



HAL
open science

Interactions entre le traitement des visages et le traitement du langage durant le rétrécissement perceptif

Olivier Clerc

► **To cite this version:**

Olivier Clerc. Interactions entre le traitement des visages et le traitement du langage durant le rétrécissement perceptif. Psychologie. Université Grenoble Alpes [2020-..], 2022. Français. NNT : 2022GRALS017 . tel-03828533

HAL Id: tel-03828533

<https://theses.hal.science/tel-03828533v1>

Submitted on 25 Oct 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

École doctorale : ISCE - Ingénierie pour la Santé la Cognition et l'Environnement

Spécialité : PCN - Sciences cognitives, psychologie et neurocognition

Unité de recherche : Laboratoire de Psychologie et Neuro Cognition

Interactions entre le traitement des visages et le traitement du langage durant le rétrécissement perceptif

Interactions between face processing and speech during the perceptual narrowing

Présentée par :

Olivier CLERC

Direction de thèse :

Olivier PASCALIS

Université Grenoble Alpes

Hélène LOEVENBRUCK

Chercheuse, Université Grenoble Alpes

Directeur de thèse

Co-directrice de thèse

Rapporteurs :

BENOIST SCHAAL

Directeur de recherche, CNRS DELEGATION CENTRE-EST

PASCAL BARONE

Directeur de recherche, CNRS DELEGATION OCCITANIE OUEST

Thèse soutenue publiquement le **25 mai 2022**, devant le jury composé de :

OLIVIER PASCALIS

Directeur de recherche, CNRS DELEGATION ALPES

BENOIST SCHAAL

Directeur de recherche, CNRS DELEGATION CENTRE-EST

PASCAL BARONE

Directeur de recherche, CNRS DELEGATION OCCITANIE OUEST

BAHIA GUELLAI

Maître de conférences, UNIVERSITE PARIS 10 - NANTERRE

KARINE MAZENS

Maître de conférences HDR, UNIVERSITE GRENOBLE ALPES

Directeur de thèse

Rapporteur

Rapporteur

Examinatrice

Présidente

Invités :

HELENE LOEVENBRUCK

Directeur de recherche, CNRS DELEGATION ALPES



REMERCIEMENTS

Financements

Cette thèse a été financé par l'ANR-DFG BabyFaceSpeech (ANR-17-FRAL-0014).

RESUME DE THESE

Dès la naissance, les nourrissons sont exposés à des visages qui parlent. Afin de pouvoir correctement interagir avec leurs congénères, les nouveau-nés vont devoir apprendre à traiter l'information provenant de ceux-ci. Le traitement des visages et le traitement du langage se développent ainsi rapidement durant la première année de vie des nourrissons. Cependant, que ce soit pour les visages ou pour le langage, beaucoup de nourrissons ont un biais d'exposition : ils sont presque exclusivement exposés aux visages de leur type et à leur langue maternelle. Une conséquence de ce biais d'exposition est que les nourrissons vont développer des capacités de discrimination plus fines pour traiter les stimuli natifs que les stimuli non-natifs. Dans la littérature scientifique, ce phénomène appelé rétrécissement perceptif a été mis en évidence de nombreuses fois dans le cadre du développement du langage et dans le cadre du développement du traitement des visages. La trajectoire développementale commune de ces deux systèmes cognitifs durant la première année de vie suggère des interactions entre ces deux systèmes. Cependant, ces interactions sont encore peu étudiées.

Le but de la thèse présentée ici était d'étudier les interactions entre les traitements du langage et des visages durant la première année de vie.

Dans une première étude, nous avons voulu étudier l'impact du type de visage sur une tâche de correspondance phonémique, sur des nourrissons de 3 et 9 mois. Les nourrissons de 3 mois ne semblent pas faire correspondre une voyelle avec la vidéo d'une locutrice si celle-ci n'est pas d'un type familier. Les résultats de cette étude nous indiquent que dès 3 mois, les nourrissons traitent différemment le signal audio-visuel selon le type du visage qui le produit. Dans une deuxième étude, nous avons voulu évaluer l'impact du type de visage sur la perception de l'effet McGurk, sur des nourrissons de 6, 9 et 12 mois. De plus, nous avons souhaité voir la robustesse de cet effet en l'étudiant de manière interculturelle (en France et au Japon). Nous montrons que la sensibilité à cette illusion audio-visuelle semble dépendante du type de visage. De plus, mis en commun avec nos collègues japonais, nos résultats montrent que la sensibilité à l'effet McGurk peut être conditionné par la culture dans laquelle grandissent les nourrissons. Dans une troisième étude, nous nous sommes intéressés à l'impact des associations entre types de visages et types de langues sur l'attention visuelle des nourrissons de 6, 9 et 12 mois. Cette étude montre qu'à 3 mois, certaines associations de langues et de visages semblent attendus par les nourrissons et plus regardées. Ces associations sont considérées comme congruentes puisqu'elles ne vont pas à l'encontre de ce que les nourrissons rencontrent habituellement dans

leur environnement. Dans une quatrième étude, nous avons testé l'impact de ces associations sur la reconnaissance d'individus par des nourrissons de 9 et 12 mois. Nous montrons que les associations congruentes aident la reconnaissance des individus, tandis que les associations incongruentes perturbent la reconnaissance des individus.

Ces études renforcent l'idée que d'étroites interactions lient le traitement du langage et le traitement des visages durant la petite enfance. De plus, nous montrons de nouveaux marqueurs du rétrécissement perceptif avant 9 mois. Nous montrons aussi un nouveau moyen expérimental permettant de moduler l'impact du rétrécissement perceptif. Ces travaux de thèse permettent d'élargir nos connaissances concernant le rétrécissement perceptif et ainsi d'en affiner notre compréhension.

Mot clés : traitement des visages, langage, rétrécissement perceptif, nourrissons, effet d'autre type, audio-visuel, psychologie développementale

ABSTRACT

From birth, infants are exposed to talking faces. Newborns will have to learn to process the information from them to be able to properly interact with their social partners. Face processing and language processing therefore develop quickly during infants' first year of life. However, many infants have an exposure bias, whether for faces or for language: they are almost exclusively exposed to faces of their own type and to their native language. A consequence of this exposure bias is that infants will develop finer discrimination skills to process native stimuli than non-native stimuli. In the scientific literature, this phenomenon called perceptual narrowing has been highlighted many times in the context of language development and in the context of the development of face processing. The similar developmental trajectory of these two cognitive systems during the first year of life suggests interactions between these two systems. However, these interactions have not been investigated in detail.

The aim of this thesis was to study the interactions between language and face processing during the first year of life.

In a first study, we wanted to study the impact of face's familiarity (established on race) on a phonemic matching task, on 3- and 9- month-old infants. 3-month-old infants do not seem to match a vowel with a corresponding speaker's video if the speaker is not of their own race. The results of this study indicate that from 3 months, infants process the audio-visual signal differently depending on the race of the face that produces it. In a second study, we wanted to evaluate the impact of face's familiarity on the perception of the McGurk effect, on 6-, 9- and 12-month-old infants. Moreover, we wanted to see the robustness of this effect by studying it cross-culturally (in France and in Japan). We show that the sensitivity to this audio-visual illusion seems to be dependent on the race of the face. Moreover, our results and the Japanese's results differ, showing that sensitivity to the McGurk effect may be conditioned by the culture in which infants grow up. In a third study, we were interested in the impact of associations between face types and language types on the visual attention of 6-, 9- and 12-month-old infants. This study shows that at 3 months, some associations of languages and faces seem expected by infants and are consequently looked at longer. These associations are considered congruent since they do not disrupt the type of associations infants usually encounter in their environment. In a fourth study, we tested the impact of these associations on other-race face recognition by 9- and 12-month-old infants. We found that congruent associations help the

recognition of individuals, while incongruent associations disturb the recognition of individuals.

These studies support the idea that close interactions exist between language processing and face processing during infancy. Moreover, we show new perceptual narrowing's consequences in infants under 9 months. We also show a new experimental way to modulate the impact of the perceptual narrowing. This thesis broadens our knowledge of perceptual narrowing and helps us to better understand it.

Keywords: face processing, language, perceptual narrowing, infants, other-race effect, audio-visual, developmental psychology

TABLE DES MATIERES

Liste des Abréviations	16
Table des Figures	17
Table des tableaux	19
Partie théorique	20
Chapitre 1 : Le traitement des visages chez l'adulte et le nourrisson.....	21
I. Un traitement spécialisé.....	21
I.1. Les visages : des stimuli uniques ?.....	21
I.2. L'effet d'inversion.....	22
I.3. Traitement configural et componentiel	22
I.4. Traitement holistique.....	24
I.5. L'effet d'expertise	27
II. L'apport de la neurophysiologie.....	29
II.1. La Fusiform Face Area.....	29
II.2. La réponse neuronale	31
II.3. Les patients cérébro-lésés.....	31
III. Le système visuel et son développement	32
III.1. Le système visuel	32
III.2. Développement des capacités, de <i>in utero</i> jusqu'à la naissance	35
III.3. Développement de l'acuité et de la capacité d'accommodation	37
III.4. Développement de la saccade oculaire.....	38
III.5. Développement de la perception des couleurs	39
III.6. Développement de la sensibilité aux mouvements	39
III.7. Développement de la perception de profondeur.....	40
IV. Modèles du traitement des visages.....	41
IV.1. Le modèle de Bruce & Young.....	41
IV.2. Les modèles plus récents.....	43
V. L'espace de représentation des visages	45
V.1. L'effet de distinctivité	45
V.2. L'effet d'autre type.....	46
V.3. Le modèle de Valentine.....	47
VI. Perception des visages chez le nouveau-né.....	48
VI.1. Des mécanismes spécifiques : Conspec/Conlern	49

VI.2. Des mécanismes sensoriels dépendant de l'expérience	50
VI.3. Des capacités plus poussées : la reconnaissance de la mère	52
Chapitre 2 : Le rétrécissement perceptif	55
I. Rétrécissement perceptif pour le traitement des visages	55
I.1. Au niveau de l'espèce	55
I.2. Au niveau du type	56
II. Rétrécissement perceptif pour le traitement du langage.....	60
II.1. Diminution de la capacité à discriminer les contrastes phonémiques non-natifs	60
II.2. Amélioration de la capacité à discriminer les contrastes phonémiques natifs ...	62
II.3. Un modèle explicatif : le Native Language Magnet model.....	64
III. Rétrécissement perceptif inter-sensoriel	66
IV. Implications sur la cognition sociale précoce	68
V. Moduler le rétrécissement perceptif	70
V.1. Par l'expérimentation	70
V.1.a. Après l'émergence du rétrécissement perceptif.....	70
V.1.b. Avant l'émergence du rétrécissement perceptif.....	71
V.1. Par l'environnement	73
V.1.a. Par l'exposition à différents types de visages	73
V.1.b. Par l'exposition à différentes langues	77
VI. La nature du rétrécissement perceptif	80
VI.1. Un phénomène inter-sensoriel.....	80
VI.2. Un processus non spécifique à l'Humain.....	83
VI.3. L'apprentissage statistique et la mise en place du rétrécissement perceptif	85
VI.3.a. L'apprentissage statistique.....	85
VI.3.b. L'apprentissage statistique et son influence sur le rétrécissement perceptif pour les langues	86
VI.3.c. L'apprentissage statistique et son influence sur le rétrécissement perceptif pour les visages.....	91
VI.4. Attentes et apprentissage	94
VI.5. Les mécanismes neuronaux à l'origine du rétrécissement perceptif.....	96
Chapitre 3 : Intégration multisensorielle	99
I. Apport de la modalité visuelle sur la compréhension de la parole	100
II. Intégration multisensorielle chez le nourrisson	101
II.1. Redondance inter-sensorielle	101
II.2. Le cas de l'intégration audio-visuelle	103
Chapitre 4 : Méthodologie de l'expérimentation sur le nourrisson	107

I.	Un traitement spécialisé.....	107
I.1.	Le regard préférentiel	107
I.2.	Habituation/déshabituation.....	108
II.	L'expérimentation au Babylab de Grenoble.....	109
II.1.	Recrutement des participants	109
II.2.	Accueil et recueil de données	110
II.3.	Mesures oculométriques avec un appareil EyeLink.....	110
II.4.	Mesure de temps de regard.....	111
II.5.	Codage des vidéos	113
	Problématique.....	114
	Partie expérimentale	116
	Chapitre 5 : Correspondance audio-visuelle de voyelles sur des visages de même type et d'autre type	117
I.	Introduction	117
I.1.	Capacités de distinction des sons de langues	117
I.2.	Sensibilité à la redondance multisensorielle	118
I.3.	La tâche de correspondance audio-visuelle de voyelle	118
I.4.	Un rétrécissement perceptif multisensoriel ?	122
I.5.	Objectif et hypothèses de l'étude 1	126
II.	Matériel et méthodes	128
II.1.	Participants	128
II.2.	Stimuli	128
II.3.	Procédure.....	129
II.4.	Codage et analyse.....	130
III.	Résultats.....	131
III.1.	Phase de familiarisation	131
III.2.	Phase de test	131
III.3.	Résultats pour les nourrissons de 3 mois	132
III.3.a.	Pour le groupe exposé aux visages de même type	133
III.3.b.	Pour le groupe exposé aux visages d'un autre type	133
III.4.	Résultats pour les nourrissons de 9 mois	135
III.4.a.	Pour le groupe exposé aux visages de même type	135
III.4.b.	Pour le groupe exposé aux visages d'un autre type	136
IV.	Discussion.....	137
IV.1.	Un impact du type de visage précoce sur la capacité de correspondance phonémique	138

IV.2. Une tâche peu adaptée pour les nourrissons de 9 mois ?	140
Chapitre 6 : Effet McGurk sur des visages de même type et d'autre type	143
I. Introduction	143
I.1. L'effet McGurk	143
I.1.a. Paramètres jouant sur la sensibilité à l'effet	143
I.1.b. L'effet McGurk chez le nourrisson	146
I.2. Objectif et hypothèses de l'étude 2	149
II. Matériel et méthode	152
II.1. Participants	152
II.2. Stimuli	152
II.3. Procédure	153
III. Résultats	153
III.1. Phase de familiarisation	153
III.2. Phase de test	155
III.2.a. Modèle ART	155
III.2.b. Analyse comparative avec Ujiie et al. (2020)	156
III.2.c. Analyse sur le 1 ^{er} test avec temps de regard pondérés	156
III.2.d. Analyse des taux de déshabituations	159
IV. Discussion	162
IV.1. L'effet McGurk présenté avec un visage de même type	162
IV.2. L'effet McGurk présenté avec un visage d'autre type	164
IV.3. Limites du protocole	165
IV.4. La tâche de McGurk comme proxy de l'intégration audio-visuelle	167
Chapitre 7 : Association langage/visage	169
I. Introduction	169
I.1. Associations voix/visages	169
I.2. Protocole d'association intermodale	172
I.3. Objectif et hypothèses de l'étude 3	173
II. Matériels et méthodes	174
II.1. Participants	174
II.2. Stimuli	175
II.3. Procédure	175
II.4. Analyse	176
III. Résultats	176
III.1. Temps de regard avant association	177

III.2. Temps de regard après association.....	177
IV. Discussion.....	181
IV.1. Baisse générale du temps de regard	182
IV.2. Nature de l'association langage/visage	182
IV.3. Des associations plus précoces ?.....	183
IV.4. Catégorisation de la langue	183
Chapitre 8 : Influence de la langue sur la reconnaissance de visages d'autre type	189
I. Introduction	189
I.1. L'effet de familiarité du langage.....	189
I.2. Objectif et hypothèses de l'étude 4	190
II. Matériel et méthode	192
II.1. Participants	192
II.2. Stimuli	192
II.3. Procédure.....	193
III. Résultats.....	194
III.1. Tests de familiarisation	194
III.2. Préférence lors du test de reconnaissance	194
IV. Discussion.....	196
IV.1. La langue et le visage en tant que catégoriseurs sociaux	196
IV.2. Redondance inter-sensoriel et incongruence.....	197
IV.3. Reconsidérer le rétrécissement perceptif pour les visages	198
IV.4. Conclusions	199
Chapitre 9 : Discussion générale	201
I. Récapitulatif de l'objectif de la thèse et des résultats principaux.....	201
II. Remise en contexte dans le cadre du modèle <i>face-space</i>	203
III. Reconsidérer le rétrécissement perceptif	207
III.1. Des marqueurs plus précoces	207
III.2. Des conséquences différentes avec des stimuli multisensoriels.....	208
IV. Le cas des milieux hétérogènes.....	210
V. Conclusion	211
Références	213

LISTE DES ABBREVIATIONS

ASL : langue des signes américaine

CGL : Corps Géniculé Latéral

CV : Consonne-Voyelle

EEG : Electroencéphalographie

FFA : Fusiform face Area

IRMf : imagerie par résonance magnétique fonctionnelle

ISPIf : imagerie spectroscopique proche infrarouge fonctionnelle

LEF : Language Familiarity Effect

M : Magnocellulaire

MEG : Magnétoencéphalographie

NLMm : Native Language Magnet model

OFA : Occipital Face Area

ORE : Other Race Effect

P : Parvocellulaire

PE : Potentiel Evoqué

PME : Perceptual Magnet Effect

STS : Sillon Temporal Supérieur

V1 : Aire visuel primaire

V2 : Aire visuel secondaire

TABLE DES FIGURES

Figure 1. Illusion « Thatcher », reproduit d'après Thompson (1980).....	25
Figure 2. Exemple d'éléments (haut), visages entiers (milieu), et visages mélangés (bas) présenté par Tanaka et Farah (1993).....	26
Figure 3. Stimuli utilisés par Reid et al. (2017).	36
Figure 4. Modèle de reconnaissance des visages, tiré de Bruce & Young (1986).....	43
Figure 5. Modèle de distribution neuronale de la perception des visages. Tiré de Haxby et al (2000).....	45
Figure 6. Représentation en deux dimensions de l'espace de représentation des visages de Valentine (1991).	48
Figure 7. Stimuli utilisés par Johnson et al. (1991) : un schéma de visage avec ses éléments constitutifs, un schéma de visage avec les éléments constitutifs mélangé, et un schéma de visage sans ses éléments constitutifs.....	50
Figure 8. Exemple de stimuli via lequel l'ORE est mis en évidence chez les nourrissons.	59
Figure 9. Capacité de discrimination de contrastes phonémiques de l'hindi et du salish par des nourrissons monolingues d'environnements anglais, hindi ou salish. Tiré et adapté de Werker & Tees (1984).	62
Figure 10. Evolution de l'amplitude de la Mismatch Negativity au cours de la 1 ^{ère} année de vie chez des nourrissons finnois et estoniens de l'étude de Cheour et al. (1998).....	64
Figure 11. Illustration du PME.....	66
Figure 12. Exemple de stimuli vidéo et audio utilisés dans le test d'intégration multisensorielle de Lewkowicz & Ghazanfar (2006).....	67
Figure 13. Nombre de présentation des huit stimuli lors de la familiarisation en fonction des deux conditions de l'étude de Maye et al. (2022).	87
Figure 14. Schéma du fonctionnement du modèle de Schatz et al. (2021) lors de la phase d'entraînement (a) et lors de la phase de test (b).	89
Figure 15. Intensité relative de la réponse <i>thêta</i> par conditions de l'étude de Bosseler et al. (2013).	91
Figure 16. Nombre de présentation des 21 différents stimuli lors de la familiarisation en fonction des deux conditions de l'étude de Altvater-Mackensen et al. (2022).....	93
Figure 17. Stimuli utilisés dans l'étude de Krasotkina et al. (2021) pour les différents degrés de morphing. Tiré de Krasotkina et al. (2021).	94
Figure 18. Différence de temps de regard proportionnel vers les yeux et la bouche au cours de la 1 ^{ère} année de vie et à l'âge adulte. Tiré de Lewkowicz & Hansen-Tift (2012).	104
Figure 19. Dispositif oculométrique au Babylab. Repris de Damon (2015).	111
Figure 20. Schéma en vue du dessus de la salle d'expérimentation du Babylab.	112
Figure 21. Protocole mis en place par Pons et al. (2009) pour tester la capacité des nourrissons à faire correspondre des phonèmes non-natifs avec des visèmes non-natifs. Tiré de Pons et al. (2009).	123
Figure 22. Protocole expérimental utilisé dans l'étude 1.	130
Figure 23. Pourcentage de temps passé à regarder le visage qui correspond à la voyelle jouée par groupe d'âge et par type de visage.....	132
Figure 24. Pourcentage de temps passé à regarder le visage correspondant à la voyelle jouée pour les nourrissons de 3 mois par type de visage et par voyelle.	134

