language development. *Cognitive Science*, *47*(7), e13307 (cit. on p. 162).

Curtin, S., Fennell, C., & Escudero, P. (2009). Weighting of vowel cues explains patterns of word–object associative learning. *Developmental Science*, *12*(5), 725–731 (cit. on pp. 80, 164, 172, 185).

Custodero, L. A., Britto, P. R., & Brooks-Gunn, J. (2003). Musical lives: a collective portrait of american parents and their young children. *Journal of Applied Developmental Psychology*, *24*(5), 553–572 (cit. on p. 62).

Cychosz, M., Romeo, R., Soderstrom, M., Scaff, C., Ganek, H., Cristia, A., Casillas, M., De Barbaro, K., Bang, J. Y., & Weisleder, A. (2020). Longform recordings of everyday life: ethics for best practices. *Behavior research methods*, *52*, 1951–1969 (cit. on p. 146).

Dale, R., Warlaumont, A. S., & Johnson, K. L. (2023). The fundamental importance of method to theory. *Nature Reviews Psychology*, *2*(1), 55–66 (cit. on p. 91).

Darwin, C. (1877). A biographical sketch of an infant [Publisher: JSTOR]. *Mind*, *2*(7), 285–294 (cit. on p. 28).

Dautriche, I., & Chemla, E. (2014). Cross-situational word learning in the right situations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(3), 892 (cit. on p. 190).

Dawson, G. (1991). Viii a psychobiological perspective on the early socio-emotional development of children with autism. *Models Integr, 3*, 207 (cit. on p. 30).

De Saussure, F. (1989). *Cours de linguistique générale* (Vol. 1). Otto Harrassowitz Verlag. (Cit. on pp. 69, 70).

De Vries, P. (2021). Music at home with the under-fives: what is happening? In *Music in the lives of young children* (pp. 59–72). Routledge. (Cit. on p. 62).

de Boysson-Bardies, B. (1993). Ontogeny of language-specific syllabic productions. In *Developmental neurocognition: speech and face processing in the first year of life* (pp. 353–363). Springer. (Cit. on p. 47).

DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: newborns prefer their mothers' voices. *Science, 208*(4448), 1174–1176 (cit. on p. 40).

Dehaene-Lambertz, G. (2000a). Cerebral specialization for speech and non-speech stimuli in infants. *Journal of Cognitive Neuroscience, 12*(3), 449–460 (cit. on p. 30).

Dehaene-Lambertz, G. (2000b). Cerebral specialization for speech and non-speech stimuli in infants. *Journal of Cognitive Neuroscience, 12*(3), 449–460 (cit. on pp. 44, 54, 104).

Dehaene-Lambertz, G., & Dehaene, S. (1994). Speed and cerebral correlates of syllable discrimination in infants. *Nature, 370*(6487), 292–295. https://doi.org/10.1038/370292a0 (cit. on p. 104).

Dehaene-Lambertz, G., & Gliga, T. (2004). Common neural basis for phoneme processing in infants and adults. *Journal of cognitive neuroscience, 16*(8), 1375–1387 (cit. on p. 53).

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 (cit. on p. 47).

Dehaene-Lambertz, G., & Pena, M. (2001). Electrophysiological evidence for automatic phonetic processing in neonates. *Neuroreport, 12*(14), 3155–3158 (cit. on p. 40).

Delattre, P. C., Liberman, A. M., & Cooper, F. S. (1955). Acoustic loci and transitional cues for consonants. *The journal of the acoustical society of America, 27*(4), 769–773 (cit. on p. 46).

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 (cit. on p. 109).

Diehl, R. L., Lotto, A. J., & Holt, L. L. (2004). Speech perception. *Annu. Rev. Psychol.*, *55*(1), 149–179 (cit. on pp. 47, 48).

Draper, C. E., Barnett, L. M., Cook, C. J., Cuartas, J. A., Howard, S. J., McCoy, D. C., Merkley, R., Molano, A., Maldonado-Carreño, C., Obradović, J., et al. (2023). Publishing child development research from around the world: an unfair playing field resulting in most of the world's child population under-represented in research. *Infant and Child Development*, *32*(6), e2375 (cit. on p. 198).

Duff, F. J., Reen, G., Plunkett, K., & Nation, K. (2015). Do infant vocabulary skills predict school-age language and literacy outcomes? *Journal of Child Psychology and Psychiatry*, *56*(8), 848–856 (cit. on pp. 92, 197).

Duffy, K. R., & Mitchell, D. E. (2013). Darkness alters maturation of visual cortex and promotes fast recovery from monocular deprivation. *Current biology*, *23*(5), 382–386 (cit. on p. 104).

Dupoux, E., Pallier, C., Sebastian, N., & Mehler, J. (1997). A destressing "deafness" in french? *Journal of memory and language*, *36*(3), 406–421 (cit. on p. 189).

Dupoux, E., Peperkamp, S., & Sebastián-Gallés, N. (2001). A robust method to study stress "deafness". *The Journal of the Acoustical Society of America*, *110*(3), 1606–1618 (cit. on p. 189).

Eggermont, J. J., & Moore, J. K. (2012). Morphological and functional development of the auditory nervous system. In *Brain development and cognition: a reader* (pp. 181–201). John Wiley & Sons. (Cit. on p. 136).

Eibl-Eibesfeldt, I., & Kramer, S. (1958). Ethology, the comparative study of animal behavior. *The Quarterly review of biology*, *33*(3), 181–211 (cit. on p. 28).

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, *171*(3968), 303–306 (cit. on pp. 40, 43, 113).

Ellis, E. M., Borovsky, A., Elman, J. L., & Evans, J. L. (2015). Novel word learning: an eye-tracking study. are 18-month-old late talkers really different from their typical peers? *Journal of communication disorders*, *58*, 143–157 (cit. on pp. 79, 82, 84–86, 166, 172, 174).

Elmer, S., Hausheer, M., Albrecht, J., & Kühnis, J. (2017). Human brainstem exhibits higher sensitivity and specificity than auditory-related cortex to short-term phonetic discrimination learning. *Scientific reports*, *7*(1), 7455 (cit. on p. 105).

Escera, C. (2023). Contributions of the subcortical auditory system to predictive coding and the neural encoding of speech. *Current Opinion in Behavioral Sciences* (cit. on p. 105).

Eyben, F., Scherer, K. R., Schuller, B. W., Sundberg, J., André, E., Busso, C., Devillers, L. Y., Epps, J., Laukka, P., Narayanan, S. S., et al. (2015). The geneva minimalistic acoustic parameter set (gemaps) for voice research and affective computing. *IEEE transactions on affective computing*, *7*(2), 190–202 (cit. on p. 154).

Eyben, F., Wöllmer, M., & Schuller, B. (2010). Opensmile: the munich versatile and fast open-source audio feature extractor. *Proceedings of the 18th ACM international conference on Multimedia*, 1459–1462 (cit. on p. 154).

Fais, L., Werker, J. F., Cass, B., Leibowich, J., Barbosa, A. V., & Vatikiotis-Bateson, E. (2012). Here's looking at you, baby: what gaze and movement reveal about minimal pair word-object association at 14 months. *Laboratory Phonology*, *3*(1), 91–124 (cit. on p. 188).

Fellman, V., Kushnerenko, E., Mikkola, K., Ceponiené, 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 (cit. on p. 87).

Fennell, C., & Byers-Heinlein, K. (2014). You sound like mommy: bilingual and monolingual infants learn words best from speakers typical of their language environments. *International Journal of Behavioral Development*, *38*(4), 309–316 (cit. on p. 190).

Fennell, C. T. (2011). Habituation procedures. *Research methods in child language: A practical guide*, 1–16 (cit. on p. 85).

Fennell, C. T., & Waxman, S. R. (2010). What paradox? referential cues allow for infant use of phonetic detail in word learning. *Child development*, *81*(5), 1376–1383 (cit. on pp. 81, 83, 164, 172).

Fennell, C. T., & Waxman, S. (2006). Infants of 14 months use phonetic detail in novel words embedded in naming phrases. *Proceedings of the 30th annual Boston University conference on language development*, 178–189 (cit. on p. 188).

Fenson, L., Dale, P. S., Thal, D., Bates, E., Hartung, J. P., Pethick, S., & Reilly, J. S. (1993). *Macarthur communicative development inventories: user's guide and technical manual*. San Diego: Singular. (Cit. on pp. 86, 171).

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., Tomasello, M., Mervis, C. B., & Stiles, J. (1994). Variability in early communicative development. *Monographs of the society for research in child development*, i–185 (cit. on p. 87).

Ferguson, C. A. (1964). Baby talk in six languages. *American anthropologist*, *66*(6), 103–114 (cit. on p. 60).

Ferguson, C. A. (1994). Phonological development: models, research, implications. *Language and Speech*, *37*(3), 299–314 (cit. on p. 40).

Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: is the melody the message? *Child development*, 1497–1510 (cit. on p. 60).

Fernald, A., Marchman, V. A., & Weisleder, A. (2013). Ses differences in language processing skill and vocabulary are evident at 18 months. *Developmental science*, *16*(2), 234–248 (cit. on p. 191).

Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding: speech processing efficiency and vocabulary growth across the 2nd year. *Developmental psychology*, *42*(1), 98 (cit. on pp. 79, 87, 165, 191).

Fernald, A., Pinto, J. P., Swingley, D., Weinberg, A., & McRoberts, G. W. (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological science*, *9*(3), 228–231 (cit. on pp. 79, 83, 85, 165, 169, 191).

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental psychology*, *20*(1), 104 (cit. on p. 60).

Floccia, C., Nazzi, T., Austin, K., Arreckx, F., & Goslin, J. (2011). Lexical stress and phonetic processing in word learning in20-to 24-month-old english-learning children. *Developmental Science*, *14*(3), 602–613 (cit. on pp. 168, 189).

Font-Alaminos, M., Ribas-Prats, T., Gorina-Careta, N., & Escera, C. (2021). Emergence of prediction error along the human auditory hierarchy. *Hearing Research*, *399*, 107954 (cit. on p. 133).

Ford, M., Baer, C. T., Xu, D., Yapanel, U., & Gray, S. (2008). The lena^{TM} language environment analysis system. *The LENA^{TM} Language Environment Analysis System* (cit. on p. 153).

Foundation, L. (2015). Lena foundation [https://www.lena.org/ [Accessed: (July 2024)]]. (Cit. on pp. 63, 151, 153).

Fowler, C. A. (1996). Listeners do hear sounds, not tongues. *The Journal of the Acoustical Society of America*, *99*(3), 1730–1741 (cit. on pp. 47, 48).

François, C., Rodriguez-Fornells, A., Cerda-Company, X., Agut, T., & Bosch, L. (2024). Impact of late to moderate preterm birth on minimal pair word-learning. *Child Development* (cit. on pp. 166, 168, 172).

François, C., Rodriguez-Fornells, A., Teixidó, M., Agut, T., & Bosch, L. (2021). Attenuated brain responses to speech sounds in moderate preterm infants at term age. *Developmental Science*, *24*(1), e12990 (cit. on pp. 44, 54, 104, 105, 109, 112).

François, C., Teixidó, M., Takerkart, S., Agut, T., Bosch, L., & Rodriguez-Fornells, A. (2017). Enhanced neonatal brain responses to sung streams predict vocabulary

outcomes by age 18 months. *Scientific reports*, *7*(1), 12451 (cit. on pp. 54, 112, 139).

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2021). *Variability and consistency in early language learning: the wordbank project*. MIT Press. (Cit. on p. 27).

Freitas, L. B., Palhares, F., Cao, H., Liang, Y., Zhou, N., Mokrova, I. L., Lee, S., Payir, A., Kiang, L., Mendonça, S. E., et al. (2022). How weird is the development of children's gratitude in the united states? cross-cultural comparisons. *Developmental psychology*, *58*(9), 1767 (cit. on p. 198).

Friederici, A. D. (2005). Neurophysiological markers of early language acquisition: from syllables to sentences. *Trends in Cognitive Sciences*, *9*(10), 481–488 (cit. on pp. 137–139).

Friederici, A. D., Friedrich, M., & Weber, C. (2002). Neural manifestation of cognitive and precognitive mismatch detection in early infancy. *Neuroreport*, *13*(10), 1251–1254 (cit. on p. 140).

Friedrich, M., & Friederici, A. D. (2010). Maturing brain mechanisms and developing behavioral language skills. *Brain and language*, *114*(2), 66–71 (cit. on p. 30).

Friedrich, M., & Friederici, A. D. (2011). Word learning in 6-month-olds: fast encoding–weak retention. *Journal of cognitive neuroscience*, *23*(11), 3228–3240 (cit. on p. 78).

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic bulletin & review*, *13*(3), 361–377 (cit. on p. 46).

Galbraith, G. C., & Kane, J. M. (1993). Brainstem Frequency-Following Responses and Cortical Event-Related Potentials during Attention. *Perceptual and Motor Skills*, *76*(3_suppl), 1231–1241. https://doi.org/10.2466/pms.1993.76.3c.1231 (cit. on p. 105).

Galbraith, G. C., Threadgill, M. R., Hemsley, J., Salour, K., Songdej, N., Ton, J., & Cheung, L. (2000). Putative measure of peripheral and brainstem frequency-following in humans. *Neuroscience letters*, *292*(2), 123–127 (cit. on p. 134).

Ganek, H., & Eriks-Brophy, A. (2018a). Language environment analysis (lena) system investigation of day long recordings in children: a literature review. *Journal of Communication Disorders*, *72*, 77–85 (cit. on pp. 63, 146, 147).

Ganek, H., & Eriks-Brophy, A. (2018b). Language ENvironment analysis (LENA) system investigation of day long recordings in children: A literature review [Publisher: Elsevier]. *Journal of Communication Disorders*, *72*, 77–85 (cit. on p. 64).

Ganek, H. V., & Eriks-Brophy, A. (2017). A concise protocol for the validation of language environment analysis (lena) conversational turn counts in vietnamese. *Communication Disorders Quarterly*, *39*(2), 371–380 (cit. on p. 146).

Garcia-Sierra, A., Ramírez-Esparza, N., & Kuhl, P. K. (2016). Relationships between quantity of language input and brain responses in bilingual and monolingual infants. *International Journal of Psychophysiology*, *110*, 1–17 (cit. on p. 142).

Garcia-Sierra, A., Rivera-Gaxiola, M., Percaccio, C. R., Conboy, B. T., Romo, H., Klarman, L., Ortiz, S., & Kuhl, P. K. (2011). Bilingual language learning: An ERP study relating early brain responses to speech, language input, and later word production. *Journal of Phonetics*, *39*(4), 546–557. https://doi.org/10.1016/j.wocn.2011.07.002 (cit. on pp. 104, 141).

Gautheron, L., Rochat, N., & Cristia, A. (2023). Managing, storing, and sharing long-form recordings and their annotations. *Language Resources and Evaluation*, *57*(1), 343–375 (cit. on pp. 64, 149).

Gelman, S. A., & Taylor, M. (1984). How two-year-old children interpret proper and common names for unfamiliar objects. *Child development*, 1535–1540 (cit. on p. 76).

Gershkoff-Stowe, L., & Hahn, E. R. (2007). Fast mapping skills in the developing lexicon. *Journal of Speech, Language, and Hearing Research*, *50*, 682–697 (cit. on p. 79).

Gervain, J., & Werker, J. F. (2013). Prosody cues word order in 7-month-old bilingual infants. *Nature communications*, *4*(1), 1490 (cit. on pp. 104, 189).

Gilkerson, J., & Richards, J. A. (2008a). The lena natural language study. *Boulder, CO: LENA Foundation. Retrieved March*, *3*(2009), 15–17 (cit. on pp. 64, 146–148, 151).

Gilkerson, J., & Richards, J. A. (2008b). The LENA natural language study. *Boulder, CO: LENA Foundation. Retrieved March*, *3*(2009), 15–17 (cit. on pp. 64, 152).

Gilkerson, J., & Richards, J. A. (2009a). The power of talk. *Impact of adult talk, conversational turns and TV during the critical 0-4 years of child development: Boulder, CO: LENA Foundation* (cit. on pp. 64, 66, 67, 87).

Gilkerson, J., & Richards, J. A. (2009b). The power of talk. *Impact of adult talk, conversational turns and TV during the critical 0-4 years of child development: Boulder, CO: LENA Foundation* (cit. on pp. 66, 146).

Gilkerson, J., & Richards, J. A. (2020). A Guide to Understanding the Design and Purpose of the LENA® System. *LENA Foundation: Boulder, CO*, 15 (cit. on pp. 151, 153).

Gilkerson, J., Richards, J. A., Warren, S. F., Montgomery, J. K., Greenwood, C. R., Kimbrough Oller, D., Hansen, J. H., & Paul, T. D. (2017a). Mapping the early language

environment using all-day recordings and automated analysis. *American journal of speech-language pathology*, *26*(2), 248–265 (cit. on p. 148).

Gilkerson, J., Richards, J. A., Warren, S. F., Montgomery, J. K., Greenwood, C. R., Kimbrough Oller, D., Hansen, J. H., & Paul, T. D. (2017b). Mapping the early language environment using all-day recordings and automated analysis [Publisher: ASHA]. *American journal of speech-language pathology*, *26*(2), 248–265 (cit. on p. 153).

Gilkerson, J., Richards, J. A., Warren, S. F., Oller, D. K., Russo, R., & Vohr, B. (2018a). Language experience in the second year of life and language outcomes in late childhood [Publisher: American Academy of Pediatrics]. *Pediatrics*, *142*(4) (cit. on pp. 66, 67).

Gilkerson, J., Richards, J. A., Warren, S. F., Oller, D. K., Russo, R., & Vohr, B. (2018b). Language experience in the second year of life and language outcomes in late childhood. *Pediatrics*, *142*(4) (cit. on p. 148).

Gilkerson, J., Zhang, Y., Xu, D., Richards, J. A., Xu, X., Jiang, F., Harnsberger, J., & Topping, K. (2015). Evaluating language environment analysis system performance for chinese: a pilot study in shanghai. *Journal of Speech, Language, and Hearing Research*, *58*(2), 445–452 (cit. on pp. 146, 148).

Gleitman, L. R., Cassidy, K., Nappa, R., Papafragou, A., & Trueswell, J. C. (2005). Hard words. *Language learning and development*, *1*(1), 23–64 (cit. on p. 77).

Gogate, L. J., & Bahrick, L. E. (1998). Intersensory redundancy facilitates learning of arbitrary relations between vowel sounds and objects in seven-month-old infants. *Journal of experimental child psychology*, *69*(2), 133–149 (cit. on p. 74).

Gogate, L. J., & Bahrick, L. E. (2001). Intersensory redundancy and 7-month-old infants' memory for arbitrary syllable-object relations. *Infancy*, *2*(2), 219–231 (cit. on p. 78).

Goldfield, B. A., & Reznick, J. S. (1990a). Early lexical acquisition: Rate, content, and the vocabulary spurt [Publisher: Cambridge University Press]. *Journal of child language*, *17*(1), 171–183 (cit. on pp. 79, 165, 189).

Goldfield, B. A., & Reznick, J. S. (1990b). Early lexical acquisition: rate, content, and the vocabulary spurt. *Journal of child language*, *17*(1), 171–183 (cit. on pp. 79, 146).

Golinkoff, R. M., Can, D. D., Soderstrom, M., & Hirsh-Pasek, K. (2015). (baby) talk to me: the social context of infant-directed speech and its effects on early language acquisition. *Current Directions in Psychological Science*, *24*(5), 339–344 (cit. on pp. 61, 198).

Golinkoff, R. M., Hirsh-Pasek, K., Cauley, K. M., & Gordon, L. (1987). The eyes have it: lexical and syntactic comprehension in a new paradigm. *Journal of child language*, *14*(1), 23–45 (cit. on pp. 83, 165).

Golinkoff, R. M., & Hirsh-Pasek, K. (2006). Baby wordsmith: from associationist to social sophisticate. *Current directions in psychological science*, *15*(1), 30–33 (cit. on pp. 77, 190).

Golinkoff, R. M., Hoff, E., Rowe, M. L., Tamis-LeMonda, C. S., & Hirsh-Pasek, K. (2019). Language matters: denying the existence of the 30-million-word gap has serious consequences. *Child development*, *90*(3), 985–992 (cit. on p. 60).

Golinkoff, R. M., Ma, W., Song, L., & Hirsh-Pasek, K. (2013). Twenty-five years using the intermodal preferential looking paradigm to study language acquisition: what have we learned? *Perspectives on Psychological Science*, *8*(3), 316–339 (cit. on p. 83).

Golinkoff, R. M., Mervis, C. B., & Hirsh-Pasek, K. (1994). Early object labels: the case for a developmental lexical principles framework [*]. *Journal of child language*, *21*(1), 125–155 (cit. on pp. 71, 72).

Goren, C. C., Sarty, M., & Wu, P. Y. (1975). Visual following and pattern discrimination of face-like stimuli by newborn infants. *Pediatrics*, *56*(4), 544–549 (cit. on p. 40).

Gorga, M. P., Worthington, D. W., Reiland, J. K., Beauchaine, K. A., & Goldgar, D. E. (1985). Some comparisons between auditory brain stem response thresholds, latencies, and the pure-tone audiogram. *Ear and hearing*, *6*(2), 105–112 (cit. on p. 109).

Gorina-Careta, N., Arenillas-Alcón, S., Puertollano, M., Mondéjar-Segovia, A., Ijjou-Kadiri, S., Costa-Faidella, J., Gómez-Roig, M. D., & Escera, C. (2024). Exposure to bilingual or monolingual maternal speech during pregnancy affects the neurophysiological encoding of speech sounds in neonates differently. *Frontiers in Human Neuroscience*, *18*, 1379660 (cit. on p. 57).

Gorina-Careta, N., Kurkela, J. L., Hämäläinen, J., Astikainen, P., & Escera, C. (2021). Neural generators of the frequency-following response elicited to stimuli of low and high frequency: a magnetoencephalographic (meg) study. *Neuroimage*, *231*, 117866 (cit. on pp. 56, 196).

Gorina-Careta, N., Ribas-Prats, T., Arenillas-Alcón, S., Puertollano, M., Gómez-Roig, M. D., & Escera, C. (2022). Neonatal Frequency-Following Responses: A Methodological Framework for Clinical Applications. *Seminars in Hearing*, *43*(03), 162–176. https://doi.org/10.1055/s-0042-1756162 (cit. on p. 105).

Gottlieb, G. (1976). The roles of experience in the development of behavior and the nervous system. In *Studies on the development of behavior and the nervous system* (pp. 25–54, Vol. 3). Elsevier. (Cit. on p. 44).

Graf Estes, K., & Bowen, S. (2013). Learning about sounds contributes to learning about words: effects of prosody and phonotactics on infant word learning. *Journal of experimental child psychology*, *114*(3), 405–417 (cit. on p. 189).

Graf Estes, K., & Hurley, K. (2013). Infant-directed prosody helps infants map sounds to meanings. *Infancy*, *18*(5), 797–824 (cit. on pp. 185, 189).

Gredebäck, G., Stasiewicz, D., Falck-Ytter, T., Rosander, K., & von Hofsten, C. (2009). Action type and goal type modulate goal-directed gaze shifts in 14-month-old infants. *Developmental psychology*, *45*(4), 1190 (cit. on p. 86).

Grieser, D. L., & Kuhl, P. K. (1988). Maternal speech to infants in a tonal language: support for universal prosodic features in motherese. *Developmental psychology*, *24*(1), 14 (cit. on p. 60).

Guthrie, E. R. (1930). Conditioning as a principle of learning. *Psychological review*, *37*(5), 412 (cit. on p. 28).

Háden, G. P., Honing, H., Török, M., & Winkler, I. (2015). Detecting the temporal structure of sound sequences in newborn infants. *International Journal of Psychophysiology*, *96*(1), 23–28 (cit. on p. 54).

Háden, G. P., Németh, R., Török, M., & Winkler, I. (2015). Predictive processing of pitch trends in newborn infants. *Brain research*, *1626*, 14–20 (cit. on p. 54).

Haenschel, C., Vernon, D. J., Dwivedi, P., Gruzelier, J. H., & Baldeweg, T. (2005). Event-related brain potential correlates of human auditory sensory memory-trace formation. *Journal of Neuroscience*, *25*(45), 10494–10501 (cit. on p. 52).

Hall, E. R., Esty, E. T., & Fisch, S. M. (1990). Television and children's problem-solving behavior: a synopsis of an evaluation of the effects of square one tv.. *The Journal of Mathematical Behavior* (cit. on p. 61).

Hämäläinen, J., Landi, N., Loberg, O., Lohvansuu, K., Pugh, K., & Leppänen, P. H. (2018). Brain event-related potentials to phoneme contrasts and their correlation to reading skills in school-age children. *International Journal of Behavioral Development*, *42*(3), 357–372 (cit. on pp. 54, 141).

Han, M., De Jong, N. H., & Kager, R. (2022). Prosodic input and children's word learning in infant-and adult-directed speech. *Infant Behavior and Development*, *68*, 101728 (cit. on p. 189).

Hart, B., & Risley, T. R. (1992a). American parenting of language-learning children: Persisting differences in family-child interactions observed in natural home

environments. [Publisher: American Psychological Association]. *Developmental psychology*, *28*(6), 1096 (cit. on pp. 59, 62).

Hart, B., & Risley, T. R. (1992b). American parenting of language-learning children: persisting differences in family-child interactions observed in natural home environments. *Developmental psychology*, *28*(6), 1096 (cit. on pp. 60, 62, 146).

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. [Pages: xxiii, 268]. Paul H Brookes Publishing. (Cit. on pp. 62, 67).

Harwood, R. L., Schoelmerich, A., Schulze, P. A., & Gonzalez, Z. (1999). Cultural differences in maternal beliefs and behaviors: a study of middle-class anglo and puerto rican mother-infant pairs in four everyday situations. *Child development*, *70*(4), 1005–1016 (cit. on p. 62).

He, C., Hotson, L., & Trainor, L. J. (2007). Mismatch responses to pitch changes in early infancy. *Journal of cognitive neuroscience*, *19*(5), 878–892 (cit. on pp. 54, 104, 140).

He, C., Hotson, L., & Trainor, L. J. (2009). Maturation of cortical mismatch responses to occasional pitch change in early infancy: effects of presentation rate and magnitude of change. *Neuropsychologia*, *47*(1), 218–229 (cit. on pp. 53, 54, 104).

Heilmann, J., Weismer, S. E., Evans, J., & Hollar, C. (2005). Utility of the macarthur—bates communicative development inventory in identifying language abilities of late-talking and typically developing toddlers. *American Journal of Speech and Language Pathology*, *14*, 40–51 (cit. on p. 87).

Hensch, T. K. (2005). Critical period plasticity in local cortical circuits. *Nature Reviews Neuroscience*, *6*(11), 877–888 (cit. on p. 104).

Hervé, E., François, C., & Prevot, L. (2024). Daily auditory environments in french-speaking infants: a longitudinal dataset. *Workshop on Cognitive Modeling and Computational Linguistics*, 132–151 (cit. on p. 145).

Hervé, E., Mento, G., Desnous, B., & François, C. (2022). Challenges and new perspectives of developmental cognitive EEG studies [Publisher: Elsevier]. *NeuroImage*, *260*, 119508. https://doi.org/10.1016/j.neuroimage.2022.119508 (cit. on pp. 52, 102, 104, 134, 259, 275).

Hilton, C. B., Moser, C. J., Bertolo, M., Lee-Rubin, H., Amir, D., Bainbridge, C. M., Simson, J., Knox, D., Glowacki, L., Alemu, E., et al. (2022). Acoustic regularities in infant-directed speech and song across cultures. *Nature Human Behaviour*, *6*(11), 1545–1556 (cit. on p. 62).

Hippe, L., Hennessy, V., Ramirez, N. F., & Zhao, T. C. (2024). Comparison of speech and music input in north american infants' home environment over the first 2 years of life. *Developmental Science*, e13528 (cit. on pp. 62, 65).

Hirsh-Pasek, K., Adamson, L. B., Bakeman, R., Owen, M. T., Golinkoff, R. M., Pace, A., Yust, P. K., & Suma, K. (2015). The contribution of early communication quality to low-income children's language success. *Psychological science*, *26*(7), 1071–1083 (cit. on p. 67).

Hockett, C. F., & Hockett, C. D. (1960). The origin of speech. *Scientific American*, *203*(3), 88–97 (cit. on p. 69).

Hoff, E. (2003). The specificity of environmental influence: socioeconomic status affects early vocabulary development via maternal speech. *Child development*, *74*(5), 1368–1378 (cit. on p. 66).

Höhle, B., Bijeljac-Babic, R., Herold, B., Weissenborn, J., & Nazzi, T. (2009). Language specific prosodic preferences during the first half year of life: evidence from german and french infants. *Infant Behavior and Development*, *32*(3), 262–274 (cit. on p. 76).

Höhle, B., Fritzsche, T., Boll-Avetisyan, N., Hullebus, M., & Gafos, A. (2021). Respect the surroundings: effects of phonetic context variability on infants' learning of minimal pairs. *JASA Express Letters*, *1*(2) (cit. on pp. 81–83, 85, 164–166, 185).

Höhle, B., Fritzsche, T., Meß, K., Philipp, M., & Gafos, A. (2020). Only the right noise? effects of phonetic and visual input variability on 14-month-olds' minimal pair word learning. *Developmental Science*, *23*(5), e12950 (cit. on pp. 78, 80, 164, 187).

Holland, D., Chang, L., Ernst, T. M., Curran, M., Buchthal, S. D., Alicata, D., Skranes, J., Johansen, H., Hernandez, A., Yamakawa, R., et al. (2014). Structural growth trajectories and rates of change in the first 3 months of infant brain development. *JAMA neurology*, *71*(10), 1266–1274 (cit. on p. 104).

Hollich, G., Hirsh-Pasek, K., & Golinkoff, R. M. (2000a). I. what does it take to learn a word. *Monographs of the Society for Research in Child Development*, *65*(3), 1–16 (cit. on pp. 72, 77, 163).

Hollich, G., Hirsh-Pasek, K., & Golinkoff, R. M. (2000b). Ii. the emergentist coalition model. *Monographs of the Society for Research in Child Development*, *65*(3), 17–29 (cit. on pp. 74–76, 163, 189, 190).

Hornickel, J., Skoe, E., Nicol, T., Zecker, S., & Kraus, N. (2009). Subcortical differentiation of stop consonants relates to reading and speech-in-noise perception. *Proceedings of the National Academy of Sciences*, *106*(31), 13022–13027 (cit. on pp. 107, 138, 139).

Horst, J. S., Samuelson, L. K., Kucker, S. C., & McMurray, B. (2011). What's new? children prefer novelty in referent selection. *Cognition*, *118*(2), 234–244 (cit. on p. 186).

Houston-Price, C., & Nakai, S. (2004). Distinguishing novelty and familiarity effects in infant preference procedures. *Infant and Child Development: An International Journal of Research and Practice*, *13*(4), 341–348 (cit. on p. 186).

Hurtado, N., Marchman, V. A., & Fernald, A. (2008). Does input influence uptake? links between maternal talk, processing speed and vocabulary size in spanish-learning children. *Developmental science*, *11*(6), F31–F39 (cit. on p. 66).

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: relation to language input and gender. *Developmental psychology*, *27*(2), 236 (cit. on p. 66).

Ilari, B. S. (2021). On musical parenting of young children: musical beliefs and behaviors of mothers and infants. In *Music in the lives of young children* (pp. 17–32). Routledge. (Cit. on p. 62).

Imada, T., Hari, R., Loveless, N., McEvoy, L., & Sams, M. (1993). Determinants of the auditory mismatch response. *Electroencephalography and clinical Neurophysiology*, *87*(3), 144–153 (cit. on p. 52).

Imada, T., Zhang, Y., Cheour, M., Taulu, S., Ahonen, A., & Kuhl, P. K. (2006). Infant speech perception activates broca's area: a developmental magnetoencephalography study. *Neuroreport*, *17*(10), 957–962 (cit. on p. 47).

Intartaglia, B. (2016). *Effet de l'expertise et de l'attention sur l'encodage neuronal des sons de parole* [Doctoral dissertation, Aix-Marseille]. (Cit. on pp. 39, 56).

Javitt, D. C., Grochowski, S., Shelley, A.-M., & Ritter, W. (1998). Impaired mismatch negativity (mmn) generation in schizophrenia as a function of stimulus deviance, probability, and interstimulus/interdeviant interval. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *108*(2), 143–153 (cit. on p. 52).

Jeng, F.-C., Hu, J., Dickman, B., Montgomery-Reagan, K., Tong, M., Wu, G., & Lin, C.-D. (2011). Cross-linguistic comparison of frequency-following responses to voice pitch in american and chinese neonates and adults. *Ear and hearing*, *32*(6), 699–707 (cit. on p. 135).

Jeng, F.-C., Lin, C.-D., Chou, M.-S., Hollister, G. R., Sabol, J. T., Mayhugh, G. N., Wang, T.-C., & Wang, C.-Y. (2016). Development of Subcortical Pitch Representation in Three-Month-Old Chinese Infants. *Perceptual and Motor Skills*, *122*(1), 123–135. https://doi.org/10.1177/0031512516631054 (cit. on pp. 57, 105).

Jeng, F.-C., Schnabel, E. A., Dickman, B. M., Hu, J., Li, X., Lin, C.-D., & Chung, H.-K. (2010). Early Maturation of Frequency-Following Responses to Voice Pitch in Infants with Normal Hearing. *Perceptual and Motor Skills*, *111*(3), 765–784. https://doi.org/10.2466/10.22.24.PMS.111.6.765-784 (cit. on pp. 57, 105, 107, 134–136, 196).

Johnson, K. L., Nicol, T., Zecker, S. G., Bradlow, A. R., Skoe, E., & Kraus, N. (2008). Brainstem encoding of voiced consonant–vowel stop syllables. *Clinical Neurophysiology*, *119*(11), 2623–2635 (cit. on pp. 105, 138).

Johnson, M. H. (1998). The neural basis of cognitive development. *Handbook of child psychology: Vol. 2. Cognition, perception, and language*, 1–49 (cit. on p. 30).

Johnson, M. H., Dziurawiec, S., Ellis, H., & Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition*, *40*(1-2), 1–19 (cit. on p. 40).

Jolly, E. (2018). Pymer4: connecting r and python for linear mixed modeling. *Journal of Open Source Software*, *3*(31), 862 (cit. on p. 156).

Jusczyk, P. W., & Hohne, E. A. (1997). Infants' memory for spoken words. *Science*, *277*(5334), 1984–1986 (cit. on p. 78).

Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The beginnings of word segmentation in english-learning infants. *Cognitive psychology*, *39*(3-4), 159–207 (cit. on p. 188).

Kahng, J. (2018). The effect of pause location on perceived fluency. *Applied Psycholinguistics*, *39*(3), 569–591 (cit. on p. 69).

Kailaheimo-Lönnqvist, L., Virtala, P., Fandakova, Y., Partanen, E., Leppänen, P. H., Thiede, A., & Kujala, T. (2020). Infant event-related potentials to speech are associated with prelinguistic development. *Developmental cognitive neuroscience*, *45*, 100831 (cit. on pp. 54, 139, 141).

Kalashnikova, M., & Burnham, D. (2018). Infant-directed speech from seven to nineteen months has similar acoustic properties but different functions. *Journal of child language*, *45*(5), 1035–1053 (cit. on p. 61).

Kappenman, E. S., & Luck, S. J. (2011). Erp components: the ups and downs of brainwave recordings. *The Oxford handbook of event-related potential components*, 3–30 (cit. on p. 25).

Kartushina, N., & Mayor, J. (2019). Word knowledge in six-to nine-month-old norwegian infants? not without additional frequency cues. *Royal Society open science*, *6*(9), 180711 (cit. on p. 74).

Katz, N., Baker, E., & Macnamara, J. (1974). What's in a name? a study of how children learn common and proper names. *Child development*, 469–473 (cit. on p. 76).

Kekoni, J., Hämäläinen, H., Saarinen, M., Gröhn, J., Reinikainen, K., Lehtokoski, A., & Näätänen, R. (1997). Rate effect and mismatch responses in the somatosensory system: erp-recordings in humans. *Biological psychology, 46*(2), 125–142 (cit. on p. 53).

Keller, H., Hentschel, E., Yovsi, R. D., Lamm, B., Abels, M., & Haas, V. (2004a). The psycholinguistic embodiment of parental ethnotheories: A new avenue to understanding cultural processes in parental reasoning [Publisher: Sage Publications Sage CA: Thousand Oaks, CA]. *Culture & Psychology, 10*(3), 293–330 (cit. on pp. 62, 85).

Keller, H., Hentschel, E., Yovsi, R. D., Lamm, B., Abels, M., & Haas, V. (2004b). The psycho-linguistic embodiment of parental ethnotheories: a new avenue to understanding cultural processes in parental reasoning. *Culture & Psychology, 10*(3), 293–330 (cit. on p. 146).

Kelso, J. S., Tuller, B., Vatikiotis-Bateson, E., & Fowler, C. A. (1984). Functionally specific articulatory cooperation following jaw perturbations during speech: evidence for coordinative structures. *Journal of Experimental Psychology: Human Perception and Performance, 10*(6), 812 (cit. on p. 47).

Kemp, N., Scott, J., Bernhardt, B. M., Johnson, C. E., Siegel, L. S., & Werker, J. F. (2017). Minimal pair word learning and vocabulary size: links with later language skills. *Applied Psycholinguistics, 38*(2), 289–314 (cit. on pp. 79, 83, 85).

Klatt, D. H. (1980). Software for a cascade/parallel formant synthesizer. *the Journal of the Acoustical Society of America, 67*(3), 971–995 (cit. on p. 106).

Kleiner, K. A. (1987). Amplitude and phase spectra as indices of infants' pattern preferences. *Infant Behavior and Development, 10*(1), 49–59 (cit. on p. 40).

Knickmeyer, R. C., Gouttard, S., Kang, C., Evans, D., Wilber, K., Smith, J. K., Hamer, R. M., Lin, W., Gerig, G., & Gilmore, J. H. (2008). A structural mri study of human brain development from birth to 2 years. *Journal of neuroscience, 28*(47), 12176–12182 (cit. on p. 38).

Koenig, M., & Cole, C. (2013, March). 492 Early Word Learning. In *The Oxford Handbook of Cognitive Psychology*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195376746.013.0031 (cit. on p. 82).

Koerner, E. K. (1992). The sapir-whorf hypothesis: a preliminary history and a bibliographical essay. *Journal of Linguistic Anthropology, 2*(2), 173–198 (cit. on p. 70).

Kral, A., & Sharma, A. (2012). Developmental neuroplasticity after cochlear implantation. *Trends in Neurosciences*, *35*(2), 111–122 (cit. on p. 136).

Kraus, N., & Nicol, T. (2005). Brainstem origins for cortical input. *Nature Neuroscience*, *8*(10), 1300–1301 (cit. on p. 138).

Kraus, N., Anderson, S., & White-Schwoch, T. (2017). The Frequency-Following Response: A Window into Human Communication [Series Title: Springer Handbook of Auditory Research]. In N. Kraus, S. Anderson, T. White-Schwoch, R. R. Fay, & A. N. Popper (Eds.), *The Frequency-Following Response* (pp. 1–15, Vol. 61). Springer International Publishing. https://doi.org/10.1007/978-3-319-47944-6_1 (cit. on pp. 56, 105, 196).

Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature reviews neuroscience*, *11*(8), 599–605 (cit. on p. 105).

Krizman, J., & Kraus, N. (2019). Analyzing the FFR: A tutorial for decoding the richness of auditory function. *Hearing Research*, *382*, 107779. https://doi.org/10.1016/j.heares.2019.107779 (cit. on pp. 56, 110–112, 134).

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & psychophysics*, *50*(2), 93–107 (cit. on pp. 25, 49).

Kuhl, P. K. (1999). The role of experience in early language development: linguistic experience alters the perception and production of speech. *The role of early experience in infant development*, 101–125 (cit. on p. 40).

Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences*, *97*(22), 11850–11857 (cit. on pp. 49, 50, 141).

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nature Reviews Neuroscience*, *5*(11), 831–843. https://doi.org/10.1038/nrn1533 (cit. on pp. 40, 68, 92, 104, 136, 197, 198).

Kuhl, P. K. (2010). Brain mechanisms in early language acquisition. *Neuron*, *67*(5), 713–727 (cit. on p. 30).

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., Stolyarova, E. I., Sundberg, U., & Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science*, *277*(5326), 684–686 (cit. on pp. 60, 61).

Kuhl, P. K., Coffey-Corina, S., Padden, D., & Dawson, G. (2005). Links between social and linguistic processing of speech in preschool children with autism: behavioral and electrophysiological measures. *Developmental science*, *8*(1), F1–F12 (cit. on pp. 40, 92, 142).