Figure 25. Evolution du pourcentage de temps passé à regarder chaque visage, pour les nourrissons de 3 mois.....	135
Figure 26. Pourcentage de temps passé à regarder le visage correspondant à la voyelle jouée pour les nourrissons de 9 mois par type de visage et par son joué.	136
Figure 27. Evolution du pourcentage de temps passé à regarder chaque visage, pour les nourrissons de 9 mois.....	137
Figure 28. Temps de regard des nourrissons de 8-9 mois dans l'étude de Ujiie et al. (2020) vers le stimulus issu de la fusion (/ta/) et le stimulus audio (/pa/) durant la phase de test, après familiarisation à l'effet McGurk.	150
Figure 29. Temps de regard des nourrissons durant la phase de familiarisation, en fonction des essais, de l'âge et des conditions.....	154
Figure 30. Temps de regard pondérés des nourrissons durant la première phase de test.....	158
Figure 31. Temps de regard durant chaque phase de familiarisation et durant la première phase de test (après la ligne verticale de pointillée).....	160
Figure 32. Temps de regard pondéré durant la première phase de test par type de visage (à gauche, d'autre type ; à droite, de même type) en fonction de l'âge des nourrissons et du stimulus de test (en rouge, /pa/ ; en bleu, /ta/).	162
Figure 33. Exemple de stimuli utilisé par Vouloumanos et al. (2009).....	171
Figure 34. Protocole mise en place par Xiao et al. (2016). Tiré de Xiao et al. (2016).....	173
Figure 35. Schéma de la procédure de l'étude 3.	176
Figure 36. Temps de regard vers les visages durant le premier essai (avant association avec une langue) en fonction de son type. La p-value correspond à un test de Wilcoxon.	177
Figure 37. Evolution du temps de regard moyen par essai (en secondes) vers les visages en fonction de l'âge (en jours), pour chaque condition.	178
Figure 38. Temps de regard moyen (en secondes) vers les visages par âge et par condition.	179
Figure 39. Temps de regard (en secondes) vers le visage en fonction de l'essai, par groupe d'âge et par condition.....	181
Figure 40. Evolution du temps de regard (en secondes) pour le dernier essai vers le visage en fonction de l'âge (en jours), pour chaque condition.	185
Figure 41. Temps de regard (en secondes) lors du dernier essai vers le visage par âge et par condition.	186
Figure 42. Stimuli utilisés. Les stimuli étaient en couleurs et présentés sur un écran 22 pouces.	193
Figure 43. Score de préférence à la nouveauté en fonction de la familiarité de la langue et de l'âge.....	195
Figure 44. Représentation arbitraire du développement de l'espace de distribution des visages.	206
Figure 45. Les individus métis, situés à mi-distance entre la périphérie du nuage de points central et d'un nuage de points périphérique, sont plus proche du centre du nuage périphérique (soit du prototype de visage d'un autre type) ; ils sont catégorisés comme appartenant au groupe de visages d'un autre type	207

TABLE DES TABLEAUX

Tableau 1. Moyenne du temps de regard pondéré vers le stimulus /pa/ durant la phase de test pour les nourrissons de 6, 9 et 12 mois, pour la condition avec visage de même type et pour la condition avec visages d'autres types. Les p-values sont issues de test de Wilcoxon comparant le temps de regard durant la présentation du stimulus /pa/ contre le temps de regard durant la présentation du stimulus /ta/.....	158
Tableau 2. Moyenne du temps de regard proportionnel vers les stimuli /pa/ et /ta/ durant la phase de test pour les nourrissons de 6, 9 et 12 mois, pour la condition avec visage de même type et pour la condition avec visages d'autres types. Les p-values sont issues de test de Wilcoxon comparant le temps de regard durant la présentation du premier stimulus de test contre le temps de regard durant la présentation de la dernière phase de familiarisation.	161
Tableau 3. Répartition des nourrissons par groupes d'âges et par condition (type de visage et type de langue). Entre parenthèse : la moyenne d'âge et l'écart type de l'âge par sous-groupe.	174

PARTIE THEORIQUE

Chapitre 1 : Le traitement des visages chez l'adulte et le nourrisson

I. Un traitement spécialisé

I.1. Les visages : des stimuli uniques ?

Dès la naissance (9 minutes après pour être précis), les nourrissons présentent une inclinaison naturelle pour les visages (Goren et al., 1975). Les nouveau-nés regardent plus longtemps des schémas représentant des visages que des schémas de visages dont les éléments internes ont été mélangés ou que des schémas de visages vidés de leurs éléments internes. Un attrait très précoce, qualifié d'inné par certains, semble donc exister pour les structures schématiques rappelant les visages. Une découverte comme celle-ci nous amène à penser que les visages sont une classe de stimuli visuels à part et que nous pourrions posséder dès la naissance un système de traitement et de reconnaissance particulier pour ceux-ci.

Au-delà de l'aspect perceptif, on peut en effet considérer les visages comme occupant une place particulière dans notre vie. Parmi la multitude de stimuli auxquels le nouveau-né va être exposé dès la naissance, les visages ont une importance singulière. Traiter ces stimuli est crucial pour gérer efficacement ses futures interactions sociales ainsi que pour le développement des comportements d'*attachement* avec ses parents (Bowlby, 1982 ; voir Tereno et al., 2007 pour une revue). Les visages nous apportent en effet des informations plutôt stables sur nos partenaires sociaux, comme leur âge, leur sexe, leur origine, ainsi que des informations plus temporaires comme l'état émotionnel ou attentionnel. Une déficience dans le traitement des visages amène généralement une incapacité à gérer correctement les interactions sociales, comme c'est le cas pour les personnes présentant un trouble autistique, bien que d'autres facteurs puissent entrer en jeu (Hobson et al., 1988 ; Lord et al., 1999). Les visages vont aussi être le vecteur par lesquels les nouveau-nés vont être exposés au langage, renforçant ainsi leur importance sociale. Nos interactions langagières prennent généralement place face à face et c'est d'autant plus vrai pour les bébés. La communication verbale est de ce fait facilitée par la lecture labiale, qui améliore la compréhension auditive (Erber, 1969 ; Sumbly & Pollack, 1954).

Les visages sont une catégorie de stimuli très homogènes. Ils sont composés des mêmes éléments (nez, bouche, yeux...) organisés les uns par rapport aux autres de manière similaire. Ils sont un moyen privilégié (avec la voix) pour identifier une personne. Identifier un visage se fait très rapidement, en moins de 200 ms (Jacques & Rossion, 2006), et une exposition de

seulement 20 ms à ce visage est suffisante pour cette performance (Rizzolatti & Buchtel, 1977). Ces compétences impressionnantes en dépit des fortes ressemblances entre les visages ont amené les chercheurs à postuler l'existence d'un système de traitement pour les visages différent de celui utilisé pour les autres catégories de stimuli visuels, ainsi qu'à chercher à en expliquer ses mécanismes tant sur le plan comportemental que neuronal (Bruce & Young, 1986 ; Farah, 1996 ; McCarthy et al., 1997 ; Ellis & Lewis, 2001).

I.2. L'effet d'inversion

Robert Yin figure parmi les premiers chercheurs ayant tenté de démontrer l'existence de mécanismes dédiés à la reconnaissance des visages. Son étude de 1969 (Yin, 1969), pierre angulaire dans la recherche sur le traitement des visages, met en évidence la spécificité des visages par rapport aux autres objets visuels via un effet emblématique : l'effet d'inversion. L'effet d'inversion est mis en évidence lorsqu'un visage est présenté au sujet renversé à 180°. Dans le protocole de Yin, des sujets étaient d'abord familiarisés à des images soit à l'endroit, soit retournées. Les images étaient soit des visages, soit des maisons, soit des avions, soit des hommes en mouvement. Une seconde phase consistait en la présentation d'une image de la phase de familiarisation ainsi que d'une nouvelle image de la même catégorie. Les deux étaient présentées dans le sens de la familiarisation, et le sujet devait retrouver l'image de la familiarisation. Le temps moyen d'identification était plus long dans la condition à l'envers qu'à l'endroit, quel que soit le type d'image. Cependant, alors que les visages étaient le type d'images le plus rapidement identifié lorsqu'ils étaient présentés à l'endroit, ils devenaient la catégorie la plus lentement identifiée lorsqu'ils étaient présentés à l'envers. Un effet d'orientation existe donc pour toutes les images, les objets orientés de manière inhabituelle étant plus difficiles à identifier, mais il y a aussi un effet d'inversion spécifique aux visages, ceux-ci étant plus difficilement identifiés que les autres catégories.

I.3. Traitement configural et componentiel

L'effet d'inversion spécifique aux visages a depuis été répliqué de nombreuses fois (pour une revue, voir Valentine, 1988 et Rossion, 2008). Des questionnements surviennent alors quant à l'origine de cet effet. Bien que ne répondant pas à la question de manière empirique, Yin récolta dans cette même étude des informations, via un simple questionnaire, concernant la stratégie adoptée par les sujets pour identifier les différents stimuli. Les réponses firent émerger deux méthodes distinctes : la première consistait à identifier un trait caractéristique de l'objet permettant de le reconnaître via ce trait, la deuxième consistait à se faire une idée générale de

l'objet. Alors que la première méthode était employée de manière dominante pour les objets non-visages, les visages étaient plutôt reconnus via la deuxième méthode. De manière très intéressante, quand les visages étaient présentés en configuration inversée, 100% des sujets ont trouvé inefficace d'employer la deuxième stratégie, c'est-à-dire l'analyse globale plutôt que la recherche d'un trait caractéristique. Carey & Diamond (1977) ont proposé une hypothèse pour expliquer cet effet d'inversion. Leur postulat est le suivant : alors que les traits du visage ne sont pas impactés par l'effet d'inversion, les relations entre les différents traits est quant à elle complètement différente. C'est pourquoi les sujets ne peuvent plus s'appuyer sur celle-ci pour identifier un visage présenté à l'envers. Puisque la capacité à identifier un visage présenté à l'envers est fortement réduite, les relations entre les composants du visage doivent jouer un rôle essentiel pour sa reconnaissance en configuration canonique. Les auteurs distinguent donc plusieurs types de traitement permettant différents niveaux d'analyse (Diamond & Carey, 1986). Ils distinguent tout d'abord ce que l'on nomme les relations configurales de premier ordre, qui comprennent les relations constantes entre tous les visages, comme le fait que les yeux soient situés au-dessus du nez, que le nez soit au-dessus de la bouche, etc. Le traitement configural de premier ordre a pour fonction d'identifier le schéma caractéristique du visage le distinguant des stimuli ne partageant pas la même configuration et va ainsi permettre la catégorisation du stimulus en tant que visage. Ce système cognitif permet de détecter rapidement des visages au milieu d'une scène visuelle, sans se soucier des détails les constituant. C'est pourquoi il est courant d'identifier des visages là où il n'y en a pas, par phénomène de paréidolie. Ce type d'analyse n'apparaît cependant pas suffisant pour accéder à l'identité. Mais alors comment identifie-t-on un visage ? Pour individualiser les visages, une possibilité est de prendre en compte les traits de ceux-ci. Il est en effet probable que les visages diffèrent par la couleur des yeux, la taille du nez, ou encore par la forme de la bouche. C'est ce que l'on appelle des variations de composants, donnant lieu au traitement componentiel, ou analytique (en anglais, généralement retrouvé sous le terme « featural » ou des traits en français). L'autre type de distinction que l'on peut établir pour différencier des visages est dans la relation entre ces différents traits, ou plus précisément la distance entre ceux-ci. En effet, l'espacement des yeux ou la distance entre la bouche et le nez vont varier entre individus ; c'est ce que l'on nomme les relations configurales de second ordre. Les propriétés relationnelles de second ordre correspondent à l'extraction des particularités relationnelles au sein d'éléments (ici les visages) qui partagent une même configuration globale. C'est aussi les relations configurales de second ordre qui permettent la détection des changements d'états émotionnels ;

en effet, une personne passant de la joie à la tristesse a toujours les mêmes traits, mais l'agencement de ceux-ci va se modifier.

Les résultats des sujets de l'étude de Yin nous montrent que le traitement componentiel est la seule stratégie utilisée pour reconnaître des visages inversés, signifiant que l'effet d'inversion semble rendre le traitement configural de second ordre décrit par Carey & Diamond très difficile. D'autres études ont depuis confirmé que l'effet d'inversion impacte plus fortement le traitement configural de second ordre que le traitement componentiel. Sergent (1984) a montré que des modifications des propriétés relationnelles de second ordre du visage (par exemple la distance entre les yeux) sont beaucoup moins perceptibles lorsque le visage est présenté inversé. La modification de traits (par exemple enlever des dents ou supprimer la pupille) se remarque quant à elle dans les deux sens (Searcy & Bartlett, 1996). Puisque le traitement du visage est beaucoup plus difficile lorsqu'il repose principalement sur le traitement componentiel (comme c'est le cas d'une présentation inversée), la conclusion fut que le traitement des visages s'appuie principalement sur le traitement configural de second ordre.

I.4. Traitement holistique

L'illusion de « Thatcher » est une expérience classique de cette littérature ayant contribué à notre compréhension actuelle du traitement des visages (Thompson, 1980). Cette illusion tire son nom de Margaret Thatcher, ancienne premier ministre anglaise, puisque c'est son visage qui a été utilisé pour mettre en évidence cet effet. Dans le cadre de cette illusion, les yeux et la bouche sont retournés mais toujours situés à leur emplacement d'origine. Si le visage est présenté dans le sens normal, le visage est perçu comme monstrueux (voir Figure 1). La seule inversion des yeux est suffisante pour produire cet effet (Lewis & Johnston, 1997). En revanche, s'il est présenté inversé, l'effet se remarque beaucoup moins (Bartlett & Searcy, 1993 ; Searcy & Bartlett, 1996). Bien que l'on puisse tenter d'expliquer ces résultats via le traitement configural ou componentiel, de telles interprétations trouvent vite leurs limites. En effet, on retrouve bien dans cette illusion une bouche et des yeux à l'endroit où ils se trouvent normalement sur un visage, et les distances séparant les éléments n'ont pas été modifiées. De plus, bien que retournés, les composants du visage sont les mêmes et n'ont pas subi de modifications. Le traitement configural de premier ordre et de second ordre ainsi que le traitement componentiel ne suffiraient pas à expliquer ce phénomène. Une solution possible amène à définir un nouveau traitement, le traitement « holistique ». Celui-ci consiste en l'appréhension du visage dans sa globalité, en « gestalt » (terme allemand signifiant « forme »)

et est donc un traitement configural. Puisque ce traitement est très fortement impacté lorsque le visage est présenté à l'envers (le caractère monstrueux n'est presque pas perçu) et que l'analyse componentielle n'est que peu affectée par l'effet d'inversion, l'effet d'inversion observé par Yin pourrait être causé par une disruption du traitement holistique. Il serait cependant probablement plus juste de conclure que l'effet d'inversion impacte les traitements configuraux, qu'ils soient de second ordre ou holistiques.



Figure 1. Illusion « Thatcher », reproduit d'après Thompson (1980)

Le traitement holistique est notamment mis en évidence par l'effet composite (Young et al., 1987). Pour produire l'effet composite, la moitié supérieure d'un visage familier est aligné avec la moitié inférieure d'un autre visage. Il devient alors plus difficile d'identifier cette partie supérieure que lorsqu'elle est présentée dans son contexte originel. En revanche, si les deux parties ne sont pas alignées, l'effet composite disparaît. Lorsque les parties sont alignées, le visage est donc perçu comme un nouveau tout, qui est différent de la somme de ses parties. Rompre l'intégralité du visage permet de ré-accéder à l'identité, ce qui signifie que le traitement n'est plus holistique. De plus, l'effet composite est aussi annulé lorsque les visages sont présentés inversés. Le traitement holistique, comme le traitement configural de second ordre, est rendu difficile voire impossible par l'effet d'inversion, le traitement componentiel devenant alors le seul moyen d'identification. On pourrait cependant arguer que le traitement configural de second ordre est lui aussi impacté lors de l'effet composite ; en effet, les relations entre les différents composants du visage ne sont forcément pas les mêmes dans le visage composite que dans le visage d'origine. Il est donc difficile de réellement évaluer la part du traitement holistique dans le cadre de cette étude. Une illustration peut-être plus pertinente du traitement holistique fut présentée par Tanaka & Farah (1993) à travers ce que l'on appelle le « whole-part effect ». Après apprentissage de différents visages ainsi que des noms qui leur sont associés,

un composant est présenté soit isolément soit dans son contexte d'origine. Concrètement, dans le premier cas, le nez de Larry sera présenté sans le reste du visage de Larry, côte à côte avec un autre nez. Dans le deuxième cas, le visage de Larry avec son nez d'origine sera présenté côte à côte avec le visage de Larry avec un autre nez. On reconnaît alors plus facilement le composant (le nez de Larry) dans son contexte initial (le visage de Larry) qu'isolément. Pour s'assurer que cet effet n'est pas dû aux autres éléments du visage, les auteurs ont aussi entraînés et testés des sujets avec des « scrambled faces » (visages avec les différents composants mélangés) ou des visages inversés, ce qui dans les deux cas perturbe les relations configurales. Ils ont trouvé que le nez de Larry était moins bien identifié dans un visage mélangé ou inversé que dans le visage d'origine, confirmant que le traitement holistique est responsable du « whole part effect » (voir Figure 2). Les auteurs ont aussi voulu voir si cet effet se retrouve sur des stimuli autres que des visages et ont donc testé les sujets avec des maisons (à la place des visages) et des portes (à la place des nez). La reconnaissance fut semblable que la porte soit présentée isolément ou dans son contexte, le traitement holistique semblant donc spécifique aux visages.