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1493), 979–1000. https://doi.org/10.1098/rstb.2007.2154 (cit. on pp. 32, 40, 47, 51, 54, 140, 141, 164).

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic Experience Alters Phonetic Perception in Infants by 6 Months of Age. *Science, New Series*, *255*(5044), 606–608. http://www.jstor.org/stable/2876832 (cit. on pp. 41, 49, 140, 141).

Kuhlman, K. A., Burns, K. A., Depp, R., & Sabbagha, R. E. (1988). Ultrasonic imaging of normal fetal response to external vibratory acoustic stimulation. *American journal of obstetrics and gynecology*, *158*(1), 47–51 (cit. on p. 39).

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 (cit. on pp. 113, 123, 124, 138).

Kushnerenko, E., Ceponiene, R., Balan, P., Fellman, V., & Näätänen, R. (2002). Maturation of the auditory change detection response in infants: a longitudinal erp study. *Neuroreport*, *13*(15), 1843–1848 (cit. on p. 54).

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203–205. https://doi.org/10.1126/science.7350657 (cit. on p. 30).

Lavechin, M., Bousbib, R., Bredin, H., Dupoux, E., & Cristia, A. (2020). An open-source voice type classifier for child-centered daylong recordings. *Interspeech 2020 - Conference of the International Speech Communication Association*. https://hal.science/hal-02989487 (cit. on pp. 148, 149, 151, 154).

Lavechin, M., de Seyssel, M., Métais, M., Metze, F., Mohamed, A., Bredin, H., Dupoux, E., & Cristia, A. (2024). Modeling early phonetic acquisition from child-centered audio data. *Cognition*, *245*, 105734 (cit. on p. 161).

Leblanc, A. (2017). Le bébé, la télé, la tablette et le smartphone. *Enfances*, *74*(1), 6–10 (cit. on p. 61).

Lee, J. (2011). Size matters: early vocabulary as a predictor of language and literacy competence. *Applied Psycholinguistics*, *32*(1), 69–92 (cit. on p. 92).

Legacy, J., Zesiger, P., Friend, M., & Poulin-Dubois, D. (2018). Vocabulary size and speed of word recognition in very young french–english bilinguals: a longitudinal study. *Bilingualism: Language and Cognition*, *21*(1), 137–149 (cit. on p. 191).

Lehongre, K., Ramus, F., Villiermet, N., Schwartz, D., & Giraud, A.-L. (2011). Altered low-gamma sampling in auditory cortex accounts for the three main facets of dyslexia. *Neuron*, *72*(6), 1080–1090 (cit. on p. 37).

Lemos, F. A., da Silva Nunes, A. D., de Souza Evangelista, C. K., Escera, C., Taveira, K. V. M., & Balen, S. A. (2021). Frequency-following response in newborns and infants: a systematic review of acquisition parameters. *Journal of Speech, Language, and Hearing Research*, *64*(6), 2085–2102 (cit. on p. 57).

Leopold, W. (1939). *Speech development of a bilingual child. vol i: vocabulary growth in the first two years*. Northwestern University. (Cit. on p. 28).

Leppänen, P. H., Eklund, K. M., & Lyytinen, H. (1997). Event-related brain potentials to change in rapidly presented acoustic stimuli in newborns. *Developmental Neuropsychology*, *13*(2), 175–204 (cit. on p. 140).

Leppänen, P. H., Hämäläinen, J. A., Salminen, H. K., Eklund, K. M., Guttorm, T. K., Lohvansuu, K., Puolakanaho, A., & Lyytinen, H. (2010). Newborn brain event-related potentials revealing atypical processing of sound frequency and the subsequent association with later literacy skills in children with familial dyslexia. *Cortex*, *46*(10), 1362–1376 (cit. on p. 54).

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological review*, *74*(6), 431 (cit. on p. 46).

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*(1), 1–36 (cit. on pp. 46, 47).

Liberman, A. M., & Mattingly, I. G. (1989). A specialization for speech perception. *Science*, *243*(4890), 489–494 (cit. on p. 46).

Liebig, J., Friederici, A. D., Neef, N. E., Friederici, A., Emmrich, F., Brauer, J., Wilcke, A., Neef, N., Boltze, J., Skeide, M., et al. (2020). Auditory brainstem measures and genotyping boost the prediction of literacy: a longitudinal study on early markers of dyslexia. *Developmental Cognitive Neuroscience*, *46*, 100869 (cit. on p. 142).

Linebarger, D. L., & Vaala, S. E. (2010). Screen media and language development in infants and toddlers: an ecological perspective. *Developmental Review*, *30*(2), 176–202 (cit. on p. 61).

Linebarger, D. L., & Walker, D. (2005). Infants' and toddlers' television viewing and language outcomes. *American behavioral scientist*, *48*(5), 624–645 (cit. on p. 61).

Linnavalli, T., Putkinen, V., Huotilainen, M., & Tervaniemi, M. (2017). Phoneme processing skills are reflected in children's mmn responses. *Neuropsychologia*, *101*, 76–84 (cit. on pp. 54, 141).

Litovsky, R. (2015). Development of the auditory system. *Handbook of clinical neurology*, *129*, 55–72 (cit. on pp. 37, 38).

Liu, H.-M., Kuhl, P. K., & Tsao, F.-M. (2003). An association between mothers' speech clarity and infants' speech discrimination skills. *Developmental science*, *6*(3), F1–F10 (cit. on pp. 61, 139).

Liu, L., Peter, V., & Tyler, M. D. (2023). Understanding the neural mechanisms for infants' perception of native and non-native speech. *Brain and Language*, *242*, 105279. https://doi.org/10.1016/j.bandl.2023.105279 (cit. on pp. 49, 104, 105).

Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y., & Yamada, T. (1994). Training japanese listeners to identify english/r/and/l/. iii. long-term retention of new phonetic categories. *The Journal of the acoustical society of America*, *96*(4), 2076–2087 (cit. on p. 45).

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 (cit. on p. 37).

Lyytinen, H., Erskine, J., Ahonen, T., Aro, M., Eklund, K., Guttorm, T., Hintikka, S., Hämäläinen, J., Ketonen, R., & Laakso, M.-L. (2008). Early identification and prevention of dyslexia: results from a prospective follow-up study of children at familial risk for dyslexia. In *The sage handbook of dyslexia* (pp. 121–146). Sage. (Cit. on p. 200).

Ma, W., Golinkoff, R. M., Houston, D. M., & Hirsh-Pasek, K. (2011). Word learning in infant-and adult-directed speech. *Language Learning and Development*, *7*(3), 185–201 (cit. on pp. 168, 189).

MacBeth, T. M. (1996). *Tuning in to young viewers: social science perspectives on television*. Calif Sage. (Cit. on p. 61).

MacWhinney, B. (2000). The childes project. *Computational Linguistics*, *26*(4), 657–657 (cit. on pp. 148, 149, 162).

Mani, N., & Plunkett, K. (2008). Fourteen-month-olds pay attention to vowels in novel words. *Developmental science*, *11*(1), 53–59 (cit. on p. 84).

Marchman, V. A., & Fernald, A. (2008). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental science*, *11*(3), F9–F16 (cit. on p. 191).

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of eeg-and meg-data. *Journal of neuroscience methods*, *164*(1), 177–190 (cit. on p. 115).

Martin, A., Schatz, T., Versteegh, M., Miyazawa, K., Mazuka, R., Dupoux, E., & Cristia, A. (2015). Mothers speak less clearly to infants than to adults: a comprehensive test

of the hyperarticulation hypothesis. *Psychological science*, *26*(3), 341–347 (cit. on p. 61).

Mather, E., & Plunkett, K. (2011). Mutual exclusivity and phonological novelty constrain word learning at 16 months. *Journal of Child Language*, *38*(5), 933–950 (cit. on p. 186).

Mattingly, I. G., & Liberman, A. M. (1988). Specialized perceiving systems for speech and other biologically significant sounds. *Auditory function*, 775–793 (cit. on p. 46).

Maurer, D., & Werker, J. F. (2014). Perceptual narrowing during infancy: A comparison of language and faces. *Developmental Psychobiology*, *56*(2), 154–178. https://doi.org/10.1002/dev.21177 (cit. on pp. 44, 104).

McCarthy, D. (1946). Language development in children. In L. Carmichael (Ed.), *Manual of child psychology* (pp. 476–581). John Wiley & Sons. (Cit. on p. 29).

McMahon, E., Wintermark, P., & Lahav, A. (2012). Auditory brain development in premature infants: the importance of early experience. *Annals of the New York Academy of Sciences*, *1252*(1), 17–24 (cit. on p. 39).

McMurray, B., Horst, J. S., & Samuelson, L. K. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological review*, *119*(4), 831 (cit. on p. 74).

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, *29*(2), 143–178 (cit. on pp. 40, 188).

Mehr, S. A., Singh, M., Knox, D., Ketter, D. M., Pickens-Jones, D., Atwood, S., Lucas, C., Jacoby, N., Egner, A. A., Hopkins, E. J., et al. (2019). Universality and diversity in human song. *Science*, *366*(6468), eaax0868 (cit. on p. 62).

Meints, K., & Woodford, A. (2008). Lincoln infant lab package 1.0: a new programme package for ipl, preferential listening, habituation and eyetracking.[www document: computer software & manual]. *URL: http://www. lincoln. ac. uk/psychology/babylab. htm* (cit. on p. 84).

Mendoza, J. K., & Fausey, C. M. (2021). Everyday music in infancy. *Developmental Science*, *24*(6), e13122 (cit. on pp. 62, 65, 147).

Mento, G., Suppiej, A., Altoè, G., & Bisiacchi, P. S. (2010). Functional hemispheric asymmetries in humans: electrophysiological evidence from preterm infants. *European Journal of Neuroscience*, *31*(3), 565–574 (cit. on p. 54).

Meyer, C., Witte, J., Hildmann, A., Hennecke, K.-H., Schunck, K.-U., Maul, K., Franke, U., Fahnenstich, H., Rabe, H., Rossi, R., et al. (1999). Neonatal screening for hearing

disorders in infants at risk: incidence, risk factors, and follow-up. *Pediatrics*, *104*(4), 900–904 (cit. on p. 39).

Meylan, S. C., & Bergelson, E. (2022). Learning through processing: toward an integrated approach to early word learning. *Annual review of linguistics*, *8*, 77–99 (cit. on p. 76).

Miller, N. E., & Dollard, J. (1941). *Social learning and imitation*. Yale University Press. (Cit. on p. 28).

Miyawaki, K., Jenkins, J. J., Strange, W., Liberman, A. M., Verbrugge, R., & Fujimura, O. (1975). An effect of linguistic experience: the discrimination of [r] and [l] by native speakers of japanese and english. *Perception & Psychophysics*, *18*(5), 331–340 (cit. on p. 40).

Mollica, F., & Piantadosi, S. T. (2017). How data drive early word learning: a cross-linguistic waiting time analysis. *Open Mind*, *1*(2), 67–77 (cit. on p. 74).

Moon, C., Cooper, R. P., & Fifer, W. P. (1993). Two-day-olds prefer their native language. *Infant behavior and development*, *16*(4), 495–500 (cit. on p. 40).

Moore, J. K., & Linthicum, F. H. (2007). The human auditory system: a timeline of development. *International journal of audiology*, *46*(9), 460–478 (cit. on pp. 37, 38, 136, 196).

Moore, J. K., Perazzo, L. M., & Braun, A. (1995). Time course of axonal myelination in the human brainstem auditory pathway. *Hearing research*, *87*(1-2), 21–31 (cit. on pp. 38, 104).

Mueller, J. L., Friederici, A. D., & Männel, C. (2012). Auditory perception at the root of language learning. *Proceedings of the National Academy of Sciences*, *109*(39), 15953–15958 (cit. on p. 54).

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*, *118*(12), 2544–2590. https://doi.org/10.1016/j.clinph.2007.04.026 (cit. on pp. 49, 52, 53, 104, 112).

Narayan, C. R. (2019). An acoustic perspective on 45 years of infant speech perception, part 1: consonants. *Language and Linguistics Compass*, *13*(10), e12352 (cit. on p. 44).

Narayan, C. R., Werker, J. F., & Beddor, P. S. (2010). The interaction between acoustic salience and language experience in developmental speech perception: evidence from nasal place discrimination. *Developmental science*, *13*(3), 407–420 (cit. on pp. 81, 164, 166).

Navarrete-Arroyo, S., Virtala, P., Laasonen, M., & Kujala, T. (2024). Infant neural speech encoding is associated with pre-reading skill development. *Clinical Neurophysiology* (cit. on p. 200).

Nazzi, T., & Bertoncini, J. (2003). Before and after the vocabulary spurt: two modes of word acquisition? *Developmental Science*, *6*(2), 136–142 (cit. on pp. 77, 146).

Nazzi, T., & Bertoncini, J. (2009). Phonetic specificity in early lexical acquisition: new evidence from consonants in coda positions. *Language and Speech*, *52*(4), 463–480 (cit. on p. 164).

Nazzi, T., Iakimova, G., Bertoncini, J., Frédonie, S., & Alcantara, C. (2006). Early segmentation of fluent speech by infants acquiring french: emerging evidence for crosslinguistic differences. *Journal of Memory and Language*, *54*(3), 283–299 (cit. on p. 188).

Nazzi, T., Jusczyk, P. W., & Johnson, E. K. (2000). Language discrimination by english-learning 5-month-olds: effects of rhythm and familiarity. *Journal of Memory and Language*, *43*(1), 1–19 (cit. on p. 188).

Nazzi, T., Poltrock, S., & Von Holzen, K. (2016). The developmental origins of the consonant bias in lexical processing. *Current Directions in Psychological Science*, *25*(4), 291–296 (cit. on p. 137).

Nelson, K. (1973). Structure and strategy in learning to talk. *Monographs of the society for research in child development*, 1–135 (cit. on pp. 79, 165).

Newman, R. S., Rowe, M. L., & Ratner, N. B. (2016). Input and uptake at 7 months predicts toddler vocabulary: the role of child-directed speech and infant processing skills in language development. *Journal of child language*, *43*(5), 1158–1173 (cit. on p. 61).

Newport, E., Gleitman, H., & Gleitman, L. (1977). Mother, id rather do it myself: some effects and non-effects of maternal speech style. In C. E. Snow & C. A. Ferguson (Eds.), *Talking to children: language input and acquisition* (pp. 109–149). Cambridge University Press. (Cit. on p. 61).

Nikolaus, M., Agrawal, A., Kaklamanis, P., Warstadt, A., & Fourtassi, A. (2024). Automatic annotation of grammaticality in child-caregiver conversations. (Cit. on p. 162).

Nikolaus, M., Prévot, L., & Fourtassi, A. (2022). Communicative feedback as a mechanism supporting the production of intelligible speech in early childhood. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *44*(44) (cit. on pp. 157, 159, 160).

Nikolaus, M., Prévot, L., & Fourtassi, A. (2023). Communicative feedback in response to children's grammatical errors. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *45*(45) (cit. on pp. 161, 162).

Nishibayashi, L.-L., & Nazzi, T. (2016). Vowels, then consonants: early bias switch in recognizing segmented word forms. *Cognition*, *155*, 188–203 (cit. on p. 137).

Novitskiy, N., Maggu, A. R., Lai, C. M., Chan, P. H., Wong, K. H., Lam, H. S., Leung, T. Y., Leung, T. F., & Wong, P. C. (2022). Early development of neural speech encoding depends on age but not native language status: evidence from lexical tone. *Neurobiology of Language*, *3*(1), 67–86 (cit. on pp. 133, 136).

Nurmsoo, E., & Bloom, P. (2008). Preschoolers' perspective taking in word learning: do they blindly follow eye gaze? *Psychological Science*, *19*(3), 211–215 (cit. on p. 190).

Oakes, L. M. (2012). Advances in eye tracking in infancy research. *Infancy* (cit. on p. 85).

O'Callaghan, C. (2015). Speech perception. *Handbook of the philosophy of perception*, 475–494 (cit. on p. 37).

Oceák, A., Winkler, I., Sussman, E., & Alho, K. (2006). Loudness summation and the mismatch negativity event-related brain potential in humans. *Psychophysiology*, *43*(1), 13–20 (cit. on p. 54).

Oller, D. K., Niyogi, P., Gray, S., Richards, J. A., Gilkerson, J., Xu, D., Yapanel, U., & Warren, S. F. (2010a). Automated vocal analysis of naturalistic recordings from children with autism, language delay, and typical development [Publisher: National Acad Sciences]. *Proceedings of the National Academy of Sciences*, *107*(30), 13354–13359 (cit. on pp. 63, 152).

Oller, D. K., Niyogi, P., Gray, S., Richards, J. A., Gilkerson, J., Xu, D., Yapanel, U., & Warren, S. F. (2010b). Automated vocal analysis of naturalistic recordings from children with autism, language delay, and typical development. *Proceedings of the National Academy of Sciences*, *107*(30), 13354–13359 (cit. on pp. 65, 66, 147, 148, 159).

Orena, A. J., Byers-Heinlein, K., & Polka, L. (2019a). Reliability of the language environment analysis recording system in analyzing french–english bilingual speech. *Journal of Speech, Language, and Hearing Research*, *62*(7), 2491–2500 (cit. on pp. 65, 146, 147).

Orena, A. J., Byers-Heinlein, K., & Polka, L. (2019b). Reliability of the language environment analysis recording system in analyzing French–English bilingual speech [Publisher: ASHA]. *Journal of Speech, Language, and Hearing Research*, *62*(7), 2491–2500 (cit. on p. 63).

Oudesluys-Murphy, A., Van Straaten, H., Bholasingh, R., & Van Zanten, G. (1996). Neonatal hearing screening. *European journal of pediatrics*, *155*, 429–435 (cit. on p. 39).

Pae, S., Yoon, H., Seol, A., Gilkerson, J., Richards, J. A., Ma, L., & Topping, K. (2016a). Effects of feedback on parent–child language with infants and toddlers in korea. *First Language*, *36*(6), 549–569 (cit. on pp. 64, 65, 146–148).

Pae, S., Yoon, H., Seol, A., Gilkerson, J., Richards, J. A., Ma, L., & Topping, K. (2016b). Effects of feedback on parent–child language with infants and toddlers in Korea [Publisher: Sage Publications Sage UK: London, England]. *First Language*, *36*(6), 549–569 (cit. on pp. 63, 64).

Pang, E., Edmonds, G., Desjardins, R., Khan, S., Trainor, L., & Taylor, M. (1998). Mismatch negativity to speech stimuli in 8-month-old infants and adults. *International Journal of Psychophysiology*, *29*(2), 227–236. https://doi.org/10.1016/S0167-8760(98)00018-X (cit. on pp. 44, 54, 105).

Partanen, E., Pakarinen, S., Kujala, T., & Huotilainen, M. (2013). Infants' brain responses for speech sound changes in fast multifeature MMN paradigm. *Clinical Neurophysiology*, *124*(8), 1578–1585. https://doi.org/10.1016/j.clinph.2013.02.014 (cit. on pp. 54, 109).

Pasman, J. W., Rotteveel, J. J., de Graaf, R., Stegeman, D. F., & Visco, Y. M. (1992). The effect of preterm birth on brainstem, middle latency and cortical auditory evoked responses (bmc aers). *Early human development*, *31*(2), 113–129 (cit. on pp. 39, 104).

Pater, J., Stager, C., & Werker, J. (2004). The perceptual acquisition of phonological contrasts. *Language*, 384–402 (cit. on pp. 187, 188).

Pavlov, I. P. (1927). *Conditioned reflexes: an investigation of the physiological activity of the cerebral cortex*. Oxford University Press. (Cit. on p. 28).

Pena, M., Maki, A., Kovacić, D., Dehaene-Lambertz, G., Koizumi, H., Bouquet, F., & Mehler, J. (2003). Sounds and silence: an optical topography study of language recognition at birth. *Proceedings of the National Academy of Sciences*, *100*(20), 11702–11705 (cit. on p. 40).

Penman, R., Cross, T., Milgrom-Friedman, J., & Meares, R. (1983). Mothers' speech to prelingual infants: a pragmatic analysis. *Journal of Child Language*, *10*(1), 17–34 (cit. on p. 61).

Perrachione, T. K., Lee, J., Ha, L. Y., & Wong, P. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *The Journal of the Acoustical Society of America*, *130*(1), 461–472 (cit. on p. 45).

Petanjek, Z., Judaš, M., Šimić, G., Rokočević, Ž., Rašin, M. R., Uylings, H. B., & Kostović, I. (2011). Extraordinary neoteny of synaptic spines in the human prefrontal cortex. *Proceedings of the National Academy of Sciences*, *108*(32), 13281–13286 (cit. on p. 136).

Picton, T. W., & Taylor, M. J. (2007). Electrophysiological evaluation of human brain development. *Developmental neuropsychology*, *31*(3), 249–278 (cit. on p. 53).

Plessinger, M. A., & Woods, J. R. (1986). Fetal auditory brain stem response: external and intrauterine auditory stimulation. *Integrative and Comparative Physiology*, *250*(1), R137–R141 (cit. on p. 39).

Poeppel, D., Idsardi, W. J., & van Wassenhove, V. (2008). Speech perception at the interface of neurobiology and linguistics. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1493), 1071–1086 (cit. on p. 137).

Politimou, N., Stewart, L., Müllensiefen, D., & Franco, F. (2018). Music@ home: a novel instrument to assess the home musical environment in the early years. *PloS one*, *13*(4), e0193819 (cit. on p. 62).

Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human perception and performance*, *20*(2), 421 (cit. on pp. 41, 74, 104).

Ponton, C., Eggermont, J. J., Khosla, D., Kwong, B., & Don, M. (2002). Maturation of human central auditory system activity: separating auditory evoked potentials by dipole source modeling. *Clinical Neurophysiology*, *113*(3), 407–420 (cit. on pp. 37, 38).

Punamäki, R.-L., Diab, S. Y., Drosos, K., Qouta, S. R., & Vänskä, M. (2024). The role of acoustic features of maternal infant-directed singing in enhancing infant sensorimotor, language and socioemotional development. *Infant Behavior and Development*, *74*, 101908 (cit. on p. 61).

Puolakanaho, A., Ahonen, T., Aro, M., Eklund, K., Leppänen, P. H., Poikkeus, A.-M., Tolvanen, A., Torppa, M., & Lyytinen, H. (2007). Very early phonological and language skills: estimating individual risk of reading disability. *Journal of Child Psychology and Psychiatry*, *48*(9), 923–931 (cit. on p. 200).

Quine, W., & Van, O. (1960). *Word and object: an inquiry into the linguistic mechanisms of objective reference*. John Wiley. (Cit. on p. 69).

Quinn, P. C., & Eimas, P. D. (1996). Perceptual cues that permit categorical differentiation of animal species by infants. *Journal of experimental child psychology*, *63*(1), 189–211 (cit. on p. 70).

Quinn, P. C., Eimas, P. D., & Rosenkrantz, S. L. (1993). Evidence for representations of perceptually similar natural categories by 3-month-old and 4-month-old infants. *Perception*, *22*(4), 463–475 (cit. on p. 70).

Ramírez, N. F., & Hippe, D. S. (2024). Estimating infants' language exposure: a comparison of random and volume sampling from daylong recordings collected in a bilingual community. *Infant Behavior and Development*, *75*, 101943 (cit. on pp. 65, 147).

Räsänen, O., Seshadri, S., Lavechin, M., Cristia, A., & Casillas, M. (2021). Alice: an open-source tool for automatic measurement of phoneme, syllable, and word counts from child-centered daylong recordings. *Behavior Research Methods*, *53*, 818–835 (cit. on pp. 148, 149, 162).

Reznick, J. S., & Goldfield, B. A. (1992). Rapid change in lexical development in comprehension and production. *Developmental psychology*, *28*(3), 406 (cit. on p. 79).

Ribas-Prats, T., Almeida, L., Costa-Faidella, J., Plana, M., Corral, M., Gómez-Roig, M. D., & Escera, C. (2019). The frequency-following response (FFR) to speech stimuli: A normative dataset in healthy newborns. *Hearing Research*, *371*, 28–39. https://doi.org/10.1016/j.heares.2018.11.001 (cit. on pp. 57, 105, 109–111, 135).

Ribas-Prats, T., Arenillas-Alcón, S., Lip-Sosa, D. L., Costa-Faidella, J., Mazarico, E., Gómez-Roig, M. D., & Escera, C. (2022). Deficient neural encoding of speech sounds in term neonates born after fetal growth restriction. *Developmental Science*, *25*(3), e13189. https://doi.org/10.1111/desc.13189 (cit. on pp. 57, 105, 109–112, 116, 135).

Ribas-Prats, T., Cordero, G., Lip-Sosa, D. L., Arenillas-Alcón, S., Costa-Faidella, J., Gómez-Roig, M. D., & Escera, C. (2023). Developmental Trajectory of the Frequency-Following Response During the First 6 Months of Life. *Journal of Speech, Language, and Hearing Research*, *66*(12), 4785–4800. https://doi.org/10.1044/2023_JSLHR-23-00104 (cit. on pp. 57, 105, 109, 110, 134, 136, 196).

Richard, C., Neel, M. L., Jeanvoine, A., Connell, S. M., Gehred, A., & Maitre, N. L. (2020). Characteristics of the frequency-following response to speech in neonates and potential applicability in clinical practice: a systematic review. *Journal of Speech, Language, and Hearing Research*, *63*(5), 1618–1635 (cit. on p. 134).

Ridgeway, D., Waters, E., & Kuczaj, S. A. (1985). Acquisition of emotion-descriptive language: receptive and productive vocabulary norms for ages 18 months to 6 years. *Developmental Psychology*, *21*(5), 901 (cit. on p. 29).

Rindermann, H., & Baumeister, A. E. (2015). Parents' ses vs. parental educational behavior and children's development: a reanalysis of the hart and risley study. *Learning and individual differences*, *37*, 133–138 (cit. on p. 60).

Rivera-Gaxiola, M., Johnson, M. H., Csibra, G., & Karmiloff-Smith, A. (2000). Electrophysiological correlates of category goodness. *Behavioural Brain Research*, *112*(1-2), 1–11 (cit. on p. 45).

Rivera-Gaxiola, M., Silva-Pereyra, J., & Kuhl, P. K. (2005). Brain potentials to native and non-native speech contrasts in 7-and 11-month-old american infants. *Developmental science*, *8*(2), 162–172 (cit. on pp. 43–45, 54, 139, 141).

Rivera-Gaxiola, M., Csibra, G., Johnson, M. H., & Karmiloff-Smith, A. (2000). Electrophysiological correlates of cross-linguistic speech perception in native english speakers. *Behavioural Brain Research*, *111*(1-2), 13–23 (cit. on pp. 45, 104, 105, 196).

Robinson, T. N. (1999). Reducing children's television viewing to prevent obesity: a randomized controlled trial. *Jama*, *282*(16), 1561–1567 (cit. on p. 61).

Romeo, R. R., Leonard, J. A., Robinson, S. T., West, M. R., Mackey, A. P., Rowe, M. L., & Gabrieli, J. D. (2018). Beyond the 30-million-word gap: children's conversational exposure is associated with language-related brain function. *Psychological science*, *29*(5), 700–710 (cit. on p. 67).

Rost, G. C., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental science*, *12*(2), 339–349 (cit. on pp. 80, 164, 187).

Rost, G. C., & McMurray, B. (2010). Finding the signal by adding noise: the role of noncontrastive phonetic variability in early word learning. *Infancy*, *15*(6), 608–635 (cit. on pp. 82, 165, 188).

Rowe, M. L. (2008). Child-directed speech: relation to socioeconomic status, knowledge of child development and child vocabulary skill. *Journal of child language*, *35*(1), 185–205 (cit. on p. 66).

Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child development*, *83*(5), 1762–1774 (cit. on p. 66).

Rowe, M. L., & Snow, C. E. (2020). Analyzing input quality along three dimensions: interactive, linguistic, and conceptual. *Journal of child language*, *47*(1) (cit. on p. 60).

Saayman, G., Ames, E. W., & Moffett, A. (1964). Response to novelty as an indicator of visual discrimination in the human infant. *Journal of Experimental Child Psychology*, *1*(2), 189–198 (cit. on p. 74).

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928 (cit. on p. 76).

Sams, M., Alho, K., & Näätänen, R. (1984). Short-term habituation and dishabituation of the mismatch negativity of the erp. *Psychophysiology*, *21*(4), 434–441 (cit. on p. 52).

Samuelson, L. K. (2021). Toward a precision science of word learning: understanding individual vocabulary pathways. *Child Development Perspectives*, *15*(2), 117–124 (cit. on pp. 73, 74, 76, 163, 190).

Santolin, C., Russo, S., Calignano, G., Saffran, J. R., & Valenza, E. (2019). The role of prosody in infants' preference for speech: a comparison between speech and birdsong. *Infancy*, *24*(5), 827–833 (cit. on p. 76).

Schaadt, G., Werwach, A., Obrig, H., Friederici, A. D., & Männel, C. (2023). Maturation of consonant perception, but not vowel perception, predicts lexical skills at 12 months. *Child Development*, *94*(3). https://doi.org/10.1111/cdev.13892 (cit. on pp. 54, 92, 141, 197).

Schröger, E., & Winkler, I. (1995). Presentation rate and magnitude of stimulus deviance effects on human pre-attentive change detection. *Neuroscience letters*, *193*(3), 185–188 (cit. on p. 53).

Semenzin, C., Hamrick, L., Seidl, A., Kelleher, B., & Cristia, A. (2021). Towards large-scale data annotation of audio from wearables: validating zooniverse annotations of infant vocalization types. *2021 IEEE Spoken Language Technology Workshop (SLT)*, 1079–1085 (cit. on p. 150).

Shiga, T., Althen, H., Cornella, M., Zarnowiec, K., Yabe, H., & Escera, C. (2015). Deviance-related responses along the auditory hierarchy: combined ffr, mlr and mmn evidence. *PloS one*, *10*(9), e0136794 (cit. on p. 133).

Shneidman, L. A., & Goldin-Meadow, S. (2012). Language input and acquisition in a mayan village: how important is directed speech? *Developmental science*, *15*(5), 659–673 (cit. on p. 60).

Shukla, M., White, K. S., & Aslin, R. N. (2011). Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-mo-old infants. *Proceedings of the National Academy of Sciences*, *108*(15), 6038–6043 (cit. on p. 189).

Skinner, B. F. (1938). *The behavior of organisms: an experimental analysis*. Appleton-Century. (Cit. on p. 28).

Skinner, B. F. (1953). *Science and human behavior*. The Macmillan Company. (Cit. on p. 28).

Skinner, B. F. (1957). *Verbal behavior*. Appleton-Century-Crofts. (Cit. on p. 28).

Skoe, E., & Kraus, N. (2010). Auditory Brain Stem Response to Complex Sounds: A Tutorial. *Ear & Hearing*, *31*(3), 302–324. https://doi.org/10.1097/AUD.0b013e3181cdb272 (cit. on pp. 106, 109–111, 134).

Skoe, E., Krizman, J., Anderson, S., & Kraus, N. (2015). Stability and plasticity of auditory brainstem function across the lifespan. *Cerebral Cortex*, *25*(6), 1415–1426 (cit. on p. 105).

Sloutsky, V. M. (2003). The role of similarity in the development of categorization. *Trends in cognitive sciences*, *7*(6), 246–251 (cit. on p. 71).

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558–1568 (cit. on p. 190).

Snedeker, J., & Gleitman, L. R. (2004). Why it is hard to label our concepts. *Weaving a Lexicon*, 257–294 (cit. on p. 77).

Snowling, M. J., Gallagher, A., & Frith, U. (2003). Family risk of dyslexia is continuous: individual differences in the precursors of reading skill. *Child development*, *74*(2), 358–373 (cit. on p. 200).

Snowling, M. J., & Melby-Lervåg, M. (2016). Oral language deficits in familial dyslexia: a meta-analysis and review. *Psychological bulletin*, *142*(5), 498 (cit. on p. 200).

Soderstrom, M., Blossom, M., Foygel, R., & Morgan, J. L. (2008). Acoustical cues and grammatical units in speech to two preverbal infants. *Journal of Child Language*, *35*(4), 869–902 (cit. on p. 61).

Soderstrom, M., & Wittebolle, K. (2013). When do caregivers talk? the influences of activity and time of day on caregiver speech and child vocalizations in two childcare environments. *PloS one*, *8*(11), e80646 (cit. on p. 147).

Spence, M. J., & DeCasper, A. J. (1987). Prenatal experience with low-frequency maternal-voice sounds influence neonatal perception of maternal voice samples. *Infant Behavior and Development*, *10*(2), 133–142 (cit. on p. 40).

Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*, *388*(6640), 381–382 (cit. on pp. 80, 81, 84, 164, 165).

Stern, C., & Stern, W. (1907). Die kindersprache [children's speech]. *Leipzig: Barth* (cit. on p. 28).

Stone, W. L., & Yoder, P. J. (2001). Predicting spoken language level in children with autism spectrum disorders. *Autism*, *5*(4), 341–361 (cit. on p. 87).

Su, P., Kuan, C.-C., Kaga, K., Sano, M., & Mima, K. (2008). Myelination progression in language-correlated regions in brain of normal children determined by quantitative mri assessment. *International journal of pediatric otorhinolaryngology*, *72*(12), 1751–1763 (cit. on pp. 38, 104).

Swingley, D. (2011). The looking-while-listening procedure. *Research methods in child language: A practical guide*, 29–42 (cit. on pp. 83, 84, 165).

Swingley, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition*, *76*(2), 147–166 (cit. on pp. 165, 197).

Sy, Y., Havard, W., Lavechin, M., Dupoux, E., & Cristia, A. (2023). Measuring language development from child-centered recordings. *Interspeech 2023*, 4618–4622 (cit. on p. 148).

Takada, A. (2005). Early vocal communication and social institution: appellation and infant verse addressing among the central kalahari san. *Crossroads of Language, Interaction, and Culture*, *6*, 80–108 (cit. on p. 61).

Takegata, R., Paavilainen, P., Näätänen, R., & Winkler, I. (1999). Independent processing of changes in auditory single features and feature conjunctions in humans as indexed by the mismatch negativity. *Neuroscience Letters*, *266*(2), 109–112 (cit. on p. 53).

Teixidó, M., François, C., Bosch, L., & Männel, C. (2018). The role of prosody in early speech segmentation and word-referent mapping. *The development of prosody in first language acquisition*, *23*, 79 (cit. on pp. 76, 188, 190).

Templin, M. C. (1957). *Certain language skills in children; their development and interrelationships*. University of Minnesota Press. (Cit. on p. 29).

Thiessen, E. D. (2007). The effect of distributional information on children's use of phonemic contrasts. *Journal of Memory and Language*, *56*(1), 16–34 (cit. on pp. 82, 164).

Thorson, J. C. (2018). The role of prosody in early word learning. *The development of prosody in first language acquisition*, 59–77 (cit. on p. 189).

Tinbergen, N. (1963). On aims and methods of ethology. *Zeitschrift für Tierpsychologie*, *20*(4), 410–433. https://doi.org/10.1111/j.1439-0310.1963.tb01161.x (cit. on p. 31).

Tincoff, R., & Jusczyk, P. W. (1999). Some beginnings of word comprehension in 6-month-olds. *Psychological science*, *10*(2), 172–175 (cit. on pp. 77, 163).

Topping, K., Dekhinet, R., & Zeedyk, S. (2011). Hindrances for parents in enhancing child language. *Educational Psychology Review*, *23*, 413–455 (cit. on p. 67).

Trainor, L., McFadden, M., Hodgson, L., Darragh, L., Barlow, J., Matsos, L., & Sonnadara, R. (2003). Changes in auditory cortex and the development of mismatch negativity between 2 and 6 months of age. *International Journal of Psychophysiology*, *51*(1), 5–15 (cit. on pp. 54, 104, 140).

Trainor, L. J. (2012). Musical experience, plasticity, and maturation: issues in measuring developmental change using eeg and meg. *Annals of the New York Academy of Sciences*, *1252*(1), 25–36 (cit. on p. 104).

Trehub, S. E. (1976). The discrimination of foreign speech contrasts by infants and adults. *Child development*, 466–472 (cit. on p. 43).

Trevarthen, C. (1977). Descriptive analyses of infant communicative behaviour. *Studies in mother-infant interaction* (cit. on p. 59).

Tsao, F.-M., Liu, H.-M., & Kuhl, P. K. (2004). Speech perception in infancy predicts language development in the second year of life: a longitudinal study. *Child development*, *75*(4), 1067–1084 (cit. on p. 32).

Tsui, A. S. M., Byers-Heinlein, K., & Fennell, C. T. (2019). Associative word learning in infancy: a meta-analysis of the switch task. *Developmental psychology*, *55*(5), 934 (cit. on pp. 77, 164).

Tsushima, T., Takizawa, O., Sasaki, M., Shiraki, S., Nishi, K., Kohno, M., Menyuk, P., & Best, C. (1994). Discrimination of English /r-l/ and /w-y/ by Japanese infants at 6-12 months: language-specific developmental changes in speech perception abilities. *Proc. 3rd International Conference on Spoken Language Processing (ICSLP 1994)*, 1695–1698. https://doi.org/10.21437/ICSLP.1994-438 (cit. on pp. 43, 104).

Tummeltshammer, K. S., Mareschal, D., & Kirkham, N. Z. (2014). Infants' selective attention to reliable visual cues in the presence of salient distractors. *Child Development*, *85*(5), 1981–1994 (cit. on p. 74).

Tyler, M. D., Best, C. T., Faber, A., & Levitt, A. G. (2014). Perceptual assimilation and discrimination of non-native vowel contrasts. *Phonetica*, *71*(1), 4–21 (cit. on p. 49).

Tyrrell, D. J., Stauffer, L. B., & Snowman, L. G. (1991). Perception of abstract identity/difference relationships by infants. *Infant Behavior and Development*, *14*(1), 125–129 (cit. on p. 75).

Valenza, E., Simion, F., Cassia, V. M., & Umiltà, C. (1996). Face preference at birth. *Journal of experimental psychology: Human Perception and Performance*, *22*(4), 892 (cit. on p. 40).

van der Leij, A., Van Bergen, E., van Zuijen, T., De Jong, P., Maurits, N., & Maassen, B. (2013). Precursors of developmental dyslexia: an overview of the longitudinal dutch dyslexia programme study. *Dyslexia*, *19*(4), 191–213 (cit. on pp. 54, 141).

Van Dyke, K. B., Lieberman, R., Presacco, A., & Anderson, S. (2017). Development of phase locking and frequency representation in the infant frequency-following response. *Journal of Speech, Language, and Hearing Research*, *60*(9), 2740–2751 (cit. on p. 136).

Virtala, P., Putkinen, V., Kailaheimo-Lönnqvist, L., Thiede, A., Partanen, E., & Kujala, T. (2022). Infancy and early childhood maturation of neural auditory change detection and its associations to familial dyslexia risk. *Clinical Neurophysiology*, *137*, 159–176. https://doi.org/10.1016/j.clinph.2022.03.005 (cit. on pp. 53, 54, 104).

Vlach, H. A., & Sandhofer, C. M. (2011). Developmental differences in children's context-dependent word learning. *Journal of Experimental Child Psychology*, *108*(2), 394–401 (cit. on p. 190).

Von Holzen, K., & Nazzi, T. (2020). Emergence of a consonant bias during the first year of life: new evidence from own-name recognition. *Infancy*, *25*(3), 319–346 (cit. on p. 137).

Von Holzen, K., Nishibayashi, L.-L., & Nazzi, T. (2018). Consonant and vowel processing in word form segmentation: an infant erp study. *Brain Sciences*, *8*(2), 24 (cit. on pp. 54, 141).

Vouloumanos, A., Hauser, M. D., Werker, J. F., & Martin, A. (2010). The tuning of human neonates' preference for speech. *Child development*, *81*(2), 517–527 (cit. on p. 40).

Vouloumanos, A., & Werker, J. F. (2007). Listening to language at birth: evidence for a bias for speech in neonates. *Developmental science*, *10*(2), 159–164 (cit. on pp. 40, 104).

Vouloumanos, A., & Werker, J. F. (2009). Infants' learning of novel words in a stochastic environment. *Developmental psychology*, *45*(6), 1611 (cit. on p. 190).

Wagner, L., Ladek, A. S., Plontke, S. K., & Rahne, T. (2023). Electrically evoked mismatch negativity responses to loudness and pitch cues in cochlear implant users. *Scientific Reports*, *13*(1), 2413 (cit. on p. 54).

Warlaumont, A. S., Sobowale, K., & Fausey, C. M. (2022). Daylong mobile audio recordings reveal multitimescale dynamics in infants' vocal productions and auditory experiences. *Current directions in psychological science*, *31*(1), 12–19 (cit. on pp. 65, 146, 147).