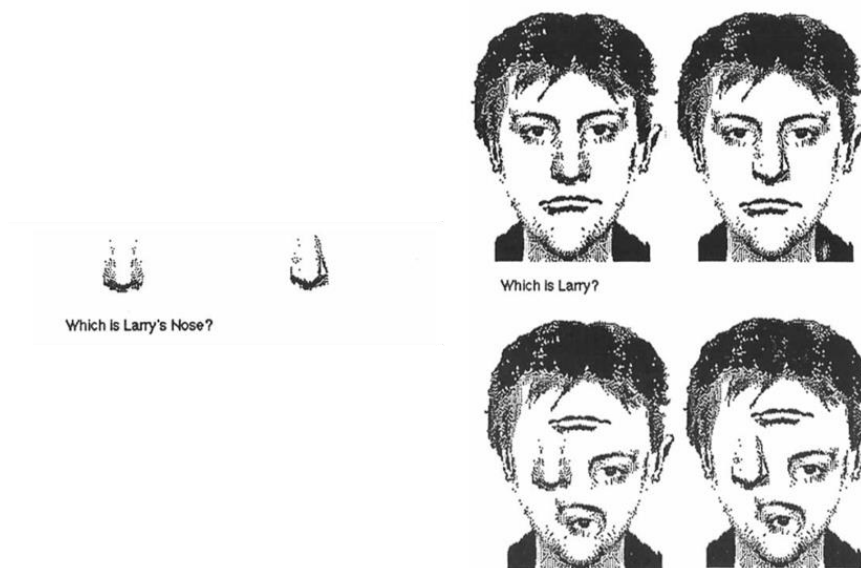


Figure 2. Exemple d'éléments (haut), visages entiers (milieu), et visages mélangés (bas) présenté par Tanaka et Farah (1993).

Un élément (le nez par exemple) est plus facilement reconnu lorsqu'il est présenté dans le visage entier que lorsqu'il est présenté de manière isolé ou dans un visage mélangé. Tiré de Tanaka et Farah (1993).

I.5. L'effet d'expertise

Pourquoi alors cette spécificité pour les visages humains ? Ont-ils une place particulière, en quelque sorte inscrite dans notre code génétique ? Ou bien cette spécificité résulte-t-elle tout simplement de notre exposition accrue aux visages humains comparativement à d'autres stimuli ?

Pour vérifier cette dernière hypothèse, Diamond et Carey (1986) ont cherché à savoir si nous pouvions recourir à des traitements spécifiques aux visages pour des stimuli non-visages auxquels nous serions très fortement exposés. Elles ont donc testé l'effet d'inversion avec des silhouettes de chiens, chez des experts canins et chez des sujets non-experts. Chez les novices, bien qu'un effet délétère de l'inversion sur le traitement des silhouettes de chiens soit trouvé, il est plus faible que celui mis en évidence pour les visages humains. Chez les experts canins, les résultats montrent un effet d'inversion pour les visages de chiens tout comme pour les visages d'humains. L'exposition répétée à un certain type de stimuli conduit donc à traiter celui-ci de manière différente.

Pour que l'effet d'inversion émerge, il faudrait en fait 3 paramètres. Les stimuli à traiter doivent avoir des configurations globales similaires, il doit être possible de les traiter sur la base des relations configurales, et le sujet doit avoir une expertise pour ces stimuli. Il faut cependant grandement nuancer l'interprétation de ces résultats. Dans l'étude originale de Carey and Diamond, bien que les experts canins soient plus sensibles à l'effet d'inversion pour les silhouettes de chien que les novices, les experts ne reconnaissent pas mieux que les novices les silhouettes de chiens lorsqu'elles sont présentées à l'endroit. Le fait que les participants de la catégorie « expert » n'aient pas de meilleures capacités de reconnaissance que les participants de la catégorie « novice » remet en doute le fait que l'expertise soit ici responsable de l'effet d'inversion. Cependant, si l'interprétation de Diamond & Carey est juste, cela voudrait donc dire que d'autres stimuli que les visages humains et les silhouettes de chiens peuvent générer cet effet d'inversion.

Les Greebles sont des petites figurines similaires à des visages sous plusieurs aspects, ayant été conçues pour tester si les processus dits « spécifiques » aux visages peuvent s'appliquer à d'autres stimuli (Gauthier & Tarr, 1997). Les Greebles ont en effet une configuration commune mais se distinguent par l'agencement et la forme de leurs différents traits. Selon leurs caractéristiques, les Greebles sont regroupées en deux genres différents ainsi que 5 différentes familles. Des sujets ont été entraînés pour devenir expert à la reconnaissance de ces Greebles.

Une fois ce niveau d'expertise atteint, les chercheurs ont trouvé que les sujets étaient aussi sensibles à l'effet d'inversion sur des Greebles que sur des visages humains (Ashworth III et al., 2008). Leur conclusion fut donc que l'effet d'inversion n'est pas spécifique aux visages mais spécifique à des propriétés du stimulus.

Cependant, là aussi l'interprétation est sujette à débat. Les Greebles sont justement des stimuli créés pour partager des propriétés avec les visages humains. Si l'on dispose d'un système de traitement spécifique pour les visages capable de traiter les caractéristiques du visage humain, il apparaît donc logique que ce système soit recruté et capable de traiter des stimuli créés pour être semblables aux visages humains. De plus, l'effet d'expertise pour les Greebles survient seulement après quelques heures d'entraînements. L'effet d'expertise chez les experts canins requiert une dizaine d'années d'expérience, ce qui correspond aussi au temps qu'il faut pour qu'un enfant traite les visages de manière similaire à un adulte (Carey & Diamond, 1994). Si l'effet d'expertise pour les Greebles apparaît aussi rapidement, n'est-ce pas justement parce qu'un effet d'expertise est déjà présent de manière sous-jacente ? D'autres études cherchant à mettre en évidence un effet d'expertise pour des stimuli non-visages viennent de plus nuancer ces résultats. On pourra notamment citer Yue et al. (2006), ayant entraîné des sujets à devenir expert dans la reconnaissance de « blob », dans des laps de temps similaires à ceux de l'étude de Gauthier & Tarr (1997) sur les Greebles. A la fois sur le plan comportemental et neuronal, ceux-ci ne trouvent pas de résultats similaires pour les visages et pour les blobs. Une autre étude, de Robbins & McKone (2007), tenta de répliquer les observations de Carey & Diamond sans toutefois trouver des résultats allant dans le sens d'un traitement configural similaire à celui des visages pour les experts canins identifiant des silhouettes de chien. Ces résultats viendraient confirmer l'idée que le traitement des Greebles, semblable au traitement configural des visages, tire justement profit de l'expertise dont nous disposons déjà pour les visages. De nos jours, il est de plus en plus admis que l'expertise que l'on possède pour traiter les visages peut être utilisée pour distinguer d'autres stimuli très similaires entre eux. Un mécanisme de recyclage neuronal (Dehaene & Cohen, 2007) permettrait d'utiliser des régions du cerveau spécialiste dans un domaine à contribution d'une nouvelle tâche. Par exemple, le langage écrit est trop récent dans l'histoire de l'humanité pour que des régions du cerveau soient spécialisées dans son traitement. Pinel et al., 2015, montrent ainsi qu'une région du cerveau généralement considéré comme spécifique de la reconnaissance des visages s'active au moins en partie lors du traitement du langage écrit. De plus, il existe une différence essentielle entre l'expertise pour les visages et l'expertise acquise pour d'autres stimuli (les Greebles ou les chiens par exemple).

La première se développe de manière spontanée chez le nourrisson par simple exposition et sans entraînement particulier. Si l'exposition à une catégorie de stimulus était un critère suffisant pour acquérir une expertise, alors nous devrions être expert pour traiter beaucoup d'autres stimuli. Cela ne semble pas être le cas.

II. L'apport de la neurophysiologie

II.1. La Fusiform Face Area

Il fallut attendre les progrès de la neuro-imagerie pour préciser les mécanismes et les régions cérébrales impliquées lors du traitement des visages. Une étude pionnière en neuro-imagerie concernant le traitement des visages chez l'humain est celle de Haxby et al. (1991). En utilisant la tomographie par émission de positons (TEP ou PET scan en anglais), ces chercheurs ont tenté d'identifier les régions actives lors de la présentation de certains stimuli, et notamment des visages. Les résultats montrèrent l'implication de régions différentes pour les visages et les stimuli non-visages (ici des points) et notamment les régions occipito-temporale et occipito-latérale. L'année suivante, Sergent et al., (1992), décrivent sans la nommer une zone impliquée dans le traitement des visages que l'on connaît désormais sous le nom de « Fusiform Face Area » (FFA). Il faudra toutefois attendre l'essor de l'imagerie par résonance magnétique fonctionnelle (IRMf) pour obtenir des données de neuro-imagerie plus précises. Puce et al. (1996) ont ainsi présenté soit des visages, soit des lettres, soit des motifs de textures à 12 participants, dans le cadre d'un protocole en IRMf. L'activation cérébrale observée différait selon la nature des stimuli, les visages induisant notamment une plus forte implication des gyri fusiformes droit et gauche, du sillon occipital temporal droit et d'une région du cortex latéral centrée dans le gyrus temporal moyen. Le profil d'activation ne peut cependant que difficilement être décrit comme spécifique aux visages, compte tenu de la non-homogénéité des stimuli utilisés. Kanwisher et al. (1996, 1997) ont donc exploré la réponse de participants regardant des visages ou des objets, stimuli plus équivalents en termes d'agencement que les lettres ou les textures. Dans cette étude, les zones cérébrales activées étaient assez similaires pour la plupart des participants et montraient une implication supérieure de la zone qu'ils nommèrent FFA, pour signifier qu'elle est activée en réponse à des visages. De plus, les auteurs ont également constaté une activation du sillon temporal supérieur (STS) chez certains participants. L'activation de la FFA et du STS trouvée par Kanwisher et al. était relativement topographiquement conformes aux régions actives lors du traitement des visages chez les macaques (Gross et al., 1972 ; Perret et al., 1982 ; Desimone, 1991 ; activations confirmés par des recherches plus récentes : Tsao et al., 2008). Chez les macaques, les neurones de la région

homologue à la FFA semblent plutôt réagir à l'identité (Freiwald et al., 2009) tandis que les neurones de la région STS sont plutôt activés en réponse aux expressions faciales, à la direction du regard et à l'orientation de la tête (Hasselmo et al., 1989 ; Perrett et al., 1991) bien que les deux régions soient en interactions. Les études IRMf des années suivantes chez des participants humains ont confirmé l'implication de trois régions cérébrales dans la perception des visages (voir Haxby et al., 2000, pour un modèle neuro-anatomique intégrant les trois régions) : une zone bilatérale du gyrus fusiforme, plus fortement active dans l'hémisphère droit (la FFA) (Halgren et al 1999; Haxby et al, 1999; Hoffman & Haxby 2000; Ishai et al 1999), la partie postérieure du STS, ainsi qu'une région du gyrus occipital inférieure (« occipital face area » ou OFA en anglais) (Gauthier et al., 2000b ; pour une revue, voir Pitcher et al., 2011). On peut cependant de nouveau critiquer la pertinence des stimuli de l'étude de Kanwisher et al. L'implication de la FFA lors du traitement des visages n'indique pas une spécificité de cette zone, d'autant plus que celle-ci s'active aussi (bien que moins intensément) lors de la présentation d'objets. La FFA pourrait s'activer en fonction de la complexité des stimuli ; les visages étant des stimuli particulièrement complexes, il semble naturel qu'ils recrutent plus fortement cette zone. Pour examiner cette possibilité, Kanwisher et al. (1998) ont donc testé des participants avec des stimuli de complexité strictement similaire à celle des visages, les visages inversés. Les résultats mettent en évidence que l'inversion des visages réduit drastiquement l'implication de la FFA chez les participants. Ainsi, il est intéressant de noter que les visages inversés n'activent pas cette zone cérébrale et ne sont pas, d'un point de vue comportemental, traités comme des visages. Une autre possibilité est que le niveau d'expertise pour un stimulus soit corrélé avec l'activation de la FFA. Nous avons précédemment discuté des Greebles, stimuli qui pourraient générer un effet d'expertise après entraînement. Les études IRMf présentent des résultats quelques peu divergents. Gauthier et al. (1999) obtiennent une activation de la FFA après entraînement pour les Greebles et un effet d'inversion, c'est-à-dire moins d'activation de la FFA pour des Greebles inversés. Ils trouvent des résultats similaires pour des participants experts en reconnaissance d'oiseaux ou de voitures (Gauthier et al., 2000a). Cependant, ces résultats sont discutables. L'effet d'expertise serait synonyme d'un traitement configural (de second ordre ou holistique). Les résultats de Gauthier nous amènent à penser que si la FFA est active pour tous types de stimuli associés à une forte expertise, ce n'est donc pas qu'elle répond spécifiquement aux visages mais plutôt qu'elle intervient lors d'un traitement configural. Or, Yovel & Kanwisher (2004) ont examiné l'activation de la FFA en fonction de changements configuraux ou componentiels pour des visages et des maisons. L'activation de la FFA n'était pas plus importante lors de la tâche configurale, et était plus

faible pour le traitement des maisons, que la tâche soit configurale ou componentielle, montrant ainsi que la FFA n'est pas spécifique d'un type de traitement mais bien d'un type de stimulus, en l'occurrence les visages. De plus, Brants et al. (2011) ont tenté de répliquer les résultats de l'étude de Gauthier et al. (1999) avec plus de participants, cependant sans succès. La réaction de novices à la présentation de Greebles montre, avant tout entraînement, des ressemblances avec la réaction suite à la présentation de visages. L'activation de la FFA est déjà présente chez les novices, et ne l'est pas plus chez les experts, sous-tendant que ceux-ci sont traités de facto de manière similaire à des visages.

II.2. La réponse neuronale

Si l'IRMf permet de distinguer des zones du cerveau répondant spécifiquement aux visages, l'électroencéphalographie (EEG) permet de distinguer des réponses neuronales spécifiques et nous renseigne de plus sur la temporalité de celles-ci. Bentin et al. (1996) ont mis en évidence le potentiel évoqué (PE) N170, pic négatif survenant approximativement 170 ms après l'apparition d'un visage, mais pas d'un objet. Ce PE a été détecté dans la région postérieure latérale (électrodes T5 et T6) et de manière plus intense à droite, correspondant relativement bien avec la localisation de la FFA. Cette réponse a été également confirmée par la magnétoencéphalographie (MEG) avec le PE M170 (Liu et al., 2000). Une réaction rapide et spécifique survient donc en réaction aux visages et nous permet de distinguer un visage d'autres types de stimuli en moins de 200 ms. L'identification d'un visage familier induirait une réponse EEG un peu plus tardive, de l'ordre de 250 ms. La présentation récurrente d'un visage non connu n'engendre initialement pas de N250 mais finit par le faire après quelques présentations (Tanaka et al., 2006).

II.3. Les patients cérébro-lésés

Il est toujours tentant mais cependant souvent inexact d'établir des liens de causalité entre les observations cérébrales (neuro-imagerie et PE) et les manifestations comportementales. Un moyen plus direct pour connaître le rôle des différentes régions du cerveau est d'examiner des personnes chez qui ces zones sont déficientes, endommagées voire absentes ou bien des personnes chez qui le traitement des visages est déficient (l'idéal pour confirmer le rôle d'une ou plusieurs régions étant que ces deux ensembles de personnes se recoupent). Pendant longtemps, les seules données disponibles sur les lésions cérébrales étaient obtenues *post-mortem*. Ce type de données a permis, par exemple, de mettre en lumière le lien entre une lésion du gyrus fusiforme et la prosopagnosie (Damasio et al., 1982). Les personnes atteintes de

prosopagnosie sont incapables de reconnaître un visage, même familial, et ce malgré une vision tout à fait normale. La technique d'IRM a cependant permis l'avantage de ne pas attendre le décès d'un patient pour étudier sa neuroanatomie, et donc d'examiner plus précisément les hypothèses sur l'implication d'une ou plusieurs régions via des tests comportementaux ou l'IRMf. Les patients prosopagnosiques présentent pour la plupart des lésions des régions occipito-temporales ventrales et plus précisément le gyrus lingual, fusiforme et le cortex parahippocampique (De Renzi et al., 1994, pour une revue récente, voir Rossion & Retter, 2020). Wada & Yamamoto (2001) ont testé les performances d'un homme de 67 ans présentant une prosopagnosie suite à une hémorragie cérébrale. Alors qu'il ne présentait aucun déficit cognitif ou visuel et aucune difficulté à reconnaître des objets ou des lieux publics, il lui était impossible de reconnaître des visages familiers y compris le sien. Un scan IRMf mit en évidence des lésions cérébrales au niveau du gyrus fusiforme droit et dans la région occipitale latérale, c'est-à-dire les régions habituellement actives lors de la présentation de visages.

Toutes ces observations, tant comportementales que neuroanatomiques, ont permis aux chercheurs d'établir des modèles explicatifs du traitement des visages. Avant de les présenter, nous allons dans un premier temps établir un bref rappel des mécanismes physiologiques et cérébraux de la vision, puisque ces modèles découlent de nos connaissances sur la vision humaine. De plus, si nous souhaitons étudier la reconnaissance des visages chez le nourrisson, il faut tenir compte de la vue imparfaite du nouveau-né. Celle-ci, bien que difficile à décrire, n'est en effet pas équivalente à celle de l'adulte. Nous verrons ensuite comment la reconnaissance des visages peut être modélisée chez l'adulte. Enfin, nous examinerons quelles sont les capacités initiales de traitement des visages chez le nouveau-né et comment ces compétences se développent par la suite.

III. Le système visuel et son développement

III.1. Le système visuel

L'œil est un organe ayant pour principal fonction de transformer une information lumineuse en information nerveuse, permettant ainsi la vision. Ceci est rendu possible par les différentes structures dont est composé l'œil, dont nous ne détaillerons que très succinctement les rôles ici. L'information lumineuse se présente sous forme de photons, renvoyés par les objets qui nous entourent. Les photons traversent tout d'abord la cornée, qui permet une première focalisation des rayons lumineux. L'iris s'ouvre et se referme en fonction de la quantité lumineuse reçue. Le cristallin se contracte ou se relâche et forme ainsi l'image sur la rétine, composé entre autres

des photorécepteurs. Les photorécepteurs captent la lumière et la transforment en information électrique et la transmettent au cerveau via le nerf optique. Ces photorécepteurs sont de deux formes : les bâtonnets, plus présents en périphérie de la rétine, et les cônes, plus présents en son centre. La sensibilité des bâtonnets à la lumière est plus forte que celles des cônes ; les premiers, plus nombreux (environ 100 millions de cellules) participent plus fortement à la vision nocturne (scotopique) tandis que les seconds (environ sept millions de cellules) participent à la vision diurne (photopique). C'est le système photopique qui permet la distinction des couleurs, les cônes étant divisés en trois sortes, selon la longueur d'onde qu'ils captent (et donc la couleur qui sera perçue) : les cônes bleu (S), vert (M) et rouge (L). Les photorécepteurs transmettent ensuite le signal via les neurones bipolaires aux cellules ganglionnaires. Une cellule ganglionnaire peut ainsi être reliée à plusieurs photorécepteurs, constituant ce que l'on appelle son champ récepteur. Plus les champs récepteurs sont petits, plus la discrimination des détails sera importante ; à l'inverse, plus il y a de photorécepteurs communiquant avec une seule et même cellule ganglionnaire, plus celle-ci résumera l'information provenant d'une grande zone et moins bonne sera la résolution. De manière générale, les cellules ganglionnaires sont reliées à plus de bâtonnets que de cônes. Nous avons précédemment mentionné que les bâtonnets sont plus présents en périphérie de la rétine tandis que les cônes sont présents dans la région centrale aussi appelée fovéale. Les conséquences de cette répartition sont une acuité maximale au niveau de la fovéa, ainsi qu'une moins bonne discrimination des couleurs à mesure que l'on s'éloigne de son centre (puisque la périphérie rétinale est plus concentrée en bâtonnets, dénués de sensibilité aux couleurs). On distingue deux types de cellules ganglionnaires, les magnocellulaires (cellules M) et les parvocellulaires (cellules P). Les cellules M ont un champ récepteur plus grand, c'est-à-dire qu'elles relayent l'information venant de plus de photorécepteurs. Les axones des cellules ganglionnaires forment le nerf optique qui assure la fonction de transmettre l'information nerveuse jusqu'au cerveau. Dans le cerveau se déroule alors une seconde phase, l'analyse du signal nerveux. Le nerf optique va se terminer au niveau du corps géniculé latéral (CGL) droit ou gauche qui d'après nos connaissances actuelles, sert principalement de relai entre la rétine et le cortex visuel primaire. Il est cependant important de noter que le CGL reçoit plus *d'input* nerveux provenant du cerveau que provenant des yeux. Il est ainsi très probable que le CGL présente un rôle filtrant sur l'information qui va être transmise de l'œil jusqu'au cortex visuel. Le CGL droit reçoit l'information venant de l'hémichamp visuel gauche tandis que le CGL gauche reçoit celle de l'hémichamp visuel droit. Chaque œil perçoit une grande partie du même champ visuel (contribuant ainsi à la perception des distances), avec une contribution un peu plus importante de l'œil gauche pour l'hémichamp

gauche et de l'œil droit pour l'hémichamp droit. Le nerf optique de chaque œil se divise donc pour rejoindre les deux CGL, les cellules recevant de l'information du champ visuel gauche projetant dans le CGL droit et les cellules recevant de l'information du champ visuel droit projetant dans le CGL gauche. Le CGL est organisé en six couches : les deux couches inférieures reçoivent l'information des cellules M tandis que les quatre couches supérieures reçoivent celles de la voie P formant ainsi les deux voies du traitement visuel (Lee, 1996). La voie magnocellulaire répond aux hautes fréquences (véhiculant plus les informations relatives aux détails, aux contours...) tandis que la voie P répond aux basses fréquences (véhiculant plus les informations globales). Ces réponses spécifiques à différentes fréquences ont été mises en évidence via des lésions de ces régions chez des singes (Merigan et al., 1991 ; Merigan et Maunsell, 1990). Ensuite, l'information va être transmise du CGL au cortex visuel primaire V1, situé dans la partie postérieure du lobe occipital. L'aire V1, aussi constituée de six couches (définies selon la taille et le type des neurones ainsi que leur densité), reçoit sur sa couche 4C beta les projections de la voie P, sur la couche 4C alpha celles de la voie M. Ces couches vont alors projeter sur la couche 4B et 2/3 de cette même aire avant que l'information ne soit relayée au cortex visuel secondaire V2 puis d'autres aires visuelles. Une caractéristique importante de V1 et V2 et que l'on retrouve dans d'autres aires visuelles de bas niveau est que les relations spatiales que les cellules ganglionnaires entretiennent entre elles au sein de la rétine sont conservées dans le cortex. Ce particularisme, appelé rétinotopie, a comme souvent d'abord été mis en évidence chez le macaque (Tootell et al., 1982) et confirmé en IRMf chez l'humain (Engel et al., 1997). Un autre élément important concernant V1 est son organisation en colonnes neuronales. Chaque colonne neuronale répond à un type de stimulation précis, présenté sur un champ récepteur précis. Ces colonnes peuvent répondre à des angles particuliers, des couleurs ou une certaine profondeur tandis que d'autres s'activent lorsqu'un stimulus est présenté pour un œil mais pas pour l'autre. Si V1 et V2 sont les premières aires par lequel le signal visuel transite, on a identifié près d'une trentaine d'aires corticales impliquées dans le traitement du signal visuel. De cette organisation ressort deux voies de traitement différentes, œuvrant en parallèle : la voie dorsale dans le lobe pariétal et la voie ventrale le long du lobe temporal. La voie dorsale, longtemps surnommée la voie du « où », aurait pour principale fonction d'identifier les informations relatives à la localisation, i.e. à la position des objets perçus. La voie ventrale, la voie du « quoi », servirait principalement à l'identification de l'objet et ses propriétés intrinsèques comme la couleur ou la taille (Livingstone et Hubel, 1988 ; DiCarlo et al., 2012). Nous développerons par la suite comment les chercheurs ont tenté de modéliser le traitement de l'information au sein de ces différentes aires visuelles.

III.2. Développement des capacités, de *in utero* jusqu'à la naissance

L'œil se développe *in utero* dès la 3^{ème} semaine avec la constitution du globe oculaire. On observe ensuite une différenciation de la rétine à la 4^{ème} semaine et la formation du nerf optique durant les deux premiers mois de gestation. La zone fovéale se met en place dès la 24^{ème} semaine, bien qu'elle continue de se développer jusqu'au 15^{ème} mois de vie du nourrisson (Kovarski, 2014). Les différents types de récepteurs rétiniens sont présents à partir de 26-28 semaines mais leur répartition n'est pas semblable à celle de l'adulte, la périphérie étant développée plus tôt que la zone fovéale (Hendrickson, 1992). Les fonctions visuelles ne sont donc pas matures dès la naissance, le développement de l'œil continuant pendant des années encore. De plus, si le fœtus peut déjà avoir une certaine expérience auditive, proprioceptive ou bien gustative *in utero* permettant l'observation, dès la naissance, de compétences découlant d'un apprentissage, son expérience visuelle est quasi inexistante. Toutefois, les études sur le modèle animal montrent que la lumière est transmise *in utero* et que cette transmission augmente au cours de la gestation avec l'étirement des tissus abdominaux (Bearden, Wilson et al., 2001 ; Fulford et al., 2003 ; Del Giudice, 2011). La luminosité perçue avoisine les 50 lux chez la brebis (Parraguez, Sales et al., 1998) et la transmission serait meilleure pour les longueurs d'onde proche de 650 nm, soit les rouges (Jacques et al., 1987). Des modélisations estiment qu'entre 0.1 et 1% de la luminosité extérieure parvient au fœtus (Del Giudice, 2011). Cependant, le système visuel encore immature du fœtus est le plus sensible pour les fréquences correspondant aux rouges (les mieux transmises) et leur acuité est meilleure pour des niveaux de luminosité avoisinant les 50 lux (Adams et al., 1994 ; Banks & Shannon, 1993 ; Glass, 2002), correspondant donc à l'*input* visuel présent *in utero*. Des stimulations lumineuses grâce à une lumière introduite lors d'amnioscopies induisent une accélération du rythme cardiaque des fœtus (entre 38 et 43 semaines de gestation) indiquant une réaction du fœtus à celles-ci (Peleg & Goldman, 1980). Des données en MEG fœtale mettent en évidence des réponses évoquées à des flashes visuels dès la 28^{ème} semaine de gestation (Eswaran et al., 2004 ; McCubbin et al., 2007). *In utero*, le nourrisson a donc principalement une (faible) expérience aux changements de luminosité mais en aucun cas à la discrimination d'objets. Cependant, une préférence pour des sources lumineuses reprenant l'organisation schématique des visages est déjà présente chez le fœtus (Reid et al., 2017). En effet, Reid et al. ont créé des stimuli représentant très schématiquement la configuration d'un visage en configuration canonique ou en configuration inverse, en disposant de manière triangulaire des diodes lumineuses (voir Figure 3). Les résultats montrent que les fœtus tournent significativement plus la tête vers le stimulus

représentant la configuration canonique des visages que vers le stimulus inverse. Ce résultat n'est pour l'instant pas répliqué mais suggère qu'il existe une représentation (très schématique) des visages avant la naissance.

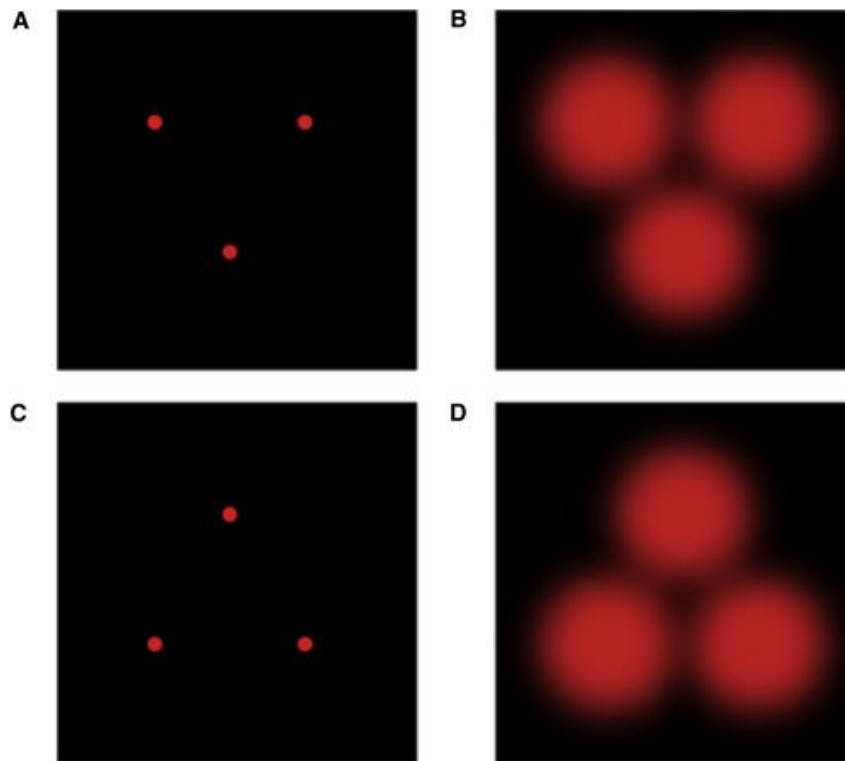


Figure 3. Stimuli utilisés par Reid et al. (2017).

Les stimuli A et B représente la configuration canonique des visages tandis que les stimuli C et D représente la configuration inversée. Les stimuli A et C sont une représentation fidèle des stimuli. Les stimuli B et D sont une estimation de ce à quoi ressemblent les stimuli une fois en contact avec le ventre de la mère, en prenant compte de la diffusion lumineuse. Tiré de Reid et al. (2017).

A la naissance, bien qu'il soit difficile de savoir précisément ce que voient les nourrissons, il est communément admis que leur acuité, leur perception des couleurs ainsi que leur champ visuel sont peu développés. Les pupilles ne se dilatent pas encore complètement, le cristallin est plus sphérique qu'ovale et le globe oculaire n'a pas sa taille définitive. Ces différences physiques ont des conséquences fonctionnelles qui vont progressivement s'améliorer au cours du développement (Odom & Green, 1984). Le globe oculaire va notamment croître, d'abord rapidement pour atteindre une taille adulte vers 2 ans mais son développement continue, plus lentement, pendant une vingtaine d'années encore. Le nerf optique va se myéliniser principalement durant les deux premières années (Friede & Hu, 1967 ; Magoon & Robb, 1981). La capacité d'accommodation du cristallin va se renforcer durant les premiers mois améliorant

ainsi la convergence des rayons lumineux sur la rétine (Green et al., 1980 ; Lotmar, 1976). Ces développements physiologiques vont améliorer la qualité optique, mais l'évolution de la densité des photorécepteurs contribue aussi en grande partie à améliorer la vue du nourrisson. En terme quantitatif, les photorécepteurs sont déjà présents. En revanche, les cônes, essentiels pour la vision diurne, présentent dans la partie fovéale une densité deux fois moins importante que chez l'adulte. Initialement, le centre du champ visuel est donc plus en retard par rapport à l'état adulte que la périphérie (Lewis & Maurer, 1992). Les nouveau-nés sont donc plus habiles pour détecter un stimulus en périphérie de leur champ visuel, bien que celui-ci soit plus petit que celui de l'adulte (40° environ contre 80° chez l'adulte). De plus, leur vision des détails et des couleurs est réduite, puisque c'est le centre de la rétine et en particulier la fovéa qui en est responsable (ainsi que cela a été montré chez l'adulte). Des changements se font aussi au niveau des structures cérébrales impliquées dans la vision. L'organisation en six couches des CGL est présente dès la naissance mais leurs tailles adultes ne sont atteintes qu'après 6 mois et les cellules restent immatures jusqu'à 9 mois (Garey & De Courten, 1983). Les aires visuelles connaissent une maturation qui dure jusqu'à l'adolescence (Klaver et al., 2011).

III.3. Développement de l'acuité et de la capacité d'accommodation

Au cours du développement, la réduction de la fovéa et la migration des cônes entraîne une amélioration de l'acuité visuelle et une meilleure perception des couleurs. L'acuité correspond à la plus petite distance perçue entre deux points. Pour mesurer l'acuité des nourrissons, une méthode est de leur présenter des réseaux de différentes fréquences spatiales soit des successions de barres alternativement claires et sombres (Courage & Adams 1990). Lorsque les nourrissons ne sont plus capables de distinguer la distance entre les bandes claires et les bandes sombres, le réseau apparaît comme une masse homogène grise. Les nourrissons ont naturellement tendance à orienter leur regard vers le stimulus visuel qu'ils trouvent le plus intéressant, en l'occurrence un stimulus hétérogène plutôt qu'un stimulus homogène. Par une technique de préférence visuelle, on présente une cible constituée d'une alternance de bande noires et blanches à côté d'une cible homogène grise. Si les nourrissons sont sensibles à la fréquence spatiale de la cible, alors ils la percevront comme un stimulus hétérogène et leurs temps de fixation devraient être supérieurs à ceux observés pour une cible homogène. Si les nourrissons ne sont pas sensibles à cette fréquence spatiale, ils percevront un stimulus homogène et leurs temps de regard ne devraient pas varier significativement d'un côté ou de l'autre. L'acuité correspond à la plus grande fréquence spatiale que le sujet aura détectée (plus petite alternance de barres noires et blanches). On trouve que l'acuité visuelle des nourrissons

est réduite par rapport à celle des adultes, étant environ 30 fois inférieure à la naissance, 10 fois inférieure à 3 mois, mais seulement 2.5 fois moins précise à 12 mois et atteignant le niveau de l'adulte vers 3-5 ans (Banks & Salapatek, 1978 ; Birch et al., 1983). On peut aussi présenter le même type de cible, mais en faisant varier la luminance des bandes noires et blanches pour évaluer la sensibilité au contraste en fonction de la fréquence spatiale. Un niveau adulte est observable vers 3-4 ans pour des stimuli de haute fréquence tandis qu'il faut attendre 9 ans pour les basses fréquences (Adams & Courage, 2002). La capacité d'accommodation aussi est différente. Celle-ci correspond à la capacité de faire la mise au point sur des objets situés à différentes distances. Cette capacité est limitée chez le nouveau-né. Les nourrissons accommodent mal jusqu'à l'âge de 2 mois. Cependant, bien que réduite, il semble que la capacité d'accommoder du nouveau-né est plutôt adaptée à son environnement. En effet, la distance de meilleure visibilité se situe aux alentours de 20-50cm, ce qui correspond à la distance qui le sépare du visage de ses parents lorsqu'il est dans leurs bras ou en interaction avec eux. A 6 mois, le pouvoir accommodateur du bébé est identique à celui de l'adulte. Pour le champ visuel, celui-ci atteint 90% de celui des adultes vers 6-7 mois (Mayer et al., 1988).

III.4. Développement de la saccade oculaire

Discerner un objet de la meilleure manière possible se fait en plaçant la réflexion de celui-ci au sein de la fovéa, la zone d'acuité maximale. Cette orientation du regard vers une cible se fait au moyen des saccades oculaires, des mouvements très rapides (entre 20 et 50 ms). Dès le premier mois de vie, le nourrisson est capable d'effectuer des saccades vers un stimulus apparaissant soudainement en périphérie de son regard (Aslin & Salapatek, 1975). Cependant, le temps de réaction avant de déclencher une saccade est plus long que chez l'adulte, ce temps de réaction saccadique diminuant jusqu'à l'âge de 14-15 ans, en lien avec la myélinisation des structures impliquées dans le contrôle oculomoteur (Fischer et al., 1997 ; Fukushima et al., 2000 ; Irving et al., 2006 ; Klein & Foerster, 2001 ; Munoz et al., 1998). Une caractéristique des saccades est qu'elles sont souvent légèrement hypométriques, c'est-à-dire que le regard s'arrête un peu avant la cible. Cette hypométrie est encore plus marquée chez les nourrissons, ce qui les amène généralement à faire plusieurs saccades successives avant de fixer correctement la cible (Aslin & Salapatek, 1975 ; Salapatek et al., 1980). C'est aux alentours de 8 ans que les amplitudes des saccades atteignent des valeurs comparables à celles des adultes (Cohen & Ross, 1978 ; Salman et al., 2006).