Warlaumont, A. S., Warlaumont, A. S., Richards, J. A., Richards, J. A., Gilkerson, J., Gilkerson, J., Oller, D. K., & Oller, D. K. (2014a). A social feedback loop for speech development and its reduction in autism. *Psychological Science*. https://doi.org/10.1177/0956797614531023 (cit. on pp. 64, 65, 146–148, 156, 157, 159–161).

Warlaumont, A. S., Warlaumont, A. S., Richards, J. A., Richards, J. A., Gilkerson, J., Gilkerson, J., Oller, D. K., & Oller, D. K. (2014b). A Social Feedback Loop for Speech Development and Its Reduction in Autism. *Psychological Science*. https://doi.org/10.1177/0956797614531023 (cit. on pp. 66, 94, 196).

Warren, S. F., Gilkerson, J., Richards, J. A., Oller, D. K., Xu, D., Yapanel, U., & Gray, S. (2010). What automated vocal analysis reveals about the vocal production and language learning environment of young children with autism. *Journal of autism and developmental disorders*, *40*, 555–569 (cit. on pp. 64, 65, 146, 148).

Watson, J. B. (1913). Psychology as the behaviorist views it. *Psychological Review*, *20*(2), 158–177. https://doi.org/10.1037/h0074428 (cit. on p. 28).

Waxman, S. R., & Markow, D. B. (1995). Words as invitations to form categories: evidence from 12-to 13-month-old infants. *Cognitive psychology*, *29*(3), 257–302 (cit. on p. 76).

Weikum, W. M., Oberlander, T. F., Hensch, T. K., & Werker, J. F. (2012). Prenatal exposure to antidepressants and depressed maternal mood alter trajectory of infant speech perception. *Proceedings of the National Academy of Sciences*, *109*(supplement_2), 17221–17227 (cit. on pp. 43, 104).

Weisleder, A., & Fernald, A. (2013a). Talking to children matters: Early language experience strengthens processing and builds vocabulary [Publisher: Sage Publications Sage CA: Los Angeles, CA]. *Psychological science*, *24*(11), 2143–2152 (cit. on p. 60).

Weisleder, A., & Fernald, A. (2013b). Talking to children matters: early language experience strengthens processing and builds vocabulary. *Psychological science*, *24*(11), 2143–2152 (cit. on pp. 66, 146–148).

Werker, J. F., Cohen, L. B., Lloyd, V. L., Casasola, M., & Stager, C. L. (1998). Acquisition of word–object associations by 14-month-old infants. *Developmental psychology*, *34*(6), 1289 (cit. on pp. 78, 80, 84, 164, 185, 190).

Werker, J. F., Fennell, C. T., Corcoran, K. M., & Stager, C. L. (2002). Infants' ability to learn phonetically similar words: effects of age and vocabulary size. *Infancy*, *3*(1), 1–30 (cit. on pp. 79, 80, 164, 167, 188, 191, 192).

Werker, J. F., Gilbert, J. H., Humphrey, K., & Tees, R. C. (1981). Developmental aspects of cross-language speech perception. *Child development*, 349–355 (cit. on pp. 43, 113).

Werker, J. F., & Hensch, T. K. (2015). Critical periods in speech perception: new directions. *Annual review of psychology*, *66*(1), 173–196 (cit. on pp. 38, 92, 104).

Werker, J. F., & Tees, R. C. (1984a). Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant behavior and development*, *7*(1), 49–63 (cit. on pp. 146, 164).

Werker, J. F., & Tees, R. C. (1984b). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*(1), 49–63. https://doi.org/10.1016/s0163-6383(84)80022-3 (cit. on pp. 40, 43, 104, 113, 138–140).

Werker, J. F., & Tees, R. C. (2005). Speech perception as a window for understanding plasticity and commitment in language systems of the brain. *Developmental Psychobiology*, *46*(3), 233–251 (cit. on p. 137).

Werner, L. A., & Kitzes, L. M. (2011). Auditory development: a path from child to adult hearing. *Developmental Cognitive Neuroscience*, *1*(4), 371–385 (cit. on p. 136).

Werwach, A., Männel, C., Obrig, H., Friederici, A. D., & Schaadt, G. (2022). Longitudinal trajectories of electrophysiological mismatch responses in infant speech discrimination differ across speech features. *Developmental Cognitive Neuroscience*, *56*, 101127. https://doi.org/10.1016/j.dcn.2022.101127 (cit. on pp. 54, 55, 113, 115, 138, 140, 196).

White-Schwoch, T., Woodruff Carr, K., Thompson, E. C., Anderson, S., Nicol, T., Bradlow, A. R., Zecker, S. G., & Kraus, N. (2015). Auditory processing in noise: a preschool biomarker for literacy. *PLoS biology*, *13*(7), e1002196 (cit. on pp. 138, 142).

Wong, P., & Ng, K. W. S. (2018). Testing the hyperarticulation and prosodic hypotheses of child-directed speech: insights from the perceptual and acoustic characteristics of child-directed cantonese tones. *Journal of Speech, Language, and Hearing Research*, *61*(8), 1907–1925 (cit. on p. 61).

Woods, J. R., Plessinger, M. A., & Mack, C. E. (1984). Fetal auditory brainstem evoked response (abr). *Pediatric research*, *18*(1), 83–85 (cit. on p. 39).

Woodward, J. Z., & Aslin, R. N. (1990). Segmentation cues in maternal speech to infants. *7th biennial meeting of the International Conference on Infant Studies, Montreal, Quebec, Canada* (cit. on p. 69).

Xu, D., Yapanel, U., & Gray, S. (2009a). Reliability of the LENA Language Environment Analysis System in young children's natural home environment. *Boulder, CO: Lena Foundation*, 1–16 (cit. on p. 63).

Xu, D., Yapanel, U., & Gray, S. (2009b). Reliability of the lena language environment analysis system in young children's natural home environment. *Boulder, CO: Lena Foundation*, 1–16 (cit. on p. 148).

Xu, D., Yapanel, U., Gray, S., & Baer, C. T. (2008a). The lena language environment analysis system: the interpreted time segments (its) file. *Boulder, CO: Lena Foundation*, 1–7 (cit. on p. 148).

Xu, D., Yapanel, U., Gray, S., & Baer, C. T. (2008b). The LENA Language Environment Analysis System: The Interpreted Time Segments (ITS) File. *Boulder, CO: Lena Foundation*, 1–7 (cit. on p. 151).

Xu, D., Yapanel, U., Gray, S., Gilkerson, J., Richards, J., & Hansen, J. (2008). Signal processing for young child speech language development. *First Workshop on Child, Computer and Interaction* (cit. on pp. 146, 148).

Yoshida, K. A., Fennell, C. T., Swingley, D., & Werker, J. F. (2009). Fourteen-month-old infants learn similar-sounding words. *Developmental science*, *12*(3), 412–418 (cit. on pp. 82, 84, 165, 169, 172, 184, 187, 188).

Young, S. (2008). Lullaby light shows: everyday musical experience among under-two-year-olds. *International Journal of Music Education*, *26*(1), 33–46 (cit. on p. 62).

Yu, C., & Smith, L. B. (2011). What you learn is what you see: using eye movements to study infant cross-situational word learning. *Developmental science*, *14*(2), 165–180 (cit. on p. 190).

Yu, C., Zhong, Y., & Fricker, D. (2012). Selective attention in cross-situational statistical learning: evidence from eye tracking. *Frontiers in psychology*, *3*, 148 (cit. on p. 190).

Zamuner, T. S. (2006). Sensitivity to word-final phonotactics in 9-to 16-month-old infants. *Infancy*, *10*(1), 77–95 (cit. on pp. 81, 188).

Zimmerman, F. J., & Christakis, D. A. (2005). Children's television viewing and cognitive outcomes: a longitudinal analysis of national data. *Archives of pediatrics & adolescent medicine*, *159*(7), 619–625 (cit. on p. 61).

Zimmerman, F. J., Gilkerson, J., Richards, J. A., Christakis, D. A., Xu, D., Gray, S., & Yapanel, U. (2009). Teaching by listening: the importance of adult-child conversations to language development. *Pediatrics*, *124*(1), 342–349 (cit. on pp. 66, 67, 147, 148).

ANNEXES

A. Recruitment documents

In this Annex are shown the documents given to parents during recruitment:

1. The information notice that described the entire project, and participants' rights.
2. The consent form.
3. The flyer with contact information and a QR code to an explanatory video.

NOTICE D'INFORMATION

De la perception de la parole à l'acquisition du vocabulaire chez le bébé

Promoteur de la recherche : *CNRS*

Investigateur principal : *Clément François, Dr. en Neurosciences*

Madame, Monsieur,

Nous vous proposons de faire participer l'enfant pour lequel vous disposez de l'autorité parentale à une recherche impliquant la personne humaine relevant du *2°* de l'article L 1121-1 du code de la santé publique, intitulée *De la perception de la parole à l'acquisition du vocabulaire chez le bébé*.

Conformément aux dispositions du titre II de la première partie du code de la santé publique relatif aux recherches impliquant la personne humaine, cette recherche a obtenu un avis favorable du Comité de Protection des Personnes SUD-EST VI le 6 Janvier 2023.

Le CNRS, promoteur de cette recherche est représenté par André LE BIVIC, directeur de l'institut des sciences biologiques, 3 rue Michel-Ange, 75794 - PARIS Cedex 16. Il a souscrit une assurance de responsabilité civile auprès de la société HDI GLOBAL SE (contrat N°0100630314037).

Il est important que vous lisiez attentivement la présente notice d'information qui vous apportera les informations nécessaires concernant les différents aspects de la participation à cette recherche. N'hésitez pas à poser toutes les questions que vous jugerez utiles à *Clément François*, investigateur principal de cette recherche.

Le choix d'accepter la participation de l'enfant pour lequel vous disposez de l'autorité parentale est entièrement volontaire.

Quel est l'objectif de cette recherche?

L'objectif principal de l'étude LANGDEV est de mieux comprendre les mécanismes cérébraux impliqués dans le développement du langage chez les enfants entre 3 et 24 mois grâce à un suivi longitudinal. Le neurodéveloppement est l'ensemble des mécanismes qui, dès le plus jeune âge structurent la mise en place des réseaux du cerveau impliqués dans la motricité, la vision, l'audition, le langage ou les interactions sociales.

Ainsi, nous cherchons à comprendre le lien entre le traitement cérébral des syllabes et les capacités d'apprentissage de nouveaux mots au cours des 2 premières années de vie. Nous cherchons d'autre part à comprendre la relation entre l'environnement linguistique de l'enfant (comme les mesures de la voix, des mots utilisés par l'adulte interagissant avec l'enfant) et le niveau langagier à 24 mois.

Quels sont les bénéfices et les risques associés à cette recherche ?

Bénéfices attendus : Il n'y a aucun bénéfice individuel direct à la participation à la recherche. Il s'agit de bénéfices attendus en termes de développement des connaissances biologiques ou médicales.

Risques liés à la participation de votre enfant :

• L'électro-encéphalographie (EEG) permet de mesurer l'activité du cerveau en réponse à des stimulations sensorielles comme des sons ou des images. L'EEG réalisée dans le cadre de cette recherche nécessite seulement la pose d'un bonnet adapté au périmètre crânien de votre enfant ainsi qu'un gel électroconducteur. L'EEG ne représente aucun risque pour votre enfant. Aucun moyen de fixation des électrodes n'est nécessaire. Toutes les précautions seront prises afin de s'assurer que votre bébé ne ressente aucun inconfort. Votre enfant pourra être assis dans une chaise haute ou sur les genoux d'un membre du personnel. Des écouteurs intra-auriculaires seront utilisés pour la présentation des stimuli auditifs. Il est toutefois possible que le temps de mesure (environ 45 minutes incluant la pose du bonnet) soit perçu comme contraignant et puisse induire une gêne pendant l'expérimentation. Même si très peu probable, il est également possible que

le gel électroconducteur ne soit pas bien toléré par certains participants provoquant une légère dermatite chez les peaux très sensibles. Cet examen n'est pas réalisé dans un but de diagnostic médical. Il peut toutefois arriver que soient détectées des anomalies du traitement de l'information auditive, dans ce cas vous serez informés et conseillés par l'investigateur principal. Ces résultats seront transmis au médecin de votre choix. Ce médecin ou l'investigateur pourra décider d'interrompre la participation de votre enfant à la recherche. Les titulaires de l'autorité parentale, qui ne souhaitent pas être informés des résultats de l'EEG ne pourront pas participer ou accepter que leur enfant participe à cette recherche.

• Des tâches comportementales reposant sur l'analyse du regard de l'enfant permettent de mesurer les capacités d'apprentissage de nouveaux mots (association mot-image). Aucun risque n'est attendu pour la réalisation des tâches. Il est toutefois possible que le temps de mesure (15 min) soit vu comme un inconvénient et puisse entraîner une fatigue du participant.

• Concernant l'évaluation neurodeveloppementale, nous n'anticipons aucun risque relié à cette évaluation. Cet examen n'est pas réalisé dans un but de diagnostic médical. Il peut toutefois arriver que soient détectées des anomalies du neurodéveloppement. Dans ce cas, vous serez informés et conseillés par l'investigateur principal afin d'être orientés avec votre enfant vers un orthophoniste travaillant en libéral ou vers une structure rééducative pour une prise en charge adaptée. Les titulaires de l'autorité parentale qui ne souhaitent pas être informés des résultats de cette évaluation ne pourront pas accepter que leur enfant participe à cette recherche.

• Concernant les enregistrements acoustiques : vous réaliserez les enregistrements de l'environnement acoustiques de votre enfant. Pour cela, vous utiliserez le dispositif LENA, attaché aux vêtements de votre enfant, que vous activerez vous-même, un jour par semaine pendant 16 heures maximum (capacité maximum du dispositif). Vous vous verrez expliquer comment l'activer/désactiver afin d'utiliser votre droit de retrait et vous permettre de l'activer uniquement pendant les plages horaires que vous désirez. En outre, vous pourrez également spécifier les plages horaires que vous ne voulez pas voir incluses dans les données et/ou effacées une fois le dispositif rendu aux chercheurs. Vous pourrez également spécifier que vous ne voulez pas voir inclus la totalité d'un enregistrement. Vous retournerez le dispositif aux chercheurs par l'intermédiaire d'une corbeille maintenue dans une armoire fermée à clé dans le bureau de la direction de la crèche après chaque enregistrement. Chaque semaine, un des chercheurs participant au projet passera dans les crèches afin de récupérer les enregistreurs, les vider de leurs données, les charger avant de les disposer dans une deuxième corbeille pour vous l'être restitué. Nous n'anticipons donc aucun risque relié à la réalisation de ces enregistrements. Il est toutefois possible que le temps de mesure (16h) soit vu comme un inconvénient. L'enregistrement audio d'un tiers non consentant constitue une atteinte au respect de sa vie privée conformément à l'article 9 du code civil. Il vous est donc demandé de procéder aux enregistrements uniquement au sein de votre domicile et d'en informer préalablement les éventuels visiteurs (notamment par un panneau sur la porte d'entrée « enregistrement en cours »).

<u>Comment va se dérouler cette recherche?</u>

Il vous a été délivré la présente notice d'information. Après avoir lu et compris les informations décrites, vous avez disposé d'un délai de réflexion de 15 jours mininum entre la réception du document et la visite d'inclusion. Lors de la visite d'inclusion (V0, Figure 1), les critères d'inclusion ont été vérifiés par Clément François.

Si vous acceptez que votre enfant participe à cette étude longitudinale, vous signerez le formulaire de consentement avec le chercheur en charge de l'étude, Clément François.

Chaque bébé sera recruté en amont de ses 3 mois et participera à l'étude au travers de six visites réalisées à 3, 6, 8, 10, 18 et 24 mois. Parallèlement à ces visites, les titulaires de l'autorité parentale réaliseront des enregistrements de l'environnement acoustique de leur enfant de 3 à 24 mois à raison d'un enregistrement par semaine d'une durée de 16 heures maximum (voir Figure 1).

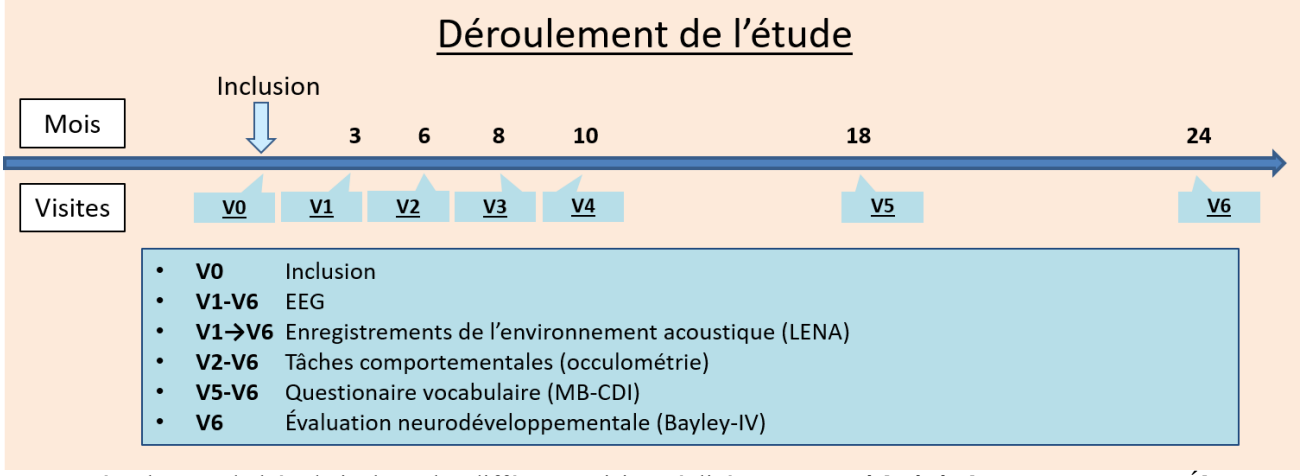


Figure 1: Déroulement de l'étude incluant les différentes visites réalisées et actes réalisés à chaque visite. EEG : Électro-encéphalographie. V : Visite

La 1ère visite sera réalisée à 3 mois (V1, +/- 15 jours). Cette visite aura lieu soit dans les locaux de la plateforme expérimentale dédiée à l'étude du développement de la maison de la recherche, soit dans les locaux de la plateforme expérimentale dédiée à l'étude du développement du Laboratoire de Neurosciences Cognitives, soit au sein même de la crèche Babilou dans laquelle votre enfant est inscrit.

Au cours de cette première visite, un EEG sera réalisé grâce à l'utilisation d'un bonnet de taille adaptée au périmètre crânien de votre enfant (voir Figure 2). Au total, 20 électrodes pourront être apposées sur le bonnet. Cet EEG ne nécessite pas de préparation particulière. La durée de cette visite est de 45 minutes environ, incluant le temps de pose du bonnet et la réalisation de l'EEG à proprement parlé. Votre enfant sera éveillé, assis sur une chaise, dans un siège incliné, assis sur les genoux d'un membre du personnel.

Des stimulations auditives seront réalisées : ces dernières ne représentent aucun risque pour votre enfant. Des écouteurs intra-auriculaires adaptés à l'oreille des enfants seront utilisés.

Cet EEG permettra de mesurer les temps de réponses du cerveau aux stimuli auditifs.

Par la suite, 5 autres visites seront planifiées et seront réalisées soit dans les locaux de la plateforme expérimentale dédiée à l'étude du développement de la maison de la recherche (Université Aix-Marseille), soit dans les locaux de la plateforme expérimentale dédiée à l'étude du développement du laboratoire de Neurosciences Cognitives (Université Aix-Marseille), soit au sein même de la crèche Babilou dans laquelle votre enfant est inscrit.

Au cours des visites 2 à 6:

- Un enregistrement EEG similaire à celui réalisé lors de la première visite (V1) sera effectué.



Figure 2 : Illustration du matériel EEG utilisé dans cette étude (BIOSEMI ActiveTwo)

- Des tâches expérimentales d'apprentissage de nouveaux mots reposant sur la présentation de sons et d'images seront réalisées afin de mesurer les temps et l'orientation du regard du bébé vers les images grâce à un système

d'occulométrie (Figure 3). Cette mesure dure approximativement 15 minutes. Dans le cas où cette mesure serait trop longue à mettre en œuvre pour vous et votre enfant, elle sera stoppée. Les mesures seront soit effectuées en temps réel via l'utilisation d'un occulomètre, soit effectuées après coup uniquement en cas de défaillance du matériel d'occulométrie, sur la base d'un enregistrement vidéo du visage du bébé pendant le test. Conformément à la méthodologie de référence MR-001, ces vidéos seront d'emblée pseudonymisées. Avant stockage, les visages apparaissant sur la vidéo seront masqués grâce à l'utilisation de rectangles noirs pour masquer les zones allant du nez de l'enfant à la partie inférieure de son visage et de la base des cheveux jusqu'aux yeux. Seuls seront conservés les yeux afin de réaliser la mesure. Les données vidéos brutes seront détruites dans le cas où un contrôle manuel n'est pas nécessaire.



Figure 3 : Exemple du matériel d'occulométrie utilisé pendant la tâche comportemental d'association mot-image.

- Une évaluation du vocabulaire de votre enfant sera réalisée au moyen du questionnaire MacArthur-Bates Communicative Development Inventories (MB-CDI) à 18 et 24 mois (V5 et V6). Vous remplirez vous-même ce questionnaire dans la salle d'attente, dans la crèche ou à votre domicile. La durée approximative de remplissage de ce questionnaire est de 20 minutes environ.

- Une évaluation neurodéveloppementale sera réalisée à 24 mois (V6) à l'aide de l'échelle de Bayley-IV (BSID-IV) par le chercheur ou un membre de son équipe. Cette évaluation vise à mesurer les capacités de votre enfant au niveau moteur, cognitif et langagier. La durée d'évaluation est d'une heure et demi.

- Le recueil des données d'environnement acoustique de votre enfant sera réalisé de 3 à 24 mois, une fois par semaine sur une durée de 16h maximum. Cet enregistrement sera effectué par vos soins à l'aide du dispositif LENA (Figure 4). Spécifiquement, votre enfant sera invité à porter un dispositif LENA pendant une journée par semaine. L'appareil enregistre des données audio pendant une durée maximale de 16 heures et s'arrête automatiquement une fois sa mémoire pleine ou sa batterie déchargée. Vous vous verrez expliquer comment l'activer/désactiver afin d'utiliser votre droit de retrait. En outre, vous pourrez spécifier les plages horaires que vous ne voulez pas voir incluses dans les données et/ou effacées. L'enregistrement audio d'un tiers non consentant constitue une atteinte au respect de sa vie privée conformément à l'article 9 du code civil. Il vous est donc demandé de n'effectuer les enregistrements qu'à votre domicile et de mettre un panneau sur la porte d'entrée indiquant « enregistrement en cours » pour avertir les éventuels visiteurs.

Vous retournerez le dispositif aux chercheurs chaque semaine par l'intermédiaire d'une corbeille maintenue dans une armoire fermée à clé située dans le bureau de la direction de la crèche dans laquelle votre enfant est inscrit. Chaque semaine, un des chercheurs participant au projet passera dans les crèches afin de récupérer l'enregistreur, le vider de ses données, le charger avant de le disposer dans une deuxième corbeille pour retour. Le dispositif ne sera jamais restitué par courrier postal ni par vous ni par les chercheurs. Les données brutes sont accessibles uniquement aux chercheurs participant au projet de recherche et ayant validé la formation aux bonnes pratiques de recherche. L'accès aux données brutes ne peut se faire uniquement qu'au travers du programme LENAHub (connexion sécurisée par mot de passe et nom d'utilisateur) installé sur un ordinateur portable chiffré. Dès l'enregistrement, les données brutes sont systématiquement et directement cryptées sur le dispositif et ne sont donc jamais accessibles par un tiers, par vous-mêmes, ou quiconque trouverait le dispositif en cas de perte ou de vol.

Les données brutes sont ensuite sauvegardées sur un disque dur chiffré lui aussi et entreposé dans un placard fermé à clé au laboratoire Parole et Langage. Étant donné la longueur de ces enregistrements, les données ne seront jamais étudiées ni écoutées dans leur intégralité par un chercheur humain. En revanche, elles seront traitées par un système automatisé qui attribue des sections des enregistrements à l'enfant portant le

dispositif, à d'autres enfants, à des adultes masculins ou féminins. Une étape de traitement automatisé supplémentaire classe ensuite les vocalisations comme étant dirigées par l'enfant ou entendues par l'enfant (c'est-à-dire dirigées à quelqu'un d'autre), estime le nombre de sons, de syllabes et de mots dans un tour de parole; et/ou décide si une vocalisation est un pleur, un rire ou d'autres types de voix. Le contenu de ce qui est dit n'est donc pas inspecté manuellement. Cependant, un sous-ensemble de courts segments de 10 min d'enregistrements réalisés chez 5 participants sera sélectionné aléatoirement et utilisé pour une analyse manuelle visant à valider la qualité et la fiabilité du traitement automatique. Les plages de signal analysées seront sélectionnées aléatoirement et les analyses manuelles effectuées seront restreintes aux objectifs de recherche (nombre de vocalisations de l'enfant, nombre de tours de parole en conversation et nombre de mots prononcés par les adultes en conversation avec l'enfant) et réalisées par les personnes participant au projet. Dans tous les cas, les données générées et analysées automatiquement et manuellement ne sont que des valeurs numériques (valeurs résumant l'activité sonore des passages) et des étiquettes ne concernant que les dimensions linguistiques des données.



Figure 4 : Boîtier d'enregistrement LENA (taille réelle ; https://www.lena.org/wp-content/uploads/2016/07/LTR-03-2_Audio_Specifications.pdf).

Le tableau ci-après résume les différentes visites prévues dans le cadre de ce protocole et les différents actes prévus.

Visites	***V0 Inclusion***	***V1***	***V2***	***V3***	***V4***	***V5***	***V6***
		M3	**M6**	**M8**	**M10**	**M18**	**M24**
		(+/- 14 jours)	**(+/- 14 jours)**	**(+/- 14 jours)**	**(+/- 14 jours)**	**(+/- 14 jours)**	**(+/- 14 jours)**
Recueil du consentement éclairé /inclusion	X						
Enregistrements EEG		X	X	X	X	X	X
Enregistrements de l'environnement acoustique (LENA)		X	X	X	X	X	X
Tâches comportementales (mesure temps de regard)			X	X	X	X	X
Questionnaire de vocabulaire MB-CDI						X	X
Évaluation neurodéveloppementale Bayley IV							X

Qui peut participer ?

Critères d'inclusion :
- âge : de 3 mois à 24 mois
- né à terme (> 37 semaines d'aménorrhée)

- en bonne santé (pas de troubles de l'audition ou de la vision connus, pas de troubles développementaux connus, pas de symptômes du COVID19)
- affilié à un régime de sécurité sociale
- les deux titulaires de l'autorité parentale signent le consentement de participation à la recherche.

Critères de non-inclusion :
- prématurité : nés à moins de 37 semaines d'aménorrhée
- présentant des troubles développementaux connus
- refus d'être informé des anomalies qui pourraient être fortuitement décelées
- présentant des symptômes du COVID ou qui ont côtoyé des personnes positives au COVID ou qui font partie de la catégorie des personnes à risque pour le COVID.
- Non compréhension par les titulaires de l'autorité parentale des informations communiquées au cours du recrutement

Lieu de la recherche ?

La recherche se fera soit (i) dans les locaux de la plateforme expérimentale dédiée à l'étude du développement de la maison de la recherche (Université Aix-Marseille) Agrément ARS PACA portant autorisation d'un lieu de recherche biomédical n°: 2019-04 du 8 janvier 2020, soit (ii) dans les locaux de la plateforme expérimentale dédiée à l'étude du développement du Laboratoire de Neurosciences Cognitives (Université Aix-Marseille) Agrément ARS PACA portant autorisation d'un lieu de recherche biomédical n°: 2021-07 du 4 décembre 2021, soit (iii) au sein même de la crèche Babilou dans laquelle votre enfant est inscrit. Tous les lieux précédemment mentionnés disposent des moyens humains, matériels et techniques adaptés à la recherche et compatibles avec les impératifs de rigueur scientifique et de sécurité des personnes qui s'y prêtent (Article L1121-13 CSP).

Durée de la recherche ?

La durée de participation est d'une heure maximum par visite, en comptant le temps d'explication des détails de la procédure au début, un temps de jeu pour l'enfant afin qu'il se sente en confiance et un temps pour répondre aux questions sur nos recherches en général, à la fin. Les enregistrements EEG durent 45 minutes (incluant la pose du bonnet) et les tâches comportementales durent 15 minutes environ (correspondant à la durée de l'enregistrement vidéo et des enregistrements occulométriques). Le suivi est réalisé pendant une période de 21 mois après inclusion (de 3 à 24 mois). La durée totale de la recherche est de 5 ans.

La possibilité de participer à une autre recherche ?

La participation de votre enfant à cette recherche n'est assortie d'aucune interdiction de participation à une autre étude simultanément, sans période d'exclusion.

Quelle est la compensation des contraintes subies et des frais exposés ?

Les familles qui participent recevront un « diplôme de jeune scientifique » avec le prénom de l'enfant et un t-shirt (valeur maximale 5€) pour l'enfant.
Pas de dédommagement des frais de transport n'est prévu dans le cas où les expériences sont réalisées au sein des crèches Babilou dans une pièce désignée.
Dans le cas où vous choisissez d'effectuer une des visites dans le babylab d'Aix-en-Provence ou de Marseille, un remboursement des frais exposés sur justificatif vous sera proposé.

Quels sont vos droits ?

- Information et Consentement

  - Vous pouvez poser toutes les questions que vous souhaitez avant le début de la recherche et au cours de celle-ci à *Clément François.*
  - Vous avez le droit de disposer d'un délai de réflexion raisonnable avant de consentir à participer à cette recherche

- Vous avez la possibilité de refuser de refuser de faire participer votre enfant à cette recherche, ou de retirer votre consentement à tout moment au cours de celle-ci sans avoir à le justifier et sans que vous puissiez encourir aucune responsabilité ni aucun préjudice de ce fait.

- Utilisation de l'image de votre enfant

Au cours de la présente recherche votre enfant sera *filmé.e/pris.e en photo.* L'utilisation de son image sera limitée au besoin du présent projet de recherche. Vous avez la possibilité de vous opposer à l'utilisation de son image, dans ce cas, votre enfant ne pourra pas participer à la présente recherche.

- Utilisation de votre voix et celle de votre enfant

Au cours de la présente recherche votre voix et celle de votre enfant sera enregistré. L'utilisation des enregistrements vocaux de votre voix et celle de votre enfant sera limitée au besoin du présent projet de recherche. Vous avez la possibilité de vous opposer à l'utilisation de votre voix et celle de votre enfant, dans ce cas votre enfant ne pourra pas participer à la présente recherche.

- Traitement des données à caractère personnel

Dans le cadre de la recherche impliquant la personne humaine à laquelle le CNRS (représenté par M. André LE BIVIC, Institut des sciences biologiques, 3 Rue Michel-Ange, 75794 – Paris Cedex 16) vous propose de participer ainsi que votre enfant, un traitement de vos données personnelles et de celles de votre enfant va être mis en œuvre (date de naissance (jour, mois, année), nom, prénom, sexe, réponses au questionnaire de vocabulaire MB-CDI, réponse lors de la visite d'inclusion, résultats ou données issues des enregistrements EEG, de l'occulométrie, données issues des enregistrements de l'environnement acoustiques, donnée issues des réponses aux comportements durant ces examens).

Le responsable de traitement est *Christine Meunier*, *christine.meunier@univ-amu.fr*, directrice du laboratoire *Parole et Langage.*

La déléguée à la protection des données (DPD) compétente est *Gaëlle Bujan, dpd.demandes@cnrs.fr*.

Le traitement de vos données personnelles est nécessaire à la réalisation de la recherche. La finalité du traitement est de mieux comprendre les mécanismes cérébraux impliqués dans le développement du langage chez les enfants entre 3 et 24 mois.

La base légale du traitement est l'exécution d'une mission d'intérêt public dont est investi le responsable du traitement (article 6 du RGPD). Le responsable du traitement des données doit mettre en œuvre des mesures appropriées permettant de garantir vos droits et libertés, notamment le seul recueil de données strictement nécessaires à la recherche

Conformément aux dispositions de la loi n°78-17 du 6 janvier 1978 modifiée relative à l'informatique, aux fichiers et aux libertés, ce traitement a fait l'objet d'un engagement de conformité à la MR-001.

Vos données et celles de votre enfant seront pseudonymisées par un numéro de code pour les besoins de la recherche, sans mention de vos noms et prénoms, ni ceux de votre enfant.

Ces données codées seront accessibles à l'investigateur principal, ainsi qu'aux professionnels intervenant dans la recherche et aux personnels agissant sous leur responsabilité ou autorité. Le cas échéant, des experts indépendants chargés de ré-analyser les données pour vérifier les résultats de l'étude, en vue de leur publication, dans des conditions strictes de sécurité, auront également accès à ces données codées.

Les informations concernant votre identité et celle de votre enfant ne seront connues que par l'équipe scientifique recueillant ces données, ainsi que par les professionnels de la recherche. En cas de contrôle et dans le strict respect des règles de confidentialité, les autorités sanitaires ou publiques, le personnel habilité du promoteur ou de l'organisme d'assurance souscrite, ainsi que la DPD, si vous la contactez, auront également accès à ces données.

Le cas échéant, vos données d'identification correspondant à vos nom, prénom, coordonnées postales et bancaires pourront être transmises au service financier de la délégation régionale du CNRS aux seules fins de permettre le remboursement des frais exposés et/ou l'indemnisation des contraintes subies.

L'ensemble de ces personnes, soumises au secret professionnel, auront accès à ces données dans le cadre de leur fonction et en conformité avec la réglementation.

La publication des résultats de la recherche ne comportera aucun résultat individuel identifiant. Les résultats pourront faire l'objet de présentations scientifiques ou de publications dans la presse médicale et scientifique, mais en aucun cas votre nom, celui de votre enfant ou toute autre donnée identifiante n'apparaîtra.

Les données recueillies seront conservées jusqu'à *deux ans après la dernière publication des résultats de la recherche ou, en cas d'absence de publication, jusqu'à la signature du rapport final de la recherche*, puis seront archivées conformément à la réglementation en vigueur.

Vous disposez d'un droit d'accès, de rectification et d'opposition au traitement de vos données et de celles de votre enfant et un droit d'effacement. Vous pouvez retirer votre consentement au traitement de vos données et de celles de votre enfant à tout moment. Vous avez également la possibilité de demander la limitation de l'usage de vos données et de celles de votre enfant et d'en obtenir la portabilité lorsque cela est possible et de définir le sort des données en cas de décès.

Ces droits s'exercent auprès du *responsable scientifique Clément François (coordonnées : clement.francois@univ-amu.fr, 06 69 59 15 19).*

En cas de retrait de votre consentement et/ou d'exercice du droit à l'effacement de vos données et celles de votre enfant, ce retrait n'a pas d'incidence sur les activités menées et sur l'utilisation des données obtenues sur la base du consentement éclairé exprimé avant que celui-ci n'ait été retiré. Les données antérieurement collectées pourront être conservées et utilisées dans le cadre de la présente recherche.

Dans l'hypothèse où les critères d'inclusion ne sont plus remplis au cours de l'étude ou en cas de changement de situation susceptible de constituer une contre-indication à la recherche, vous êtes informé.e que les données antérieurement collectées peuvent être conservées et utilisées dans le cadre de la présente recherche.

Enfin, si vous y consentez, vos données et celles de votre enfant pourront être utilisées ultérieurement pour d'autres projets exclusivement à des fins scientifiques. Vous pouvez vous opposer à la réutilisation de vos données et celles de votre enfant à tout moment.

Vous disposez d'un droit d'introduire une réclamation auprès de la CNIL (https://www.cnil.fr/fr/plaintes) à l'adresse suivante : Commission nationale de l'informatique et des libertés, Service des plaintes, 3 Place de Fontenoy, TSA 80715, 75334 PARIS CEDEX 07.

- Accès aux résultats globaux de la recherche :

Lorsque cette recherche sera terminée, vous serez tenu informé.e personnellement des résultats globaux dès que ceux-ci seront disponibles, si vous le souhaitez sur demande à *Clément François.*

Votre participation à la présente recherche est totalement libre. Vous pouvez quitter l'étude à tout moment, sans encourir aucune responsabilité ni subir aucun préjudice. Pour cela, il vous suffira d'en informer *Clément François.*

Après avoir lu cette notice d'information, n'hésitez pas à poser à *Clément François* toutes les questions que vous désirez. Après un délai de réflexion, si vous acceptez de participer à cette recherche, vous devez compléter et signer le formulaire de consentement de participation. Un exemplaire du document complet (notice d'information et formulaire de consentement complété et signé) vous sera remis.

FORMULAIRE DE CONSENTEMENT DESTINE AUX TITULAIRES DE L'AUTORITE PARENTALE

De la perception de la parole à l'acquisition du vocabulaire chez le bébé

Promoteur de la recherche : ***CNRS***
Investigateur principal : ***Clément François (Chargé de Recherche, CNRS)***

Nous soussignés

.....

titulaires de l'autorité parentale à l'égard de .....

déclarons avoir lu et compris la notice d'information qui nous a été remise, et avoir pu poser toutes les questions que nous souhaitions à Clément François (Chargé de Recherche, CNRS), qui nous a expliqué les objectifs et la méthodologie de la recherche, nos droits et ceux de notre enfant ainsi que les risques prévisibles et les contraintes liées à la participation de notre enfant à cette recherche.

Nous avons été informés que cette recherche a reçu l'avis favorable du CPP SUD-EST VI, le 6 janvier 2023.

Le CNRS, promoteur de cette recherche est représenté par André LE BIVIC, directeur de l'institut des sciences biologiques, 3 rue Michel-Ange, 75794 - PARIS Cedex 16. Il a souscrit une assurance de responsabilité civile auprès de la société HDI Global SE (contrat N°0100630314037).

Il nous a été précisé que :

▪ Nous sommes libres d'accepter ou de refuser ainsi que d'arrêter à tout moment la participation de notre enfant sans avoir à nous justifier et sans encourir aucune responsabilité ni aucun préjudice de ce fait.

▪ Nous avons le droit d'avoir accès aux données de notre enfant

▪ Nous pouvons être tenus au courant des résultats globaux de la recherche en nous adressant à Clément François (clement.francois@univ-amu.fr)

▪ Nous pouvons être remboursé de nos frais de transport (taxi, ou transport en commun) sur présentation des justificatifs dans le cas où une des visites s'effectue dans le babylab d'Aix-en-Provence ou Marseille.

▪ Notre enfant recevra un « diplôme de jeune scientifique » avec son prénom comme remerciement et un t-shirt d'une valeur maximale de 5€.

Nous déclarons :

▪ Avoir pris connaissance des mesures de sécurité sanitaire en vigueur sur le lieu de la recherche et nous engager à les respecter et à les faire respecter par notre enfant.

▪ Que notre enfant est affilié.e à un régime de sécurité sociale ou bénéficiaire d'un tel régime

Nous consentons :

▪ Au traitement des données personnelles qui nous concernent et qui concernent notre enfant en conformité avec les dispositions de la loi n°78-17 du 6 janvier 1978 modifiée relative à l'informatique, aux fichiers et aux libertés et du règlement européen n°2016/679 sur la protection des données.

Le responsable de traitement est Christine Meunier (joignable au 04 13 55 36 33/ christine.meunier@univ-amu.fr). Nous pourrons à tout moment demander des informations à Clément François (joignable au 06 69 59 15 19/ clement.francois@univ-amu.fr) et exercer auprès de lui notre droit d'accès, de rectification, d'opposition et d'effacement ainsi que notre droit de demander la limitation de l'usage de nos données et celles de notre enfant, la portabilité et de définir le sort de ces données en cas de décès. Nous pourrons retirer notre consentement à tout moment auprès de lui.

En cas de retrait de notre consentement et/ou d'exercice du droit à l'effacement de nos données et celles de notre enfant, nous avons été informés que ce retrait n'a pas d'incidence sur les activités menées et sur l'utilisation des données obtenues sur la base du consentement éclairé exprimé avant que celui-ci n'ait été retiré. Nous consentons ainsi à ce que les données antérieurement collectées puissent être conservées et utilisées dans le cadre de la présente recherche.

Nous consentons également à ce traitement dans l'hypothèse où les critères d'inclusion ne sont plus remplis au cours de l'étude ou en cas de changement de situation susceptible de constituer une contre-indication à la recherche (tel que la découverte d'anomalies du traitement de l'information auditive).

Nous consentons :
▪ à être tenus informés des informations relatives à l'état de santé de notre enfant et des éventuelles anomalies qui pourraient être décelées à l'occasion de la recherche en cas d'examen susceptible de déceler des anomalies (ex : Electro-encéphalographie, évaluation neurodéveloppementale):
Oui ☐ / Non ☐

▪ à ce que l'image de notre enfant soit enregistrée et utilisée pour les besoins de la présente recherche :
Oui ☐ / Non ☐

▪ à ce que notre voix ainsi que celle de notre enfant soient enregistrées et utilisées pour les besoins de la présente recherche : Oui ☐ / Non ☐

▪ à ce que nos données ainsi que celle de notre enfant puissent être utilisées ultérieurement pour d'autres projets exclusivement à des fins scientifiques: Oui ☐ / Non ☐

La publication des résultats de la recherche ne comportera aucun résultat individuel identifiant.

Notre consentement ne décharge pas les organisateurs de la recherche de leurs responsabilités. Nous conservons tous nos droits garantis par la loi.

Ayant disposé d'un temps de réflexion suffisant avant de prendre notre décision, nous acceptons librement et volontairement que notre enfant participe à la recherche :

Titulaire de l'autorité parentale 1 : .....

☐ **Oui /** ☐ **Non**

Signature:

Fait à ....……..…….,
le .....……

Titulaire de l'autorité parentale 2 : .....

☐ **Oui /** ☐ **Non**

Signature:

Fait à ....……..…….,
le .....……

Nom et signature de l'investigateur principal

Fait à ....……..…….,
le .....……

Comment bébé acquiert-il ses premiers mots ?

Vous voulez nous aider à répondre à cette question ? Participez à notre recherche !

POUR PLUS D'INFORMATIONS :

https://amubox.univ-amu.fr/s/rpKCycTSZLcypS8

Contact :

Estelle Hervé : estelle.herve@univ-amu.fr / 06.43.69.13.59

Babilou cnrs PAROLE ET LANGAGE Aix*Marseille université Socialement engagée

B. List of data collected for each participant

Participant code	Nursery / Babylab	EEG age (months)	Eye-tracking age (months)	LENA	BSID	MCDI
001	None	-	-	No	No	No
002	None	-	-	Yes	No	No
003	Schuman	3, 6, 8, 10, 18	18	Yes	No	No
004	Rép. Dames	6, 8, 10, 18, 24	18, 24	Yes	Yes	Yes
005	Schuman	18	-	No	No	No
006	Duranne	10, 18, 24	18, 24	Yes	Yes	Yes
007	Duranne	8, 10, 18, 24	18, 24	No	Yes	Yes
008	Schuman	10, 18, 24	18, 24	Yes	Yes	No
009	Schuman	6, 8, 10	-	Yes	No	No
010	Schuman	18, 24	24	No	Yes	Yes
011	Schuman	10, 18, 24	18, 24	Yes	Yes	Yes
012	Schuman	10, 18, 24	18, 24	Yes	Yes	Yes
013	Duranne	8, 10, 24	24	Yes	Yes	Yes
014	Schuman	24	24	Yes	Yes	Yes
015	Schuman	18, 24	24	Yes	Yes	Yes
016	Duranne	18, 24	24	No	Yes	Yes
017	Duranne	18, 24	24	Yes	No	No
018	Duranne	6, 8, 10, 18, 24	18, 24	Yes	Yes	Yes
019	Duranne	24	24	Yes	Yes	Yes
020	Schuman	24	24	Yes	Yes	Yes
021	Duranne	18	-	Yes	No	No
022	Duranne	18	-	No	No	No
023	Duranne	24	24	Yes	Yes	No
024	Duranne	6	-	Yes	No	No
025	None	-	24	Yes	Yes	Yes
026	Schuman	18, 24	24	Yes	Yes	No
027	Rép. Dames	18, 24	18, 24	Yes	Yes	Yes
028	Duranne	24	24	Yes	No	Yes
029	Duranne	10, 18, 24	18, 24	Yes	Yes	Yes
030	Schuman	10, 18, 24	18, 24	Yes	Yes	Yes
031	Schuman	18, 24	18, 24	Yes	Yes	Yes

Participant code	Nursery / Babylab	EEG age (months)	Eye-tracking age (months)	LENA	BSID	MCDI
032	Schuman	10, 18, 24	18, 24	Yes	Yes	Yes
033	Duranne	18	24	Yes	Yes	No
034	Rép. Dames	18, 24	18, 24	Yes	Yes	Yes
035	Rép. Dames	18	18	Yes	No	No
036	Duranne	18, 24	18	Yes	Yes	Yes
037	Schuman	6, 8, 10, 18, 24	18	Yes	Yes	Yes
038	Rép. Dames	8, 10	18, 24	Yes	Yes	Yes
039	Schuman	24	24	Yes	Yes	Yes
040	Schuman	18, 24	18, 24	No	Yes	Yes
041	Rép. Dames	24	18, 24	Yes	Yes	Yes
042	Duranne	18, 24	18, 24	Yes	Yes	Yes
043	Rép. Dames	3, 6, 8, 10	-	Yes	No	No
044	Duranne	8, 10	-	No	No	No
045	Duranne	8, 10	-	No	No	No
046	Schuman	18, 24	18, 24	No	Yes	No
047	Labo Minots	8, 18	18	No	No	No
048	Duranne	8	-	Yes	No	No
049	Schuman	18	18, 24	Yes	Yes	No
050	Duranne	24	24	No	No	No
051	Labo Minots	8, 10	-	No	No	No
052	Labo Minots	6, 10	-	No	No	No
053	Labo Minots	8	-	No	No	No
054	Labo Minots	10	-	No	No	No
055	Labo Minots	18	-	No	No	No
056	Labo Minots	10	-	No	No	No
057	Labo Minots	8	-	No	No	No
058	Labo Minots	8	-	No	No	No
059	Labo Minots	8	-	No	No	No
060	Labo Minots	8, 10	-	No	No	No
061	Schuman	3, 6, 8, 10	-	Yes	No	No
062	Labo Minots	18	18	No	No	No
063	Rép. Dames	8, 10	-	Yes	No	No
064	Labo Minots	10	-	No	No	No
065	Duranne	10	-	No	No	No
066	Labo Minots	8, 10	-	No	No	No
067	Labo Minots	10	-	No	No	No
068	Labo Minots	8	-	No	No	No
069	Labo Minots	8, 10	-	No	No	No
070	Labo Minots	10	-	No	No	No
071	Labo Minots	10	-	No	No	No
072	Labo Minots	10	-	No	No	No
073	Labo Minots	10	-	No	No	No

C. Challenges and new perspectives of developmental cognitive EEG studies: full text

Full text of Hervé et al. (2022):

Hervé, E., Mento, G., Desnous, B., & François, C. (2022). Challenges and new perspectives of developmental cognitive EEG studies [Publisher: Elsevier]. NeuroImage, 260, 119508. https://doi.org/10.1016/j.neuroimage.2022.119508 (cit. on pp. 4, 40, 41, 88).

NeuroImage 260 (2022) 119508



Contents lists available at ScienceDirect

NeuroImage

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



Challenges and new perspectives of developmental cognitive EEG studies



Estelle Hervé[a], Giovanni Mento[b,c], Béatrice Desnous[d,e], Clément François[a,*]

[a] CNRS, LPL, Aix-Marseille University, 5 Avenue Pasteur, Aix-en-Provence 13100, France
[b] Department of General Psychology, University of Padova, Padova 35131, Italy
[c] Padua Neuroscience Center (PNC), University of Padova, Padova 35131, Italy
[d] APHM, Reference Center for Rare Epilepsies, Timone Children Hospital, Aix-Marseille University, Marseille 13005, France
[e] Inserm, INS, Aix-Marseille University, Marseille 13005, France

A B S T R A C T

Despite shared procedures with adults, electroencephalography (EEG) in early development presents many specificities that need to be considered for good quality data collection. In this paper, we provide an overview of the most representative early cognitive developmental EEG studies focusing on the specificities of this neuroimaging technique in young participants, such as attrition and artifacts. We also summarize the most representative results in developmental EEG research obtained in the time and time-frequency domains and use more advanced signal processing methods. Finally, we briefly introduce three recent standardized pipelines that will help promote replicability and comparability across experiments and ages. While this paper does not claim to be exhaustive, it aims to give a sufficiently large overview of the challenges and solutions available to conduct robust cognitive developmental EEG studies.

1. Introduction

From a biological perspective, studying early brain development and functioning has substantial theoretical implications because it provides an appropriate level of analysis to understand human cognitive functions (Tinbergen, 1963; MacDougall-Shackleton, 2011; Hladký and Havlíček, 2013). In particular, electro-encephalographic recordings (EEG) applied to developmental research represent an irreplaceable source of knowledge to understand the temporal dynamics of brain activity and, consequently, the timing of cognitive processes in typical and atypical populations. Noteworthy, under specific technical and methodological conditions, EEG can provide reliable spatial information about the neural generators underlying scalp-level activity in infants (Reynolds and Richard, 2005, 2009; Reynolds et al., 2010) and children (Bathelt et al., 2014; Timeo et al., 2019). Further, EEG is an essential tool for studying brain activity in newborns (0 to 1 month old), infants (2 to 24 months old), and very young children (2 to 4 years old) because it presents undeniable advantages compared to other neuroimaging techniques. Compared to magnetic resonance imaging (MRI), EEG is more affordable, more tolerant to movement, more flexible (it can be used while sleeping, sitting, standing, or laying), and less ethically demanding. Nonetheless, collecting good quality EEG data in young participants is conditioned by significant methodological challenges that depend on the experimental contexts and the planned analyses.

The first challenge stands in the limited infants' attentional capabilities that effectively preclude the use of experimental paradigms widely used in preschool and school-aged children. Indeed, infants' visual attention span is limited to a few minutes within the first year of life (Atkinson and Braddick, 2012). A second, not less problematic aspect is the reduced behavioral repertoire that preverbal children can produce. Thus, experimental tasks requiring verbal responses or deliberate motor acts (e.g., pressing a key to select a choice) are impossible for young infants. These issues inevitably reduce the range of possible experimental designs, forcing researchers to select only certain types of stimuli and reduce the number of experimental conditions to maintain a good signal-to-noise ratio (SNR). These two examples are just a tiny part of many challenges experimenters have to deal with when working with infants and EEG.

In addition, data processing should be applied accordingly and adapted to developmental populations. EEG signals can be analyzed using different approaches, thus providing different insights into the human brain's functioning. A first general distinction can be made between time-locked and continuous EEG recordings. Event-related brain potentials (ERPs), the "voltage fluctuations in the ongoing EEG that are time-locked to an event, such as the onset of a stimulus" (Kappenman and Luck, 2011), allow researchers to link brain activities to specific events with precise temporal resolution. More recently, time-frequency (TF) transformations have gained significant interest in the developmental community allowing the study of brain oscillations during rest periods or in response to external events. Furthermore, although TF analyses of continuous EEG recordings were mainly used for clinical purposes (Connell et al., 1989; Hosain et al., 2003; Schumacher et al., 2011) or to describe general states of the brain (de Haan, 2013), they can also be used to study infants brain activity during more ecologically valid situations.

* Corresponding author.
E-mail address: clement.francois@univ-amu.fr (C. François).

https://doi.org/10.1016/j.neuroimage.2022.119508.
Received 31 March 2022; Received in revised form 7 July 2022; Accepted 22 July 2022
Available online 23 July 2022.
1053-8119/© 2022 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

This article aims to provide an overview of representative EEG studies tapping early cognitive development to help researchers consider each processing step before jumping in. Rather than an exhaustive review of the developmental EEG literature, we report and discuss the main challenges and specificities that any researcher may encounter when conducting an EEG study with young humans. We also provide some practical and methodological guidelines to maximize the success rate in such studies.

2. Working with young participants: specific issues

This section focuses on some crucial practical challenges allowing researchers to implement EEG studies in developmental populations successfully. We do not claim to provide an exhaustive list of practical issues, but help investigators anticipate and get efficiently through the most specific steps of developmental EEG study. Indeed, compared to studies conducted with adults, developmental studies differ in several ways and rely on specific challenges due to the characteristics of the targeted population that must be considered when designing a developmental EEG experiment.

2.1. Artifacts and attrition

Due to their young age, infants' behaviors are more challenging to control or predict, leading to poor data quality. Artifacts are widespread in infants' EEG studies, especially movement artifacts. The age of the children (from 12 months of age, infants tend to be more active and may try to grab the electrodes), the EEG device (nets tend to move a little compared to caps), the paradigm (visual paradigms tend to encourage head movements), and the experimental setup (children are more likely to move when they are on their parent's lap) can increase the risk of movements during data acquisition (de Haan et al., 2013). The general state of the children is also a key element in obtaining a good SNR: depending on the moment of the day, the quality and duration of sleep during the previous night, the emotional state of the child, the time after feeding time, and external events, the infant can be more or less fussy thus causing movement artifacts. A particularly relevant aspect to consider is the level of vigilance and the sleep stage in which the EEG recording is performed. This aspect is fundamental when collecting data from newborns who spend considerable time sleeping. In human infants, cyclic periods of active and quiet sleep are present after 35 weeks post-conception and can be easily detected by behavioural observation and EEG staging (Mercuri et al., 1995; Pressler et al., 2003). For instance, the neonatal brain is functionally more activated during active sleep, similar to wakefulness. Conversely, cortical activation decreases during quiet sleep (Graven, 2006), suggesting that the sleep stage may significantly impact spontaneous EEG and event-related activity. However, while some evidence supports this claim (Tokioka et al., 1995, Cheour et al., 2002a, Friederici et al., 2002, Friedrich et al., 2004, Suppiej et al., 2010), other studies reported no significant sleep impact (Alho et al., 1990a; Cheour-Lutanen et al., 1996, Martynova et al., 2003; Leppänen et al., 1997; Cheour et al., 1998a). Hence, a definitive consensus on the best sleep stage to collect EEG/ERP activity in newborns and young infants is still unclear. Nevertheless, as a good lab practice, we suggest carefully considering this aspect, reporting, when possible, consistency across sleep stages in EEG frequency band power and distribution or ERP amplitude and latencies.

Overall, the experiment must be well prepared to ensure optimal conditions for electrode application and data acquisition that will minimize artifacts and optimize data quality (Lloyd et al., 2015). It is also essential that the researchers adapt their behavior, tone of voice, and sensitivity to each child (Cotter et al. 2002). In addition, a pleasant and skilled staff should be selected to help the researcher (Cotter et al., 2002). Welcoming young participants to a different room from the experimental booth can facilitate acclimatization (de Haan, 2013, p.19). This room should be pleasant both for parents and children as it can impact the general feeling of the families (see Fig. 1 for an example of a babylab welcoming room). Importantly, a specific space or room should be available for parents to feed the infants or change their diapers (Hoel and Wahl, 2012).

However, some artifacts are inevitable: eye movements, blinks, facial muscle activity, cardiac activity, and respiration are non-cerebral activities that distort the EEG signal. Cardiac activity and respiration are rhythmic and thus have an easily recognizable EEG pattern that can be detected and removed from the signal (Hoehl and Wahl, 2012). While eye movements, blinks, and facial muscle activity can be avoided or strongly decreased in adults under specific instructions as they can control them, this is not the case with infants. Consequently, these artifacts are commonly observed in infants' EEG signals, although more difficult to separate from background EEG activity than in adults (Bell and Wolfe, 2008).

While sucking a pacifier can sometimes be compatible with behavioural experiments, tongue movements can induce rhythmic low-frequency EEG activity that is not easily distinguishable from the background EEG (Hoehl and Wahl, 2012; Vanhatalo et al., 2003). Nevertheless, it is worth considering that the potential drawback due to the artifactual EEG activity is usually largely overweighted by the advantage in terms of behavioural compliance, as infants are usually calmed when using the pacifier, resulting in data quality improvement. Further, non-nutritive sucking has no significant impact on EEG data in newborns compared to nutritive sucking (Lehtonen et al., 2016). Cardiac activity should be recorded, and external electrodes should be positioned on the face to detect all these non-cerebral artifacts from the scalp. For example, one generally chooses to place one electrode for ocular artifacts (near the eyes) and one or two for facial muscle movements (on a cheek and near the mouth; Picton et al., 2000). A suitable lab practice is to avoid artifacts by using external control of stimuli and experimental manipulation. Indeed, to minimize data loss, the experimenter can design the experiment so that it can be paused at any time and that some stimuli or sequences of stimuli can be repeated (de Haan, 2013). Although technically challenging, an advanced approach relies on an eye-tracker to allow baby-driven experimental control procedures (Kulke et al., 2017). Eye-tracking systems can be interfaced via software with the EEG amplifier to control the timing and type of experimental conditions based on the child's eye behavior. Importantly, this approach offers accurate online monitoring of artifacts and can lead the child to direct visual attention towards the target stimuli through reinforcement (Maguire et al., 2014). Overall, artifacts affect infants' EEG signals more than adults'. Importantly, these artifacts are often removed from the data, automatically leading to data loss and, thus, a decrease in the SNR. A recent study shows the effects of variability in data editing methods on the quality of infants' visual ERPs (Monroy et al., 2021). Based on 19 EEG datasets collected in 10 m-o infants during an action-perception paradigm, results were clear cut in showing differences between expert editors in (i) the number of participants accepted, (ii) the number of trials accepted, and (iii) the channels marked for interpolation. In this sense, the use of video-EEG to monitor the online child's behavior during the experiment should be implemented (see Fig. 2). This technique allows video synchronization with the EEG via software, making motion artifacts very easy to isolate (e.g., blinks and eye movements, head-turning, sucking, or arm movements) from other environmental artifacts (e.g., electric noise, bad quality channels). In this way, it is possible to adjust online data acquisition quality by providing feedback to the parents or directly to the baby (i.e., an attention-getter picture or sound). This procedure ultimately allows preventing the number of artifacts by monitoring the child's behavior in progress. However, when video-EEG is unavailable, an alternative option is to have a research assistant use a response button to manually insert triggers in the EEG recordings that will help spot artefactual EEG sections (see Partanen et al., 2022).

An alternative option is to correct these artifacts offline using specific signal processing algorithms such as Independent Component Analysis (ICA). The Independent Component Analysis (ICA) source separation method is particularly efficient in correcting artifacts from continuous

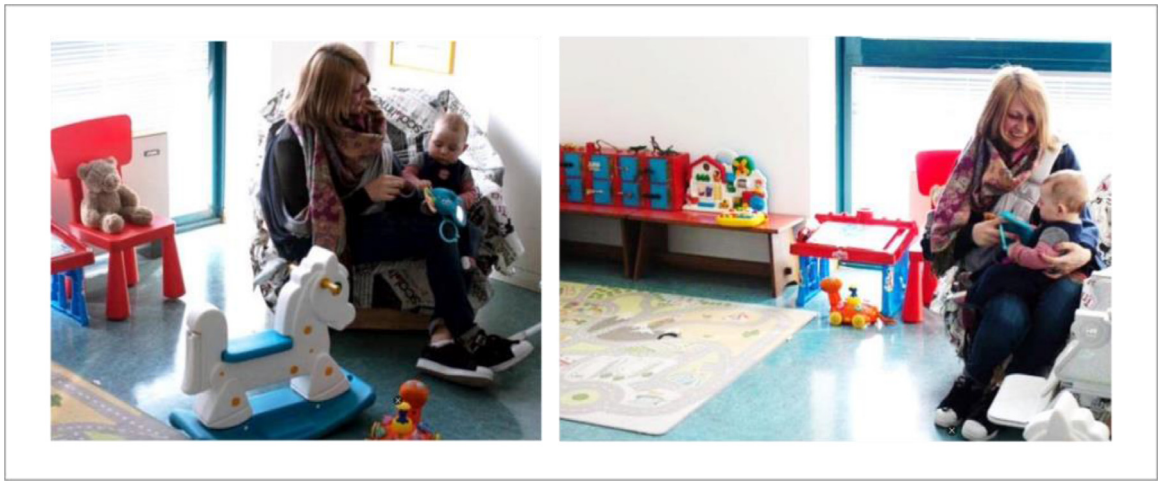


Fig. 1. Example of a waiting room to welcome young infants and their families. Note the presence of child-friendly furniture, toys, and games to habituate the infant before starting the experiment. Courtesy of the Department of Developmental and Social Psychology, University of Padova, Italy.

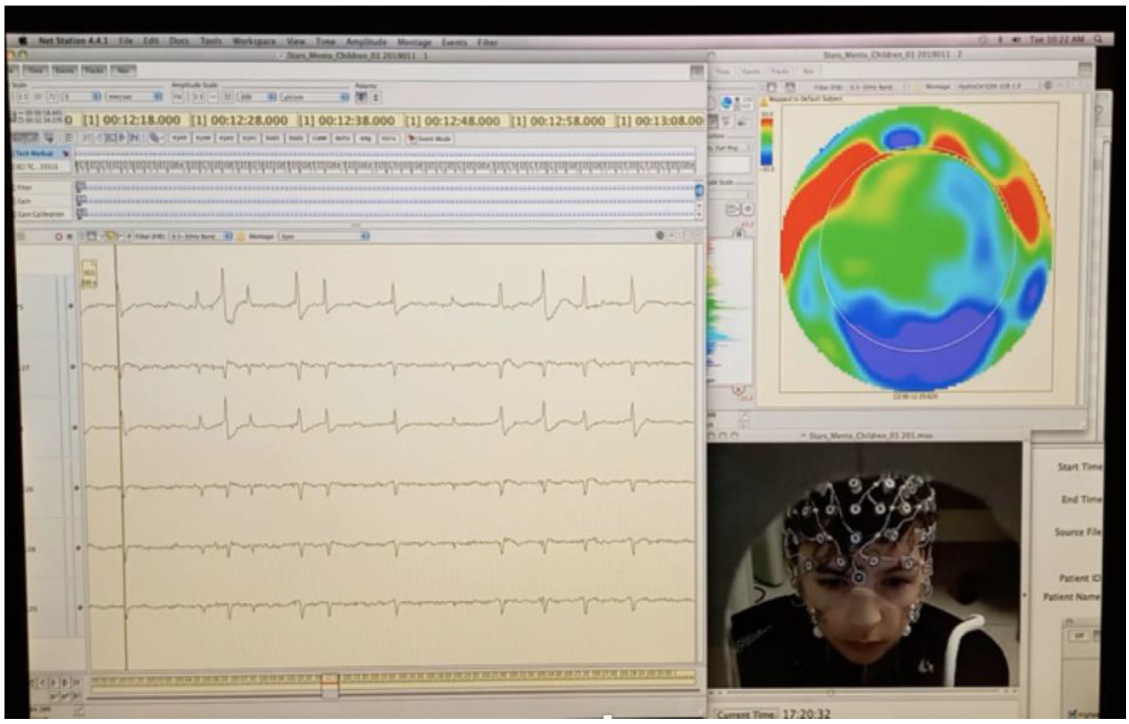


Fig. 2. Snapshot of a video-EEG recording collected in a 5-year-old child while executing a cognitive task. The left panel shows the online EEG recording with a detailed view of the electrodes around the eyes. The top-right panel displays the online topographical spline map, while the bottom-right shows the EEG-synchronized eye behaviour (EGI© system). The EEG recording is taken from a dataset originally published by Timeo, Mento, Fronza, and Farroni, 2018.

EEG (Makeig et al., 1996). Nevertheless, this method requires "the rapid and brief activation of spatially fixed brain regions" (Johnson et al., 2001) for optimal functioning so that it might be more efficient in ERP studies than in continuous EEG recording. Moreover, while ICA may perform best at separating signals after a 1–2 Hz high-pass filter (Winkler et al., 2015), such a filter may also drastically impact several ERP components. Therefore, an alternative two-step procedure might be to apply such a hard filter before ICA decomposition in one dataset, select, save the artifactual ICs, and use them for correcting another dataset with the appropriate filter settings. Although ICA-based artifact correction has traditionally been less used in the infant literature, several studies suggest it works well with participants as young as a few months old (de Haan, 2013, 2002; (Fujioka et al., 2011); Miljković et al., 2010). Fig. 3 illustrates an example of blink-related IC easily recognized and removed from the epoched EEG trace of a four-month-old infant.

Nonetheless, while component-based artifact rejection methods are viable options for high-density EEG recordings as often done with adults, it is still possible that some components may contain activity originating from the brain. Such a situation is highly likely in the case of low-density EEG recordings where the number of sensors is much smaller than the number of brain sources, which may result in activity from multiple brain sources being mixed into a single component. In such a case, rejecting the artefactual IC will likely remove data originating from brain sources as well. Furthermore, the decision about which and how many ICs to reject is not trivial because it is too operator-dependent and thus ultimately arbitrary. The rejection decision should be limited to only those ICs that unambiguously refer to recognizable artifacts such

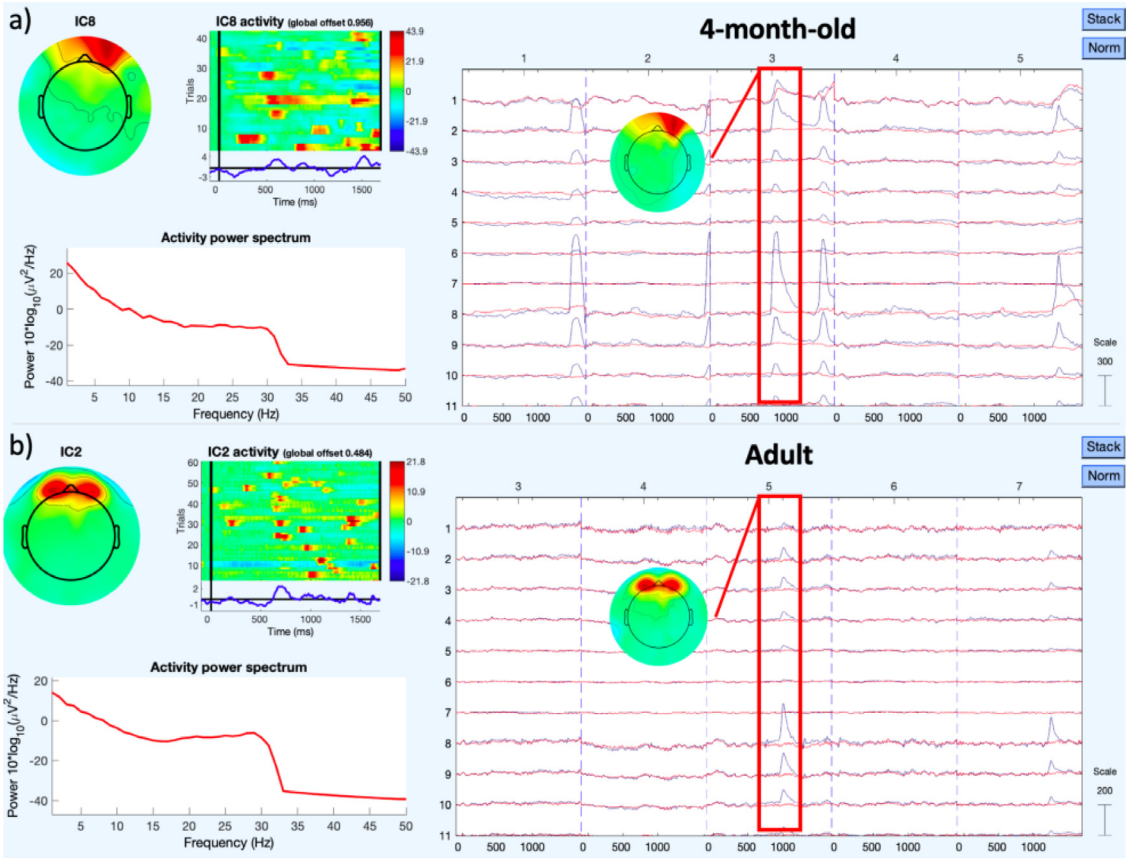


Fig. 3. Independent Component Analysis. The picture shows the details of a blink-related IC in a 4-month-old infant (a) and an adult (b) extracted from a High-Density EEG dataset (128 channels; EGI system) by using the EEGlab toolbox. The left part of the picture shows the spectral, temporal, and spatial properties of the IC. The right part displays the EEG recording with a detailed view of the electrodes around the eyes. The blue and red lines depict the EEG signal with and without the IC. Note that the rejection of this IC reliably rules out the eye-blink artifactual noise without losing much EEG signal in both the adult and infant recording. (The EEG recordings are taken from a dataset originally published by Mento and Valenza, 2016).

as blinks, heartbeat, or network noise. Although this appears relatively straightforward in the case of adults, it is not always possible to easily detect artifactual ICs in infants' EEG data.

Interestingly, the wavelet-ICA (w-ICA) is a two-step ICA implemented in the recent toolboxes developed for infants' EEG data that may be particularly relevant as it increases component separation efficiency and decreases data loss when removing artifacts (Azzerboni et al., 2005; Mammone et al., 2012). Moreover, a common problem of the ICA methods is that classical algorithms used with adults usually require long datasets for optimal functioning, whereas infants' recordings are generally shorter. Blind Source Separation (BSS) algorithms are very efficient in separating cerebral from extra-cerebral sources (Romo-Vázquez et al., 2012). However, the final result mainly depends on the type and number of artifacts and the choice of the BSS algorithm (Fitzgibbon et al., 2007). A combination of different algorithms seems to be the most powerful solution for artifact separation, as Romo-Vázquez et al. (2012) pointed out. However, data loss (and attrition) is inevitable when considering artifacts and other uncontrollable events inherent to developmental studies.

Studies involving human beings necessarily imply attrition, especially when they are very young. Data loss can be due to inter- or intra-individual differences, experimental characteristics, technical issues during data acquisition, interruption of data recording upon participant's request, or any other measure-related issue. By reducing the sample size and decreasing statistical power, attrition can "threaten the internal validity of experimental studies" (Ribisl et al., 1996). Attrition rates in infants' studies are usually higher than in adults and negatively impact data reliability. Although attrition during EEG acquisition tends to show an inverted u-shaped pattern through development (Bell and Cuevas, 2012), the factors that may increase data loss in developmental studies are not fully known. In a meta-analysis investigating attrition rate in infant ERP studies, Stets et al. (2012) found that the nature of the stimuli was the main influential factor for attrition, but that age was not. Remarkably, studies involving visual stimuli showed the highest attrition rates. However, the attrition rate may also depend on the paradigm and the age of participants (de Haan, 2013). These contradictory results reflect the lack of knowledge about the attrition rate in infants' EEG studies.

Nevertheless, attrition can be overcome by recruiting larger samples of participants, choosing adequate experimental settings, implementing recruitment strategies or databases, and piloting before actual data acquisition (Hurwitz et al., 2017; Ribisl et al., 1996). The accelerated longitudinal design is an alternative to longitudinal studies, which usually present high attrition rates (Galbraith et al., 2017). This method involves the recruitment of multiple cohorts at different ages, each covering a part of the broader targeted age range instead of a single cohort spanning all ages. In addition to decreasing attrition rates, its main advantage is that all ages are studied longitudinally for a given age range but within a shorter period than a classical longitudinal design. As attrition rates are very variable within the literature, similar studies should be compared to estimate the expected amount of data loss. In any case, while several EEG artifact management procedures are available nowadays, it remains crucial to remember that the researcher should prepare the experimental task and lab setting in the most appropriate way to minimize noisy activity. In other words, the most critical aspect of developmental EEG research remains that prevention is better than cure.

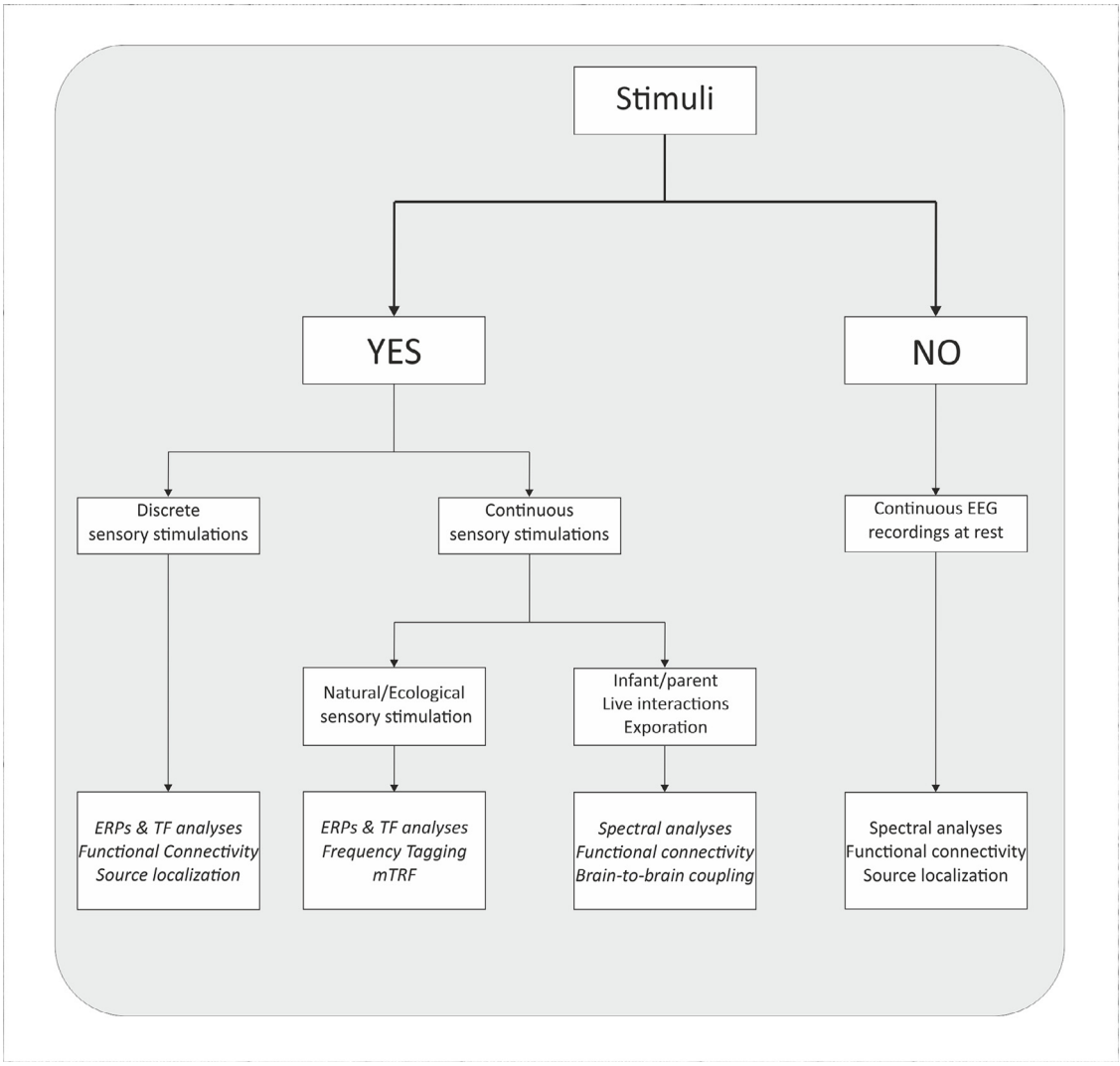


Fig. 4. Decision tree illustrating the different analyses available for current developmental EEG research. Continuous resting-state EEG data can be analyzed when no stimulation is used using advanced methods such as spectral, functional connectivity, or source localization, depending on the electrode layout. Different experimental paradigms with continuous or discrete stimulations can be analyzed using various methods depending on the electrode layout.

2.2. Classical paradigms and time-domain analyses

When designing an experiment for developmental populations, the first challenge is to provide an attractive experiment that captivates the child's attention and avoids boredom. If stimuli are presented, they must be carefully chosen, and attention-grabbers (pictures, videos, sounds) can be displayed between trials depending on the task. Because infants have limited attentional, cognitive, and behavioral abilities, the second challenge is to use specific tasks adapted to the age of the participants. Indeed, unlike adults, infants can hardly follow instructions (de Haan, 2013), which must be considered when designing the experiment. Consequently, collecting EEG during rest or passive paradigms can overcome this limitation while allowing the analysis of brain activity (see Fig. 4 for an illustration of the different options for the analysis of EEG data during development and Table 1 for an integrative summary of the main ERP results). Similarly, splitting passive tasks into short blocks may help obtain good quality data for restless babies. When the experimental setting is divided into small blocks, the changes in the behavioural state may not negatively impact the entire experiment but only a single recording block which may be particularly relevant for experimental designs interested in observing changes over the recording session.

2.2.1. Auditory paradigms

The auditory oddball paradigm is one of the most common paradigms in infants' EEG studies. When presented with auditory stimuli such as tone bips or speech sounds, a succession of waves forming the cortical auditory evoked potentials (CAEPs) can be observed over frontocentral electrodes (see Wunderlich and Cone-Wesson, 2006 for a review). Wunderlich et al. (2006) showed that the CAEPs elicited by pure tones or complex speech sounds such as a consonant-vowel-consonant word develop from early infancy showing a gradual increase in peak amplitude. Specifically, they collected CAEPs in newborns (N = 10), toddlers (N = 19), children (N = 20), and adults (N = 9) and showed that the P1 and N2 peak amplitudes decreased, whereas the N1 and P2 peak amplitudes increased with age. The oddball paradigm is also referred to as the "MMN" paradigm for "Mismatch Negativity", the brain response peaking between 170 and 250 ms after stimulus onset elicited by the preattentive detection of deviants (Näätänen et al., 2005, 2007; Kushnerenko et al., 2002). It relies on the brain sensitivity to rarely occurring deviant stimuli that are pseudo-randomly presented within a series of repeated standard stimuli. Deviant stimuli have a lower probability compared to standard ones, which generally occur with a probability comprised between 0.75 and 0.90 (Sambeth et al., 2006; Cheour et al., 2002b; Kushnerenko et al., 2002; Leppänen, Eklund and Heikki, 1997;

Table 1
Summary table showing the main ERP components obtained in classic EEG paradigms used in developmental research. Note that the adult components are also mentioned but should not be considered similar to the infant's components. Instead, ERPs obtained during early development should be seen as precursors of adult ERPs because they originate from an immature brain. For example, although the infant and adult MMNs may reflect a similar change detection process, the neural generators may differ. NI = No information. ISI: Inter-Stimulus-Interval.