III.5. Développement de la perception des couleurs

Concernant la perception des couleurs, nous avons vu que celle-ci dépend de la densité des photorécepteurs et est donc réduite au cours du premier mois de la vie postnatale. Adams & Courage (1998) ont montré que les nouveau-nés font la distinction chromatique-achromatique quand les niveaux de pureté colorimétrique (soit la proportion de lumière blanche qui est mélangée avec la lumière d'une certaine longueur d'onde, aussi appelée saturation) sont élevés, c'est-à-dire quand la couleur est plus intense. Les nouveau-nés distinguent notamment le vert à 545 nm pour un niveau d'au moins 41 % de pureté chromatique, le jaune à 585 nm pour une pureté de 65% et le rouge à 650 pour une pureté de 47% tandis que les adultes font la distinction, pour ces mêmes longueurs d'ondes, à moins de 1% de pureté chromatique. Les nouveau-nés sont donc très tôt capables de discriminer les couleurs si celles-ci sont très vives. La sensibilité augmente rapidement dans les premiers mois de vie et plus progressivement jusqu'à l'adolescence (Knoblauch et al., 2001).

III.6. Développement de la sensibilité aux mouvements

La sensibilité aux mouvements se fait précocement via des réseaux neuronaux sous-corticaux localisés dans l'aire visuelle V1. L'orientation du mouvement serait détectable à la naissance tandis que la détection de l'orientation se ferait vers 2 mois (Braddick et al., 2005). Toutefois, ce n'est encore une fois pas avant le début de l'adolescence que les capacités de détection des mouvements atteignent leur niveau « adulte » (Braddick et al., 2003). Parmi les différents types de mouvements, on remarque que notre vision est particulièrement sensible aux mouvements « biologiques » c'est-à-dire produits par un organisme vivant. Cette sensibilité se retrouve dès la naissance. Lorsqu'on présente simultanément, via une tâche de préférence visuelle, un mouvement biologique et un mouvement non biologique (déplacements aléatoires de points lumineux) à des nouveau-nés âgés de 2 jours, ils regardent plus longtemps le mouvement biologique, et ce même si celui-ci ne représente pas un humain (Bardi et al., 2011 ; Simion et al., 2008). Cette capacité a d'abord été mise en évidence chez les poussins et pourrait être partagée par de nombreuses espèces animales (Vallortigara & Rogers, 2005 ; Vallortigara & Regolin, 2006). L'origine de cette capacité commune peut être discutée, mais percevoir le mouvement d'autres animaux peut être utile chez une proie pour identifier plus facilement les prédateurs, chez un prédateur pour détecter les proies, ou simplement pour distinguer les congénères avec lesquelles on peut interagir. Cependant, dès que la structure du motif de points ne représente plus un être vivant, la préférence pour le mouvement biologique des nouveau-nés disparaît (Bardi et al., 2011). Au contraire, lorsqu'on présente un unique point réalisant soit un

mouvement biologique, soit un mouvement non biologique à des nouveau-nés (déplacement aléatoire de points lumineux), ils regardent plus longtemps le mouvement non-biologique lors des premières secondes du test. Si cela peut paraître contre-intuitif, ce temps de regard pourrait traduire un intérêt pour ce mouvement qui transgresse les attentes du nourrisson qui n'a pas encore d'expérience des mouvements non biologiques (Méary et al., 2007). Les nourrissons présentent en effet un intérêt pour les événements impossibles lors desquels leur compréhension du monde est bousculée (Hespos & Baillargeon, 2001). A partir de 3 mois, on observe cette fois-ci une préférence pour les motifs de points représentant des mouvements biologiques (Bertenthal et al., 1984 ; 1987), non observée chez les nourrissons de 2 mois (Fox & Daniels, 1982). Il n'est pas clair si ceux-ci échouent par manque d'expérience ou de maturation du système visuel.

III.7. Développement de la perception de profondeur

En ce qui concerne la perception de la profondeur, la tâche de la falaise visuelle mise en place par Gibson et Walk (1960) semble la plus emblématique. Cette tâche consiste en une table en verre transparente sur laquelle les bébés peuvent se déplacer. Celle-ci est formée de deux parties distinctes : d'un côté un panneau quadrillé collé juste sous la vitre, de l'autre, le panneau se décolle et se trouve plusieurs centimètres sous la vitre donnant ainsi l'impression que la surface tombe brusquement. Le nourrisson est posé sur la table tandis qu'un parent se trouvant de l'autre côté de la table l'appelle. Si le nourrisson veut rejoindre sa mère, celui-ci devra franchir le « gouffre » visuel. Le postulat de Gibson & Walk est que si la perception de la profondeur est présente chez le nourrisson, celui-ci sera beaucoup plus réticent à traverser la table que s'il n'a pas conscience de la profondeur. Ces auteurs ont trouvé une réticence seulement de la part des enfants ayant commencé à marcher depuis peu (indépendamment de l'âge auquel ils commencent à marcher). La peur des hauteurs surviendrait donc avec les premières chutes et serait dépendante de l'expérience locomotrice. Cependant, l'absence de réaction comportementale n'implique pas nécessairement une absence de perception de la profondeur. En effet, des nourrissons de 3 mois présentent une accélération du rythme cardiaque et respiratoire ainsi qu'une dilatation des pupilles lorsqu'ils rampent sur le « gouffre » (voir Adolph et al., 2014 pour une revue). Par leur manque d'expérience locomotrice, le « vide » n'entraîne pas un comportement d'évitement. De plus, on sait que dès la naissance, le nourrisson est en partie sensible aux distances et aux profondeurs. Ce que l'on nomme constance perceptive est la capacité à être sensible aux caractéristiques objectives de l'environnement. La constance perceptive nous permet d'avoir une vision du monde cohérente

malgré les changements permanents (de luminosité, d'angle, de taille...) de l'image des objets sur la rétine. Par exemple, on sait que la taille d'un objet reste invariable, même si son image rétinienne est modifiée en fonction de la distance à laquelle on l'observe. Slater et al. (1990) ont étudié la constance de taille chez le nouveau-né âgé de quelques jours. Ils ont d'abord procédé par une phase de familiarisation durant laquelle soit un petit cube soit un gros cube étaient présentés à des distances différentes au fil des essais. Durant une phase de test, les deux cubes étaient montrés côte à côte mais à des distances différentes de sorte que leurs images rétiniennes étaient identiques (le petit cube étant donc présenté plus proche que le gros cube). Les résultats montrent une préférence pour la nouveauté, c'est-à-dire que les nourrissons regardent plus longtemps le cube pour lequel ils n'ont pas été familiarisés. Cela indique que les nouveau-nés différencient les tailles des cubes malgré les similarités de taille rétinienne et qu'ils sont sensibles au fait qu'un objet conserve sa taille malgré les changements de distance.

On voit donc que les capacités dépendant de critères physiologiques se développent assez rapidement, principalement durant la première année de vie. Cependant, d'autres capacités, dépendant cette fois-ci de la maturation du cortex et de l'expérience, nécessitent d'attendre l'adolescence pour atteindre un niveau de performance similaire à celui de l'adulte. De plus, nous avons pu voir que des capacités qui semblent strictement visuelle (dans notre exemple, éviter des précipices) peuvent dépendre de la maturation d'autres systèmes cognitifs (en l'occurrence celle du système locomoteur).

IV. Modèles du traitement des visages

IV.1. Le modèle de Bruce & Young

Les premiers systèmes tentant d'expliquer le fonctionnement de la vision étaient fortement imprégnés du modularisme de Fodor, pour lequel le fonctionnement cognitif est divisé en modules distincts (se traduisant par différentes régions cérébrales) traitant des informations spécifiques. Marr & Poggio (1979) a proposé un modèle computationnel hiérarchisé incluant trois étapes principales (ou trois modules). La première étape, celle du « Primal sketch » permet le traitement des contours et des surfaces d'une image. La 2ème étape, « 2.5D sketch », permet l'extraction d'informations relatives à la tridimensionnalité de l'objet mais dépendante du point de vue de l'observateur, comme l'orientation et la profondeur. Enfin, une 3ème étape, « 3D representation », consiste en l'extraction des invariants de l'objet et permet donc la reconnaissance de celui-ci indépendamment de sa position.

Avant même l'essor de la neuroimagerie, le caractère automatique (par sa vitesse) et spécifique du traitement des visages laissait fortement suggérer que des zones du cerveau puissent être spécialisées pour ces tâches. L'idée qu'un module cognitif puisse être dédié au traitement des visages est étayée, comme nous l'avons précédemment évoqué, par l'exemple de patients cérébro-lésés. Certains perdent la capacité à traiter les visages et uniquement ceux-ci, ou bien à l'inverse, des patients perdent la capacité à traiter certains stimuli visuels mais conservent la capacité à traiter les visages (Campbell et al., 1986 ; Young et al., 1993 ; Moscovitch et al., 1997). De plus, des patients cérébrólésés peuvent présenter des troubles d'identification des visages (prosopagnosie) sans que la reconnaissance des émotions soit affectée (Hécaen & Angelergues, 1962). Un module de traitement des visages pourrait donc à son tour être subdivisé en différents modules plus ou moins indépendants et jouant différents rôles. C'est donc naturellement que les premiers modèles de traitement des visages ayant émergé ont conceptualisé le traitement et la reconnaissance des visages comme une suite d'étapes hiérarchisées fonctionnant de manière strictement ascendante. Le modèle proposé par Bruce & Young en 1986 est l'un des plus repris et découle directement du modèle de Marr (voir Figure 4). Selon Bruce & Young, la reconnaissance des visages commence par une étape d'encodage dépendante du point de vue de l'observateur et correspondant donc au 2.5 sketch du modèle de Marr. Cette étape va permettre, via deux autres modules, l'extraction des expressions relatives à l'état émotionnel de la personne ainsi que des informations faciales relatives au décodage de la parole. Les caractéristiques indépendantes de l'orientation du visage sont quant à elles encodées dans une autre étape similaire à l'étape 3D de Marr. C'est cette étape qui permet l'accès à l'identité du visage grâce à l'extraction des invariants faciaux (les informations du visage qui ne dépendent ni de l'orientation, ni des émotions, ni des mouvements relatifs à la parole). Si le visage est familier, le degré de ressemblance entre les invariants extraits et les représentations stockées en mémoire sera fort et le sentiment de familiarité sera perçue. Ensuite, deux modules successifs permettent d'accéder aux informations sémantiques relatives à la personne (c'est-à-dire les informations stockées sur la personne, comme son métier, ses passions...) puis à son nom. Ces modules sont bien hiérarchisés dans le sens où c'est l'accès aux informations sémantiques relatives à la personne qui permet ensuite l'accès au nom, et jamais le contraire. Un dernier module, résultant de l'encodage des invariants ainsi que des caractéristiques dépendantes du point de vue permet l'accès aux informations sémantiques accessibles indépendamment de la familiarité du visage, comme son âge, son origine, son genre. Ces étapes sont en lien direct ou indirect avec le système cognitif, mais la nature de ces liens n'est pas clairement établie.

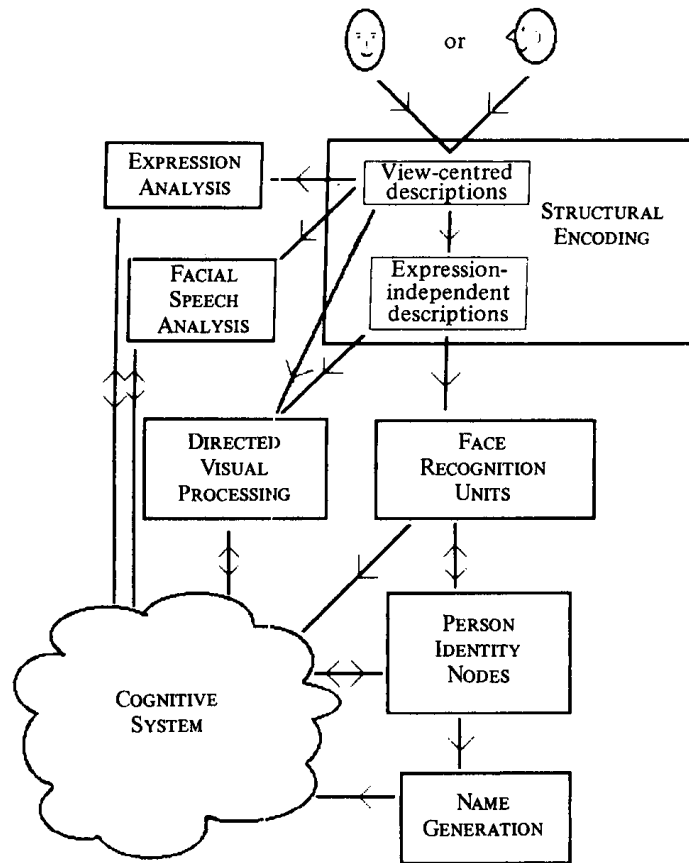


Figure 4. Modèle de reconnaissance des visages, tiré de Bruce & Young (1986).

IV.2. Les modèles plus récents

Les progrès techniques ont permis d'affiner, de confirmer ou d'infirmer les modèles théoriques de traitement de la vision et des visages présentés précédemment. Ainsi, les différentes aires visuelles (V1, V2...) peuvent être rattachées à une théorie modulaire et hiérarchisée comme celle de Marr. Selon des modèles strictement hiérarchiques (Felleman & Van Essen, 1991 ; pour une revue voir DiCarlo et al., 2012), chaque aire visuelle effectue des traitements de complexité supérieure à l'aire précédente (résultant des informations extraites de l'aire inférieure) pour arriver à la reconstruction complète de la scène visuelle. L'idée de hiérarchisation est notamment soutenue par les temps de latence des neurones, puisque chez le macaque, l'activation de chaque aire visuelle est retardée/ décalée de 10 ms par rapport à la précédente (Nowak & Bullier, 1997 ; Schmolesky et al., 1998). Cependant, d'autres modèles insistent sur l'importance d'un feedback et l'existence de mécanismes de traitement s'effectuant en parallèle, à la fois ascendants et descendants. (Bullier, 2001). Cette idée se base sur le fait qu'un modèle strictement feedforward ne permettrait pas d'identifier les éléments d'une scène visuelle aussi rapidement que nous en sommes capables, comme par exemple identifier des

objets ou des personnes dans une scène visuelle bruitée (Mumford, 1994). Des *feedbacks* provenant d'aires supérieures devraient donc aider les aires de bas niveaux à sélectionner les informations pertinentes pour fonctionner correctement.

Pour revenir au traitement des visages, les progrès de la neuro-imagerie permettent de localiser les différentes étapes de traitements au sein d'un système neuronal distribué (Haxby et al., 2000 ; Ishai et al., 2005 ; voir Figure 5). Une première perception des traits faciaux se fait au niveau des gyri occipitaux inférieurs. Les informations telles que la direction du regard, les mouvements des lèvres liés à la parole, sont intégrées dans le sillon temporal supérieur (Hoffman & Haxby 2000 ; Puce et al., 2003) tandis que les expressions faciales sont traitées dans l'amygdale et l'insula (Breiter et al., 1996 ; Morris et al., 1998 ; Ishai et al., 2004). Les invariants faciaux eux, sont traités dans le gyrus fusiforme latéral. Des régions du circuit de la récompense comme le noyau accumbens et le cortex orbitofrontal sont impliqués pour ce qui est du jugement de l'attractivité et de la pertinence en tant que partenaire sexuel (Aharon et al., 2001 ; Gottfried et al., 2003 ; Kranz & Ishai 2006). Même si certaines régions ont des rôles très spécifiques, des traitements sont effectués par un système relativement étendu. Ces découvertes tendent donc à sortir d'une conception strictement modulaire comme cela pu être le cas, via des modèles plus distribués et connexionnistes, avec notamment une importance des boucles de rétroaction. Cependant, bien que des connexions bilatérales (suggérant donc une rétroaction) entre l'arrière du gyrus fusiforme droit, le gyrus occipital moyen, le gyrus occipital inférieur droit ainsi que le gyrus temporal inférieur gauche aient été mises en évidence (Rotshtein, et al., 2007), il est impossible, de par la latence de la technique d'IRMf ainsi que la rapidité du traitement du visage, d'observer la temporalité des activations de ces régions et donc de conclure sur des processus de type purement ascendant ou rétroactifs.

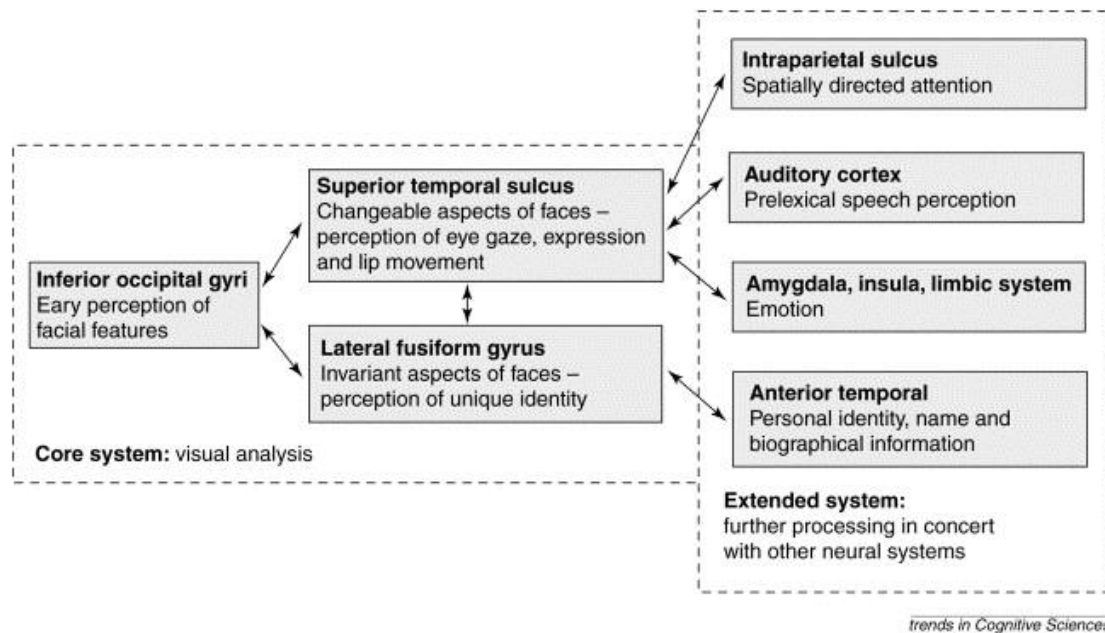


Figure 5. Modèle de distribution neuronale de la perception des visages. Tiré de Haxby et al (2000).