Component	Age	Polarity	Latency (ms)	Spatial distribution	Cognitive correlate	Studies
CAEPs						
P1	Newborns	+	74-86	NI	Automatic brain responses to tones or complex sounds	Wunderlich & Cone-Wesson, 2006; Kushnerenko et al., 2002; Barnet, 1971
	Toddlers	+	67-90	Fronto-temporal		
	Children	+	77-93	Fronto-central		
	Adults	+	46-58	Fronto-central		
N1	Newborns	-	154-186	Fronto-central		
	Toddlers	-	136-150	Distributed across frontal, central, temporal and parietal regions		
	Children	-	157-177			
	Adults	-	107-114	Fronto-central		
P2	Newborns	+	214-241	Fronto-central & temporal		
	Toddlers	+	255-297	Distributed across frontal, central, temporal and parietal regions		
	Children	+	254-287			
	Adults	+	175-215	Fronto-central		
N2	Newborns	-	373-422	Distributed across the scalp		
	Toddlers	-	373-450	Distributed across the scalp		
	Children	-	376-400	Fronto-central		
	Adults	-	290-312	Fronto-central		
MMN	Newborns	- / +	350-550	Fronto-parietal	Automatic change detection (auditory stimuli)	Partanen et al., 2013; Tanaka et al., 2001; Nätäänen, 2000; Cheour et al. 2000
	Toddlers	- / +	200-250	Fronto-parietal		
	Children	- / +	200-400	Fronto-central		
	Adults	-	100-200	Fronto-central		
N170	Newborns	NI	NI	NI	Visual face and face inversion processing	Halit et al., 2003; de Haan et al., 2002; Gliga & Dehaene-Lambertz, 2007; Kadosh & Johnson, 2007; Taylor et al., 2001
	Toddlers	- or +	220-350 or 400	Occipito-temporal		
	Children	-	170-250	Occipito-temporal		
	Adults	-	144-200	Occipito-temporal		
Pb	Newborns	NI	NI	NI	Stimulus expectancy and contextual processing	Karrer et al., 1998; Webb et al., 2005 ; Jonkman, 2006 ; Squires et al., 1976
	Toddlers	+	200-400	Occipito-temporal		
	Children	+	300	Occipital		
	Adults	+	300	Central		
PSW	Newborns	NI	NI	NI	Visual working memory updating	de Haan & Nelson, 1997 ; Duarte et al., ; 2013 ; Pelegrina et al., 2020
	Toddlers	+	1000	Fronto-central		
	Children	+	500-700	Parieto-occipital		
	Adults	+	300-700	Centro-Parietal		
Nc	Newborns	-	1000-1200	Fronto-central	Stimulus familiarity and frequency Attention allocation	DeBoer, Scott & Nelson, 2004; Nelson, 1996 ; De Lorenzo et al., 2020 ; Grune et al., 1996 ; Gomarus et al., 2006
	Toddlers	-	400-600	Fronto-central		
	Children	+	300-400	Fronto-central		
	Adults	+	300	Centro-parietal		
CNV	Newborns	NI	NI	NI	Stimulus expectancy	Mento & Valenza, 2016; Mento & Vallesi, 2016; Mento, 2013
	Toddlers	-	ISI-related	Fronto-parietal		
	Children	-	ISI-related	Fronto-parietal		
	Adults	-	ISI-related	Fronto-parietal		

Winkler et al., 2003). The oddball paradigm has been consistently used with deviants in sound duration (Ceponiene et al., 2002; Clarkson et al., 1989), frequency (Alho et al., 1990a,b; Kushnerenko et al., 2002; Leppänen et al., 2010; Bisiacchi et al., 2009; Mento et al., 2010), intensity (Kushnerenko et al., 2002; Partanen et al., 2013), phonetic features of speech sounds (Dehaene-Lambertz, 2000; Cheour et al., 1998a,b; Pang et al., 1998; François et al., 2021), complex musical or speech sequences (Háden et al., 2015a, b; François et al., 2017; Mueller et al., 2012). Importantly, the term "Mismatch Response" or MMR is often preferred in developmental studies because of its high intersubject variability. Indeed, this component tends to be positive in younger infants and sometimes up to school age, depending on the developmental status of children (Bisiacchi et al., 2009; Mento and Bisiacchi, 2012; François et al., 2017; Moser et al., 2020; Kailaheimo-Lönnqvist et al., 2020; Hoehl and Wahl, 2012; Partanen et al., 2013; Maurer et al., 2003; Putkinen et al., 2014; see also Virtala et al., 2002 for a recent discussion on positive vs. negative MMR in a large longitudinal study). Similarly, positive and negative MMRs have been reported in infants between 2 and 6 months of age, suggesting that experiments with small sample sizes might be more likely to result in erroneous results than adult studies (Trainor et al., 2003). Various versions of this paradigm have been developed depending on the aim of the study and the infant's age (see Virtala et al., 2022 for a longitudinal study). For instance, the familiarization-test oddball paradigm is adapted for more complex stimuli such as words, nonwords, or longer sequences of sounds (Bosseler et al., 2016; François et al., 2017; Kabdebon et al., 2015; Teixidó et al., 2018). Importantly, the oddball paradigm can include multiple deviant types in one experimental session (Näätänen et al., 2005; Partanen et al., 2013). Using multiple deviants inserted in the sequence (also known as *Multi-feature Oddball* or *Optimum 1;* Näätänen et al., 2004) is a methodologically sound choice for developmental EEG research. First, it allows to partial out the effect of probability from the nature of the deviance *per se*. Using three different deviants (e.g., duration, pitch, and intensity; 10% each) allows to directly compare the processing of different deviant stimuli neat of their absolute frequency while keeping the SNR constant. Second, it allows testing of up to four or five different deviants in a single session, thus providing a wide-range profile of brain responses according to stimulus deviance properties. This experimental paradigm has proven particularly useful for investigating infants' language development (Lovio et al., 2009; Partanen et al., 2013). Another modified version of the oddball paradigm, particularly useful to study speech processing in early infancy, involves multiple trains of five stimuli with four standards and a final deviant that can differ according to either local or global perceptual features (Dehaene-Lambertz and Dehaene, 1994; Dehaene-Lambertz, 2000; Bekinschtein et al., 2009). Importantly, this modified paradigm disentangles the effect of response habituation to repetition (the progressive peak amplitude reduction across the standards) from deviant-related processes. Finally, a multimodal version of the oddball paradigm that involves the simultaneous presentation of auditory and visual stimuli can also be used to evaluate audiovisual integration in 5-month-old infants (Kushnerenko et al., 2008).

Although extensively used in developmental EEG studies, the oddball paradigm is not the only auditory paradigm available. Priming paradigms are particularly adapted to study lexico-phonological processing in developmental populations (Becker, Schild, and Friedrich, 2014; Friedrich and Friederici, 2005). Such paradigms allow analyzing the effect of a "prime" stimulus on a subsequently presented target which can provide relevant information on the interaction between phonology and semantic processes early in development. Becker et al. (2014) investigated the emergence of phonological representations and predictive processing in young infants. Specifically, they presented 6-to-24-month-old infants with word onset primes that were phonologically congruent ("ma" before "Mama") or incongruent with a target word ("so" before "Mama"). As expected from adult data, the authors observed an immature N100 in the congruent condition reflecting abstract phonological processing (Friedrich et al., 2009). However, they only reported a P350-like effect in the group of 24 months old infants suggesting that phonological processing may not interact with lexical access before age two. Multimodal priming paradigms with visual stimuli are also used to study word recognition and semantic priming during development (Friedrich and Friederici, 2004; Mani and Plunkett, 2008). For example, Friedrich and Friederici reported that 19-month-old infants show an N400 to incongruent condition, suggesting that semantic processing interacts with word recognition mechanisms early in life. Other paradigms used in developmental EEG studies have evaluated speech segmentation ability based on statistical learning using ERP analyses (Bosseler et al., 2016; François et al., 2017; Kabdebon et al., 2015; Kooijman et al., 2005). These experiments rely on the human ability to detect statistical regularities to segment words contained in continuous speech (Saffran et al., 1996). For example, Bosseler et al. (2016) showed that newborns' brains are sensitive to prosodic information and syllable position for speech segmentation, indicating that statistical learning processes can already be observed from the first days of life (see also François et al., 2017). Overall, various auditory paradigms have been successfully adapted to babies, infants, and children that have provided crucial information about the early neurophysiological changes that occur during the maturation of the auditory pathway.

2.2.2. *Visual paradigms*

In the visual modality, the priming or repetition suppression paradigm has mainly been used to study infants' processing of faces, movements, or categories (Gliga and Dehaene-Lambertz, 2007; Webb and Nelson, 2001; Geangu et al., 2021; Jeschonek et al., 2010). In adults, the N170 is consistently observed in response to faces (Kappenman and Luck, 2011) and is larger for inverted than upright faces (de Haan et al., 2002). During development, the N290 may constitute a reliable precursor of the N170 as its latency decreases from 350 ms to 220 ms between 3 and 12 months (Halit et al., 2003; de Haan, 2013). The infants' N290 may assume the same functional role as in adults only after six months before this age; it is not sensitive to face inversion (de Haan et al., 2002). The P400 is another face-sensitive ERP response observed over posterior lateral electrodes between 3 and 12 months. This response has also been proposed as a precursor of the adult N170 since it shows similar face-dependent amplitude modulation (Gliga and Dehaene-Lambertz, 2007). However, these two ERP components have different morphology, polarity, and scalp distribution, suggesting that the underlying neural generators may undergo dramatic changes during the first year of life. While the exact functional meaning of the N290 and P400 has not been entirely understood yet, both ERPs have been successfully used to predict both typical and atypical development (Leppänen et al., 2007; Gredebäck et al., 2015; McCleery et al., 2009). The Pb ("positive before") develops beween 200 to 400 ms after stimulus onset and is associated with stimulus expectancy or contextual processing during visual oddball paradigms (Karrer et al., 1998). Its latency decreases, whereas its amplitude and sensitivity to experimental changes increase with age (Webb et al., 2005). The Nc is another ERP component originally associated with stimulus familiarity and frequency in visual paradigms (Courchesne et al., 1981; Karrer and Ackles, 1987; de Haan and Nelson, 1997, 1999 Nelson and Collins, 1992; Kopp and Lindenberger, 2011). However, most recent studies linked this component to attention allocation (di Lorenzo et al., 2020; Webb et al., 2005; see Conte et al., 2020 for a recent review). In newborns, the Nc shows a maximum amplitude between 1000 and 1200 ms after stimulus onset and around 500 ms in 12-month-olds (DeBoer et al., 2004; Nelson, 1997). Nonetheless, recent studies reported an Nc peaking between 400 and 600 ms even in infants younger than 12 months (di Lorenzo et al., 2020, Conte et al., 2020, Xie et al., 2019). Finally, the Positive Slow Wave (PSW, 1000 ms after stimulus onset) can occur when familiar faces are similar to unfamiliar ones and may thus represent visual working memory updating (de Haan and Nelson, 1997).

3. Working with young participants: advanced methods

EEG recordings are particularly relevant to understanding the maturational processes that support the emergence and development of early brain oscillations (Schaworonkow and Voytek, 2021). A crucial aspect of developmental EEG studies is that the structural and functional changes support the acquisition of increasingly complex behaviors (Johnson, 2001). Interestingly, learning may be supported by the synchronized activity of long-distance brain regions (Fries, 2005), and the intrinsic hierarchical organization of brain oscillations may facilitate stimulus processing (Lakatos et al., 2005). Besides, phase synchrony between different brain regions has been proposed as a mechanism for long-range communication of information (Varela et al., 2001; Lachaux et al., 1999) that is required during specific cognitive tasks (Buzsáki and Dragun, 2004). Therefore, TF analyses applied to infants' data may be particularly relevant and more informative than ERPs in studying the emergence of high-level cognitive functions.

3.1. Spectro-temporal analyses

Classical TF analyses of EEG data allow estimating the dynamic amplitude and phase modulations in different frequencies by computing the absolute power, the relative power, the coherence, or the hemispheric asymmetry in specific frequency bands. Indeed, TF transforms can be applied to resting-state or task-related data allowing the analysis of evoked and induced oscillatory brain activity (de Haan, 2013; Tallon-Baudry and Bertrand, 1999). The Fourier transform allows the decomposition of the signal in different frequency bands such as delta waves (<3.5 Hz), theta waves (4–7 Hz), alpha waves (8–13 Hz), beta waves (13–30 Hz), and gamma waves (>30 Hz). These different frequency bands described in infants' EEG research are slightly lower than those reported in adults but may support similar cognitive processes such as attention, memory, and emotion (Saby and Marshall, 2012).

Theta, alpha, beta, and gamma activity have been reported during different experimental paradigms and across different ages, with results showing modulations in power and phase synchrony throughout development (Begus and Bonawitz, 2020; Bazhenova et al., 2007; Bell and Wolfe, 2007; Bosseler et al., 2013; Musacchia et al., 2013; Orekhova et al., 2006, 2014; Ortiz-Mantilla et al., 2013; Stroganova et al., 1998; Csibra et al., 2000; Kaufmann, Csibra and Johnson, 2003). The alpha wave, with a frequency of 6-9 Hz in infants, is related to visual attention, emotional expression, working memory, and inhibitory control (Fox et al., 2001; Morasch and Bell, 2011; Orekhova et al., 2001). For example, Bell and Wolfe (2007) performed TF analyses of EEG data collected during a working-memory task in fifty infants and at two time points (8-m-o and four y-o). Results revealed that relative to baseline, working memory process was associated with an increased alpha power (6–9 Hz) over all the electrodes at eight months but over medial frontal electrodes only at four years. This more focal activity was interpreted as an increased functional specialization through development. The mu rhythm (8–13 Hz in adults) is observed in the sensorimotor cortex and associated with understanding others' actions using mirroring properties (Cuevas et al., 2014; Debnath et al., 2019).

TF analyses are particularly relevant to studying infants' cognitive repertoire during ecologically valid situations as they may help capture brain activity not observable using classic ERP analyses (Debnath et al., 2019; Morales and Bowers, 2022). In adults, increased alpha activity is observed at rest, whereas decreased alpha is triggered during cognitively demanding tasks. Besides, increased theta rhythm has been associated with high attentional and emotional loads (Klimesch, 1999). Interestingly, Orekhova et al. (2006) analyzed TF transforms obtained in 8- to 12-month-old infants during attentionally demanding situations such as interacting with an adult or exploring a new toy. Compared to baseline, results showed a clear reduction of alpha and an increase in theta during the more demanding situations. Because such live paradigms do not allow to easily spot brain activity time-locked to specific events required for ERPs, TF analyses may constitute a more appropriate tool to decipher the cognitive processes involved during ecologically valid situations. In addition, compared to passive studies, more interactive tasks such as conversations or games might allow longer recordings by enhancing children's interest. For example, toddlers' action observation during conversational turn-taking with caregivers may be associated with a suppression of mu rhythm (Liao et al., 2015; Meyer et al., 2022). Thus, TF analyses during ecologically valid paradigms may offer new perspectives into more socially-driven cognitive processes that ERP studies could hardly explore. In addition, different analyses can be combined to provide complementary information about the brain responses underlying specific cognitive processes (Isler et al., 2012). For instance, combined TF and ERP analyses have been successfully used to study brain responses from sleeping newborns during an auditory MMN paradigm (Isler et al., 2012). While TF analysis revealed a theta power increase to deviant stimuli at multiple scalp locations and latencies, ERPs failed to detect a difference between standard and deviant stimuli. Similarly, combined TF and functional connectivity analyses have been used to show that mu desynchronization during movement mirroring may occur in both motor and occipital areas, thus supporting a functional coupling between these two brain networks (Debnath et al., 2019). If classical TF analyses of EEG data can be very useful to study early brain functioning during realistic situations, new advanced analyses have been recently developed and may provide further insight into early brain organization.

3.2. Advanced spectro-temporal analyses

Derived from TF analyses, functional connectivity measures of resting-state or task-related data have also gained interest in the developmental research community (Boersma et al., 2011; He et al., 2019; Tóth et al., 2017; Tokariev et al., 2016; 2019; Tran et al., 2021; Cuevas et al., 2012). For example, Tokariev et al. (2016) studied 38 sleeping newborns during active and quiet sleep while recording EEG from 19 electrodes at two time points after birth. The results showed that vigilance states and postnatal age largely modulated functional connectivity measures. Similarly, Mariscal et al. (2021) recently examined the spectral changes of EEG data collected during resting-state in 98 children from three months to three years. The analysis of phase-amplitude-coupling between alpha, beta, and gamma frequencies showed opposite patterns of preferential coupling over anterior and posterior electrodes that increased with age, revealing the building of early regional phase preferential coupling. In another recent study (Xie et al., 2019), fifty-nine infants were evaluated cross-sectionally at 6 ($N = 15$), 8 ($N = 14$), 10 ($N = 17$), and 12 months of age ($N = 13$) during a sustained attention paradigm while EEG was recorded. The authors applied graph theory analyses to (i) weighted phase lag index measures (wPLI; Vinck et al., 2011) computed between scalp electrodes and (ii) seed-based connectivity measures computed between reconstructed sources. At the source level, seed-based connectivity analyses showed that sustained attention induced a decrease in the alpha band localized within the dorsal attention and default mode networks. Importantly, the strength of functional connectivity in this alpha band increased with age within specific brain networks only (visual, somatosensory, dorsal attention, and ventral attention networks). These results suggest that the development of early sustained attention is accompanied by distinct brain functional connectivity patterns representing different functional brain-network topologies that may go through rapid configurational changes during early development. In the case of sensor-space analyses, the wPLI can be particularly relevant to reducing the effects of volume conduction since zero-lag synchronization is left out from the analysis, thus implying that the result is not based on shared neural sources (Bastos and Schoffelen, 2016). Nonetheless, care must be taken when interpreting functional connectivity results in young participants because the number of electrodes constrains the accuracy of the measures and the duration of the recordings impacts these measures

(Tokariev et al., 2016; Haartsen et al., 2020). Moreover, synchrony analyses at the scalp level may give ambiguous information about the underlying brain networks due to the physical changes occurring during development. Therefore, translating scalp-level coherence patterns to functional connectivity in the source space requires a certain degree of cautiousness (Hu et al., 2010; Schiff, 2005; see Michel and Murray, 2012 for a review).

Continuous natural speech, music, or movies can also be used while recording EEG in infants (Begus et al., 2015; Carver and Vaccaro, 2007; Leong et al., 2017; Striano et al., 2006; Hoehl et al., 2014a, 2014b). Further, dual-EEG recordings during live interactions can allow collecting simultaneous EEG in infants and caregivers (Piazza et al., 2020; Haresign et al., 2022; see Noreika et al., 2020 for a review on dual-EEG during development). Most of these studies consist of interactive sessions between infants and caregivers, allowing the analysis of brain activity to objects or complex adult behaviors. These paradigms allow researchers to create more ecologically valid conditions but are often harder to control; consequently, data are harder to analyze.

Nonetheless, time-resolved analyses exploring the online neural tracking of speech, music, movies, or live interactions can help study brain development in an ecological context. Such analyses capitalize on the brain property to adapt its oscillatory activity to the rhythm of a visual or auditory stimulus presented periodically (Regan, 1982; Picton et al., 2003). For example, in adults, scalp or intracranial EEG collected during continuous stimuli such as speech or musical pieces has been successfully used to demonstrate that neural tuning occurs at the specific frequencies contained in the input (Nozaradan et al., 2014; Giraud and Poeppel, 2012; Teng and Poeppel, 2020; Lerousseau et al., 2021; Henin et al., 2021; Elmer et al., 2021; Ramos-Escobar et al., 2021, 2022). Typically, adults' brain activity shows a clear peak at stimulation frequency for auditory, visual, and somatosensory continuous inputs (Galambos et al., 1981; Namerow et al., 1974; Norcia et al., 2015). This type of analysis has been applied to various paradigms using auditory and visual continuous stimulations (see Kabdebon et al., 2022, in this special issue). For instance, previous studies in adults show that auditory cortical EEG activity can robustly track the speech envelope through neural synchronization to specific rates such as the syllabic and phrasal rates (Giraud and and Poeppel, 2012; Ding and Simon, 2012, 2014). Such neural tracking can be quantified with different approaches that rely on (i) the computation of the coherence between the auditory input and the EEG signal, (ii) the analysis of power and phase variability across trials, and (iii) the frequency tagging approach that quantifies the rhythmicity of the neural response in the frequency domain. Coherence relies on the linear correlation between the "entrained" EEG signal and specific features of long stimuli such as musical pieces or natural sentences (Doelling and Poeppel, 2015; Mesgarani et al., 2014; Peelle et al., 2013). This approach ignores the phase information and thus provides an overall estimation of neural tracking during the stimulation. In infants, frequency tagging has been recently used to study the development of high-level cognitive processes such as speech segmentation, grammatical rule-learning, musical beat processing or even face processing (Fló et al., 2022a; Choi et al., 2020; Kabdebon et al., 2015; Cirelli et al., 2016; Buiatti et al., 2019; de Heering and Rossion, 2015; Leleu et al., 2020; Köster, Langeloh, and Hoel, 2019).

A recent method to study the early development of brain functioning in response to continuous and ecologically valid stimuli relies on the so-called multivariate Temporal Response Function (mTRF, Crosse et al., 2016). This method uses linear regression to quantify the relationship between a continuous EEG signal and an ongoing stimulus. The mTRFs have been successfully applied to infants' EEG data to unravel the mechanisms supporting sensory processing of natural stimuli such as movies or spoken sentences (Attaheri et al., 2022; Jessen et al., 2019, 2021; Kalashnikova et al., 2018; Menn et al., 2022; see also Ashton et al., 2022). Briefly, it allows the mapping between specific features of continuous stimuli and the corresponding neural response in a forward and backward manner. The forward or encoding modeling approach predicts the neural response based on specific input features. It has been used to decipher the gradual weighting of speech auditory features during early development (Kalashnikova et al., 2018). By contrast, backward or decoding models allow the opposite estimation in a multivariate manner, predicting the sensory input based only on the neural EEG data. In the context of developmental EEG studies, decoding models can be particularly relevant to compare the amount of neural tracking depending on the attentional level or the linguistic environment.

Another advanced method in EEG research aims to estimate the cortical generators underlying the activity obtained in the sensor space. Accurate source localization of several ERP components across different ages based on distributed or dipole source models has now been made possible through the use of high-density EEG systems adapted to infants (Guy et al., 2016; Hämäläinen et al., 2011; Mento and Valenza, 2016; Ortiz-Mantilla et al., 2012, 2013, 2019; Xie and Richards, 2017; Lunghi et al., 2019). For example, Xie et al. (2019) collected cross-sectional EEG data from 5- (N = 49), 7- (N = 50), and 12-month-old (N = 51) infants during the presentation of faces. Results showed successful source reconstruction of the N290 and P400/Nc in the fusiform gyrus and posterior cingulate cortex. Similarly, Mento and Valenza (2016) reported the source reconstruction of the sustained face anticipatory contingent negative variation (CNV) in a group of fifteen 9-month-old infants. They showed that a distributed fronto-parietal cortical network was engaged about one second before the onset of faces when these were temporally predicted through a peek-a-boo play simulation. Remarkably, the presence of reliable stimulus anticipatory ERP activity in young infants highlights the importance of targeting pre-stimulus bran activity as a proxy to study predictive cognition from a developmental perspective.

Overall, the combined scalp- and source-level investigation suggests that EEG may also be an appropriate neuroimaging tool for the developmental community, as proposed for adults (Michel and Murray, 2012). Noteworthy, a direct comparison of brain source reconstruction across different ages is difficult due to brain structural differences, brain volume conduction, scalp thickness, and dipole orientation (Reynolds and Richards, 2009). Nevertheless, the use of a high-density electrode array (i.e., ≥128), together with other age-appropriate adjustments, has allowed applying this analysis to developmental research successfully (Bathelt et al., 2014). However, considerable caution is still required, especially for young children. A critical issue is that cortical generators' reconstruction should be performed using individual MRI to locate specific anatomical landmarks as precisely as possible (Kabdebon et al., 2015). Indeed, a wrong anatomical representation, as in the case of inappropriate MRI templates, could turn into aberrant source-map solutions. Further, age-related anatomical peculiarities, as in the case of newborns' fontanels, may considerably bias the inverse modeling (Lew et al., 2013; Gargiulo et al., 2015). The ideal solution to limit these concerns is to obtain single-subject MRI scans to inform the forward modeling most realistically, considering the individual differences in the cortical morphology. This issue is crucial, especially when comparing different age groups. One of the most challenging methodological goals of developmental EEG research is to build accurate models of age-related diffusion parameters that consider the changes in the physical properties of infants' developing brains. Indeed, because infants' ERPs tend to be larger than adults due to a better diffusion, fewer trials are needed to obtain clear ERPs but with lower SNR. Therefore, such models would greatly help in considering SNR variability at different ages. Acquiring anatomical MRI is not always possible; thus, one may use age-appropriate MRI age templates that are publicly available instead (Almli et al., 2007; Richards et al., 2016). Recently, several groups collected MRI scans of children of different ages cross-sectionally. In particular, Sanchez et al. (2012) provided age-specific MRI brain templates for thirteen age groups from birth through 4.3 years of age based on more than 150 scans. More recently, another research group released pediatric brain surface templates for infants between two weeks and two years of age (O'reilly et al., 2021). Importantly, O'Reilly and colleagues

evaluated the validity of the MRI templates by performing source localization of visual ERP components obtained in 100 high-density EEG datasets from this age range. Results revealed accurate source reconstruction with differences in estimated sources increasing with template age differences. Similarly, Kabdebon et al. (2014) examined the relationship between the 10–20 sensor placement and the underlying cerebral structures in 16 infants (3–17 weeks post-term). The authors quantified the variability in the brain-electrodes distance and the sensor locations between infants based on the anatomical variability of six main cortical sulci of an infant template parcelled in 94 regions. Results revealed increased scalp-brain distances from occipital to frontal and inferior to superior regions thus highlighting the importance of acquiring individual MRI cortical generators' for optimal source reconstruction.

4. Standardized processing pipelines

The infants' EEG research community is much smaller than the adult one; thus, it does not rely on a myriad of toolboxes with various standardized methods. In adults, EEG data processing can be done using EEGLAB (Delorme and Makeig, 2004), Fieldtrip (Oostenveld et al., 2011), MNE (Gramfort et al., 2013), Brainstorm (Tadel et al., 2011), or any compatible EEG data processing software. However, while these toolboxes have been successfully used to analyze developmental EEG data, to our knowledge, only three processing pipelines have been purposely developed for this kind of population so far. Of particular importance for infants' EEG research is detecting and correcting either bad channels or large artifacts most of the time obtained in short continuous time segments, thus hampering classic algorithms that operate well on adult data. Here, we briefly describe the main characteristics and advantages of each standardized pipeline found in the literature.

First, the "Maryland analysis of developmental EEG pipeline" (MADE; Debnath et al., 2020; https://github.com/ChildDevLab/MADE-EEG-preprocessing-pipeline) is a Matlab-based toolbox that combines EEGLAB functions with customizable processing parameters that are well adapted for infants EEG data such as trial-level channel interpolation (Buzzell et al., 2019). The MADE toolbox includes different preprocessing steps such as filtering, automated identification/rejection of bad channels, identification, and removal of artifacts based on ICA, epoching, artifact rejection on epoched data using simple thresholding, bad channel interpolation, and re-referencing of epoched data. In particular, the bad channel identification is based on a multivariate approach that takes into account three different parameters (Hurst exponent, correlation with other channels, and channel variance; for more details, see Debnath et al., 2020). These three parameters are combined and normalized to obtain a Z score. Values higher than 3 allow marking bad electrodes. This system may allow a more accurate screening of electrodes with artifactual activity resulting in a more conservative outcome (fewer electrodes to interpolate). Another advantage of this pipeline is that it implements a modified version of the adjust tool (Mognon et al., 2011) already implemented in EEGLAB and created explicitly for automatic labeling of artifactual ICs (eye movements, blinks, muscular activity, others). In particular, the authors added the possibility of quantifying the alpha activity of each IC using this parameter as a constraint to keep the components. This specific option further increases the reproducibility and the reliability of ICs classification by making the preprocessing procedure more operator-independent. The MADE pipeline has been validated on resting-state and task-based EEG data from three groups (infants, children, and late adolescents) and compared to two more traditional processing pipelines that included the interpolation of bad channels or not. The results showed that the MADE toolbox significantly retained more trials than the other two pipelines for the three age ranges. Therefore, the MADE provides an appropriate pipeline for infants, children, and adolescents' EEG data based on automated processing steps, thus ensuring a higher degree of objectivity and replicability than manual operator-dependent processing pipelines. However, the MADE toolbox may not be recommended for EEG data collected with low-density montages (i.e., less than 32 electrodes) as it may require modifying the optimal default parameters used in most of the processing steps implemented in the MADE toolbox.

Second, the Harvard Automated Processing Pipeline for EEG (HAPPE; Gabard-Durnam et al., 2018; https://github.com/lcnhappe/happe) has been proposed to optimize the processing of short-duration resting-state and task-related recordings obtained with different electrode layouts and systems. Importantly, the HAPPE toolbox implements a specific preprocessing step involving a w-ICA that improves the decomposition of the EEG data (Castellanos and Makarov, 2006). It also implements the Multiple Artifact Rejection Algorithm (MARA) to automatically identify artifactual ICs, thus allowing a more objective classification than a manual one. Interestingly, the toolbox provides individual post-processing reports containing data quality metrics that allow the objective classification of each dataset for further analyses. The performance of the HAPPE toolbox has been assessed on ten infants' EEG datasets and compared with seven alternative methods that rely on the same pre-processing steps except for the w-ICA and MARA. Results revealed that the HAPPE toolbox outperformed the alternative methods by removing more artifacts while preserving a similar amount of data compared to the seven alternative methods. Furthermore, the authors provided the distribution of post-processing data quality metrics obtained in 867 infant datasets from 3 to 36 months of age, thus offering a normative database useful to classify future EEG files objectively. However, most of the alternatives relied on the same pre-processing steps that started with a 1 Hz low-pass filter preventing the analysis of ERPs. Besides, even though the implementation of combined w-ICA/MARA identified 42% of ICs as artifactual, this number increased to 85% when using MARA only, thus suggesting that further improvements are needed to optimize automatic classification algorithms for infants' EEG data (Haresign et al., 2021). Interestingly, an adapted version of the HAPPE toolbox has been recently proposed for low-density EEG montages (Lopez et al., 2021). However, no clear comparison with alternative pipelines is provided in the current version of the article.

Third, Fló et al. (2022b) have recently developed the "Automated Pipeline for Infants Continuous EEG" (APICE toolbox; https://github.com/neurokidslab/eeg_preprocessing), in which automatized artifact detection is performed on the continuous data before further preprocessing. APICE is Matlab-based and builds on the EEGLAB toolbox (for importing the data, filtering, and epoching) and custom functions. Of particular interest is the implementation of different algorithms for artifact detection and correction optimized for infants' EEG data. Specifically, the APICE toolbox performs ICA combined with the iMARA algorithm, an adapted version of the original MARA algorithm optimized for infants' EEG data, for automatic component selection (Haresign et al., 2021) and Denoise Source Separation (DSS, De Cheveigné and Parra, 2014; de Cheveigné and Simon, 2008). The APICE pipeline has been validated on high-density task-based ERP data obtained in neonates and infants. Specifically, the authors compared the number of trials retained and the standardized measurement error (SME; Luck et al., 2021) obtained with the APICE toolbox, a standard, and a MADE-based processing pipeline. Results revealed that the APICE toolbox significantly retained more trials and reached lower SME values than the other two pipelines for infant ERP datasets. The standard pipeline retained significantly more trials than the APICE toolbox for the neonatal dataset, but this was accompanied by higher SME values suggesting lower data quality. Overall, the APICE toolbox seems promising as it may ensure high flexibility, good data recovery, and thus high reproducibility across populations.

These three toolboxes offer many customizable outputs for further analyses using appropriate software. Importantly, even though they do not offer direct visualization of the significant results, they offer crucial individual data quality metrics that allow the objective classification of individual datasets and automatic artifacts identification. While univariate statistics are still highly used in infant EEG research, future studies should prefer cluster-based permutation tests as done in adults

(Meyer et al., 2021). Indeed, ERP components recorded during early development may be ambiguous and traditional peak picking methods may not be optimal. Cluster-based permutation analyses can overcome this difficulty by not placing any a priori assumptions on component latencies. Multivariate analyses such as TRFs may be well adapted for analyzing EEG data gathered during continuous sensory stimulations (Jessen et al., 2019; 2021; see also Ashton et al., 2022). Spatial filters that reduce data dimensionality or multiway generalized canonical correlation analyses may also be very useful for analyzing multiple high-density infants' EEG data (De Cheveigné and Parra, 2014; de Cheveigné and Simon, 2008; Gloaguen et al., 2020). Moreover, the EEG-BIDS format (for Brain Imaging Data Structure; BIDS) has gained interest in the adult community because it provides harmonized metadata and offers a common baseline for comparisons of results from similar studies (Gorgolewski et al., 2016; Pernet et al., 2019; Niso et al., 2018), Nonetheless, the EEG-BIDS format, with a few exceptions (Meyer et al., 2021; Van Noordt et al., 2020) is still under-represented in the developmental community and not considered in the three previously mentioned toolboxes.

5. Conclusion

The recent methodological advances observed in developmental EEG and signal processing are of utmost importance to better understand brain-mind interaction from a developmental point of view, progressively reducing the gap between adult and developmental neuroimaging studies. In other words, the old Berger's promise that the EEG would provide a "window into the brain" (Berger, 1929; Michel and Murray, 2012) is also becoming more and more attainable for developmental research. Nonetheless, working with developmental populations clearly constrains most of the choices related to the experimental design, procedure, and analyses. The classic ERP approach is still essential to ensure the comparability and reliability of the results and determine the developmental continuum of specific cognitive processes from early infancy to adulthood. However, ERPs may not fully capture some oscillatory brain patterns that explain specific developmental cognitive changes (Morales and Bowers, 2022). In this sense, TF analyses applied to resting-state or task-related data are particularly relevant for studying ecologically valid paradigms involving long recordings. Until the recent development of standardized processing pipelines adapted to infants' EEG data, these choices primarily relied on the specific expertise that developmental researchers had acquired, rendering them almost exclusively operator-dependent. Moreover, even though specific designs and procedures can be more suitable than others to tackle specific cognitive questions, such choices may inherently vary with the age of the participants. For instance, the physical characteristics of different ages are still a major issue that must be considered. Indeed, one of the most challenging methodological goals of developmental EEG research is to build accurate models of age-related diffusion parameters that consider the changes in the physical properties of infants' developing brains. Because of their physical characteristics, infants' ERPs tend to be larger than adults due to a better diffusion, and fewer trials are needed to obtain clear ERPs but with lower SNR. Such models will significantly help in considering SNR variability at different ages. Furthermore, longitudinal and cross-sectional developmental studies should prefer within- instead of between-group comparisons to determine how a given effect occurs at different ages, net of obvious anatomical maturational differences. Additionally, the brain changes reported in developmental studies should be associated with later behavioral standardized outcomes to control and minimize the effects of the rapid physiological changes that occur in these populations.

Therefore, by offering fine-grained signal processing tools best adapted to each age range and data, standardized pipelines will help promote replicability and comparability across experiments and ages. While the MADE pipeline might be well adapted for standardized processing of resting-state and task-based EEG data acquired with high-density EEG systems, it may not be compatible with data acquired with low-density EEG setups. The HAPPE pipeline appears more flexible because it allows processing of short-duration resting-state and task-related recordings obtained with different electrode layouts and systems. Nonetheless, the mandatory 1 Hz High-pass filter prevents using this toolbox for developmental studies targeting low-frequency ERP components. Finally, APICE may be the more flexible pipeline for data compatibility as it offers an extensive range of denoising tools adapted to developmental data. Although challenging, the acquisition of EEG during early development offers a unique opportunity to understand better the emergence of complex cognitive functions in humans. The development of standardized pipelines, new experimental paradigms, and analysis tools pave the way to a bright future for cognitive developmental EEG studies, inevitably bringing new challenges and exciting perspectives.

Data and code availability statement

No data presented in this manuscript was newly acquired. No code was specifically written for this manuscript.

Declaration of Competing Interest

The authors declare they have no conflict of interest.

Credit authorship contribution statement

Estelle Hervé: Conceptualization, Writing – original draft. **Giovanni Mento:** Conceptualization, Writing – review & editing. **Béatrice Desnous:** Conceptualization, Writing – review & editing. **Clément François:** Conceptualization, Supervision, Writing – review & editing.

Acknowledgments

This work was funded by the CNRS "80PRIME" grant, "DevLang project", and the French National Agency for Research (ANR), "BabyLang project" (ANR-20-CE28-0017) to CF. This work, carried out within the Institut Convergence ILCB (ANR-16-CONV-0002), has benefited from support from the French government, managed by the French National Agency for Research (ANR) and the Excellence Initiative of Aix-Marseille University (A ∗ MIDEX). Finally, this work was carried out under the research program "Dipartimenti di Eccellenza" 970 (art.1, commi 314–337 legge 232/2016), which was supported by a grant from MIUR to the Department of General Psychology, University of Padova. The authors would like to thank the three anonymous reviewers for their constructive comments on a previous version of the manuscript.

References

Alho, K., Sainio, K., Sajaniemi, N., Reinikainen, K., Näätänen, R., 1990a. Event-related brain potential of human newborns to pitch change of an acoustic stimulus. Electroencephalogr. Clin. Neurophysiol. Evoked Potentials Sect. 77 (2), 151–155.

Alho, K., Sajaniemi, N.K., Niittyvuopio, T., O Sainio, K., Näätänen, R, 1990b. ERPs to an auditory stimulus change in preterm and fullterm infants. In: Psychophysiological Brain Research. Tilburg University Press, pp. 139–142.

Almli, C.R., Rivkin, M.J., McKinstry, R.C.Brain Development Cooperative Group, 2007. The NIH MRI study of normal brain development (Objective-2): newborns, infants, toddlers, and preschoolers. Neuroimage 35 (1), 308–325.

Ashton, K., Zinszer, B.D., Cichy, R.M., Nelson, C.A., Aslin, R.N., Bayet, L, 2022. Time-resolved multivariate pattern analysis of infant EEG data: a practical tutorial. Dev. Cogn. Neurosci. 54, 101094.

Attaheri, A., Choisdealbha, Á.N., Di Liberto, G.M., Rocha, S., Brusini, P., Mead, N., Olawole-Scott, H., Boutris, P., Gibbon, S., Williams, I., Grey, C., Flanagan, S., Goswami, U., 2022. Delta-and theta-band cortical tracking and phase-amplitude coupling to sung speech by infants. Neuroimage 247, 118698.

Atkinson, J., Braddick, O., 2012. Visual attention in the first years: typical development and developmental disorders. Dev. Med. Child Neurol. 54 (7), 589–595.

Azzerboni, B., Foresta, F.L., Mammone, N., Morabito, F.C., 2005. A new approach based on wavelet-ICA algorithms for fetal electrocardiogram extraction. In: Proceedings of the ESANN, pp. 193–198.

Barnet, A. B., 1971. EEG Audiometry in Children Under Three Years of Age. Acta Oto-Laryngologica 72 (1–6), 1–13. doi:10.3109/00016487109122450.

Bastos, A.M., Schoffelen, J.M., 2016. A tutorial review of functional connectivity analysis methods and their interpretational pitfalls. Front. Syst. Neurosci. 9, 175.
Bathelt, J., O'Reilly, H., de Haan, M., 2014. Cortical source analysis of high-density EEG recordings in children. J. Vis. Exp. 88, 51705.
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. Int. J. Psychophysiol. 63 (1), 64–76.
Becker, A.B., Schild, U., Friedrich, C.K., 2014. ERP correlates of word onset priming in infants and young children. Dev. Cogn. Neurosci. 9, 44–55.
Begus, K., Bonawitz, E., 2020. The rhythm of learning: theta oscillations as an index of active learning in infancy. Dev. Cogn. Neurosci. 45, 100810.
Begus, K., Southgate, V., Gliga, T., 2015. Neural mechanisms of infant learning: differences in frontal theta activity during object exploration modulate subsequent object recognition. Biol. Lett. 11 (5), 20150041.
Bekinschtein, T. A., Dehaene, S., Rohaut, B., Tadel, F., Cohen, L., Naccache, L., 2009. Neural signature of the conscious processing of auditory regularities. Proceedings of the National Academy of Sciences 106 (5), 1672–1677. doi:10.1073/pnas.0809667106.
Bell, M.A., Cuevas, K., 2012. Using EEG to study cognitive development: issues and practices. J. Cogn. Dev. 13 (3), 281–294.
Bell, M.A., Wolfe, C.D., 2007. Changes in brain functioning from infancy to early childhood: evidence from EEG power and coherence working memory tasks. Dev. Neuropsychol. 31 (1), 21–38.
Bell, M.A., Wolfe, C.D., Schmidt, L.A., Segalowitz, S.J., 2008. The use of the electroencephalogram in research on cognitive development. In: Developmental Psychophysiology: Theory, Systems, and Methods. Cambridge University Press, pp. 150–170.
Berger, H., 1929. Über das elektroenkephalogramm des menschen. Arch. Psychiatr. Nervenkrankh. 87 (1), 527–570.
Bisiacchi, P.S., Mento, G., Suppiej, A., 2009. Cortical auditory processing in preterm newborns: an ERP study. Biol. Psychol. 82 (2), 176–185.
Boersma, M., Smit, D.J.A., de Bie, H.M.A., Van Baal, G.C.M., Boomsma, D.I., de Geus, E.J.C., Delemarre-van de Waal, H.A., Stam, C.J., 2011. Network analysis of resting state EEG in the developing young brain: Structure comes with maturation. Hum. Brain Mapp. 32 (3), 413–425.
Bosseler, A.N., Taulu, S., Pihko, E., Mäkelä, J.P., Imada, T., Ahonen, A., Kuhl, P.K., 2013. Theta brain rhythms index perceptual narrowing in infant speech perception. Front. Psychol. 4, 690.
Bosseler, A.N., Teinonen, T., Tervaniemi, M., Huotilainen, M., 2016. Infant directed speech enhances statistical learning in newborn infants: an ERP study. PLoS One 11 (9), e0162177.
Buiatti, M., Di Giorgio, E., Piazza, M., Polloni, C., Menna, G., Taddei, F., Baldo, E., Vallortigara, G., 2019. Cortical route for facelike pattern processing in human newborns. Proc. Natl. Acad. Sci. 116 (10), 4625–4630.
Buzsáki, G., Draguhn, A., 2004. Neuronal oscillations in cortical networks. Science 304 (5679), 1926–1929.
Buzzell, G.A., Barker, T.V., Troller-Renfree, S.V., Bernat, E.M., Bowers, M.E., Morales, S., Bowman, L.C., Henderson, H.A., Pine, D.S., Fox, N.A., 2019. Adolescent cognitive control, theta oscillations, and social observation. Neuroimage 198, 13–30.
Carver, L.J., Vaccaro, B.G., 2007. 12-month-old infants allocate increased neural resources to stimuli associated with negative adult emotion. Dev. Psychol. 43 (1), 54–69.
Castellanos, N.P., Makarov, V.A., 2006. Recovering EEG brain signals: artifact suppression with wavelet enhanced independent component analysis. J. Neurosci. Methods 158 (2), 300–312.
Čeponiene, R., Kushnerenko, E., Fellman, V., Renlund, M., Suominen, K., Näätänen, R., 2002. Event-related potential features indexing central auditory discrimination by newborns. Cogn. Brain Res. 13 (1), 101–113.
Cheour, M., Čeponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., Näätänen, R., 1998a. Development of language-specific phoneme representations in the infant brain. Nat. Neurosci. 1 (5), 351–353.
Cheour, M., Alho, K., Čeponiené, R., Reinikainen, K., Sainio, K., Pohjavuori, M., Aaltonen, O., Näätänen, R., 1998b. Maturation of mismatch negativity in infants. Int. J. Psychophysiol. 29 (2), 217–226.
Cheour, M., Kushnerenko, E., Čeponiene, R., Fellman, V., Näätänen, R., 2002a. Electric brain responses obtained from newborn infants to changes in duration in complex harmonic tones. Dev. Neuropsychol. 22 (2), 471–479.
Cheour, M., Čëponiené, R., Leppänen, P., Alho, K., Kujala, T., Renlund, M., Felmann, V., Näätänen, R., 2002b. The auditory sensory memory trace decays rapidly in newborns. Scand. J. Psychol. 43 (1), 33–39.
Cheour, M, Leppänen, P.H, Kraus, N, 2000. Mismatch negativity (MMN) as a tool for investigating auditory discrimination and sensory memory in infants and children. Clinical neurophysiology 111 (1), 4–16.
Cheour-Luhtanen, M., Alho, K., Sainio, K., Rinne, T., Reinikainenm, K., Pohjavuori, M., Renlund, M., Aaltonen, O., Eerola, O., Näätänen, R., 1996. The ontogenetically earliest discriminative response of the human brain. Psychophysiology 33, 478–481.
Choi, D., Batterink, L.J., Black, A.K., Paller, K.A., Werker, J.F., 2020. Preverbal infants discover statistical word patterns at similar rates as adults: evidence from neural entrainment. Psychol. Sci. 31 (9), 1161–1173.
Cirelli, L.K., Spinelli, C., Nozaradan, S., Trainor, L.J., 2016. Measuring neural entrainment to beat and meter in infants: effects of music background. Front. Neurosci. 10, 229.
Clarkson, M.G., Clifton, R.K., Swain, I.U., Perris, E.E., 1989. Stimulus duration and repetition rate influence newborns' head orientation toward sound. Dev. Psychobiol. 22 (7), 683–705.
Connell, J., Oozeer, R., de Vries, L., Dubowitz, L.M., Dubowitz, V., 1989. Continuous EEG monitoring of neonatal seizures: diagnostic and prognostic considerations. Arch. Dis. Child. 64 (4 Spec No), 452–458.
Conte, S., Richards, J.E., Guy, M.W., Xie, W., Roberts, J.E., 2020. Face-sensitive brain responses in the first year of life. Neuroimage 211, 116602.
Cotter, R.B., Burke, J.D., Loeber, R., Navratil, J.L., 2002. Innovative retention methods in longitudinal research: a case study of the developmental trends study. J. Child Fam. Stud. 11 (4), 485–498.
Courchesne, E., Ganz, L., Norcia, A.M., 1981. Event-related brain potentials to human faces in infants. Child Dev. 52 (3), 804–811.
Crosse, M.J., Di Liberto, G.M., Bednar, A., Lalor, E.C., 2016. The multivariate temporal response function (mTRF) toolbox: A MATLAB toolbox for relating neural signals to continuous stimuli. Front. Hum. Neurosci. 10.
Csibra, G., Davis, G., Spratling, M.W., Johnson, M.H., 2000. Gamma oscillations and object processing in the infant brain. Science 290 (5496), 1582–1585.
Cuevas, K., Cannon, E.N., Yoo, K., Fox, N.A., 2014. The infant EEG mu rhythm: methodological considerations and best practices. Dev. Rev. 34 (1), 26–43.
Cuevas, K., Raj, V., Bell, M.A., 2012. Functional connectivity and infant spatial working memory: a frequency band analysis: Infant frequency bands and EEG coherence. Psychophysiology 49 (2), 271–280.
de Cheveigné, A., Parra, L.C., 2014. Joint decorrelation, a versatile tool for multichannel data analysis. Neuroimage 98, 487–505.
de Cheveigné, A., Simon, J.Z., 2008. Denoising based on spatial filtering. J. Neurosci. Methods 171 (2), 331–339.
de Haan, M., 2013. Infant EEG and Event-Related Potentials. Psychology Press.
de Haan, M., Nelson, C. A., 1999. Brain activity differentiates face and object processing in 6-month-old infants. Developmental Psychology 35 (4), 1113–1121. doi:10.1037/0012-1649.35.4.1113.
de Haan, M., Pascalis, O., Johnson, M.H., 2002. Specialization of neural mechanisms underlying face recognition in human infants. J. Cogn. Neurosci. 14 (2), 199–209.
de Haan, M., Nelson, C.A., 1997. Recognition of the mother's face by six-month-old infants: a neurobehavioral study. Child Dev. 68 (2), 187–210.
de Heering, A., Rossion, B., 2015. Rapid categorization of natural face images in the infant right hemisphere. ELife, 4, 1–14.
Debnath, R., Salo, V.C., Buzzell, G.A., Yoo, K.H., Fox, N.A., 2019. Mu rhythm desynchronization is specific to action execution and observation: Evidence from time-frequency and connectivity analysis. Neuroimage 184, 496–507.
Debnath, R., Buzzell, G.A., Morales, S., Bowers, M.E., Leach, S.C., Fox, N.A., 2020. The Maryland analysis of developmental EEG (MADE) pipeline. Psychophysiology 57 (6), e13580.
DeBoer, T., Scott, L., Nelson, C.A., Handy, T., 2004. Event-related potentials in developmental populations. In: Event-Related Potentials: A Methods Handbook. MIT Press, pp. 263–297.
Dehaene-Lambertz, G., Dehaene, S., 1994. Speed and cerebral correlates of syllable discrimination in infants. Nature 370 (6487), 292–295.
Dehaene-Lambertz, G., 2000. Cerebral specialization for speech and non-speech stimuli in infants. J. Cogn. Neurosci. 12 (3), 449–460.
Delorme, A., Makeig, S., 2004. EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Methods 134 (1), 9–21.
Di Lorenzo, R., van den Boomen, C., Kemner, C., Junge, C., 2020. Charting development of ERP components on face-categorization: results from a large longitudinal sample of infants. Dev. Cogn. Neurosci. 45, 100840.
Ding, N., Simon, J.Z., 2012. Emergence of neural encoding of auditory objects while listening to competing speakers. Proc. Natl. Acad. Sci. 109 (29), 11854–11859.
Ding, N., Simon, J.Z., 2014. Cortical entrainment to continuous speech: functional roles and interpretations. Front. Hum. Neurosci. 8, 311.
Doelling, K.B., Poeppel, D., 2015. Cortical entrainment to music and its modulation by expertise. Proc. Natl. Acad. Sci. 112 (45), E6233–E6242.
Duarte, A., Hearons, P., Jiang, Y., Delvin, M.C., Newsome, R.N., Verhaeghen, P., 2013. Retrospective attention enhances visual working memory in the young but not the old: an ERP study. Psychophysiology 50 (5), 465–476.
Elmer, S., Valizadeh, S.A., Cunillera, T., Rodriguez-Fornells, A., 2021. Statistical learning and prosodic bootstrapping differentially affect neural synchronization during speech segmentation. Neuroimage 235, 118051.
Fitzgibbon, S.P., Powers, D.M.W., Pope, K.J., Clark, C.R., 2007. Removal of EEG noise and artifact using blind source separation. J. Clin. Neurophysiol. 24 (3), 232–243.
Fló, A., Benjamin, L., Palu, M., Dehaene-Lambertz, G., 2022a. Sleeping neonates track transitional probabilities in speech but only retain the first syllable of words. Sci. Rep. 12 (1), 1–13.
Fló, A., Gennari, G., Benjamin, L., Dehaene-Lambertz, G., 2022b. Automated pipeline for infants continuous EEG (APICE): a flexible pipeline for developmental studies. Dev. Cogn. Neurosci., 101077.
Fox, N.A., Henderson, H.A., Rubin, K.H., Calkins, S.D., Schmidt, L.A., 2001. Continuity and discontinuity of behavioral inhibition and exuberance: psychophysiological and behavioral influences across the first four years of life. Child Dev. 72 (1), 1–21.
François, C., Rodriguez-Fornells, A., Teixidó, M., Agut, T., Bosch, L., 2021. Attenuated brain responses to speech sounds in moderate preterm infants at term age. Dev. Sci. 24 (1), e12990.
François, C., Teixidó, M., Takerkart, S., Agut, T., Bosch, L., Rodriguez-Fornells, A., 2017. Enhanced neonatal brain responses to sung streams predict vocabulary outcomes by age 18 months. Sci. Rep. 7 (1), 1–13.
Friederici, A.D., Friedrich, M., Weber, C., 2002. Neural manifestation of cognitive and precognitive mismatch detection in early infancy. Neuroreport 13 (10), 1251–1254.
Friedrich, M., Weber, C., Friederici, A.D., 2004. Electrophysiological evidence for delayed mismatch response in infants at-risk for specific language impairment. Psychophysiology 41 (5), 772–782.

Friedrich, M., Friederici, A.D., 2004. N400-like semantic incongruity effect in 19-month-olds: processing known words in picture contexts. J. Cogn. Neurosci. 16 (8), 1465–1477.
Friedrich, M., Friederici, A.D., 2005. Lexical priming and semantic integration reflected in the event-related potential of 14-month-olds. Neuroreport 16 (6), 653–656.
Friedrich, C.K., Schild, U., Röder, B., 2009. Electrophysiological indices of word fragment priming allow characterizing neural stages of speech recognition. Biol. Psychol. 80 (1), 105–113.
Fries, P., 2005. A mechanism for cognitive dynamics: Neuronal communication through neuronal coherence. Trends Cogn. Sci. 9 (10), 474–480.
Fujioka, T, Mourad, N, He, C, Trainor, L.J, 2011. Comparison of artifact correction methods for infant EEG applied to extraction of event-related potential signals. Clinical Neurophysiology 122 (1), 43–51.
Gabard-Durnam, L.J., Mendez Leal, A.S., Wilkinson, C.L., Levin, A.R., 2018. The Harvard automated processing pipeline for electroencephalography (HAPPE): standardized processing software for developmental and high-artifact data. Front. Neurosci. 12, 97.
Galambos, R., Makeig, S., Talmachoff, P.J., 1981. A 40-Hz auditory potential recorded from the human scalp. Proc. Natl. Acad. Sci. 78 (4), 2643–2647.
Galbraith, S., Bowden, J., Mander, A., 2017. Accelerated longitudinal designs: An overview of modelling, power, costs and handling missing data. Statistical methods in medical research 26 (1), 374–398. doi:10.1177/0962280214547150.
Gargiulo, P., Belfiore, P., Friðgeirsson, E.A., Vanhatalo, S., Ramon, C., 2015. The effect of fontanel on scalp EEG potentials in the neonate. Clin. Neurophysiol. 126 (9), 1703–1710.
Geangu, E., Roberti, E., Turati, C., 2021. Do infants represent human actions cross–modally? An ERP visual-auditory priming study. Biol. Psychol. 160, 108047.
Giraud, A.L., Poeppel, D., 2012. Cortical oscillations and speech processing: emerging computational principles and operations. Nat. Neurosci. 15 (4), 511–517.
Gliga, T., Dehaene-Lambertz, G., 2007. Development of a view-invariant representation of the human head. Cognition 102 (2), 261–288.
Gloaguen, A., Philippe, C., Frouin, V., Gennari, G., Dehaene-Lambertz, G., Le Brusquet, L., Tenenhaus, A., 2020. Multiway generalized canonical correlation analysis. Biostatistics 0 (0), 1–29.
Gomarus, H.K., Althaus, M., Wijers, A.A., Minderaa, R.B., 2006. The effects of memory load and stimulus relevance on the EEG during a visual selective memory search task: an ERP and ERD/ERS study. Clin. Neurophysiol. 117 (4), 871–884.
Gorgolewski, K.J., Auer, T., Calhoun, V.D., Craddock, R.C., Das, S., Duff, E.P., Flandin, G., Ghosh, S.S., Glatard, T., Halchenko, Y.O., Handwerker, D.A., Hanke, M., Keator, D., Li, X., Michael, Z., Maumet, C., Nichols, B.N., Nichols, T.E., Pellman, J., … Poldrack, R.A., 2016. The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. Sci. Data 3 (1), 1–9.
Gramfort, A., Luessi, M., Larson, E., Engemann, D.A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., Hämäläinen, M., 2013. MEG and EEG data analysis with MNE-Python. Front. Neurosci. 7, 267.
Graven, S., 2006. Sleep and brain development. Clin. Perinatol. 33 (3), 693–706.
Gredebäck, G., Kaduk, K., Bakker, M., Gottwald, J., Ekberg, T., Elsner, C., Reid, V., Kenward, B., 2015. The neuropsychology of infants' pro-social preferences. Dev. Cogn. Neurosci. 12, 106–113.
Grune, K., Metz, A.M., Hagendorf, H., Fischer, S., 1996. Information processing in working memory and event-related brain potentials. Int. J. Psychophysiol. 23 (1-2), 111–120.
Guy, M.W., Zieber, N., Richards, J.E., 2016. The cortical development of specialized face processing in infancy. Child Dev. 87 (5), 1581–1600.
Haartsen, R., van der Velde, B., Jones, E.J.H., Johnson, M.H., Kemner, C., 2020. Using multiple short epochs optimises the stability of infant EEG connectivity parameters. Sci. Rep. 10 (1), 12703.
Háden, G.P., Honing, H., Török, M., Winkler, I., 2015a. Detecting the temporal structure of sound sequences in newborn infants. Int. J. Psychophysiol. 96 (1), 23–28.
Háden, G.P., Németh, R., Török, M., Winkler, I., 2015b. Predictive processing of pitch trends in newborn infants. Brain Res. 1626, 14–20.
Halit, H., de Haan, M., Johnson, M.H., 2003. Cortical specialisation for face processing: face-sensitive event-related potential components in 3- and 12-month-old infants. Neuroimage 19 (3), 1180–1193.
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 6months of age. Neuroimage 54 (3), 1910–1918.
Haresign, I.M., Phillips, E., Whitehorn, M., Noreika, V., Jones, E.J.H., Leong, V., Wass, S.V., 2021. Automatic classification of ICA components from infant EEG using MARA. Dev. Cogn. Neurosci. 52, 101024.
Haresign, I.M., Phillips, E., Whitehorn, M., Goupil, L., Noreika, V., Leong, V., Wass, S.V., 2022. Measuring the temporal dynamics of inter-personal neural entrainment in continuous child-adult EEG hyperscanning data. Dev. Cogn. Neurosci., 101093.
He, W., Sowman, P.F., Brock, J., Etchell, A.C., Stam, C.J., Hillebrand, A., 2019. Increased segregation of functional networks in developing brains. Neuroimage 200, 607–620.
Henin, S., Turk-Browne, N.B., Friedman, D., Liu, A., Dugan, P., Flinker, A., Melloni, L., 2021. Learning hierarchical sequence representations across human cortex and hippocampus. Sci. Adv. 7 (8), eabc4530.
Hladký, Havlíček, 2013. Was Tinbergen an Aristotelian? Comparison of Tinbergen's four whys and Aristotle's four causes. Hum. Ethol. Bull. 28 (4), 3–11.
Hoehl, S., Wahl, S., 2012. Recording infant ERP data for cognitive research. Dev. Neuropsychol. 37 (3), 187–209.
Hoehl, S., Wahl, S., Pauen, S., 2014a. Disentangling the effects of an adult model's eye gaze and head orientation on young infants' processing of a previously attended object. Infancy 19 (1), 53–64.
Hoehl, S., Michel, C., Reid, V.M., Parise, E., Striano, T., 2014b. Eye contact during live social interaction modulates infants' oscillatory brain activity. Soc. Neurosci. 9 (3), 300–308.
Hurwitz, L.B., Schmitt, K.L., Olsen, M.K., 2017. Facilitating development research: suggestions for recruiting and re-recruiting children and families. Front. Psychol. 8, 1525.
Hosain, S, La Vega-Talbott, M, Solomon, G, Green, N, 2003. Apneic Seizures in Infants: Role of Continuous EEG Monitoring. Clinical Electroencephalography 34 (4), 197–200.
Hu, S., Stead, M., Dai, Q., Worrell, G.A., 2010. On the recording reference contribution to EEG correlation, phase synchrony, and coherence. IEEE Trans. Syst. Man Cybern. B Cybern. 40, 1294–1304.
Isler, J.R., Tarullo, A.R., Grieve, P.G., Housman, E., Kaku, M., Stark, R.I., Fifer, W.P., 2012. Toward an electrocortical biomarker of cognition for newborn infants. Dev. Sci. 15 (2), 260–271.
Jeschonek, S., Marinovic, V., Hoehl, S., Elsner, B., Pauen, S., 2010. Do animals and furniture items elicit different brain responses in human infants? Brain Dev. 32 (10), 863–871.
Jessen, S., Fiedler, L., Münte, T.F., Obleser, J., 2019. Quantifying the individual auditory and visual brain response in 7- month-old infants watching a brief cartoon movie. Neuroimage 202, 116060.
Jessen, S., Obleser, J., Tune, S., 2021. Neural Tracking in Infants – An Analytical Tool for Multisensory Social Processing in Development, 52. Dev. Cogn. Neurosci..
Johnson, M.H., de Haan, M., Oliver, A., Smith, W., Hatzakis, H., Tucker, L.A., Csibra, G., 2001. Recording and analyzing high-density event-related potentials with infants using the geodesic sensor net. Dev. Neuropsychol. 19 (3), 295–323.
Johnson, M.H., 2001. Functional brain development in humans. Nat. Rev. Neurosci. 2 (7), 475–483.
Jonkman, L.M., 2006. The development of preparation, conflict monitoring and inhibition from early childhood to young adulthood; a Go/Nogo ERP study. Brain Res. 1097 (1), 181–193.
Kabdebon, C., Leroy, F., Simmonet, H., Perrot, M., Dubois, J., Dehaene-Lambertz, G., 2014. Anatomical correlations of the international 10–20 sensor placement system in infants. Neuroimage 99, 342–356.
Kabdebon, C., Pena, M., Buiatti, M., Dehaene-Lambertz, G., 2015. Electrophysiological evidence of statistical learning of long-distance dependencies in 8-month-old preterm and full-term infants. Brain Lang. 148, 25–36.
Kabdebon, C., Fló, A., de Heering, A., Aslin, R., 2022. The power of rhythms: how steady-state evoked responses reveal early neurocognitive development. Neuroimage 254, 119150.
Kadosh, K. C., Johnson, M. H., 2007. Developing a cortex specialized for face perception. Trends in Cognitive Sciences 11 (9), 367–369. doi:10.1016/j.tics.2007.06.007.
Kailaheimo-Lönnqvist, L., Virtala, P., Fandakova, Y., Partanen, E., Leppänen, P.H.T., Thiede, A., Kujala, T., 2020. Infant event-related potentials to speech are associated with prelinguistic development. Dev. Cogn. Neurosci. 45, 100831.
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. Sci. Rep. 8 (1), 13745.
Kappenman, E.S., Luck, S.J., 2011. The Oxford Handbook of Event-Related Potential Components, 1st ed. Oxford University Press.
Karrer, J.H., Karrer, R., Bloom, D., Chaney, L., Davis, R., 1998. Event-related brain potentials during an extended visual recognition memory task depict delayed development of cerebral inhibitory processes among 6-month-old infants with down syndrome. Int. J. Psychophysiol. 29 (2), 167–200.
Karrer, R., Ackles, P.K., 1987. Visual event-related potentials of infants during a modified oddball procedure. Electroencephalogr. Clin. Neurophysiol. 40, 603–608.
Kaufman, J., Csibra, G., Johnson, M.H., 2003. Representing occluded objects in the human infant brain. Proc. R. Soc. Lond. B Biol. Sci. 270 (suppl_2), S140–S143.
Klimesch, W., 1999. EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. Brain Res. Rev. 29 (2-3), 169–195.
Kooijman, V., Hagoort, P., Cutler, A., 2005. Electrophysiological evidence for prelinguistic infants' word recognition in continuous speech. Cogn. Brain Res. 24 (1), 109–116.
Kopp, F., Lindenberger, U., 2011. Effects of joint attention on long-term memory in 9-month-old infants: an event-related potentials study. Dev. Sci. 14 (4), 660–672.
Köster, M., Langeloh, M., Hoehl, S., 2019. Visually entrained theta oscillations increase for unexpected events in the infant brain. Psychol. Sci. 30 (11), 1656–1663.
Kulke, L, Atkinson, J, Braddick, O, 2017. Neural mechanisms of attention become more specialised during infancy: Insights from combined eye tracking and EEG. Developmental psychobiology 59 (2), 250–260.
Kushnerenko, E., Ceponiene, R., Balan, P., Fellman, V., Näätänen, R., 2002. Maturation of the auditory change detection response in infants: a longitudinal ERP study. Neuroreport 13 (15), 1843–1848.
Kushnerenko, E., Teinonen, T., Volein, A., Csibra, G., 2008. Electrophysiological evidence of illusory audiovisual speech percept in human infants. Proc. Natl. Acad. Sci. 105 (32), 11442–11445.
Lachaux, J.P., Rodriguez, E., Martinerie, J., Varela, F.J., 1999. Measuring phase synchrony in brain signals. Hum. Brain Mapp. 8 (4), 194–208.
Lakatos, P., Shah, A. S., Knuth, K. H., Ulbert, I., Karmos, G., Schroeder, C. E., 2005. An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the the auditory cortex. Journal of neurophysiology 94 (3), 1904–1911.
Lehtonen, J., Valkonen-Korhonen, M., Georgiadis, S., Tarvainen, M.P., Lappi, H., Niskanen, J.P., Pääkkönen, A., Karjalainen, P.A., 2016. Nutritive sucking induces age-specific EEG-changes in 0-24 week-old infants. Infant behavior & development 45 (Pt A), 98–108.
Leleu, A., Rekow, D., Poncet, F., Schaal, B., Durand, K., Rossion, B., Baudouin, J., 2020. Maternal odor shapes rapid face categorization in the infant brain. Dev. Sci. 23 (2), e12877.

Leong, V., Byrne, E., Clackson, K., Georgieva, S., Lam, S., Wass, S., 2017. Speaker gaze increases information coupling between infant and adult brains. Proc. Natl. Acad. Sci. 114 (50), 13290–13295.

Leppänen, J.M., Moulson, M.C., Vogel-Farley, V.K., Nelson, C.A., 2007. An ERP study of emotional face processing in the adult and infant brain. Child Dev. 78 (1), 232–245.

Leppänen, P.H.T., Eklund, K.M., Lyytinen, H., 1997. Event-related brain potentials to change in rapidly presented acoustic stimuli in newborns. Dev. Neuropsychol. 13 (2), 175–204.

Leppänen, P.H.T., Hämäläinen, J.A., Salminen, H.K., Eklund, K.M., Guttorm, T.K., Lohvansuu, K., Puolakanaho, A., Lyytinen, H., 2010. Newborn brain event-related potentials revealing atypical processing of sound frequency and the subsequent association with later literacy skills in children with familial dyslexia. Cortex 46 (10), 1362–1376.

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

Lerousseau, J.P., Trébuchon, A., Morillon, B., Schön, D., 2021. Frequency selectivity of persistent cortical oscillatory responses to auditory rhythmic stimulation. J. Neurosci. 41 (38), 7991–8006.

Liao, Y., Acar, Z.A., Makeig, S., Deak, G., 2015. EEG imaging of toddlers during dyadic turn-taking: Mu-rhythm modulation while producing or observing social actions. Neuroimage 112, 52–60.

Lloyd, R., Goulding, R., Filan, P., Boylan, G., 2015. Overcoming the practical challenges of electroencephalography for very preterm infants in the neonatal intensive care unit. Acta Paediatr. 104 (2), 152–157.

Lopez, K. L., Monachino, A. D., Morales, S., Leach, S. C., Bowers, M. E., & Gabard-Durnam, L. J. (2021). HAPPILEE: the Harvard automated processing pipeline in low electrode electroencephalography, a standardized software for low density EEG and ERP data. bioRxiv.

Lovio, R., Pakarinen, S., Huotilainen, M., Alku, P., Silvennoinen, S., Näätänen, R., Kujalaa, T., 2009. Auditory discrimination profiles of speech sound changes in 6-year-old children as determined with the multi-feature MMN paradigm. Clinical Neurophysiology 120 (5), 916–921. doi:10.1016/j.clinph.2009.03.010.

Luck, S.J., Stewart, A.X., Simmons, A.M., Rhemtulla, M., 2021. Standardized measurement error: a universal metric of data quality for averaged event-related potentials. Psychophysiology 58 (6), e13793.

Lunghi, M., Piccardi, E.S., Richards, J.E., Simion, F., 2019. The neural correlates of orienting to walking direction in 6-month-old infants: an erp study. Dev. Sci. 22 (6), e12811.

MacDougall-Shackleton, S.A., 2011. The levels of analysis revisited. Philos. Trans. R. Soc. B Biol. Sci. 366 (1574), 2076–2085.

Makeig, S., Bell, A. J., Jung, T.P., & Sejnowski, T. J. (1996). Independent component analysis of electroencephalographic data. *Advances in neural information processing systems*, 7.

Maguire, M.J, Magnon, G, Fitzhugh, A.E, 2014. Improving data retention in EEG research with children using child-centered eye tracking. Journal of neuroscience methods 238, 78–81.

Mammone, N., La Foresta, F., Morabito, F.C., 2012. Automatic artifact rejection from multichannel scalp EEG by wavelet ICA. IEEE Sens. J. 12 (3), 533–542.

Mani, N., Plunkett, K., 2008. Fourteen-month-olds pay attention to vowels in novel words. Dev. Sci. 11 (1), 53–59.

Mariscal, M.G., Levin, A.R., Gabard-Durnam, L.J., Xie, W., Tager-Flusberg, H., Nelson, C.A., 2021. EEG phase-amplitude coupling strength and phase preference: association with age over the first three years after Birth. Eneuro 8 (3).

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

Maurer, U., Bucher, K., Brem, S., Brandeis, D., 2003. Development of the automatic mismatch response: from frontal positivity in kindergarten children to the mismatch negativity. Clin. Neurophysiol. 114 (5), 808–817.

McCleery, J. P., Akshoomoff, N., Dobkins, K. R., Carver, L. J., 2009. Atypical Face Versus Object Processing and Hemispheric Asymmetries in 10-Month-Old Infants at Risk for Autism. Biological Psychiatry 66 (10), 950–957. doi:10.1016/j.biopsych.2009.07.031.

Menn, K.H., Michel, C., Meyer, L., Hoehl, S., Männel, C., 2022. Natural infant-directed speech facilitates neural tracking of prosody. Neuroimage 251, 118991.

Mento, G., Bisiacchi, P.S., 2012. Neurocognitive development in preterm infants: Insights from different approaches. Neurosci. Biobehav. Rev. 36 (1), 536–555.

Mento, G., Valenza, E., 2016. Spatiotemporal neurodynamics of automatic temporal expectancy in 9-month old infants. Sci. Rep. 6 (1), 36525.

Mento, G., Vallesi, A., 2016. Spatiotemporally dissociable neural signatures for generating and updating expectation over time in children: a high density-ERP study. Dev. Cogn. Neurosci. 19, 98–106.

Mento, G., Suppiej, A., Altoè, G., Bisiacchi, P.S., 2010. Functional hemispheric asymmetries in humans: electrophysiological evidence from preterm infants. Eur. J. Neurosci. 31 (3), 565–574.

Mento, G., 2013. The passive CNV: carving out the contribution of task-related processes to expectancy. Front. Hum. Neurosci. 7, 827.

Mercuri, E, Von Siebenthal, K, Tutuncuoglu, S, Guzzetta, F, Casaer, P, 1995. The effect of behavioural states on visual evoked responses in preterm and full-term newborns. Neuropediatrics 26, 211–213.

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

Meyer, M., van Schaik, Poli, F., Hunnius, S., 2021. How infant-directed actions enhance infants' attention, learning, and exploration: evidence from EEG and computational modeling. Dev. Sci. e13259.

Meyer, M., Chung, H., Debnath, R., Fox, N., Woodward, A.L., 2022. Social context shapes neural processing of others' actions in 9-month-old infants. J. Exp. Child. Psychol. 213, 105260.

Michel, C.M., Murray, M.M., 2012. Towards the utilization of EEG as a brain imaging tool. Neuroimage 61 (2), 371–385.

Miljković, N., Matić, V., Van Huffel, S., Popović, M. B., 2010. Independent Component Analysis (ICA) methods for neonatal EEG artifact extraction: Sensitivity to variation of artifact properties. 10th Symposium on Neural Network Applications in Electrical Engineering 19–21. doi:10.1109/NEUREL.2010.5644041.

Mognon, A., Jovicich, J., Bruzzone, L., Buiatti, M., 2011. ADJUST: an automatic EEG artifact detector based on the joint use of spatial and temporal features. Psychophysiology 48 (2), 229–240.

Monroy, C., Domínguez-Martínez, E., Taylor, B., Marin, O.P., Parise, E., Reid, V.M., 2021. Understanding the causes and consequences of variability in infant ERP editing practices. Dev. Psychobiol. 63 (8), e22217.

Morales, S., Bowers, M.E., 2022. Time-frequency analysis methods and their application in developmental EEG data. Dev. Cogn. Neurosci. 54, 101067.

Morasch, K.C., Bell, M.A., 2011. The role of inhibitory control in behavioral and physiological expressions of toddler executive function. J. Exp. Child. Psychol. 108 (3), 593–606.

Moser, J., Schleger, F., Weiss, M., Sippel, K., Dehaene-Lambertz, G., Preissl, H., 2020. Magnetoencephalographic signatures of hierarchical rule learning in newborns. Dev. Cogn. Neurosci. 46, 100871.

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

Musacchia, G., Choudhury, N.A., Ortiz-Mantilla, S., Realpe-Bonilla, T., Roesler, C.P., Benasich, A.A., 2013. Oscillatory support for rapid frequency change processing in infants. Neuropsychologia 51 (13), 2812–2824.

Näätänen, R., Pakarinen, S., Rinne, T., Takegata, R., 2004. The mismatch negativity (MMN): towards the optimal paradigm. Clin. Neurophysiol. 115 (1), 140–144.

Näätänen, R., Jacobsen, T., Winkler, I., 2005. Memory-based or afferent processes in mismatch negativity (MMN): a review of the evidence. Psychophysiology 42 (1), 25–32.

Näätänen, R., Paavilainen, P., Rinne, T., Alho, K., 2007. The mismatch negativity (MMN) in basic research of central auditory processing: a review. Clin. Neurophysiol. 118 (12), 2544–2590.

Namerow, N.S., Sclabassi, R.J., Enns, N.F., 1974. Somatosensory responses to stimulus trains: Normative data. Electroencephalogr. Clin. Neurophysiol. 37 (1), 11–21.

Nätäänen, R., 2000. Mismatch negativity (MMN): perspectives for application. International journal of psychophysiology : official journal of the International Organization of Psychophysiology 37 (1), 3–10. doi:10.1016/s0167-8760(00)00091-x.

Nelson, C.A., 1997. Electrophysiological correlates of memory development in the first year of life. In: Reese, H.W., Franzen, M.D. (Eds.), Biological and neuropsychological mechanisms: Life-span developmental psychology. Lawrence Erlbaum Associates, Inc.31, pp. 95–131.

Nelson, C.A., Collins, P.F., 1992. Neural and behavioral correlates of visual recognition memory in 4- and 8-month-old infants. Brain Cogn. 19 (1), 105–121.

Niso, G., Gorgolewski, K.J., Bock, E., Brooks, T.L., Flandin, G., Gramfort, A., Henson, R.N., Jas, M., Litvak, V., T Moreau, J., Oostenveld, R., Schoffelen, J.M., Tadel, F., Wexler, J., Baillet, S, 2018. MEG-BIDS, the brain imaging data structure extended to magnetoencephalography. Sci. Data 5 (1), 180110.

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

Noreika, V., Georgieva, S., Wass, S., Leong, V., 2020. 14 challenges and their solutions for conducting social neuroscience and longitudinal EEG research with infants. Infant Behav. Dev. 58, 101393.

Nozaradan, S., 2014. Exploring how musical rhythm entrains brain activity with electroencephalogram frequency-tagging. Philos. Trans. R. Soc. B Biol. Sci. 369 (1658), 20130393.

Oostenveld, R., Fries, P., Maris, E., Schoffelen, J.M., 2011. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput. Intell. Neurosci. 2011, 1–9.

O'Reilly, C., Larson, E., Richards, J.E., Elsabbagh, M., 2021. Structural templates for imaging EEG cortical sources in infants. Neuroimage 227, 117682.

Orekhova, E.V., Elsabbagh, M., Jones, E.J., Dawson, G., Charman, T., Johnson, M.H., 2014. EEG hyper-connectivity in high-risk infants is associated with later autism. J. Neurodev. Disord. 6 (1), 1–11.

Orekhova, E., Stroganova, T., Posikera, I., Elam, M., 2006. EEG theta rhythm in infants and preschool children. Clin. Neurophysiol. 117 (5), 1047–1062.

Orekhova, E.V., Stroganova, T.A., Posikera, I.N., 2001. Alpha activity as an index of cortical inhibition during sustained internally controlled attention in infants. Clin. Neurophysiol. 112 (5), 740–749.

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.

Ortiz-Mantilla, S., Hamalainen, J.A., Musacchia, G., Benasich, A.A., 2013. Enhancement of gamma oscillations indicates preferential processing of native over foreign phonemic contrasts in infants. J. Neurosci. 33 (48), 18746–18754.

Ortiz-Mantilla, S., Realpe-Bonilla, T., Benasich, A.A., 2019. Early interactive acoustic experience with non-speech generalizes to speech and confers a syllabic processing advantage at 9 months. Cereb. Cortex 29 (4), 1789–1801.

Pang, E.W., Edmonds, G.E., Desjardins, R., Khan, S.C., Trainor, L.J., Taylor, M.J., 1998. Mismatch negativity to speech stimuli in 8-month-old infants and adults. Int. J. Psychophysiol. 29 (2), 227–236.

Partanen, E., Pakarinen, S., Kujala, T., Huotilainen, M., 2013. Infants' brain responses for speech sound changes in fast multifeature MMN paradigm. Clin. Neurophysiol. 124 (8), 1578–1585.