V. L'espace de représentation des visages

V.1. L'effet de distinctivité

On notera que dans le modèle de Bruce & Young ainsi que dans la majorité des modèles alternatifs, la reconnaissance de la familiarité est effectuée en faisant correspondre le percept visuel avec les représentations stockées en mémoire. Cette étape implique qu'il existerait un espace de représentation des visages rencontrés au cours de la vie d'un individu. Les chercheurs ont tenté d'expliquer la formation et la nature de cet hypothétique espace. Valentine & Bruce (1986) ont supposé l'existence d'un prototype facial pour expliquer cet effet. Chaque visage rencontré est comparé au prototype et plus le nombre de transformations nécessaires pour passer du prototype au visage est élevé, plus le visage est reconnu facilement (plus il est distinctif). Cette hypothèse apporte une réponse directe à ce que l'on appelle l'effet de distinctivité, impliquant que nous reconnaissons plus facilement les visages jugés inhabituels que les visages communs (Bartlett et al., 1984 ; Brigham, 1990 ; Bruyer & Courvoisier, 1990 ; Going & Read, 1974 ; Light et al., 1979 ; Winograd, 1981). Il est à noter que si les visages distinctifs sont plus rapidement identifiés, ils sont en revanche moins rapidement catégorisés en tant que visage que des visages se rapprochant plus du prototype. Une autre explication formulée en termes de trace mnésique a été proposée par Light et al. (1979). Selon cette hypothèse, chaque visage rencontré est encodé en une trace mnésique. L'exposition à un visage typique a plus de chance d'activer

la trace mnésique d'un autre visage typique, puisque que leurs traces mnésiques sont encodées de façons relativement proches. Il y a donc plus d'erreurs pour l'identification des visages typiques que distinctifs.

V.2. L'effet d'autre type

Appelé *other-race effect* (ORE) en anglais, l'effet d'autre type est un phénomène ayant amené des réflexions similaires à l'effet de distinctivité sur la nature de l'espace de conception des visages. L'ORE implique que nous avons plus de facilités à reconnaître des visages de notre propre type que des visages d'autres types. Cet effet a été de nombreuses fois documenté (Bothwell et al., 1989 ; Brigham, 1986 ; Carroo, 1986, 1987 ; Chance et al., 1975 ; Chiroro & Valentine, 1995 ; Lindsay et al., 1991 ; Lindsay, & Wells, 1983 ; Ng & Lindsay, 1994 ; Rhodes et al., 1990 ; Rhodes et al., 1989 ; Shapiro & Penrod, 1986 ; Shepherd & Deregowski, 1981 ; Valentine & Bruce, 1986 ; Valentine & Endo, 1992) mais la nature de celui-ci reste ouverte à débat. L'universalité de l'ORE nous amène d'emblée à rejeter l'hypothèse que certains types de visages seraient plus homogènes que d'autres ; de plus, Goldstein (1979a, b) a étudié la variabilité de visages de type européen¹ et de visages de type asiatiques et n'a pas observé de différences significatives. Une autre idée avancée par Chance & Goldstein (1981) est qu'un déficit d'attention serait à l'origine de l'ORE. Selon cette hypothèse, nous dirigerions plus notre attention vers les visages du type majoritaire dans l'environnement que vers les visages d'autres types. Si cette étude montre que le traitement est différent pour les visages d'autres types que pour les visages de même type, rien ne permet de conclure que cette différence provient d'un déficit d'attention. L'hypothèse la plus établie à ce jour est celle de contact. Le mécanisme est semblable à celui observé pour l'effet d'expertise ; une exposition accrue à un certain type de stimulus entraîne de meilleures capacités pour traiter ce type de stimulus. Dans la majorité des cas, étant principalement exposé à des visages de notre type, nous développons de meilleures capacités pour traiter ces visages-là, ce qui entraîne l'ORE. Goldstein & Chance (1980) ont proposé une explication théorique à l'ORE impliquant la création d'un schéma facial basé sur les visages rencontrés quotidiennement. Le schéma formé ressemble donc plus aux visages du type majoritaire de notre environnement qu'aux visages d'autres types. Les visages rencontrés sont ensuite comparés à ce schéma facial. Les visages d'autres types étant plus éloignés du

¹ Le terme employé dans la littérature anglo-saxonne est « caucasien », terme obsolète en français. Nous utiliserons donc le terme « de type européen » comme équivalent, sans que cela n'implique rien quant à l'origine des personnes mentionnées.

schéma type, ils seraient donc traités de manière moins efficiente que les visages de type majoritaire.

V.3. Le modèle de Valentine

Valentine (1991), voyant de grandes similitudes entre ces explications, a proposé un encodage des visages dans un espace multidimensionnel, appelé *face-space*. Les n dimensions de cet espace correspondent aux différentes caractéristiques physiologiques des visages (par exemple la couleur des yeux, la forme de la bouche etc.). Il existe par conséquent un nombre de dimensions très élevé puisqu'elles doivent permettre de représenter le visage. Cependant il est usuel de représenter ce modèle dans un espace à deux dimensions pour plus de clarté. Le centre de ce modèle représente le visage prototypique, le visage moyen sur toutes les dimensions. Les différents visages rencontrés vont être encodés et répartis dans cet espace. Les visages les plus proches de la valeur centrale seront dits « typiques » et les visages s'éloignant du prototype seront dits « atypiques ». Les visages typiques étant plus courants que les visages atypiques, la densité de point diminue à mesure que l'on s'éloigne du centre de l'espace euclidien. Lorsque l'on rencontre un visage, la reconnaissance va être établie si un point se trouve déjà dans l'espace multidimensionnel là où le visage rencontré doit être encodé. Le système est sensible à l'erreur puisqu'aucun visage ne sera encodé deux fois de la même manière. Puisque la densité de visages encodés est plus faible dans les régions de visages atypiques, les visages atypiques sont moins sensibles à l'erreur ; c'est l'effet de distinctivité, les visages atypiques sont plus facilement reconnus que les visages typiques (Going & Read, 1974 ; Light et al., 1979 ; Rhodes et al., 1997 ; Valentine, 1991 ; Valentine & Bruce, 1986). Cet effet explique aussi pourquoi on peut reconnaître des personnes connues via des caricatures, les traits caractéristiques étant rendus encore plus distinctifs et éloignant donc la représentation du centre du modèle, plus propice aux erreurs d'identification ; l'effet opposé étant obtenu avec les anti-caricatures (Rhodes et al., 1987 ; Benson & Perrett, 1991, 1994 ; Calder et al., 1996 ; Rhodes & Moody, 1990 ; Stevenage, 1995). On pourrait arguer que les traits inhabituels des visages atypiques entraînent un gain d'attention envers ceux-ci, expliquant l'effet de distinctivité. Cependant, il est montré que les visages typiques sont classés plus rapidement en tant que visages (et bénéficient donc d'une attention au moins aussi soutenue) que les visages atypiques (Valentine & Bruce, 1986). Le modèle de Valentine permet cependant de trouver une réponse à cette question. La moins bonne reconnaissance des visages d'autres type peut de premier abord paraître contre-intuitif avec l'effet de distinctivité. Les visages d'autres types sont éloignés de notre prototype et devraient donc être plus distinctifs ; c'est effectivement le cas. Cependant,

l'intra-variabilité au sein des différents types de visages est relativement constante (la variabilité entre les visages européens est semblable à la variabilité entre les visages asiatiques). Les visages d'autres types formeront un conglomérat de point, tout comme les visages typiques, mais éloignés du centre prototypique (voir Figure 6). Une même distance entre deux visages sera beaucoup plus facilement identifiable entre des visages proches du centre, car la distance angulaire sera supérieure comparativement à des visages éloignés du centre prototypique. La distance par rapport au prototype ainsi que la densité de la région où se situe le visage à identifier détermine donc sa facilité ou difficulté d'identification.

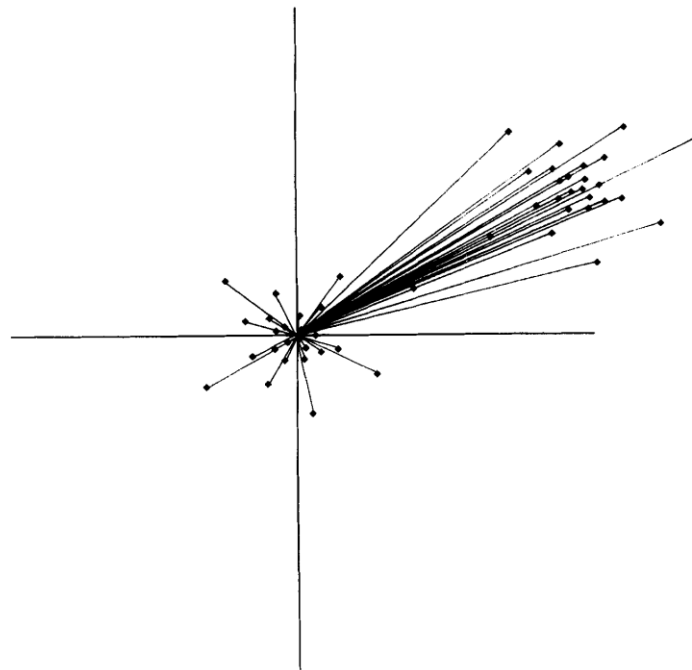


Figure 6. Représentation en deux dimensions de l'espace de représentation des visages de Valentine (1991).

Les visages de même type constituent la population centrale et les visages d'autres types constituent la population en périphérie. Tiré de Valentine (1991).

VI. Perception des visages chez le nouveau-né

Nous avons donc vu des modèles permettant d'expliquer comment une vision mature et un certain niveau d'expérience nous permettent de traiter et reconnaître les visages. Cependant, que se passe-t-il lorsque l'expérience est inexistante, et la vue ne permet pas d'accéder aux détails constituant les visages ? Autrement dit, qu'en est-il du traitement des visages par les nouveau-nés ? La préférence naturelle observée pour des schémas de visages traduit-elle l'existence d'un mécanisme inné de traitement des visages ? Si l'on suppose effectivement

qu'un espace de représentation des visages existe, est-il vierge à la naissance ? Comment se construit-il ? Voilà autant de questionnements auquel nous tenterons de répondre dans cette partie.

VI.1. Des mécanismes spécifiques : Conspec/Conlern

Deux conceptions se sont principalement opposées concernant la reconnaissance des visages chez le nouveau-né. Une première hypothèse est que la reconnaissance des visages seraient traités par un mécanisme spécifique inné. La seconde hypothèse est que les visages sont traités par les mêmes mécanismes perceptifs que les autres stimuli et qu'une spécialisation dans le traitement des visages apparaît au fil de l'expérience.

Voulant explorer l'existence ou non de mécanismes innés permettant une reconnaissance ou du moins une détection des visages, Goren et al. (1975) ont évalué la préférence pour des motifs schématiques ressemblant à des visages par rapport à d'autres motifs. Les schémas de visages utilisés comprennent trois points représentant les yeux et la bouche, disposés dans un ovale représentant le contour du visage. Cette étude et les répliques successives (Johnson et al., 1991 ; Valenza et al., 1996) ont montré que les nouveau-nés orientent plus fréquemment leur regard et regardent plus longtemps des motifs schématiques représentant la configuration d'un visage plutôt que des motifs représentatifs d'une autre configuration (voir Figure 7). Il existerait donc au moins un système de détection des stimuli ressemblant ou ayant la configuration de visage, présent quelques minutes seulement après la naissance. Il est peu probable que ces compétences puissent résulter d'une quelconque expérience, étant donné que la vue n'est quasiment pas stimulée avant la naissance. L'existence d'une capacité de détection et de traitement des visages très précoce n'est plus remise en cause, mais sa nature est matière à débat. Cette compétence précoce est possiblement interprétable comme l'existence d'un système inné qui permettrait aux nourrissons de diriger une attention toute particulière envers les visages et de rapidement apprendre à tirer de l'information de ceux-ci ; cependant, elle peut tout aussi bien résulter d'un biais attentionnel envers certaines caractéristiques visuelles plus globales qui se retrouvent dans les visages (Pascalis & Kelly, 2009, Slater & Quinn, 2001).

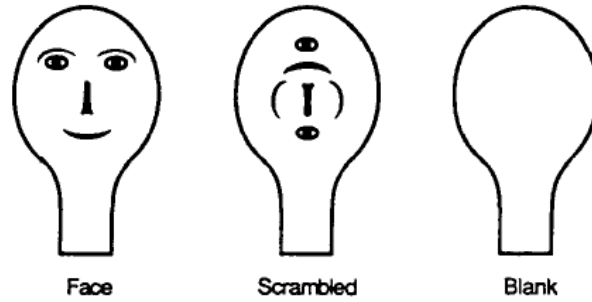


Figure 7. Stimuli utilisés par Johnson et al. (1991) : un schéma de visage avec ses éléments constitutifs, un schéma de visage avec les éléments constitutifs mélangé, et un schéma de visage sans ses éléments constitutifs.

Les nouveau-nés orientent plus le regard vers le schéma le plus à gauche soit le plus représentatif d'un visage réel. Tiré de Johnson et al. (1991).

L'hypothèse d'un système configural inné a été formulée par Morton & Johnson (1991). Se basant sur le phénomène d'empreinte mis en évidence chez les oies de Lorenz, cette théorie stipule l'existence de deux systèmes : CONSPEC et CONLEARN. CONSPEC serait un système sous-cortical et CONLEARN un système cortical. Le premier système, CONSPEC, serait dédié à la détection des visages et prédisposerait les nourrissons à tourner leurs regards vers ceux-ci. Le système sous-cortical CONSPEC serait sensible aux basses fréquences visuelles, correspondant aux aspects globaux du visage et non aux détails, permettant ainsi la détection d'un visage. Le second, CONLEARN, serait dédié aux hautes fréquences et permettrait l'identification des visages. CONSPEC agirait en quelque sorte comme un tuteur pour CONLEARN. CONSPEC serait présent dès la naissance et permettrait à CONLEARN, par le fait d'une attention accrue envers les visages, d'acquérir une expertise pour les visages via CONLEARN. CONSPEC disparaîtrait à mesure que CONLEARN émergerait. Cette transition viendrait expliquer l'absence de préférence pour les visages schématiques observée vers 5-6 semaines, avant que celle-ci ne réapparaisse vers 10 semaines.

VI.2. Des mécanismes sensoriels dépendant de l'expérience

Toutefois, si ce à quoi CONSPEC est sensible peut-être réduit à un simple schéma constitué de trois points, alors l'attraction des nouveau-nés vers les visages devrait être due à des propriétés structurelles de ceux-ci. Le biais attentionnel envers ces propriétés dépendrait du système visuel immature du nourrisson et pourrait donc s'appliquer à d'autres stimuli que des visages, accordant du crédit à la seconde hypothèse. Une étude de Simion et al. (2002) montre en effet que les nouveau-nés ont une préférence pour tous les stimuli ayant une asymétrie haut-bas, avec

une plus forte densité d'éléments dans la partie haute plutôt que dans la partie basse. De plus, nous avons précédemment vu que des colonnes corticales des aires visuelles primaires sont sensibles à différentes orientations de stimuli. Les nouveau-nés (âgés de moins d'une semaine) présentent une préférence pour les stimuli horizontaux plutôt que pour les stimuli verticaux (Slater et al., 1985 ; Farroni et al., 2000). Les cellules responsables de la sensibilité aux orientations horizontales pourraient être plus précocement actives que celles sensibles à des orientations verticales. La configuration des visages présentant ce type de configuration (une asymétrie haut-bas avec plus d'éléments dans la partie haute et organisés selon des bandes horizontales plutôt que verticales) bénéficierait de ce biais attentionnel. L'attrait des nourrissons envers les visages peut donc s'expliquer par des propriétés visuelles saillantes facilement détectables pour le nourrisson, qui comme nous l'avons vu, ne présente pas une vision complètement développée. Si l'asymétrie haut-bas ou bien les lignes horizontales semblent être des propriétés essentielles dans le traitement des visages précoces, les zones à fort contrastes comme les yeux (un rond sombre dans un rond clair) semblent aussi diriger l'attention des nourrissons. En effet, dès 2 mois, les nourrissons préfèrent regarder les yeux que les autres régions du visage (Hainline, 1978 ; Haith et al., 1977). D'autres travaux ont montré que les nouveau-nés préfèrent un visage dans lequel les yeux sont dirigés vers le bébé plutôt que détournés du bébé (Farroni et al., 2002) et un visage dans lequel les yeux sont ouverts plutôt que fermés (Batki et al., 2000). En grandissant, cette préférence semble de plus spécifique aux yeux humains, puisque des visages de singes présentés avec des yeux d'humains plutôt que les leurs sont préférés par les nourrissons dès 3 mois (Dupierrix et al., 2014). Les yeux humains possèdent une sclérotique blanche créant un fort contraste avec la pupille, une morphologie se distinguant de celle des autres primates (Kobayashi & Kohshima, 1997). Cette unicité les amène à être particulièrement bien détecté par les nourrissons (Kobayashi & Kohshima, 1997 ; Itier et al., 2011). Il est probable que ce soit l'évolution conjointe de la vue du nourrisson ainsi que son expérience répétée avec ce type de stimuli qui induise cette capacité de détection. L'importance du regard est commune au sein des primates (bien que n'ayant pas le même sens selon les espèces), jouant un rôle crucial dans les interactions sociales (Argyle & Cook, 1994 ; Kleinke, 1986 ; Emery, 2000). Une sensibilité précoce au regard mutuel apparaît donc essentielle pour le bon développement des compétences sociales. Baron-Cohen (1995) a postulé l'existence d'un « détecteur de direction de regard ». Ce système permettrait la détection du regard dans un premier temps, puis servirait à suivre le regard du congénère.