Partanen, E., Mårtensson, G., Hugoson, P., Huotilainen, M., Fellman, V., Ådén, U., 2022. Auditory processing of the brain is enhanced by parental singing for preterm infants. Front. Neurosci. 16.
Peelle, J.E., Gross, J., Davis, M.H., 2013. Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. Cereb. Cortex 23 (6), 1378–1387.
Pelegrina, S., Molina, R., Rodríguez-Martínez, E.I., Linares, R., Gómez, C.M., 2020. Age-related changes in selection, recognition, updating and maintenance information in WM. an ERP study in children and adolescents. Biol. Psychol. 157, 107977.
Pernet, C.R., Appelhoff, S., Gorgolewski, K.J., Flandin, G., Phillips, C., Delorme, A., Oostenveld, R., 2019. EEG-BIDS, an extension to the brain imaging data structure for electroencephalography. Sci. Data 6 (1), 103.
Piazza, E.A., Hasenfratz, L., Hasson, U., Lew-Williams, C., 2020. Infant and adult brains are coupled to the dynamics of natural communication. Psychol. Sci. 31 (1), 6–17.
Picton, T.W., Bentin, S., Berg, P., Donchin, E., Hillyard, S.A., Johnson, R., Miller, G.A., Ritter, W., Ruchkin, D.S., Rugg, M.D., Taylor, M.J., 2000. Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria. Psychophysiology 37 (2), 127–152.
Picton, T.W., John, M.S., Dimitrijevic, A., Purcell, D., 2003. Human auditory steady-state responses: respuestas auditivas de estado estable en humanos. Int. J. Audiol. 42 (4), 177–219.
Pressler, R., Bady, B., Binnie, C.D., Boylan, G.B., Connell, J.A., Lütschg, J., 2003. Neurophysiology of the neonatal period. In: Clinical Neurophysiology: EEG, Paediatric Neurophysiology, Special Techniques and Applications. Elsevier Health Sciences, pp. 450–506.
Putkinen, V., Tervaniemi, M., Saarikivi, K., de Vent, N., Huotilainen, M., 2014. Investigating the effects of musical training on functional brain development with a novel Melodic MMN paradigm. Neurobiol. Learn. Mem. 110, 8–15.
Ramos-Escobar, N., Segura, E., Olivé, G., Rodriguez-Fornells, A., François, C., 2021. Oscillatory activity and EEG phase synchrony of concurrent word segmentation and meaning-mapping in 9-year-old children. Dev. Cogn. Neurosci. 51, 101010.
Ramos-Escobar, N., Mercier, M., Trébuchon-Fonséca, A., Rodriguez-Fornells, A., François, C., Schön, D., 2022. Hippocampal and auditory contributions to speech segmentation. Cortex 150, 1–11.
Regan, D., 1982. Comparison of transient and steady-state methods. Ann. N. Y. Acad. Sci. 388 (1), 45–71.
Reynolds, G.D., Courage, M.L., Richards, J.E., 2010. Infant attention and visual preferences: converging evidence from behavior, event-related potentials, and cortical source localization. Dev. Psychol. 46 (4), 886–904.
Reynolds, G.D., Richards, J.E., 2005. Familiarization, attention, and recognition memory in infancy: an event-related potential and cortical source localization study. Dev. Psychol. 41 (4), 598–615.
Reynolds, G.D., Richards, J.E., 2009. Cortical source localization of infant cognition. Dev. Neuropsychol. 34 (3), 312–329.
Ribisl, K.M., Walton, M.A., Mowbray, C.T., Luke, D.A., Davidson, W.S., Bootsmiller, B.J., 1996. Minimizing participant attrition in panel studies through the use of effective retention and tracking strategies: Review and recommendations. Eval. Progr.Plan. 19 (1), 1–25.
Richards, J.E., Sanchez, C., Phillips-Meek, M., Xie, W., 2016. A database of age-appropriate average MRI templates. Neuroimage 124, 1254–1259.
Romo-Vázquez, R., Vélez-Pérez, H., Ranta, R., Louis Dorr, V., Maquin, D., Maillard, L., 2012. Blind source separation, wavelet denoising and discriminant analysis for EEG artefacts and noise cancelling. Biomed. Signal Process. Control 7 (4), 389–400.
Saby, J.N., Marshall, P.J., 2012. The utility of EEG band power analysis in the study of infancy and early childhood. Dev. Neuropsychol. 37 (3), 253–273.
Saffran, J.R., Aslin, R.N., Newport, E.L., 1996. Statistical learning by 8-month-old infants. Science 274 (5294), 1926–1928.
Sambeth, A., Huotilainen, M., Kushnerenko, E., Fellman, V., Pihko, E., 2006. Newborns discriminate novel from harmonic sounds: a study using magnetoencephalography. Clin. Neurophysiol. 117 (3), 496–503.
Sanchez, C.E., Richards, J.E., Almli, C.R., 2012. Neurodevelopmental MRI brain templates for children from 2 weeks to 4 years of age. Dev. Psychobiol. 54 (1), 77–91.
Schaworonkow, N., Voytek, B., 2021. Longitudinal changes in aperiodic and periodic activity in electrophysiological recordings in the first seven months of life. | Elsevier Enhanced Reader. Dev. Cogn. Neurosci. 47, 100895.
Schiff, S.J., 2005. Dangerous phase. Neuroinformatics 3, 315–318.
Schumacher, E.M., Westvik, A.S., Larsson, P.G., Lindemann, R., Westvik, J., Stiris, T.A., 2011. Feasibility of long-term continuous EEG monitoring during the first days of life in preterm infants: an automated quantification of the EEG activity. Pediatr. Res. 69 (5), 413–417.
Squires, K. C., Wickens, C., Squires, N. K., Donchin, E., 1976. The Effect of Stimulus Sequence on the Waveform of the Cortical Event-Related Potential. Science 193 (4258), 1142–1146. doi:10.1126/science.9598.
Stets, M., Stahl, D., Reid, V.M., 2012. A meta-analysis investigating factors underlying attrition rates in infant ERP studies. Dev. Neuropsychol. 37 (3), 226–252.
Striano, T., Reid, V.M., Hoehl, S., 2006. Neural mechanisms of joint attention in infancy. Eur. J. Neurosci. 23 (10), 2819–2823.
Stroganova, T.A., V Orekhova, E., Posikera, I.N, 1998. Externally and internally controlled attention in infants: an EEG study. Int. J. Psychophysiol. 30 (3), 339–351.
Suppiej, A., Mento, G., Zanardo, V., Franzoi, M., Battistella, P.A., Ermani, M., Bisiacchi, P.S., 2010. Auditory processing during sleep in preterm infants: an event related potential study. Early Hum. Dev. 86 (12), 807–812.
Tadel, F., Baillet, S., Mosher, J.C., Pantazis, D., Leahy, R.M., 2011. Brainstorm: a user-friendly application for MEG/EEG analysis. Comput. Intell. Neurosci. 2011, 1–13.
Tallon-Baudry, C., Bertrand, O., 1999. Oscillatory gamma activity in humans and its role in object representationactivity in humans and its role in object representation. Trends Cogn. Sci. 3 (4), 151–162.
Tanaka, M., Okubo, O., Fuchigami, T., Harada, K., 2001. A study of mismatch negativity in newborns. Pediatrics International 23 (3), 281–286. doi:10.1046/j.1442-200x.2001.01395.
Taylor, M. J., Edmonds, G. E., McCarthy, G., Allison, T., 2001. Eyes first! Eye processing develops before face processing in children. Neuroreport 12 (8), 1671–1676. doi:10.1097/00001756-200106130-00031.
Teixidó, M., François, C., Bosch, L., Männel, C., Prieto, P., Esteve-Gibert, N., 2018. The role of prosody in early speech segmentation and word-referent mapping: Electrophysiological evidence. In: The Development of Prosody in First Language Acquisition, 23. John Benjamins Publishing Company, pp. 79–100.
Teng, X., Poeppel, D., 2020. Theta and gamma bands encode acoustic dynamics over wide-ranging timescales. Cereb. Cortex 30 (4), 2600–2614.
Timeo, S., Mento, G., Fronza, E., Farroni, T., 2019. Exposure to linguistic labels during childhood modulates the neural architecture of race categorical perception. Sci. Rep. 9 (1), 17743.
Tinbergen, N., 1963. On aims and methods of Ethology. Z. Tierpsychol. 20 (4), 410–433.
Tokariev, A., Stjerna, S., Lano, A., Metsäranta, M., Palva, J.M., Vanhatalo, S., 2019. Preterm birth changes networks of newborn cortical activity. Cereb. Cortex 29 (2), 814–826.
Tokariev, A., Vanhatalo, S., Palva, J.M., 2016. Analysis of infant cortical synchrony is constrained by the number of recording electrodes and the recording montage. Clin. Neurophysiol. 127 (1), 310–323.
Tokariev, A., Videman, M., Palva, J.M., Vanhatalo, S., 2016. Functional brain connectivity develops rapidly around term age and changes between vigilance states in the human newborn. Cereb. Cortex 26 (12), 4540–4550.
Tokioka, A.B., Pearce, J.W., Crowell, D.H., 1995. Endogenous event-related potentials in term and preterm infants. J. Clin. Neurophysiol. 12 (5), 468–475 Official Publication of the American Electroencephalographic Society.
Tóth, B., Urbán, G., Haden, G.P., Márk, M., Török, M., Stam, C.J., Winkler, I., 2017. Large-scale network organization of EEG functional connectivity in newborn infants. Hum. Brain Map. 38 (8), 4019–4033.
Trainor, Laurel, McFadden, Melissa, Hodgson, Lisa, Darragh, Lisa, Barlow, Jennifer, Matsos, Laura, Sonnadara, Ranil, 2003. Changes in auditory cortex and the development of mismatch negativity between 2 and 6 months of age. International Journal of Psychophysiology 51 (1), 5–15. doi:10.1016/S0167-8760(03)00148-X.
Tran, X.A., McDonald, N., Dickinson, A., Scheffler, A., Frohlich, J., Marin, A., Kure Liu, C., Nosco, E., Şentürk, D., Dapretto, M., Spurling Jeste, S., 2021. Functional connectivity during language processing in 3-month-old infants at familial risk for autism spectrum disorder. Eur. J. Neurosci. 53 (5), 1621–1637.
van Noordt, S., Desjardins, J.A., Huberty, S., Abou-Abbas, L., Webb, S.J., Levin, A.R., Segalowitz, S.J., Evans, A.C., Elsabbagh, M., 2020. EEG-IP: An international infant EEG data integration platform for the study of risk and resilience in autism and related conditions. Mol. Med. 26, 1–11.
Vanhatalo, S., Voipio, J., Dewaraja, A., Holmes, M.D., Miller, J.W., 2003. Topography and elimination of slow EEG responses related to tongue movements. Neuroimage 20 (2), 1419–1423.
Varela, F., Lachaux, J.P., Rodriguez, E., Martinerie, J., 2001. The brainweb: phase synchronization and large-scale integration. Nat. Rev. Neurosci. 2 (4), t229–239.
Vinck, M., Oostenveld, R., Van Wingerden, M., Battaglia, F., Pennartz, C.M., 2011. An improved index of phase-synchronization for electrophysiological data in the presence of volume-conduction, noise and sample-size bias. Neuroimage 55 (4), 1548–1565.
Virtala, P., Putkinen, V., Kailaheimo-Lönnqvist, L., Thiede, A., Partanen, E., Kujala, T., 2022. Infancy and early childhood maturation of neural auditory change detection and its associations to familial dyslexia risk. Clin. Neurophysiol. 137, 159–176.
Webb, S.J., Nelson, C.A., 2001. Perceptual priming for upright and inverted faces in infants and adults. J. Exp. Child. Psychol. 79 (1), 1–22.
Webb, S.J., Long, J.D., Nelson, C.A., 2005. A longitudinal investigation of visual event-related potentials in the first year of life. Dev. Sci. 8 (6), 605–616.
Winkler, I., Debener, S., Muller, K.R., Tangermann, M., 2015. On the influence of high–pass filtering on ICA-based artifact reduction in EEG-ERP. In: Proceedings of the 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 4101–4105.
Winkler, I., Kushnerenko, E., Horváth, J., Čeponienė, R., Fellman, V., Huotilainen, M., Näätänen, R., Sussman, E., 2003. Newborn infants can organize the auditory world. In: Proceedings of the National Academy of Sciences of the United States of America, 100, pp. 11812–11815.
Wunderlich, J.L., Cone-Wesson, B.K., 2006. Maturation of CAEP in infants and children: a review. Hear. Res. 212 (1–2), 212–223.
Wunderlich, J.L., Cone-Wesson, B.K., Shepherd, R., 2006. Maturation of the cortical auditory evoked potential in infants and young children. Hear. Res. 212 (1–2), 185–202.
Xie, W., Richards, J.E., 2017. The relation between infant covert orienting, sustained attention and brain activity. Brain Topogr. 30 (2), 198–219.
Xie, W., McCormick, S.A., Westerlund, A., Bowman, L.C., Nelson, C.A., 2019. Neural correlates of facial emotion processing in infancy. Dev. Sci. 22 (3), e12758.
Xie, W., Mallin, B.M., Richards, J.E., 2019. Development of brain functional connectivity and its relation to infant sustained attention in the first year of life. Dev. Sci. 22 (1), e12703.

D. Daily auditory environments in French-speaking infants: A longitudinal dataset: full text

Full text of Hervé et al. (2022):

Hervé, E., François, C., & Prévot, L.. (2024). Daily auditory environments in French-speaking infants: A longitudinal dataset In *Workshop on Cognitive Modeling and Computational Linguistics* (pp. 132-151). Association for Computational Linguistics.

Daily auditory environments in French-speaking infants: A longitudinal dataset

Estelle Hervé
CNRS & Aix Marseille Univ
Laboratoire Parole et Langage
Aix-en-Provence, France
estelle.herve@univ-amu.fr

Clément François
CNRS & Aix Marseille Univ
Laboratoire Parole et Langage
Aix-en-Provence, France
clement.francois@cnrs.fr

Laurent Prévot
CNRS & MEAE
CEFC
Taipei, Taiwan
laurent.prevot@cnrs.fr

Abstract

Babies' daily auditory environment plays a crucial role in language development. Most previous research estimating the quantitative and qualitative aspects of early speech inputs has predominantly focused on English- and Spanish-speaking families. In addition, validation studies for daylong recordings' analysis tools are scarce on French data sets. In this paper, we present a French corpus of daylong audio recordings longitudinally collected with the LENA (Language ENvironment Analysis) system from infants aged 3 to 24 months. We conduct a thorough exploration of this data set, which serves as a quality check for both the data and the analysis tools. We evaluate the reliability of LENA metrics by systematically comparing them with those obtained from the ChildProject set of tools and checking the known dynamics of the metrics with age. These metrics are also used to replicate, on our data set, findings from Warlaumont et al. (2014) about the increase of infants' speech vocalizations and temporal contingencies between infants and caregivers with age.

1 Introduction

Infants rely on their daily auditory environment to develop language and other cognitive skills. Pioneering studies interested in these early auditory inputs used observatory experiments in laboratory settings or short recordings that were manually annotated (Hart and Risley, 1992; Keller et al., 2004). In the last decades, technological advances brought new tools that allowed the collection and analysis of more considerable and ecological datasets. Daylong recordings are now increasingly used in developmental studies (Ganek and Eriks-Brophy, 2018; Bergelson et al., 2023), especially since the release of the Language Environment Analysis (LENA) system in 2004.

Daily auditory environments have been described in a variety of populations (Christakis et al., 2009; Aragon and Yoshinaga-Itano, 2012; Caskey et al., 2014; Warren et al., 2010), highlighting the positive effects of early caregiver-infant interactions on language development (Warlaumont et al., 2014; Gilkerson and Richards, 2008; Bergelson and Aslin, 2017). Nonetheless, only two datasets were collected in French-speaking households (Canault et al., 2016; Orena et al., 2019). Here, we expand the literature by describing an original dataset of infants' daylong audio recordings gathered in twenty French-speaking families.

We focused on 3-to-24-month-old babies for several reasons: (1) it allows a direct comparison with Canault et al. (2016)'s and Warlaumont et al. (2014) results ; (2) it includes crucial steps for language development, including the emergence of phonemic categories between 6 and 10 months (Werker and Tees, 1984; Cheour et al., 1998), and the vocabulary spurt between 18 and 24 months (Benedict, 1979; Goldfield and Reznick, 1990; Nazzi and Bertoncini, 2003) ; (3) the two first years of life constitute a critical period where caregiver-infant interactions are early precursors of later language outcomes and cognitive skills (Warlaumont et al., 2022; Gilkerson and Richards, 2009; Weisleder and Fernald, 2013; Bergelson and Aslin, 2017).

LENA's output correlation with human annotations has been assessed in several languages, suggesting good reliability (Xu et al., 2008b; Weisleder and Fernald, 2013; Gilkerson et al., 2015; Busch et al., 2018; Pae et al., 2016; Ganek and Eriks-Brophy, 2017). However, only one study provided evidence for LENA system reliability in European French, yielding relatively good results (Canault et al., 2016). Moreover, LENA validation studies implied listening to the continuous raw audio recordings. In addition to being highly time-consuming, this approach raises critical ethical issues associated with data privacy (Casillas and Cristia, 2019; Cychosz et al., 2020). Here, we override these difficulties by comparing the LENA metrics outputs with other

annotation systems.
The paper has the following contributions: (i) describe a new French corpus of auditory LENA-recorded data, (ii) compare different automatic annotation tools, (iii) provide a picture of the daily auditory environment in French-speaking families in 3-to-24-months infants, (iv) show the potential of the data set by replicating daylong recordings-based results on developmental trajectories.

2 Related Work

This section is a brief overview of the existing literature regarding daylong recording studies in developmental populations. We identified two main types of studies: (i) experimental studies that used daylong recordings as a tool to answer a specific research question and (ii) validation studies that focused on assessing the reliability of the recording and analysis tools themselves.

2.1 Experimental studies

Daylong recording studies in infants often involved the LENA system. Four years after its release, the first LENA normative study, the "Natural Language Study (NLS)" was conducted by the LENA Research Foundation (Gilkerson and Richards, 2008). This report relied on the three main LENA metrics (Adult Word Count: AWC; Child Vocalization Count: CVC; Conversational Turn Counts: CTC) to describe daily auditory environments in 329 English-speaking infants aged 2 to 48 months. Then, more experimental works involving daylong recordings in infants began to emerge (see Ganek and Eriks-Brophy (2018) for a review).
Studies that focused on the characteristics of the daily auditory environment in typically developing infants revealed that children's vocalizations and child-caregiver interactions increased with age within the first two years of life (Gilkerson and Richards, 2008; Pae et al., 2016). Warlaumont et al. (2014) proposed a "social feedback loop" in which contingencies between adult-child and child-adult speech-like vocalizations contribute to increasing interactions between infants and adults through age. Additionally, a higher proportion of adult-child interactions has been associated with larger vocabulary size (Weisleder and Fernald, 2013). The impact of various factors like multilingualism (Oller et al., 2010; Orena et al., 2019; Ramírez and Hippe, 2024), socio-economic status (Bergelson et al., 2023), exposure to TV Christakis et al. (2009); Zimmerman et al. (2009), musical inputs (Mendoza and Fausey, 2021), activity during the day (Soderstrom and Wittebolle, 2013) and temporal dynamics of the surrounding sounds (Warlaumont et al., 2022) have been investigated as well. Daylong recording studies in clinical populations showed the importance of understanding infants' daily soundscape for early language intervention (Caskey et al., 2011; Warren et al., 2010; Warlaumont et al., 2014; Aragon and Yoshinaga-Itano, 2012).

Overall, age ranges, sample sizes, and recording spans varied across studies. Some infants were included as early as 2 months of age (Aragon and Yoshinaga-Itano, 2012; Bergelson and Aslin, 2017; Zimmerman et al., 2009), while others started after 12 months of age (Oller et al., 2010; Warren et al., 2010; Weisleder and Fernald, 2013). Children could be followed longitudinally within various periods (Gilkerson et al., 2018; Sy et al., 2023) but not systematically (Weisleder and Fernald, 2013; Bergelson and Aslin, 2017).
Although most studies relied on the LENA system, methodological choices regarding data collection and analysis were various. For example, some authors chose to rely on preexisting datasets that already fitted their research questions (Christakis et al., 2009; Aragon and Yoshinaga-Itano, 2012; Warren et al., 2010). For data analysis, the LENA metrics were mostly used although some preferred to develop their own tools (MacWhinney, 2000; Al Futaisi et al., 2019; Lavechin et al., 2020; Räsänen et al., 2021).

2.2 Validation studies

The first LENA validation study was led in 2008 on American English, as part of the NLS (Xu et al., 2008b). Human annotations were compared with automatic outputs provided by the LENA software to determine agreement scores, measured with Pearson's correlations. LENA's AWC and CVC reached $r = 0.82$ and $r = 0.76$ respectively, indicating reliable LENA annotations for subsequent English-speaking environment studies (Christakis et al., 2009; Warren et al., 2010; Xu et al., 2008a; Zimmerman et al., 2009; Gilkerson et al., 2017).
Later, the same validation procedure was applied to other languages, focusing on the three main LENA metrics (AWC, CVC, and CTC). Overall, the AWC metric was the most reliable, although several authors reported that, on average,

the LENA's estimations were lower than the human counts (Xu et al., 2009; Canault et al., 2016). Agreement scores for AWC were reported in Spanish ($r = 0.80$, Weisleder and Fernald (2013)), Mandarin ($r = 0.72$, Gilkerson et al. (2015)), Korean ($r = 0.72$, Pae et al. (2016)), and Dutch ($r = 0.87$, Busch et al. (2018)). CVC and CTC's reliability were not systematically assessed and yielded variable results, ranging from $r = 0.52$ (Busch et al., 2018) to $r = 0.84$ (Gilkerson and Richards, 2008) (see Table 3 in Appendices). In French, we found Canault et al. (2016)'s report as the only existing validation study so far. They manually annotated and transcribed 324 ten-minute samples recorded in 3-to-48-month-olds: Pearson's correlation scores were $r = 0.64$ for AWC and $r = 0.71$ for CVC. These results suggest good reliability for LENA metrics in French, although slightly below the abovementioned languages.

Cristia et al. (2021)'s comprehensive validation study in three different linguistic and socio-cultural environments calls for more validation studies with more detailed and systematic methods. However, the concurrent emergence of annotation tools (MacWhinney, 2000; Al Futaisi et al., 2019; Lavechin et al., 2020; Räsänen et al., 2021) tends to increase methodological variability. To converge toward a standardized pipeline for daylong data management, Gautheron et al. (2023) developed the ChildProject package. It is compatible with many existing annotation formats and allows annotation systems comparisons. Here, we relied on these tools to compare LENA's metrics with measures extracted from the `Voice Type Classifier (VTC)` from Lavechin et al. (2020) and the `VoCalisation Maturity analysis (VCM)` from Al Futaisi et al. (2019).

3 Rationale

3.1 Participants

Infants were recruited between 3 and 18 months of age in three daycare centers in south-east France. An official collaboration between our team and the daycare centers was established to facilitate both participants' recruitment and data collection. We met parents in person to communicate the project and obtain their informed consent. The French Ethics Committee Review Board approved the study (Agreement 2022-A02281-42) which was conducted according to the guidelines of the Declaration of Helsinki (World Health Organisation, 2008). Parents filled out a questionnaire to ensure that infants did not have any hearing, cognitive, or developmental disorders and that they were raised in a dominant French-speaking environment. Other metadata were gathered through this questionnaire: number of caregivers, musical practice of the caregiver(s), linguistic environment (which language(s) spoken around the child), and socio-economic status (SES) assessed via profession.

Independently of their age at the inclusion date, we followed infants until 24 months of age when possible, or as long as possible otherwise. Twenty infants were involved, with a mean age at inclusion of 12 months (m = 360 days, sd = 132.5). Six additional babies were recruited but excluded from the analysis because parents did not provide enough recordings (<5).

3.2 Procedure

As mentioned above, we collaborated with three daycare centers that became a hub for data collection. Once parents had given their informed consent, they were provided with the LENA materials: a recorder and a t-shirt with a frontal pocket. Each infant had one unique recorder that they kept until the end of data collection. Clothing size was adapted to infants and changed throughout the months when needed. To help parents get used to the LENA system, we gave them some oral instructions when possible so they could ask questions and we could make sure they understood everything. Additionally, all families were given an instruction sheet that was taken from LENA's support materials and adapted to our study. Instructions comprised information about when and how often to record, how to use the device, various recommendations, and the procedure for device deposit and pickup at daycare. Parents also had our contact information and could reach us whenever they needed.

Parents were asked to have their child wear the recorder once a week for a full day, preferentially at home or during the weekends. To limit attrition, we accepted recordings at daycare occasionally, when they could not record another day or if they forgot. The frequency of the recordings was hard to maintain for some families, so we had to send them kind reminders sometimes. But overall, all families were very involved and consistent. We recommended that during a recording day, they never turn the recorder off, limit noisy environ-

ments, and let the recorder nearby while showering and during bedtimes. Once the recording was completed, they were asked to bring the LENA recorders back to the daycare center once a week. Then, the investigator could transfer the data to the database the same day, so parents could get the recorder back and start over for a new week. At the end of data collection (when the child reached 24 months or when families decided to stop), infants were given a "baby researcher" diploma and a customized t-shirt as a reward.

4 Tools and Methods

Daylong recordings present a set of challenges in terms of processing. The first constraint is the data set size: we gathered 10^4 hours of highly heterogeneous recordings (both across and within recordings) that need to be sampled. Existing literature has used the notion of *hot spots* (areas in the recordings with a high density of speech events) as well as a method consisting of human labeling of extremely short sound events (Semenzin et al., 2021). Due to the required infrastructure, the latter approach was not considered for our work. Instead, we applied state-of-the-art computational techniques and packages to perform step-by-step reliability tests and calculate agreement scores between them. The automatic tools we used for our analyses are the LENA suite (Gilkerson and Richards, 2008) and a set of tools developed or adapted whithin the framework of ChildProject (Lavechin et al., 2020).

4.1 LENA

We used the LENA system for both data collection and analysis. For data collection, LENA provides a small digital language processor (DLP) that is easily held in a child's hand and can be directly inserted into child-adapted clothing equipped with a specific pocket on the front. The DLP can save up to 16 hours of auditory input. Recordings are then processed with the LENA software, which provides automatic annotations and quantification reports.
The annotation process starts by segmenting the continuous audio recordings based on acoustic features such as intensity and pitch. The segments are then compared to general models of eight categories (Christakis et al., 2009) to be labeled as target child (CHN), adult male (MAN), adult female (FAN), other child (CXN), TV/electronic sounds (TVN), noise (NOI), silence (SIL), or overlapping sounds (OVL). Next, the four categories CHN, MAN, FAN, and CXN are further analyzed to differentiate speech-related from non-speech vocalizations (see Figure 17). The LENA software provides estimations of the number of words produced by adults (AWC) and infants' speech-related vocalizations (CVC). In this study, we only used the raw sound event segmentation (timestamps) and labeling.

4.2 VTC and VCM

The ChildProject suite starts processing recordings with the *voice-type-classifier (VTC)* (Lavechin et al., 2020), which relies on the state-of-the-art speech diarization tool, *pyannote* (Bredin et al., 2020). *VTC* identifies sound activity segments that can be mapped to some of LENA's categories: target key child (KCHI), other children (CHI), female (FEM), and male (MAL). Another tool, *Vocalisation Maturity analysis (VCM)* (Al Futaisi et al., 2019), refines the output of *VTC*. *VCM* is grounded on the state-of-the-art signal processing and emotion recognition tool, *SMILE* (Eyben et al., 2010), and more precisely on the Geneva Minimal Acoustic Parameter Set (GeMAPS) (Eyben et al., 2015). It adds information to the labeled categories (e.g., speech from the target child) by determining whether the targeted speech is Canonical (CNS), Non-canonical (NCS), cries (CRY), or other sounds (noise, laughter). Such classification has been used in (Casillas et al., 2017), for example.

5 Data set

The corpus currently consists of 8286 hours of LENA daylong recordings. Table 1 indicates the mean, minimum, and maximum values for the recording period (age span), number of recording sessions, and length of the recordings.

	Avg	Min	Max	Sum
Age span (months)	9.85	3	18	-
# sessions	27.0	6	66	540
Duration (hours)	414	87	1022	8286

Table 1: The data set. N = 20 children. Age span: number of months between the first and the last recording.

Table 1 reflects a high variability in parents' use of the LENA device. As mentioned earlier, we asked them to turn the DLP on in the morning and leave it until it automatically turns off after 16 hours of recording. However, some families turned

the device on and off multiple times during the day or stopped the recording before reaching 16 hours. Thus, we observed variability across participants in recording length and number. Additionally, there was variability in the recording span: not all children started the recordings at the same age, and not all were followed until 24 months of age. Given these observations, we selected a sample of children that 1) had at least 9 months of recording span and 2) provided at least 10 recordings. These thresholds allowed us to focus on more representative datasets while maintaining a sufficient number of data points to observe developmental trajectories. A sample of 10 children met these two criteria and were selected for complementary analyses (see Table 2).

	Avg	Min	Max	Sum
Age span (months)	14.2	10	18	-
# sessions	40.9	12	66	409
Duration (hours)	637	192	1022	6366

Table 2: Selected children for individual longitudinal metrics and plots. N = 10 children (age recording span $\geq$ 9 months; number of recordings $\geq$ 10).

6 Investigating the data set

6.1 Testing age

Our first goal was to test the reliability of the metrics extracted with the three targeted tools (LENA, VTC, VCM) on our data. A crucial check for our dataset consisted of testing whether children's production evolved with age. We expected an increase in children's speech-related metrics (such as speaking time and ratio, vocalization counts, etc.), while adults' metrics would remain stable. Finally, the voices of other children present in the recordings were expected to increase as the siblings of the target child followed their own development. We began by examining the production time ratio (the percentage of the recording time occupied by a given category), for example, for the target child as shown in Figure 1 (See Appendix D for other categories).

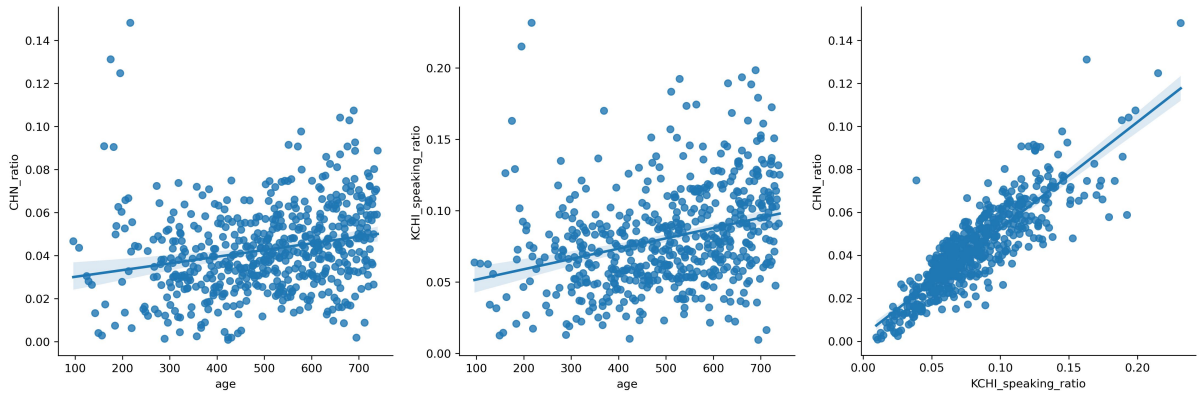
More precisely, we examined the production ratio calculated from LENA (sum of the duration of intervals labeled with *CHN* as the speaker, divided by recording duration), VTC (using the same approach with the label *KCHI*), as well as the correlation between the two measures. Importantly, the ratios obtained for the different voices from these two tools were highly correlated. More generally, all the metrics extracted with both approaches were highly correlated for all comparable categories.

We tested whether age remained a dominant factor when controlling for the available metadata. In Figure 2, we plot the production ratio alongside gender, socio-economic status (high vs. low), and linguistic context (monolingual vs. plurilingual). From these figures, a general observation is that children's speaking ratio increased with age. This was tested by conducting a linear mixed model analysis using pymer4 (Jolly, 2018). We treated *'target child speaking ratio'* as the dependent variable and *'age'* as the fixed effect, with *'child ID'* as the random effect. Only *'age'* had a significant effect on the target child production ratio ($\beta = 0.261$, $SE = 0.028$, and $p < 0.001$), controlling for *'gender'*, *'linguistic environment'*, and *'socio-economic status'* (all not significant).

VCM metrics allowed us to refine our evaluation of child production with age. We applied it to our set of selected children and found (Figure 3) that the increase in speaking ratio was due to an increase in real child speech (both canonical and non-canonical) rather than to a variation in the proportion of cries, laughter, or noise. In addition, we replicated findings from Warlaumont et al. (2014), who tested the evolution of speech-related vocalization ratio (versus non-speech-related) produced by the target child. Using our own pipeline and analysis, we also found a significant increase in this ratio with age while controlling for all metadata ($\beta = 0.219$, $SE = 0.022$, $p < 0.001$) (See Appendix C.1).

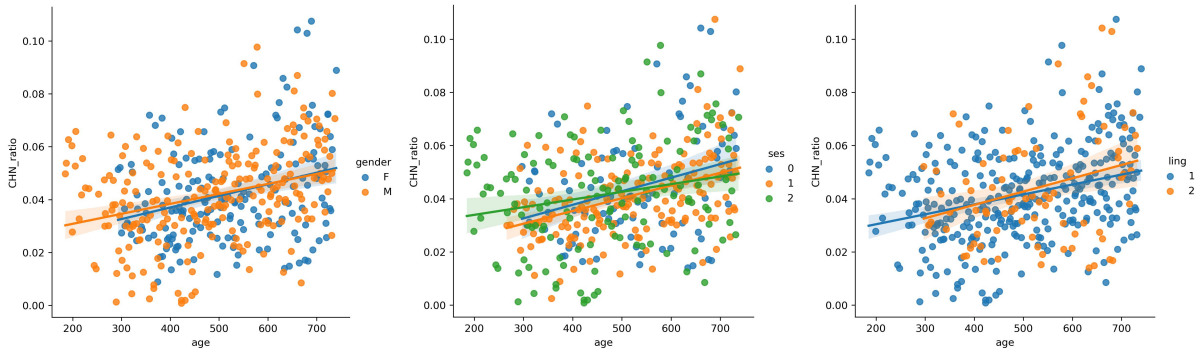
6.2 Interactional metrics

Conversations arguably constitute the most important aspect of the linguistic environment. First, we approached this by examining temporal contingencies (Bloom et al., 1987; Warlaumont et al., 2014) between the child and the other voices in the recordings. Specifically, we analyzed instances of target child productions followed by another participant, as well as target child productions preceded by other participants. Working with manually transcribed data, Nikolaus et al. (2022) explored the effect of time-contingent responses on children's intelligibility. These studies found that (1) caregivers provided more time-contingent responses to intelligible utterances from the child and (2) children



(a) LENA (b) VTC (c) Correlation ($r = 0.862$)

Figure 1: Target Child Speaking Time Ratio for (a) LENA (controlled for child id, $\beta = 0.159$; SE = 0.027; p < 0.001); (b) VTC ($\beta = 0.242$; SE = 0.029; p < 0.001) and (c) correlation plot between LENA and VTC. All children included (n=20). Age in days.



(a) Gender (b) Socio Economic Status (c) Linguistic Context

Figure 2: Target child Production Time (from LENA) differentiating for (a) gender ; (b) socio-economic status ; (c) linguistic background. Only age is significant. The three other variables are not. See Appendix D for details. Selected children. Age in days.

produced more intelligible utterances if their caregivers were responsive. Then, we investigated the "social feedback loop" as proposed by Warlaumont et al. (2014).

We employed a similar approach for both tools at our disposal. For *CHILD>ADULT* contingencies, we selected all target child productions and checked whether there was a production from an adult (MAL + FEM) participant **1 second after**. While LENA metrics extract similar information, we aimed to use the same method for both tools to enable a direct comparison. Figure 4 depicts the comparison for *CHILD>ADULT* contingencies. We also looked at ANY>CHILD contingencies by considering any activity occurring **2 seconds before**[1] before a child production. Figure 5 provides the VTC extraction for *ANY>CHILD* contingencies (additional combinations are included in Appendix E). We tested, for VTC data, the relationship between *age* and temporal contingencies for *CHILD>ADULT* ($\beta = 0.326$, $SE = 0.026$, $p < 0.001$) and *ANY>CHILD* ($\beta = 0.158$, $SE = 0.030$, $p < 0.001$), while controlling for gender, linguistic environment, and socio-economic status. This result was further refined by replicating the second finding from Warlaumont et al. (2014) on our dataset (and with LENA metrics this time). Specifically, we confirmed that children's speech-related productions tend to elicit more feedback from adults (See Appendix C.2).

Finally, we also replicated Warlaumont's result about the social loop on our data. We tested whether initial speech-related CHILD productions

[1]We considered allowing for a longer gap for children's follow-up to be appropriate. To count "turns", LENA metrics use a 5-second threshold.

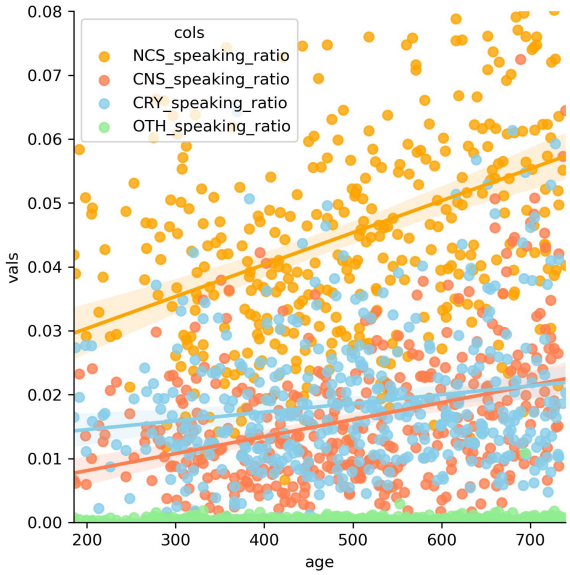


Figure 3: Speech vs. Non-speech VCM extraction for the selected children.

that were followed by an ADULT production 1 second after[2] were more likely to be followed by a speech-related CHILD production than a non-speech CHILD production within 3.5 seconds, compared to initial speech-related CHILD productions that did not receive an adult response.[3] This is illustrated in Figure 6, which shows the difference in speech-related / non-speech related ratio in child's production after an adult's response versus no response. This difference is positive, indicating that adult responses to children's speech-related productions tend to increase the proportion of child speech-related follow-ups.

7 Discussion

The collection and investigation of our original corpus in French households facilitated the comparison of different analysis tools and the replication of previous results within the same age ranges.

First, we identified a robust relationship between target child production metrics and age. As expected, we did not observe a similar significant change for adult voices. However, the evolution was also positive and significant for other children's voices ($\beta = 0.179$, $SE = 0.027$, $p < 0.001$). This can be attributed to the behavioral path of siblings as well as other children in kindergartens. All these results were obtained using both LENA and VTC pipelines. Finally, we found coherent results in line with existing literature and across the tools we used. These results strengthen our confidence in both our recording protocol and in the metrics extraction and analysis.

These comments hold for interactional metrics as well: the increases in temporal contingencies involving the children were consistent with findings from previous studies (Warlaumont et al., 2014; Nikolaus et al., 2022). We calculated temporal contingencies in a way that ensured these increases were not influenced by the overall amount of child productions.

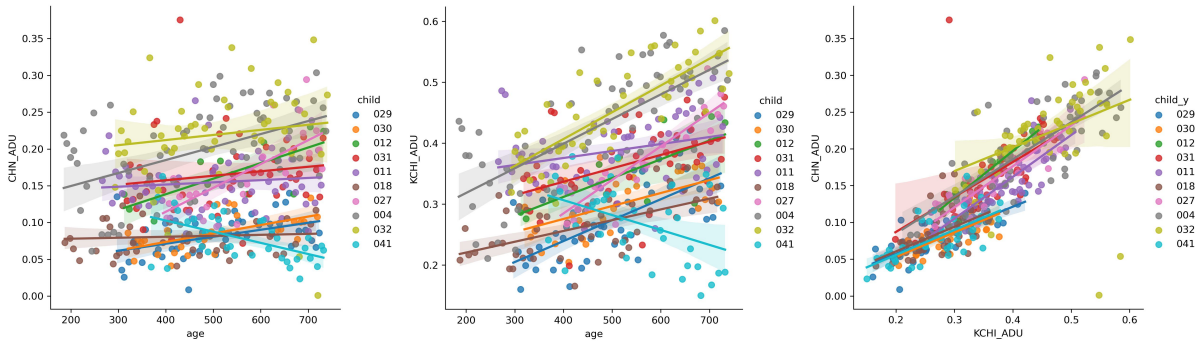
Furthermore, our more detailed analysis (depicted in Figure 3) and the replication of Warlaumont's first results, showed that the increase in child's productions with age was due to speech-related productions and not to vegetative sounds or noise. In summary, our data show that children do produce more speech while growing up in their first two years of life. These increased productions are temporally contingent on other speakers, regardless of the initiator of the interaction (target child or other speaker).

This is further elaborated by the replication of Warlaumont's third result about the social loop that showed the benefit of follow-up productions of adult feedback on children's speech. Contrary to Warlaumont et al. (2014) (but in line with Bergelson et al. (2023) we did not find any effect of parental SES, gender, or linguistic environment on children's productions. This was the same for temporal contingencies used as a proxy for measuring linguistic interaction with the child.

Other studies have attempted to dive further into linguistic metrics from large child-caregiver datasets. Some refined the speech contingencies to distinguish corrective feedback or to approach the grammaticality evaluation of the productions in these datasets (Nikolaus et al., 2023). However, those datasets have been manually transcribed, while ours did not (and will not) receive such a transcription. A second major difference is the age range of the children. Most of the existing studies involved children from 12-24 months up to

[2]Following Warlaumont et al. (2014) and Nikolaus et al. (2022), who followed Oller et al. (2010), which used a 1-second window to extract relevant vocal activity to investigate the "social feedback loop" for children between 8 to 48 months.

[3]This 3.5 second is between the end of initial CHILD production and the start of the following one. The delay is to allow for a potential ADULT response to occur in between. It is not possible to use the ADULT production for proposing a simpler time threshold for the following utterance since there is not always an ADULT production in-between.



(a) LENA (b) VTC (c) Correlation ($r = 0.831$

Figure 4: CHILD>ADULT contingencies for (a) LENA ($\beta = 0.131$; SE = 0.024; p < 0.001), (b) VTC ($\beta = 0.130$, $SE = 0.024$, $p < 0.001$) and (c) correlation plot. Selected children. Age in days.

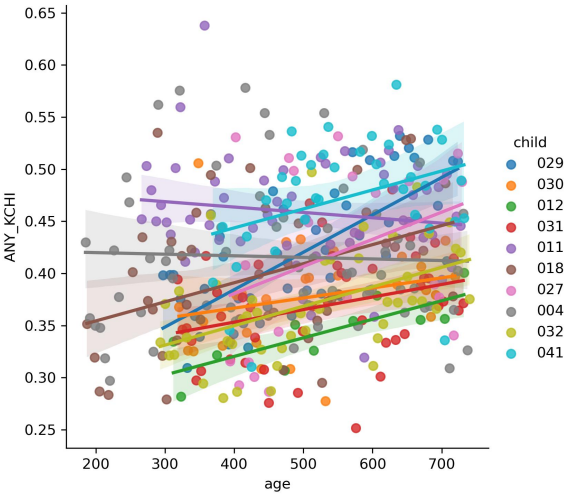


Figure 5: ANY>CHILD contingencies from VTC speech extraction. ($\beta = 0.158$, $SE = 0.030$, $p < 0.001$)

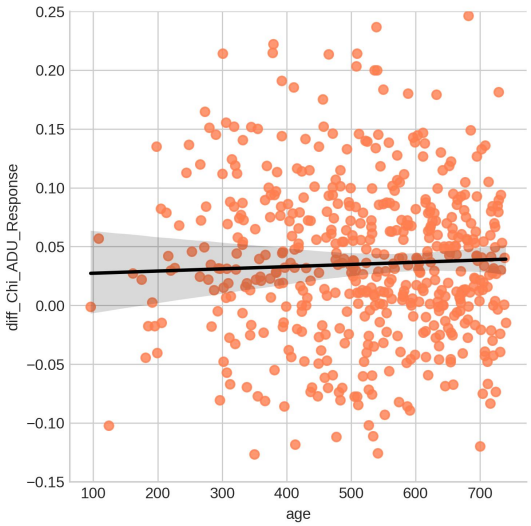


Figure 6: Child ratio difference between Child speech and non-speech productions depending on whether or not an initial child speech-related utterance was responded by an adult or not (positive mean : indicating a tendency for more speech-related responses), replication of Warlaumont et al. (2014).

later ages. This inevitably changes the nature and characteristics of the metrics that can be extracted. Another trend of work focuses on phonetic learning based on daylong recordings, such as Lavechin et al. (2024) on perceptual attunement. Finally, in a more cultural and typological direction, Bergelson et al. (2023) studied a global sample of 1,001 child-centered audio capturing 2- to 48-months-olds from many countries and various cultural backgrounds.

8 Conclusion

This work first constitutes a replication of earlier results on daylong recordings, on a completely new and independent data set, using two different tools. This contributes to answer Cruz Blandón et al. (2023)'s call for more and better meta-studies on long recordings. Indeed, despite their creation cost, daylong recordings' significance is growing in cognitive science. Showing that these data sets, despite their noisy nature, do present enough reliability to gain insights about the children's language and communicative skills development is crucial.

The second contribution consists of the characterization of our data set itself. It is a large data set (>8000 hours of recordings) that is still growing at the time of writing. It is unique by being truly longitudinal with some children's environments being recorded over a two-year span. Finally, other experimental data were collected longitudinally in the same sample of infants. These are beyond the scope of this paper but open up the possibilities of cross-analyses between the characterization of

the linguistic environment and other experimental results regarding language development. The present study, therefore, constitutes a crucial first step in this direction through the thorough exploration of the data set, and by considering individual variability.

From here, we now plan to refine the analyses by entering into more linguistic characterizations of these productions in terms of richness. We will consider tools that allow for more phonological measures such as Räsänen et al. (2021) and more content-based metrics (Nikolaus et al., 2023, 2024) that have been used so far only on transcript-based corpora from CHILDES (MacWhinney, 2000) and that can be now explored on daylong recordings thanks to the improvement of automatic speech recognition engines.

9 Limitations

One major limitation of this work is the absence of manual annotation. For legal and ethical reasons, we are not in position to perform extensive manual annotations of raw audio data, as well as sharing raw audio. We needed to find other ways to check the reliability of our dataset. By replicating previous results from the literature and comparing different computational tools, we reinforced our trust in our dataset and overcame this constraint. All metrics derived from the corpus related to this paper as well as for future work will be made available in the LLDC public repository on Ortolang institutional platform https://www.ortolang.fr/. Moreover the code for producing the analyses presented in this paper is available at : https://github.com/prevotlaurent/LENA_CMCL.

Acknowledgments

We thank all the babies and their families for participating in the study. We also thank the directors and the staff of the Babilou daycare centers where the babies were evaluated. We thank Mathilde Gaujard for her help in the recruitment and data collection. This research has been supported by an ANR grant ("BabyLang project", ANR-20-CE28-0017) to CF and a CNRS 80PRIME grant to CF and LP ("LangDev project"). This work, carried out within the Institut Convergence ILCB (ANR-16-CONV-0002), has benefited from support from the French government, managed by the French National Agency for Research (ANR) and the Excellence Initiative of Aix-Marseille University (A*MIDEX).

References

Najla Al Futaisi, Zixing Zhang, Alejandrina Cristia, Anne Warlaumont, and Bjorn Schuller. 2019. Vcmnet: Weakly supervised learning for automatic infant vocalisation maturity analysis. In *2019 International Conference on Multimodal Interaction*, pages 205–209.

Miranda Aragon and Christine Yoshinaga-Itano. 2012. Using language environment analysis to improve outcomes for children who are deaf or hard of hearing. In *Seminars in Speech and Language*, volume 33, pages 340–353. Thieme Medical Publishers.

Helen Benedict. 1979. Early lexical development: Comprehension and production. *Journal of child language*, 6(2):183–200.

Elika Bergelson and Richard N Aslin. 2017. Nature and origins of the lexicon in 6-mo-olds. *Proceedings of the National Academy of Sciences*, 114(49):12916–12921.

Elika Bergelson, Melanie Soderstrom, Iris-Corinna Schwarz, Caroline F Rowland, Nairán Ramírez-Esparza, Lisa R. Hamrick, Ellen Marklund, Marina Kalashnikova, Ava Guez, Marisa Casillas, et al. 2023. Everyday language input and production in 1,001 children from six continents. *Proceedings of the National Academy of Sciences*, 120(52):e2300671120.

Kathleen Bloom, Ann Russell, and Karen Wassenberg. 1987. Turn taking affects the quality of infant vocalizations. *Journal of child language*, 14(2):211–227.

Hervé Bredin, Ruiqing Yin, Juan Manuel Coria, Gregory Gelly, Pavel Korshunov, Marvin Lavechin, Diego Fustes, Hadrien Titeux, Wassim Bouaziz, and Marie-Philippe Gill. 2020. Pyannote. audio: neural building blocks for speaker diarization. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 7124–7128. IEEE.

Tobias Busch, Anouk Sangen, Filiep Vanpoucke, and Astrid van Wieringen. 2018. Correlation and agreement between language environment analysis (lena™) and manual transcription for dutch natural language recordings. *Behavior research methods*, 50:1921–1932.

Mélanie Canault, Marie-Thérèse Le Normand, Samy Foudil, Natalie Loundon, and Hung Thai-Van. 2016. Reliability of the language environment analysis system (lena™) in european french. *Behavior research methods*, 48:1109–1124.

Marisa Casillas, John Bunce, Melanie Soderstrom, Celia Rosemberg, Maia Migdalek, Florencia Alam, Alejandra Stein, and Hallie Garrison. 2017. Introduction: the aclew das template. https://osf.io/aknjv.

Marisa Casillas and Alejandrina Cristia. 2019. A step-by-step guide to collecting and analyzing long-format speech environment (lfse) recordings. *Collabra: Psychology*, 5(1):24.

Melinda Caskey, Bonnie Stephens, Richard Tucker, and Betty Vohr. 2011. Importance of parent talk on the development of preterm infant vocalizations. *Pediatrics*, 128(5):910–916.

Melinda Caskey, Bonnie Stephens, Richard Tucker, and Betty Vohr. 2014. Adult talk in the nicu with preterm infants and developmental outcomes. *Pediatrics*, 133(3):e578–e584.

Marie Cheour, Rita Ceponiene, Anne Lehtokoski, Aavo Luuk, Jüri Allik, Kimmo Alho, and Risto Näätänen. 1998. Development of language-specific phoneme representations in the infant brain. *Nature neuroscience*, 1(5):351–353.

Dimitri A Christakis, Jill Gilkerson, Jeffrey A Richards, Frederick J Zimmerman, Michelle M Garrison, Dongxin Xu, Sharmistha Gray, and Umit Yapanel. 2009. Audible television and decreased adult words, infant vocalizations, and conversational turns: a population-based study. *Archives of pediatrics & adolescent medicine*, 163(6):554–558.

Alejandrina Cristia, Marvin Lavechin, Camila Scaff, Melanie Soderstrom, Caroline Rowland, Okko Räsänen, John Bunce, and Elika Bergelson. 2021. A thorough evaluation of the language environment analysis (lena) system. *Behavior research methods*, 53:467–486.

María Andrea Cruz Blandón, Alejandrina Cristia, and Okko Räsänen. 2023. Introducing meta-analysis in the evaluation of computational models of infant language development. *Cognitive Science*, 47(7):e13307.

Margaret Cychosz, Rachel Romeo, Melanie Soderstrom, Camila Scaff, Hillary Ganek, Alejandrina Cristia, Marisa Casillas, Kaya De Barbaro, Janet Y Bang, and Adriana Weisleder. 2020. Longform recordings of everyday life: Ethics for best practices. *Behavior research methods*, 52:1951–1969.

Florian Eyben, Klaus R Scherer, Björn W Schuller, Johan Sundberg, Elisabeth André, Carlos Busso, Laurence Y Devillers, Julien Epps, Petri Laukka, Shrikanth S Narayanan, et al. 2015. The geneva minimalistic acoustic parameter set (gemaps) for voice research and affective computing. *IEEE transactions on affective computing*, 7(2):190–202.

Florian Eyben, Martin Wöllmer, and Björn Schuller. 2010. Opensmile: the munich versatile and fast open-source audio feature extractor. In *Proceedings of the 18th ACM international conference on Multimedia*, pages 1459–1462.

Hillary Ganek and Alice Eriks-Brophy. 2018. Language environment analysis (lena) system investigation of day long recordings in children: A literature review. *Journal of Communication Disorders*, 72:77–85.

Hillary V Ganek and Alice Eriks-Brophy. 2017. A concise protocol for the validation of language environment analysis (lena) conversational turn counts in vietnamese. *Communication Disorders Quarterly*, 39(2):371–380.

Lucas Gautheron, Nicolas Rochat, and Alejandrina Cristia. 2023. Managing, storing, and sharing long-form recordings and their annotations. *Language Resources and Evaluation*, 57(1):343–375.

Jill Gilkerson and Jeffrey A Richards. 2008. The lena natural language study. *Boulder, CO: LENA Foundation. Retrieved March*, 3(2009):15–17.

Jill Gilkerson and Jeffrey A Richards. 2009. The power of talk. *Impact of adult talk, conversational turns and TV during the critical 0-4 years of child development: Boulder, CO: LENA Foundation*.

Jill Gilkerson, Jeffrey A Richards, Steven F Warren, Judith K Montgomery, Charles R Greenwood, D Kimbrough Oller, John HL Hansen, and Terrance D Paul. 2017. Mapping the early language environment using all-day recordings and automated analysis. *American journal of speech-language pathology*, 26(2):248–265.

Jill Gilkerson, Jeffrey A Richards, Steven F Warren, D Kimbrough Oller, Rosemary Russo, and Betty Vohr. 2018. Language experience in the second year of life and language outcomes in late childhood. *Pediatrics*, 142(4).

Jill Gilkerson, Yiwen Zhang, Dongxin Xu, Jeffrey A Richards, Xiaojuan Xu, Fan Jiang, James Harnsberger, and Keith Topping. 2015. Evaluating language environment analysis system performance for chinese: A pilot study in shanghai. *Journal of Speech, Language, and Hearing Research*, 58(2):445–452.

Beverly A Goldfield and J Steven Reznick. 1990. Early lexical acquisition: Rate, content, and the vocabulary spurt. *Journal of child language*, 17(1):171–183.

Betty Hart and Todd R Risley. 1992. American parenting of language-learning children: Persisting differences in family-child interactions observed in natural home environments. *Developmental psychology*, 28(6):1096.

Eshin Jolly. 2018. Pymer4: Connecting r and python for linear mixed modeling. *Journal of Open Source Software*, 3(31):862.

Heidi Keller, Elke Hentschel, Relindis Dzeaye Yovsi, Bettina Lamm, Monika Abels, and Verena Haas. 2004. The psycho-linguistic embodiment of parental ethnotheories: A new avenue to understanding cultural processes in parental reasoning. *Culture & Psychology*, 10(3):293–330.

Marvin Lavechin, Ruben Bousbib, Hervé Bredin, Emmanuel Dupoux, and Alejandrina Cristia. 2020. An open-source voice type classifier for child-centered

daylong recordings. In *Interspeech 2020 - Conference of the International Speech Communication Association*, Shanghai / Virtual, China.

Marvin Lavechin, Maureen de Seyssel, Marianne Métais, Florian Metze, Abdelrahman Mohamed, Hervé Bredin, Emmanuel Dupoux, and Alejandrina Cristia. 2024. Modeling early phonetic acquisition from child-centered audio data. *Cognition*, 245:105734.

Brian MacWhinney. 2000. The childes project. *Computational Linguistics*, 26(4):657–657.

Jennifer K Mendoza and Caitlin M Fausey. 2021. Everyday music in infancy. *Developmental Science*, 24(6):e13122.

Thierry Nazzi and Josiane Bertoncini. 2003. Before and after the vocabulary spurt: Two modes of word acquisition? *Developmental Science*, 6(2):136–142.

Mitja Nikolaus, Abhishek Agrawal, Petros Kaklamanis, Alex Warstadt, and Abdellah Fourtassi. 2024. Automatic annotation of grammaticality in child-caregiver conversations.

Mitja Nikolaus, Laurent Prévot, and Abdellah Fourtassi. 2022. Communicative feedback as a mechanism supporting the production of intelligible speech in early childhood. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 44.

Mitja Nikolaus, Laurent Prévot, and Abdellah Fourtassi. 2023. Communicative feedback in response to children's grammatical errors. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 45.

D Kimbrough Oller, Partha Niyogi, Sharmistha Gray, Jeffrey A Richards, Jill Gilkerson, Dongxin Xu, Umit Yapanel, and Steven F Warren. 2010. Automated vocal analysis of naturalistic recordings from children with autism, language delay, and typical development. *Proceedings of the National Academy of Sciences*, 107(30):13354–13359.

Adriel John Orena, Krista Byers-Heinlein, and Linda Polka. 2019. Reliability of the language environment analysis recording system in analyzing french–english bilingual speech. *Journal of Speech, Language, and Hearing Research*, 62(7):2491–2500.

Soyeong Pae, Hyojin Yoon, Ahyoung Seol, Jill Gilkerson, Jeffrey A Richards, Lin Ma, and Keith Topping. 2016. Effects of feedback on parent–child language with infants and toddlers in korea. *First Language*, 36(6):549–569.

Naja Ferjan Ramírez and Daniel S Hippe. 2024. Estimating infants' language exposure: A comparison of random and volume sampling from daylong recordings collected in a bilingual community. *Infant Behavior and Development*, 75:101943.

Okko Räsänen, Shreyas Seshadri, Marvin Lavechin, Alejandrina Cristia, and Marisa Casillas. 2021. Alice: An open-source tool for automatic measurement of phoneme, syllable, and word counts from child-centered daylong recordings. *Behavior Research Methods*, 53:818–835.

Chiara Semenzin, Lisa Hamrick, Amanda Seidl, Bridgette Kelleher, and Alejandrina Cristia. 2021. Towards large-scale data annotation of audio from wearables: validating zooniverse annotations of infant vocalization types. In *2021 IEEE Spoken Language Technology Workshop (SLT)*, pages 1079–1085. IEEE.

Melanie Soderstrom and Kelsey Wittebolle. 2013. When do caregivers talk? the influences of activity and time of day on caregiver speech and child vocalizations in two childcare environments. *PloS one*, 8(11):e80646.

Yaya Sy, William Havard, Marvin Lavechin, Emmanuel Dupoux, and Alejandrina Cristia. 2023. Measuring language development from child-centered recordings. In *Interspeech 2023*, pages 4618–4622. ISCA.

Anne S Warlaumont, Kunmi Sobowale, and Caitlin M Fausey. 2022. Daylong mobile audio recordings reveal multitimescale dynamics in infants' vocal productions and auditory experiences. *Current directions in psychological science*, 31(1):12–19.

Anne S. Warlaumont, Anne S. Warlaumont, Jeffrey A. Richards, Jeffrey A. Richards, Jill Gilkerson, Jill Gilkerson, D. Kimbrough Oller, and D. Kimbrough Oller. 2014. A social feedback loop for speech development and its reduction in autism. *Psychological Science*.

Steven F Warren, Jill Gilkerson, Jeffrey A Richards, D Kimbrough Oller, Dongxin Xu, Umit Yapanel, and Sharmistha Gray. 2010. What automated vocal analysis reveals about the vocal production and language learning environment of young children with autism. *Journal of autism and developmental disorders*, 40:555–569.

Adriana Weisleder and Anne Fernald. 2013. Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological science*, 24(11):2143–2152.

Janet F Werker and Richard C Tees. 1984. Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant behavior and development*, 7(1):49–63.

Dongxin Xu, Umit Yapanel, and Sharmi Gray. 2009. Reliability of the lena language environment analysis system in young children's natural home environment. *Boulder, CO: Lena Foundation*, pages 1–16.

Dongxin Xu, Umit Yapanel, Sharmi Gray, and Charles T Baer. 2008a. The lena language environment analysis system: The interpreted time segments (its) file. *Boulder, CO: Lena Foundation*, pages 1–7.

Dongxin Xu, Umit Yapanel, Sharmi Gray, Jill Gilkerson, Jeff Richards, and John Hansen. 2008b. Signal processing for young child speech language development. In *First Workshop on Child, Computer and Interaction*.

Frederick J Zimmerman, Jill Gilkerson, Jeffrey A Richards, Dimitri A Christakis, Dongxin Xu, Sharmistha Gray, and Umit Yapanel. 2009. Teaching by listening: The importance of adult-child conversations to language development. *Pediatrics*, 124(1):342–349.

A Validation studies

Reference	Language(s)	r(AWC)	r(CVC)	r(CTC)
Busch et al. (2018)	Dutch	0.87	0.77	0.52
Canault et al. (2016)	European French	0.64	0.71	NI
Caskey et al. (2014)	American English and Spanish	0.93	NI	NI
Cristia et al. (2021)	American English	0.76	0.76	0.57
Ganek and Eriks-Brophy (2017)	Vietnamese	NI	NI	0.70*
Gilkerson et al. (2015)	Mandarin	0.72	0.84; 0.70**	0.72
Pae et al. (2016)	Korean	0.72	NI	0.67
Weisleder and Fernald (2013)	Spanish	0.80	NI	NI
Xu et al. (2008b)	American English	0.82	0.76	NI

Table 3: Pearson's r correlation scores between human and LENA automatic annotations for AWC, CVC and CTC in various languages. NI = No Information.*agreement score assessed via a Spearman rank correlation test. **0.84 for speech-like vocalizations, 0.70 for non-speech-like vocalizations.

B Meta-data

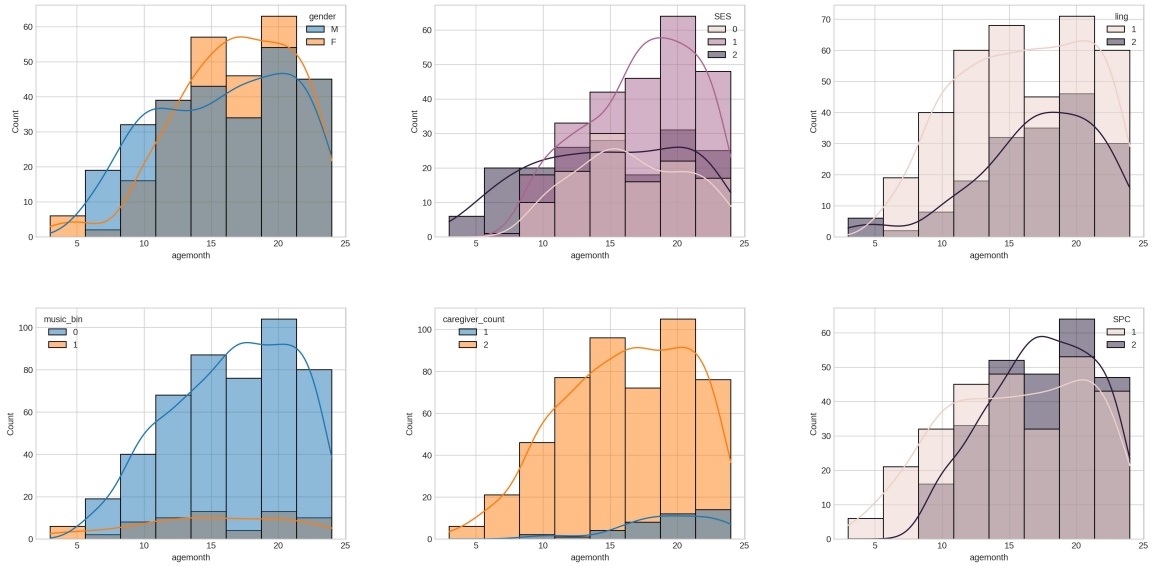
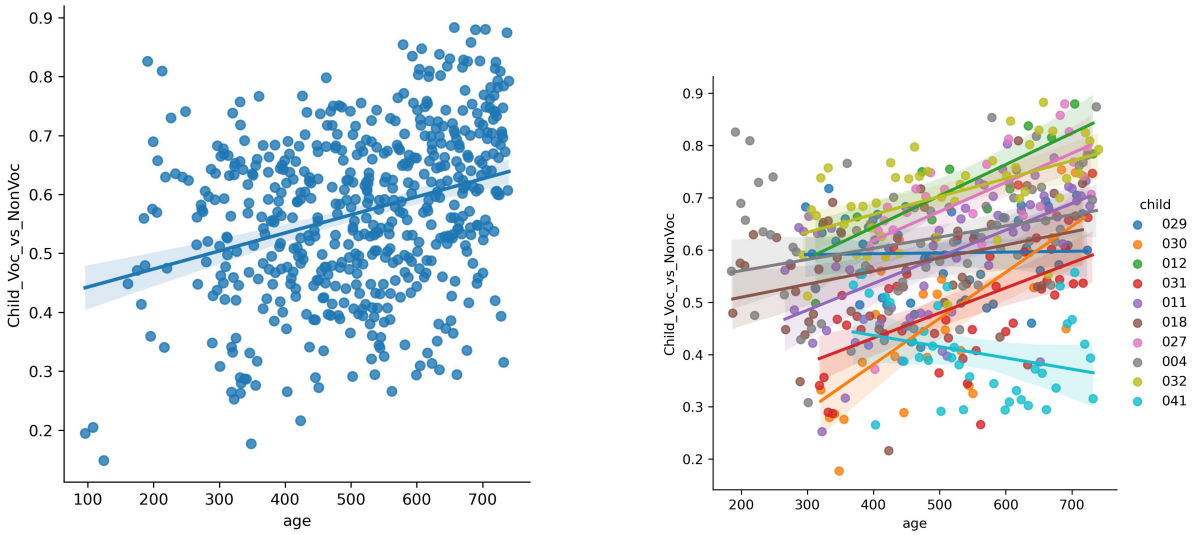


Figure 7: Distribution of session through age, depending on various demographic information: from left to right, top to bottom : gender, socio-economic status (SES: low,mid,high), linguistic environment (1:monolingual, 2: plurilingual), music practice at home(yes/no), number of cargiver, socio-professional category of the parents.

C Replication (Warlaumont et al., 2014)

C.1 Ratio of Speech Related Vocalisations



(a) All children

(b) Selected Children

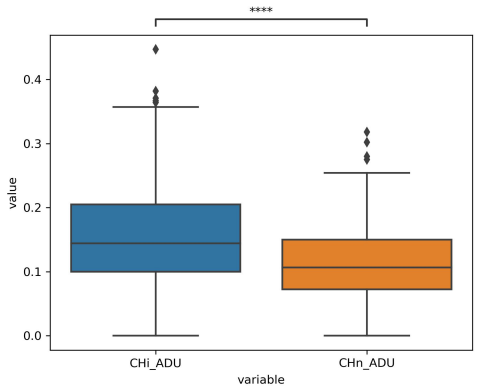
Figure 8: Ratio between the child total amount of speech related productions (`ChildVocDuration`) and all production (including Non speech related) (`ChildNonVocDuration`)(age significant $\beta = 0.219$, $p < 0.001$; `child_id` and all metadata variables not significant). Age in days.

The Figure 8 presents the first result of (Warlaumont et al., 2014), which is that the speech related percentage productions of children increase with age. See the figure caption and below for statistics.

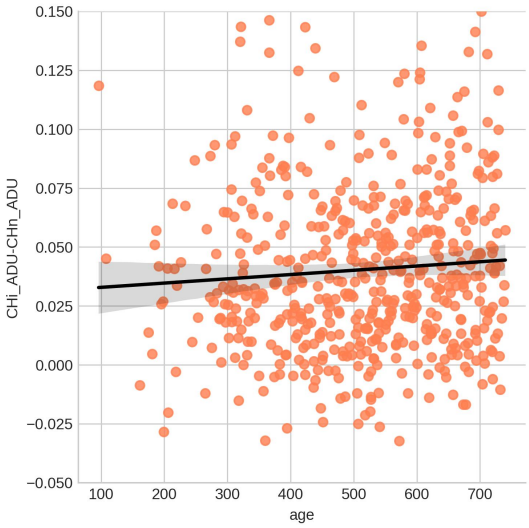
```
Formula:
Child_Voc_vs_NonVoc~age_rs+gender+ses_bin+ling_bin+gender*ses_bin
                          +gender*ling_bin+ling_bin*ses_bin+(1|child)}

Number of observations: 540  Groups: {'child': 20.0}

Random effects:
                 Name    Var    Std
child     (Intercept)  0.008  0.090
Residual               0.010  0.098

Fixed effects:
                   Estimate  2.5_ci  97.5_ci     SE       DF  T-stat  P-val  Sig
(Intercept)           0.443   0.356    0.530  0.044   17.986   9.964  0.000  ***
age_rs                0.219   0.175    0.262  0.022  532.437   9.910  0.000  ***
genderM               0.001  -0.114    0.117  0.059   13.872   0.021  0.984
ses_binL              0.028  -0.113    0.169  0.072   14.613   0.385  0.706
ling_binP            -0.111  -0.245    0.024  0.069   14.548  -1.611  0.129
genderM:ses_binL      0.104  -0.143    0.351  0.126   14.711   0.824  0.423
genderM:ling_binP     0.096  -0.109    0.301  0.104   14.535   0.917  0.374
```

C.2 Adult response to Child speech vs. non-speech productions

Figure 9 illustrates that children's speech-related productions tend to elicit more feedback from adults.



(a) Adult response to Child speech vs. non-speech productions (Mann-Whitney-Wilcoxon test two-sided, ****: p <= 1.00e-04)

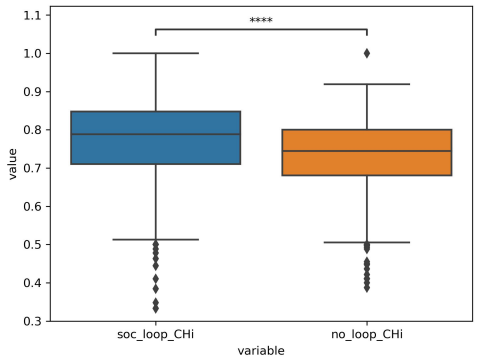


(b) Adult responses ratio difference between Child speech and non-speech productions (positive mean indicating a tendency for more responses to speech-related productions.

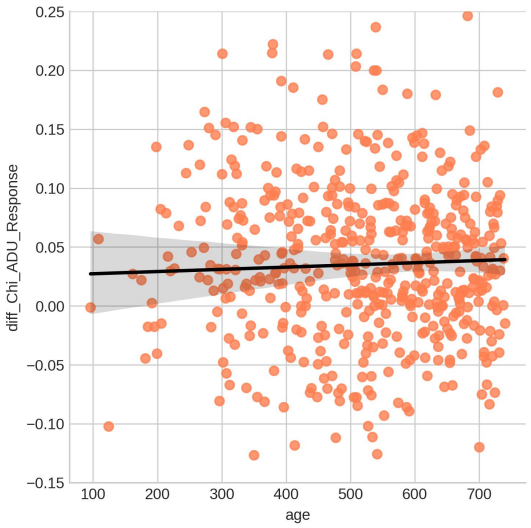
Figure 9: Adult response to Child speech vs. non-speech productions, replication of the second result of (Warlaumont et al., 2014)

C.3 Social Loop

The Figure 6 present the replication of the result on the social loop from (Warlaumont et al., 2014). More precisely it shows that given a children speech-related utterance, adult providing a response increase the proportion of speech-related (instead of non-speech related) follow-up utterance from the child.



(a) Child speech ratio (vs. non-speech) follow-up depending on whether or not an initial child speech-related utterance was responded by an adult or not (Mann-Whitney-Wilcoxon test two-sided, ****: p <= 1.00e-04)

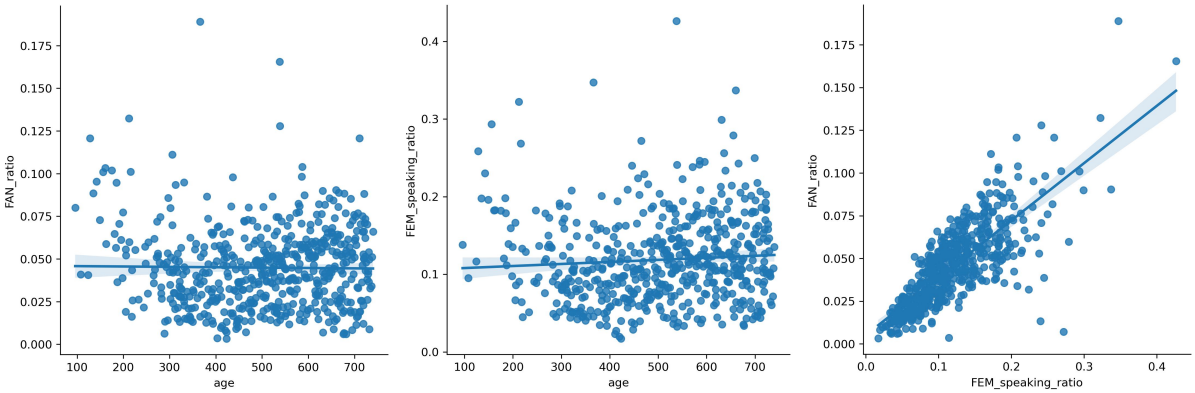


(b) Child ratio difference between Child speech and non-speech productions depending on whether or not an initial child speech-related utterance was responded by an adult or not (positive mean : indicating a tendency for more speech-related responses)

Figure 10: Adult response to Child speech vs. non-speech productions, replication of the second result of (Warlaumont et al., 2014) (left repeated from 6 in main text)

D Producing time ratios

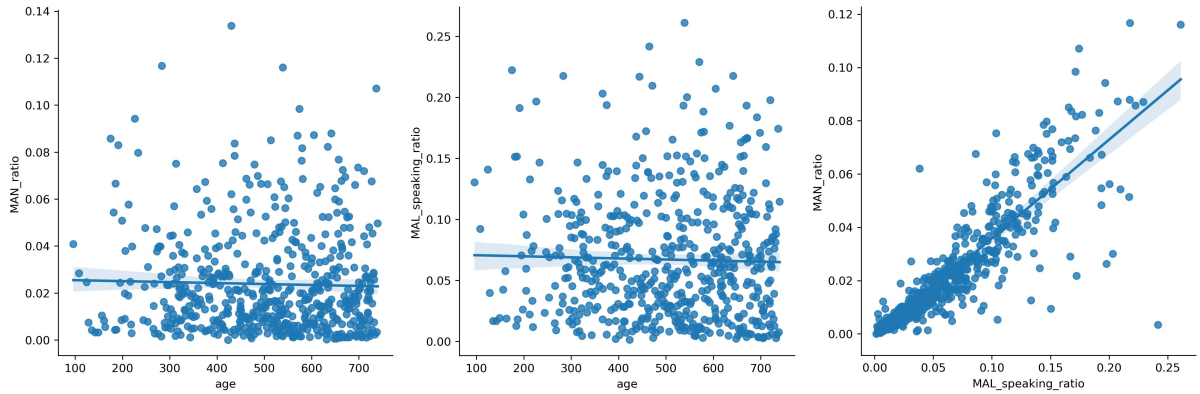
The Figures 11-13 are showing the ration of the time occupied by a category when available for both LENA and VTC as well as the correlation metrics. The Figure 14 focuses on individual children recordings from the selected data set.



(a) LENA (b) VTC (c) Correlation ($r = 0.773$)

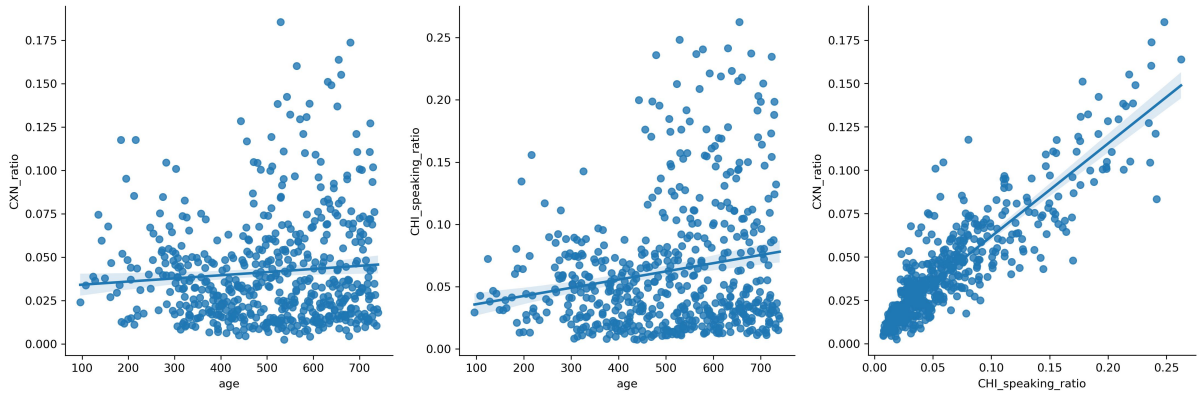
Figure 11: Female Speaking Time Ratio for LENA (a), VTC (b) and correlation plot between LENA and VTC (c). All children included (n = 20). Age in days.

```
Formula: chn_rs~age_rs+gender+ses_bin+ling_bin+gender*ses_bin+gender*ling_bin+ling_bin*ses_bin+(1

Number of observations: 540  Groups: {'child': 20.0}

Random effects:

                    Name    Var    Std
```



(a) LENA (b) VTC (c) Correlation ($r = 0.871$)

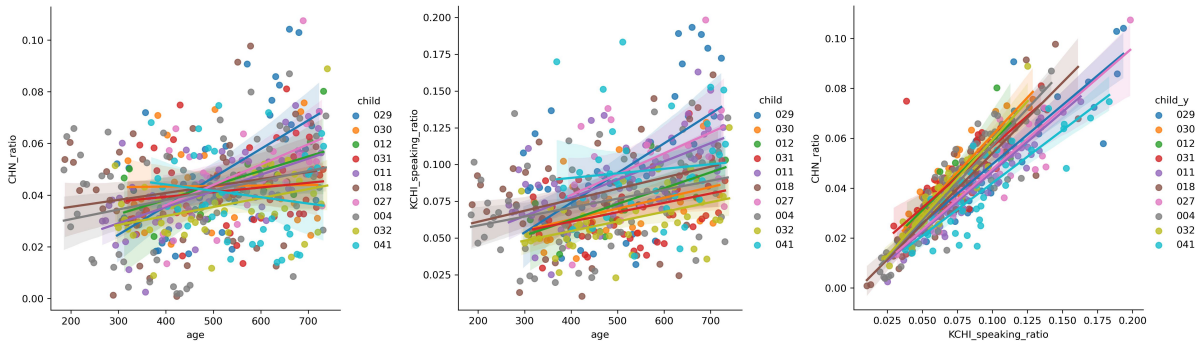
Figure 12: Male Speaking Time Ratio for LENA (a), VTC (b) and correlation plot between LENA and VTC (c). All children included (n = 20). Age in days.



(a) LENA (b) VTC (c) Correlation ($r = 0.917$)

Figure 13: Other child (than target child) Speaking Time Ratio for LENA (a), VTC (b) and correlation plot between LENA and VTC (c). All children included (n = 20). Age in days.

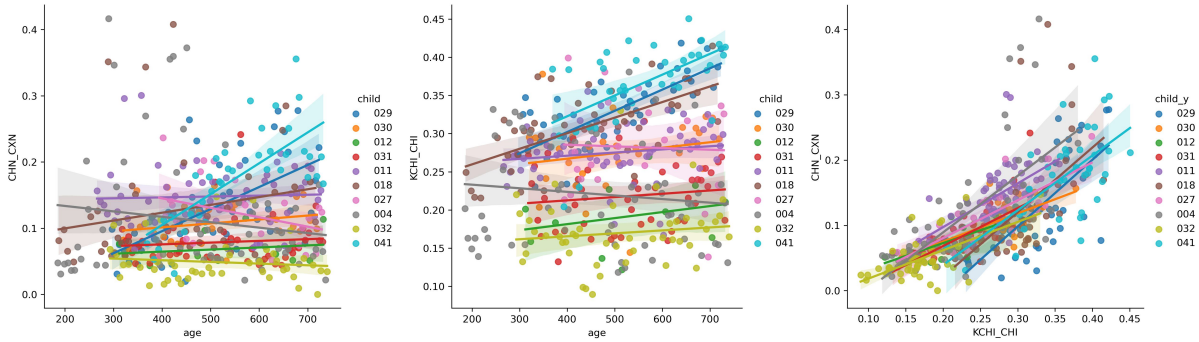
```
child     (Intercept)  0.012  0.109
Residual               0.014  0.119

Fixed effects:

                  Estimate  2.5_ci  97.5_ci     SE      DF  T-stat  P-val  Sig
(Intercept)          0.218   0.113    0.323  0.054  15.968   4.059  0.001  ***
age_rs               0.163   0.110    0.215  0.027 532.415   6.063  0.000  ***
genderM             -0.024  -0.163    0.115  0.071  12.270  -0.336  0.742
ses_binL            -0.104  -0.275    0.066  0.087  12.936  -1.201  0.251
ling_binP            0.046  -0.117    0.208  0.083  12.878   0.549  0.593
genderM:ses_binL     0.104  -0.194    0.402  0.152  13.024   0.684  0.506
genderM:ling_binP   -0.040  -0.288    0.207  0.126  12.866  -0.319  0.755
```



(a) LENA (b) VTC (c) Corelation

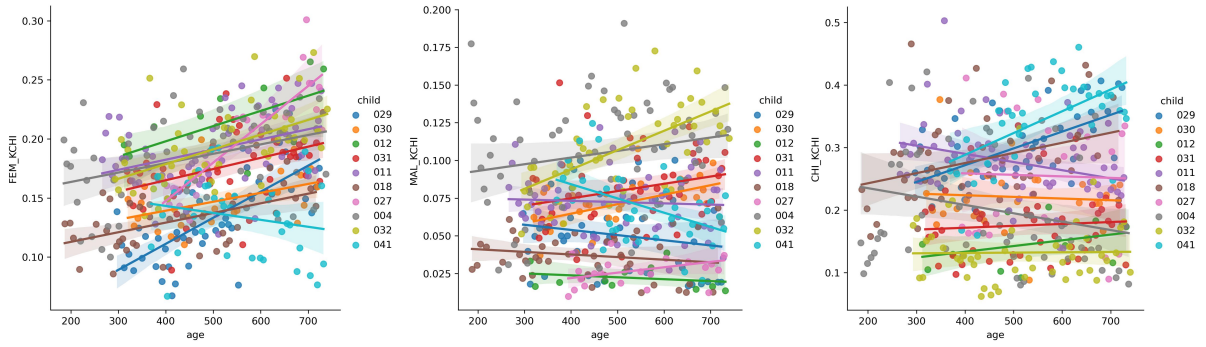
Figure 14: Target Child Speaking Time Ratio for LENA (a), VTC (b) and correlation plot (c). Selected children (n = 10, age span $\geq$ 9 months). Age in days.

E Temporal Contingencies Plots



(a) LENA (b) VTC (c) Correlation

Figure 15: CHILD>OTHER CHILD contingencies for LENA (a), VTC (b) and correlation plot (c). Selected children. Age in days.



(a) Female>Child (b) Male>Child (c) Other Child > Child

Figure 16: FEM>CHILD, MAL>CHILD, and OTHER CHILD > CHILD contingencies with VTC. Selected children. Age in days.

F Lena Annotation Process

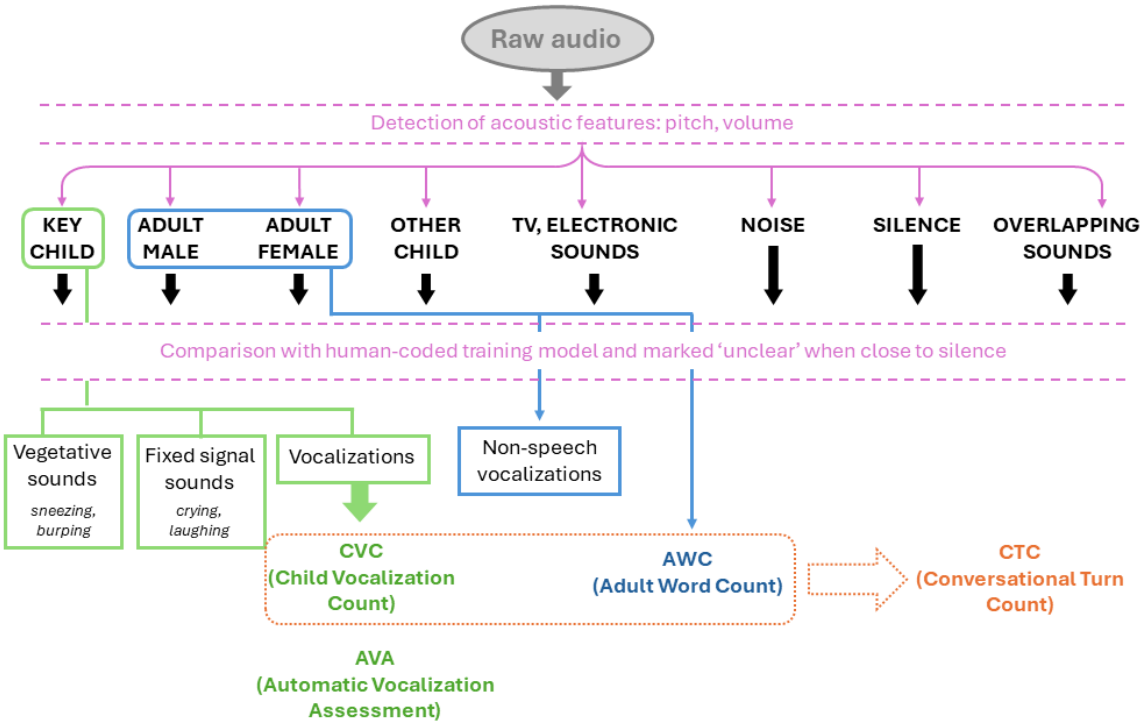


Figure 17: LENA annotation process.

G Instructions given to the families to use LENA

The instructions that were given to families were the following:

1. *When to record: once a week for a full day, until the child is 24 months old. Please prefer a day when you spend some time with your child (on the weekends for example). If you have no other choice, you can activate the device at daycare occasionally. We recommend that you record always on the same day of the week, to create a routine. Keep in mind that any day is a good day to record!*

2. *How to record: the instructions were kept identical to those provided by the LENA team. 1) Switch it on by pressing POWER; the screen should display "Paused". 2) Press RECORD for about 4 seconds; the screen should display "Recording". 3) Put the device in your child's shirt, the screen facing out, and close the pocket. 4) Leave it until the device turns off on its own at the end of the day.*

3. *Some various recommendations: never put the device out of the shirt; don't cover it with too many clothing layers; avoid noisy places as much as possible; remove the shirt (but leave the recorder inside) and keep it nearby during bath or nap times.*

4. *What to do after recording: bring the device back to the daycare center before a specific day of the week. The recorder will be ready for another week at the end of this day.*