VI.3. Des capacités plus poussées : la reconnaissance de la mère

Il faut toutefois être prudent quant aux capacités observées dans les premiers mois de vie. Il est envisageable que la vision immature des nouveau-nés masque des capacités cognitives de traitement de visages qui pourraient être plus développées que l'on ne l'observe dans la plupart des études. D'autres recherches que celles citées précédemment tendent en effet à montrer que le traitement des visages à la naissance n'est pas uniquement schématique et mettent en évidence un système de reconnaissance plus poussé que celui que le seul système CONSPEC permettrait. Heron-Delaney et al. 2011 ont montré que des nouveau-nés (âgés de seulement 3 jours) présentent une préférence (et sont par conséquent capables de différencier) pour les visages humains par rapport à des visages de macaques. Si l'on se limite à des critères comme la dominance d'éléments dans le haut et la présence d'axe horizontaux, les visages humains et de primates non-humains présentent pourtant une configuration très similaire. Dans le cadre du modèle CONSPEC/CONLEARN, CONLEARN nécessite plusieurs semaines voire plusieurs mois avant d'être efficace et ne peut donc être responsable de la distinction observée par Heron-Delaney et al. De plus, la reconnaissance des visages familiers et particulièrement celui de la mère après la naissance témoigne aussi de capacités plus poussées que ne le permettrait le système CONSPEC défini par Morton & Johnson. En effet, des nouveau-nés âgés de quelques jours seulement sont capables d'identifier le visage de leur mère et ont une préférence pour celui-ci lorsqu'il est présenté avec le visage d'une femme inconnue, que les visages soient présentés physiquement (Bushnell, 2001 ; Pascalis et al., 1995) ou via une vidéo (Walton et al., 1992). On pourrait arguer que la reconnaissance de la mère ne passe pas uniquement par la modalité visuelle, mais peut aussi se faire grâce aux signaux olfactifs (Cernoch & Porter, 1985) ou auditifs (DeCasper & Fifer, 1980). C'est pourquoi dans l'étude de Pascalis et al. (1995) les habits de la mère et de l'inconnue étaient masqués avec un drap blanc et un parfum répandu pour s'assurer que l'odeur de la mère ne joue pas un rôle dans la reconnaissance. De plus, il était demandé aux femmes de ne pas parler. Dans l'étude de Walton et al., la modalité olfactive ne pouvait évidemment pas jouer et les mères étaient aussi silencieuses. Pour savoir quels éléments visuels permettent cette discrimination par le nourrisson et sa vue immature, Pascalis et al. (1995) ont, dans la deuxième partie de leur étude, masqué avec un foulard les contours extérieurs du visage de la mère et de l'inconnue. Ainsi, lorsque la frontière front/cheveux est masquée, les nourrissons ne présentent plus de reconnaissance du visage de la mère. La reconnaissance sans la présence des traits extérieurs ne se fait que vers 2 mois. De plus, les nourrissons sont, dès 3 mois, capables de reconnaître un nouveau visage auquel ils ont été

familiarisés 2 minutes ou 24 heures auparavant, et sont de plus capables de le reconnaître même s'il est présenté sous un angle nouveau (Pascalis et al., 1998). L'exposition à la voix de la mère se fait déjà *in utero* et facilitera l'apprentissage du visage de la mère à la naissance par association de la modalité audio et visuelle. Une étude de Sai et al. (2005) montre l'importance de la voix dans la reconnaissance du visage de la mère. Dans cette étude, juste après la naissance, les mères ont été encouragées à interagir visuellement avec leur bébé mais sans parler ni faire de son. Après quelques heures, des paires mère-inconnue ont été présentées aux nouveau-nés, sans qu'une reconnaissance puisse cette fois être établie. Cela confirme donc un apprentissage rapide du visage de la mère après la naissance, facilité par une mémorisation à long terme de la voix de la mère apprise *in utero*.

Chapitre 2 : Le rétrécissement perceptif

On peut établir des parallèles entre le développement du langage et du traitement des visages durant la première année de vie. Dans la grande majorité des cas, les nourrissons sont bien plus exposés à leurs langues maternelles qu'à des langues étrangères et sont bien plus exposés à des visages de leurs types plutôt qu'à des visages d'autres types (Rennels & Davis, 2008 ; Sugden et al., 2014). Dans ce chapitre, nous tâcherons dans un premier temps de décrire le phénomène que l'on nomme communément le rétrécissement perceptif (ou « perceptual narrowing » dans la littérature anglo-saxonne) : une exposition fréquente à un type de stimuli entraîne un rétrécissement des capacités à percevoir les contrastes fins à l'extérieur de l'espace des stimuli fréquents. Il faut noter que bien que le terme de « perceptual narrowing » soit le plus utilisé pour décrire ce phénomène, celui-ci peut laisser supposer une pure et simple régression des capacités du nourrisson au cours de son développement ; les termes « attunement », « tuning », « reorganisation » (réglage, ajustement, réorganisation) sont donc parfois utilisés dans la littérature anglo-saxonne, transmettant plutôt l'idée d'un réajustement des capacités perceptives du nourrisson sur son environnement. Nous décrirons donc son émergence chez le nourrisson à la fois pour le traitement des visages et pour le traitement du langage. Nous verrons ensuite l'implication du rétrécissement perceptif sur la cognition sociale du nourrisson. Cette réorganisation perceptive entraîne en effet chez les nourrissons une différence dans la façon d'interagir avec les individus en fonction de la langue qu'ils parlent et de leur type de visage. Puis, nous verrons comment des modifications expérimentales ou des facteurs environnementaux peuvent moduler l'intensité du rétrécissement perceptif. Ce phénomène est souvent mis en évidence via des études en laboratoire sur des populations de nourrissons monolingues et vivant dans des milieux homogènes en termes de type de visages. Cependant, le type de tâche présentée ou l'environnement dans lequel grandit le nourrisson sont connus pour impacter les résultats observés. En effet, puisque le rétrécissement perceptif est l'ajustement des capacités perceptives du nourrisson par rapport à ce qui est vu/entendu dans l'environnement, un nourrisson grandissant dans une famille bilingue ou ayant des interactions récurrentes avec des personnes de types différents présentera un ajustement perceptif différent.

I. Rétrécissement perceptif pour le traitement des visages

I.1. Au niveau de l'espèce

A la naissance, le nouveau-né n'a pas encore d'expérience visuelle concernant les visages. Il n'a donc pas vu plus de visages de son type que d'autres types, ou bien plus de visages féminins

que de visages masculins, ou encore plus de visages d'adultes que de visages d'enfants. Si l'on se réfère à l'espace de représentation des visages de Valentine (1991), celui-ci est donc vierge à la naissance et les entrées qu'il va recevoir de l'environnement vont contribuer à le façonner. Malgré leur peu d'expérience des visages durant leur premiers mois de vie, les nourrissons présentent vis-à-vis des visages nous fait observer de surprenantes capacités. En effet, les nourrissons de moins de 6 mois sont capables de reconnaître au niveau individuel différents visages de singes de la même manière que des visages humains (Pascalis et al., 2002). Ces résultats ont été répliqués avec des visages de singes (Fair et al., 2012) et même des visages d'une espèce non-primate, le mouton (Simpson et al., 2011). Ces résultats sont d'autant plus intéressants que les adultes n'ont pas cette capacité de discrimination pour des visages d'autres espèces (Pascalis & Bachevalier, 1998 ; Pascalis et al., 2002), l'ayant « perdu » au cours du développement. Les nourrissons humains étant rarement exposés à des visages d'autres espèces vont en effet graduellement perdre leur capacité à les discriminer. Il est fort à parier qu'un nourrisson élevé au contact de visages de singes garderait cette capacité ; une telle étude n'existe fort heureusement pas, mais des nourrissons fréquemment exposés à des photographies de visages de singes gardent la capacité de les discriminer après 9 mois (Pascalis et al., 2005 ; Scott & Monesson, 2009). De la même manière, Sugita (2008) a montré que des macaques rhésus élevés sans contact avec des visages d'autres macaques mais uniquement à des visages humains montrent une préférence pour les visages humains. Il n'y a cependant pas ici d'évaluation de la capacité à discriminer les visages.

I.2. Au niveau du type

On observe un phénomène similaire pour le type de visage. Les nourrissons de moins de 3 mois ne présentent initialement pas de préférence spontanée pour un type de visage plutôt qu'un autre (Kelly et al., 2005, 2007a ; Quinn et al., 2008). Cependant, ils vont avoir une exposition asymétrique aux différents types de visages ; dans la majorité des cas ils vont être exposés beaucoup plus régulièrement (96% du temps) à des visages de leur type qu'à des visages d'autres types (Sudgen et al., 2014). La surreprésentation des visages de même type comparativement aux visages d'autres types dans l'environnement du nourrisson va entraîner une différence dans la façon de regarder les visages de même et d'autres types. Kelly et al. (2005) ont testé la préférence – inférée par un temps de regard plus long – pour les visages de même et d'autre types chez le nouveau-né et le nourrisson de 3 mois de type européen. Si les nouveau-nés ne présentent pas de différence, les nourrissons européens de 3 mois passent plus de temps à regarder les visages de leur propre type. Ces résultats ont été répliqués pour des

nourrissons de types africain et asiatique (Bar-Haim et al., 2006 ; Kelly, et al., 2007b). Remarquons ici que la préférence pour les visages de même type est en fait une préférence pour les visages majoritaires dans l'environnement ; dans l'étude de Bar-Haim, des nourrissons éthiopiens vivants dans des camps israéliens sont exposés à la fois à des visages de type africains et de type européens, et ne présentent de préférence ni pour les uns ni pour les autres. Liu et al. (2015) se sont intéressés à l'évolution des préférences après 3 mois. Si leurs résultats confirment une préférence pour les visages de même type à 3 mois, ils ne trouvent en revanche plus de préférence vers 6 mois et une préférence cette fois-ci pour les visages d'autres types à 9 mois. En plus des préférences, on observe aussi une différence en termes de balayage visuel (*scanning*) des visages selon leurs types. Des études oculométriques (Wheeler et al., 2011 ; Liu et al., 2011 ; Xiao et al., 2013) nous montrent que pour les nourrissons entre 4 et 10 mois, bien que les temps de fixation soient équivalents pour des visages de même type ou d'autres types, les régions d'intérêts ne sont pas les mêmes à mesure que les nourrissons grandissent. Le fait de ne pas trouver de différence en termes de temps de fixation pour les visages de même et d'autres types peut de prime abord sembler contradictoire avec les résultats de Liu et al., 2015, montrant une préférence pour les visages d'autres types (et donc plus de temps passer à regarder ces visages). Il faut revenir sur la méthodologie de chaque étude pour mieux comprendre les résultats. La préférence est testée en présentant deux visages côte à côte, un de même type et un d'un autre type. L'analyse des temps de regard des études de Wheeler et al. (2011), Liu et al. (2011) et Xiao et al. (2013) se fait en présentant un visage à la fois. On peut donc tout à fait trouver un intérêt similaire lorsque les visages sont présentés seuls mais trouver une préférence pour l'un des deux visages lorsqu'ils sont présentés simultanément. Pour ce qui est des patterns de fixation, les nourrissons, à partir de 6 mois, vont progressivement passer plus de temps à regarder les yeux et moins de temps à regarder la bouche des visages de leur type tandis que les patterns de fixation envers les visages d'autres types semblent peu évoluer.

On observe une reconnaissance amoindrie des visages d'autres types qui pourrait être due à cette différence de pattern de fixation. Concrètement, un nourrisson grandissant dans un environnement majoritairement européen verra sa capacité à discriminer des visages de types asiatique ou africain régresser mais gardera et améliorera la capacité à discriminer des visages de type européen. Les nourrissons présentent ainsi de meilleures capacités pour discriminer les visages de leur type après 9 mois (Kelly et al., 2007a, 2009) via un protocole d'habituation (voir Figure 8). Cependant, Sangrigoli & De Schonen (2004) trouvent une émergence de cet effet plus précoce, dès 3 mois, en utilisant des temps d'habituation et de test plus longs que ceux

utilisés par Kelly et al. Une différence méthodologique notable est que les visages utilisés par Sangrigoli & De Schonen sont en nuances de gris (et non en couleur, tel que ceux présentés par Kelly et al.), ce qui diminue potentiellement la quantité d'informations permettant la distinction des visages. Sangrioli & De Schonen trouvent aussi que lorsque les nourrissons sont habitués à trois visages plutôt qu'un seul, les nourrissons de 3 mois (de type européens) sont à nouveau capables de distinguer les visages d'un autre type (ici des visages asiatiques). Hayden et al. (2007) mettent aussi en évidence une baisse des capacités discriminatives pour les visages d'autres types avant 9 mois, chez des nourrissons de 3.5 mois. Pour ce faire, ils ont habitué des nourrissons de type européen à des visages de type européen ou asiatique puis ont présenté côte à côte le visage de l'habitué avec un visage issu du *morphing* (transformation morphologique) entre le visage européen et asiatique. Si le visage de l'habitué était le visage européen, le visage issu du *morphing* était composé à 70% du visage européen et à 30% du visage asiatique ; si le visage de l'habitué était le visage asiatique, le visage issu du *morphing* était composé à 70% du visage asiatique et à 30% du visage européen. Les nourrissons ont passé plus de temps à regarder le nouveau visage uniquement lorsqu'ils étaient habitués au visage européen, cette préférence pour la nouveauté signifiant qu'ils ont reconnu le visage de l'habitué uniquement lorsqu'il était européen. Une nuance à faire est que le « nouveau » visage n'est pas entièrement nouveau puisqu'il est composé à 70% du visage de l'habitué et seulement à 30% du nouveau visage. La distinction à faire est donc beaucoup plus fine que pour un visage entièrement nouveau, comme c'est le cas dans l'étude de Kelly et al. Si l'ORE semble vraiment robuste vers 9 mois, il n'est pas surprenant de pouvoir déjà observer des capacités réduites pour la discrimination de visages d'autres types vers 3 mois, l'émergence de l'ORE ne survenant pas brutalement mais se faisant de manière graduelle. De plus, il est possible que la perte de capacité de discrimination des visages d'autres types ne suive pas exactement la même temporalité selon les types de visages. Bien que nous ayons mentionné que les différents types de visages ne présentent pas de différence en termes de variabilité intra-groupe (Goldstein, 1979a, b), il est possible que certains autres types de visages soient physiquement plus proches du type local que d'autres, et soient, pour reprendre l'espace de représentation des visages de Valentine (1991), plus éloignés du centre et donc plus difficilement discriminables. En effet, Kelly et al. (2007a) trouvent que les nourrissons de type européen discriminent encore les visages asiatiques à 6 mois (mais plus à 9 mois) alors qu'ils ne discriminent déjà plus les visages de type africain à 6 mois.



Figure 8. Exemple de stimuli via lequel l'ORE est mis en évidence chez les nourrissons. Le visage du haut est présenté pendant une habituation. Puis, ce même visage est présenté à côté d'un nouveau visage d'autre type. L'angle de prise de vue de la photographie peut être différent pour s'assurer que la tâche teste bien une reconnaissance individuelle et non la reconnaissance d'une photographie. Si le temps de regard des nourrissons est assimilable au niveau de hasard pour chacun des visages, on ne pourra pas conclure que les nourrissons reconnaissent le visage de la familiarisation. Tiré de Kelly et al. (2007a).

De manière intéressante, la catégorisation que les nourrissons font des visages d'autres types va elle aussi évoluer. Dans le cadre d'une procédure de familiarisation, Quinn et al. (2015) ont d'abord présenté des visages d'un autre type que celui du nourrisson. Par exemple, trois visages de type asiatique sont présentés à tour de rôle. Ensuite, un test consiste à présenter un nouveau visage de cette catégorie (dans notre exemple un nouveau visage asiatique) à côté d'un nouveau visage d'autre type d'une autre catégorie que celle de la familiarisation (un visage africain). Si les nourrissons ont formé une catégorie « visages asiatiques » durant la familiarisation, le nouveau visage asiatique sera perçu comme appartenant à cette catégorie. Selon le principe de préférence pour la nouveauté, les nourrissons regarderont donc plus le visage africain qui appartient à une nouvelle catégorie. Les résultats montrent que les nourrissons de 6 mois regardent plus le visage de la nouvelle catégorie lors de la phase de test, que ce visage soit de type asiatique, africain ou européen et que les nourrissons aient été habitués à la catégorie « visages asiatiques » ou à la catégorie « visages africains » durant la familiarisation. A 9 mois en revanche, les nourrissons présentent une préférence pour la nouveauté uniquement lorsque

le visage d'autre type est présenté à côté d'un visage européen lors du test. Si un visage asiatique est présenté à côté d'un visage africain, les nourrissons de 9 mois ne regardent pas plus le visage correspond à la catégorie à laquelle ils n'ont pas été familiarisé. Cela signifie que les nourrissons de 6 mois séparent donc les différents types de visages en autant de catégories, c'est-à-dire qu'ils vont former la catégorie « visages européens », la catégorie « visages asiatiques » et la catégorie « visages africains ». Les nourrissons de 9 mois en revanche semblent former deux catégories, les visages de leur groupe social (endo-groupe), qu'ils voient fréquemment, et les visages extérieurs à leur groupe social (exo-groupe), qu'ils ne rencontrent que très rarement.

Le développement de l'ORE se fait donc de manière graduelle durant la première année de vie. L'exposition prédominante aux visages de même type entraîne une familiarité avec ce type et une préférence pour ces visages familiers. Cette préférence est potentiellement responsable d'un changement dans la façon de regarder ces visages familiers par rapport à des visages non familiers. Il est à son tour possible que la différence de balayage visuel entraîne de meilleures capacités pour individualiser les visages faisant partie du groupe familial, les nourrissons étant plus à même de détecter les différences pertinentes pour distinguer ces visages.

II. Rétrécissement perceptif pour le traitement du langage

II.1. Diminution de la capacité à discriminer les contrastes phonémiques non-natifs

Concernant le développement du langage, on observe une dynamique comparable à celle que l'on vient de décrire pour le traitement des visages. D'un point de vue historique, c'est d'ailleurs pour le langage que le rétrécissement perceptif a d'abord été mis en évidence. Il est depuis longtemps établi que les adultes ont des difficultés pour différencier des sons de parole qui ne sont pas utilisés dans leur langue maternelle (Singh & Black, 1966 ; Goto, 1971 ; Miyawaki et al., 1975 ; Trehub, 1976). Ces résultats ont pu amener à penser que les nourrissons apprennent progressivement à devenir sensibles spécifiquement aux sons de leurs langues. Cependant, la littérature sur les nourrissons de moins de 6 mois montrent qu'ils peuvent faire la distinction entre la plupart des phonèmes des langues du monde, bien qu'ils ne soient pas exposés à celles-ci (Eimas et al., 1971 ; Trehub 1976 ; Werker et al., 1981). Cette aptitude surprenante, non présente chez les adultes, amène à postuler un modèle de développement inverse : les nourrissons seraient sensibles à tous les sons de langues mais ne conserveraient en grandissant que la sensibilité aux contrastes phonétiques de leur langue. Werker & Tees (1984) figurent parmi les premiers à avoir montré cette transition d'une sensibilité générale à une sensibilité

spécifique aux phonèmes natifs. Leur étude de 1984 montre notamment qu'un nourrisson de 6 mois baignant dans un environnement anglophone peut, tout comme un nourrisson de 6 mois exposé à l'hindi, faire la distinction entre des contrastes phonémiques hindi non-natifs (c'est-à-dire entre l'alvéolaire rétroflexe dans la syllabe [ʈa] et la consonne dentale dans la syllabe [ʈa] de l'hindi). En revanche, si les nourrissons de langue maternelle hindi sont encore capables de faire cette différence à 10-12 mois, les nourrissons anglophones en sont devenus incapables, tout comme les adultes monolingues anglophones. Werker & Tees ont testé la capacité des nourrissons anglophones à distinguer deux autres sons formant un contraste non-natif, la vélaire éjective dans la syllabe [k'i] et l'uvulaire éjective dans la syllabe [q'i] de la langue salish (parlée au Canada). Ils ont trouvé des résultats similaires, c'est-à-dire que les nourrissons anglophones sont sensibles à ce contraste à 6-8 mois mais ne le sont plus à 10-12 mois (voir Figure 9). Des nourrissons ont aussi été testés à un âge intermédiaire, à 8-10 mois, et les résultats montrent des capacités de discrimination moins élevées qu'à 6-8 mois, même si environ la moitié des nourrissons continue de faire la distinction. La perte de sensibilité aux phonèmes non-natifs semble donc se faire graduellement après 8 mois, pour être bien établie vers les 10-12 mois. D'autres contrastes non-natifs ont été testés sur des nourrissons de différentes langues et ont produit des résultats similaires. Les nourrissons japonais sont par exemple capables de différencier les sons [r] et [l] (non-contrastifs en japonais) à 6-8 mois mais ne le sont plus à 10-12 mois (Tsushima et al., 1994). De même, les nourrissons anglais ont plus de mal à distinguer des contrastes entre fricatives et affriquées utilisés dans le mandarin à 10-12 mois qu'à 6-8 mois (Tsao et al., 2006). Les résultats comportementaux sont de plus congruents avec les réponses électrophysiologiques, Rivera-Gaxiola et al. (2005) ayant montré que les PE propres à la discrimination de sons sont observés à 7 mois pour des nourrissons anglais exposés à des contrastes natifs et non-natifs, mais ne sont plus présents que pour les contrastes natifs à 11 mois. Eilers et al. (1982) trouvent cependant des résultats un peu divergents. Ils ont testé la distinction de contrastes phonémiques anglais, espagnols et tchèques par des nourrissons de langue anglaise et espagnole de 6-8 mois. Leurs résultats montrent que si les nourrissons espagnols arrivent à distinguer le contraste natif et les deux contrastes non-natifs, les nourrissons anglais distinguent le contraste natif mais seulement le contraste non-natif tchèque. Une première différence qui se dégage de cette étude par rapport à celles citées plus haut est la temporalité, l'absence de discrimination de contrastes non-natifs étant ici mise en évidence dès 6-8 mois. Un deuxième point intéressant est la différence entre les capacités des nourrissons espagnols et anglais, les premiers étant sensibles à un contraste phonémique de la langue des deuxièmes et les deuxièmes n'étant pas sensibles à un contraste phonémique de la langue des

premiers. A noter cependant que la population de nourrissons de langue espagnole est issue de la communauté hispanophone de Miami, et est probablement plus exposée à l'anglais que les nourrissons anglais de l'étude ne sont exposés à l'espagnol. Comme pour les visages, il est plus probable qu'en fonction de leur proximité avec la langue maternelle des nourrissons, certains contrastes soient plus ou moins difficilement reconnus, et que la temporalité de la perte de sensibilité dépende donc à la fois du contraste étudié et de la langue maternelle du nourrisson.

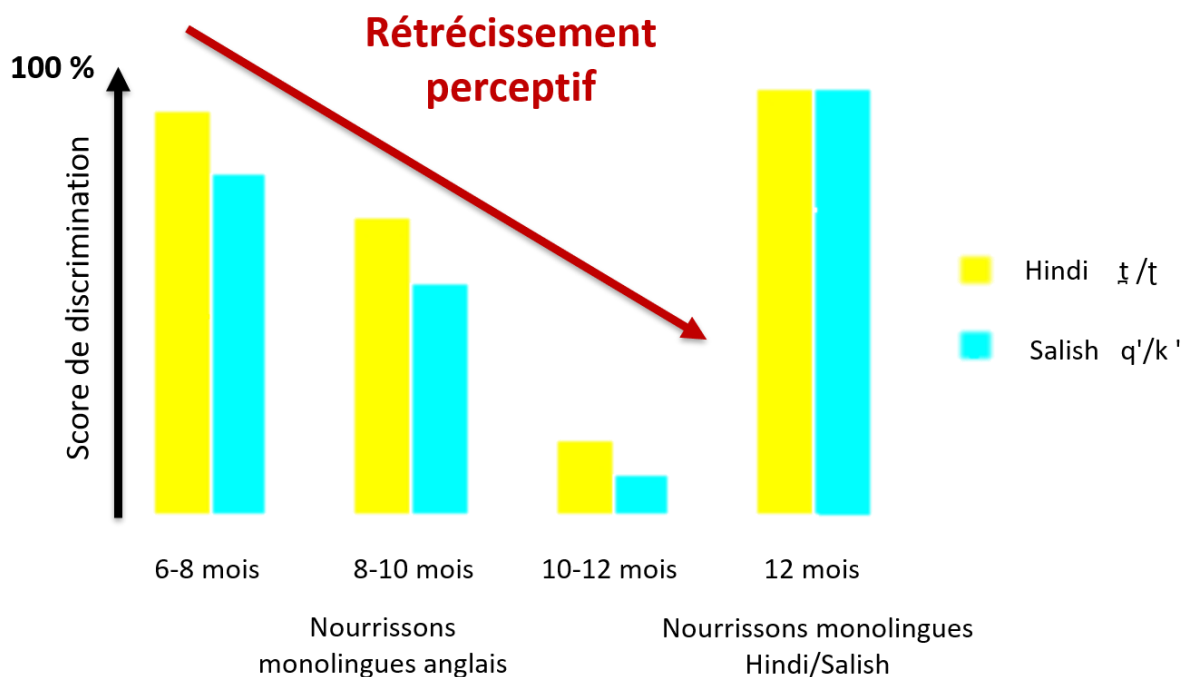


Figure 9. Capacité de discrimination de contrastes phonémiques de l'hindi et du salish par des nourrissons monolingues d'environnements anglais, hindi ou salish. Tiré et adapté de Werker & Tees (1984).

II.2. Amélioration de la capacité à discriminer les contrastes phonémiques natifs

Nous avons mentionné en introduction de ce chapitre que le rétrécissement est en fait un ajustement perceptif, et qu'en parallèle de la baisse de capacité à discriminer certains stimuli, on observe une hausse de la capacité à distinguer d'autres stimuli. Pour le langage, cela signifie que si les contrastes phonémiques non-natifs vont graduellement ne plus être correctement distingués, la distinction des contrastes phonémiques natifs va elle s'améliorer. Les données électrophysiologiques de Cheour et al. (1998) nous indiquent qu'une adaptation aux sons de la langue maternelle a lieu durant la première année de vie. Leurs résultats montrent que l'intensité de la *mismatch negativity* (une réponse cérébrale automatique suite à un stimulus inattendu)

diminue entre 6 et 12 mois suite à la présentation d'un contraste non-natif tandis qu'elle augmente suite à la présentation d'un contraste natif (voir Figure 10). Ces données sont confirmées d'un point de vue comportemental par plusieurs études. Kuhl et al. (2006) ont testé les capacités de nourrissons anglais et japonais à distinguer [ra] et [la], distinction non pertinente en japonais. A 6-8 mois, on observe des pourcentages de discrimination similaires pour les nourrissons anglais et japonais, indiquant qu'ils arrivent à faire la distinction dans environ 65% des cas. A 10-12 mois, les performances des nourrissons japonais chutent, conformément à ce qui est attendu, à environ 60% de distinction. En revanche, les nourrissons anglais de 10-12 mois présentent une amélioration de la capacité à distinguer ce contraste natif, leurs performances indiquant environ 74% de bonnes réponses. De la même manière, Tsao et al. (2006) montrent que la distinction d'un contraste phonémique utilisé en mandarin par des nourrissons exposés au mandarin est meilleure à 10-12 mois qu'à 6-8 mois et que la distinction d'un contraste phonémique utilisé en anglais est meilleure dans le groupe de nourrissons anglophones à 10-12 mois qu'à 6-8 mois. Si nous observons que les phonèmes deviennent plus finement reconnus après 7 mois, certains contrastes ne sont initialement pas distingués par les nourrissons qui ont besoin d'expérience langagière pour les distinguer correctement. Narayan et al. (2010) ont présenté les syllabes [na]–[ŋa], correspondant au contraste [n]–[ŋ] utilisé en philippin, à des nourrissons anglais et philippins de 4-5, 6-8 et 10-12 mois. Alors qu'au vu des résultats cités précédemment, on pourrait s'attendre à ce que les nourrissons de langue philippine discriminent ces sons à tous âges tandis que les nourrissons de langue anglaise ne le discriminent qu'aux plus jeunes âges, il n'en est rien. En effet, seuls les nourrissons de langue philippine de 10-12 mois sont capables de faire la distinction entre ces deux sons. De même, Sato et al. (2010) montrent que les nourrissons japonais sont capables de distinguer /mana/ de /ma:na/ (le deuxième présentant une voyelle longue, utilisée en japonais) à 9.5 mois mais pas à 4 et 7.5 mois. On voit donc que certains contrastes plus « difficiles » ne sont initialement pas distingués par les nourrissons mais que l'exposition répétée à une langue pour laquelle ces contrastes sont pertinents va permettre leur reconnaissance.

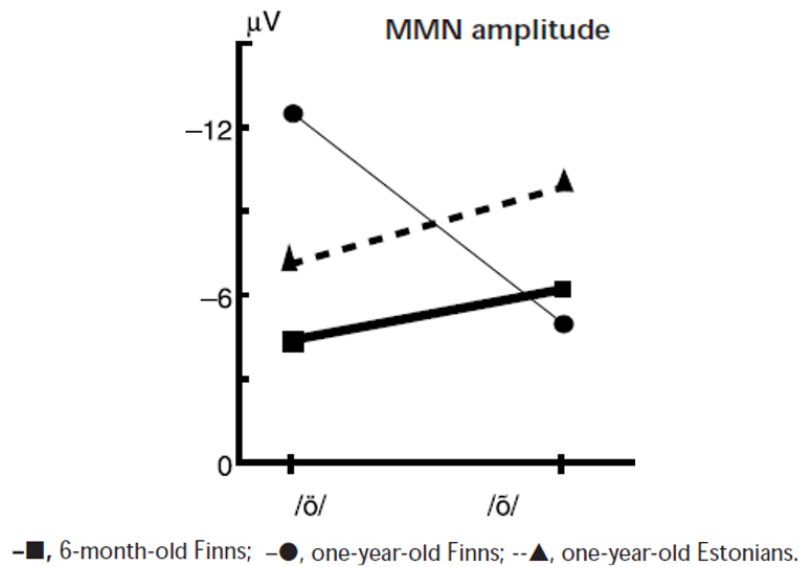


Figure 10. Evolution de l'amplitude de la Mismatch Negativity au cours de la 1^{ère} année de vie chez des nourrissons finnois et estoniens de l'étude de Cheour et al. (1998).

Les nourrissons sont exposés à une voyelle commune aux deux langues, /e/, puis à un stimulus déviant, soit présent dans les deux langues, /ö/, soit uniquement présent dans la langue estonienne, /õ/. Tiré et adapté de Cheour et al. (1998).

II.3. Un modèle explicatif : le Native Language Magnet model

Le *Native Language Magnet model* (NLMm) est un modèle avancé par Kuhl et al. (1994 ; 2000) pour tenter d'expliquer le rétrécissement perceptif pour le langage. Kuhl et al. proposent qu'un espace des sons prototypiques se forme au fur et à mesure de l'exposition à la langue ambiante. Les sons perçus sont rattachés au son prototypique le plus proche, qui fonctionne comme un aimant (d'où le terme *magnet* ; voir Figure 11). Le NLMm identifie trois étapes lors du développement (voir Kuhl et al., 1994, pour une explication plus approfondie). Durant une première étape, juste après la naissance, les catégories phonémiques ne sont pas encore formées permettant aux nourrissons de différencier la majorité des phonèmes des langues du monde. Cette distinction ne se fait pas encore sur la base de l'expérience linguistique du nourrisson mais simplement sur les propriétés acoustiques des sons perçus, grâce à des capacités auditives générales. Dans une deuxième étape, autour de 6 mois, des représentations de phonèmes prototypiques commencent à se former sur la base des propriétés distributionnelles de la langue ambiante. Si la capacité à séparer les sons sur la base de leurs propriétés acoustiques est globalement la même pour tous les nourrissons, l'espace des représentations phonémiques diffère selon la/les langues maternelles des nourrissons. On observe ainsi une distorsion de l'espace perceptif, la sensibilité perceptive étant diminuée au sein des catégories et étant

augmentée à proximité des frontières entre les catégories. Au fur et à mesure que l'expérience s'accumule, les représentations les plus souvent activées (prototypes) commencent à fonctionner comme des aimants perceptifs pour les autres membres de la catégorie, augmentant ainsi la similarité perçue entre les membres d'une même catégorie. Dans une troisième étape, à un âge non précisé par Kuhl mais après 6 mois (après le rétrécissement perceptif), les représentations phonémiques sont bien formées et agissent comme des attracteurs (voir Kuhl et al., 1992 pour une explication détaillée du *Perceptual Magnet Effect*, PME). L'effet d'aimant perceptif produit ainsi une facilitation des capacités de discrimination phonétique en langue maternelle et une réduction pour les contrastes phonétiques en langue étrangère. Une même différence acoustique n'est pas équivalente en fonction de sa distance par rapports aux attracteurs, les distances les plus proches des attracteurs étant minimisées et les distances aux frontières étant amplifiées. C'est ce mécanisme qui explique pourquoi nous pouvons entendre plusieurs instances d'un même phonème variant en termes de durée, d'intensité et de fréquence et toujours parvenir à l'identifier. Par exemple, le son [a] n'est pas prononcé de façon acoustiquement similaire par tout le monde, ni même n'est prononcé pareil par la même personne à chaque énoncé ; malgré ces fluctuations, si un [a] perçu est suffisamment proche du [a] prototypique, le son sera identifié comme un [a]. En revanche, un contraste entre deux phonèmes acoustiquement proches mais non présents dans les langues ambiantes ne donnera pas lieu à un contraste perceptif ; ces phonèmes non-natifs seront rattaché au phonème prototypique existant le plus proche et ne seront pas perçus comme des sons différents. On pourra mentionner une exception notable, les « clics », présents dans les langues khoïsan ou bantoues (dont le zoulou), phonèmes consonantiques non-natifs pour les locuteurs anglophones, qui continuent d'être discriminés après 12 mois par les nourrissons anglais (Best et al., 1988). Ces sons sont très éloignés des sons de la langue anglaise, et l'absence de concurrence articulaire et acoustique avec les phonèmes anglais peut expliquer l'absence de perte de sensibilité aux clics même après l'âge d'un an. Best & Avery (1999) semblent même montrer que ces clics sont tellement éloignés des autres phonèmes de la plupart des langues qu'ils ne sont pas traités en tant que phonème mais en tant que sons non langagiers par les auditeurs non-natifs.

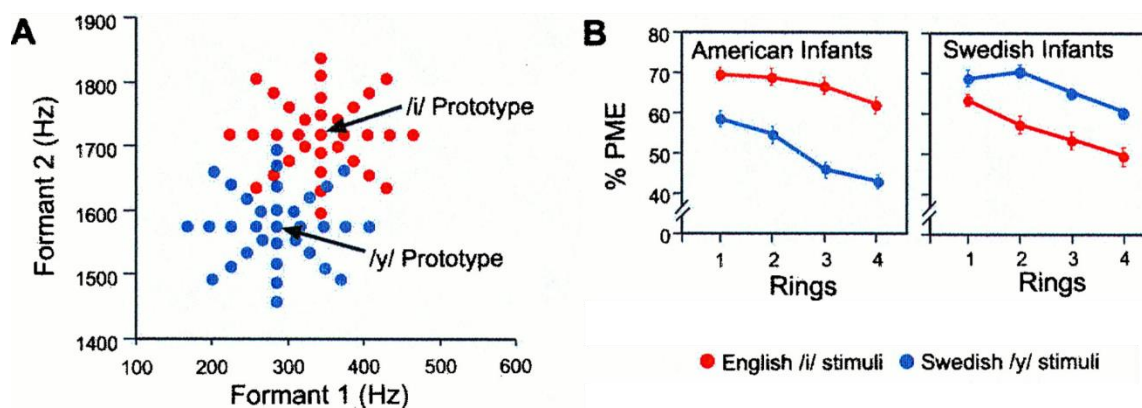


Figure 11. Illustration du PME.

Les stimuli intermédiaires sont assimilés aux prototypes de chaque langue maternelle. Un même stimulus (à équidistance des deux prototypes, sur le 3^{ème} anneau (*ring*)) sera ainsi perçu comme /i/ par les nourrissons américains et comme /y/ par les nourrissons suédois. Tiré de Kuhl (2000).

III. Rétrécissement perceptif inter-sensoriel

Nous avons pu voir que le rétrécissement perceptif est observé pour la perception du langage ainsi que des visages. La nature multisensorielle de notre environnement nous amène à nous demander si le rétrécissement perceptif affecte aussi des percepts multisensoriels. Lewkowicz & Ghazanfar (2006) ont testé des nourrissons sur une tâche inter-sensorielle d'intégration audio-visuelle. Pour ce faire, ils ont présenté à des nourrissons des vidéos de têtes de macaques en train de vocaliser, côte à côte. Il y avait deux types de vocalisation distincts, un « grunt » et un « coo », ne présentant pas la même dynamique visuelle (voir Figure 12). En synchronie avec les images vidéo, un audio correspondant à l'une des vocalisations (soit un « coo » soit un « grunt ») était joué. Les résultats montrent que les nourrissons de 4 et 6 mois regardaient plus longtemps le visage correspondant effectivement au son diffusé. En revanche, à 8 ou 10 mois, les nourrissons ne regardaient pas plus la tête de macaque correspondant au son joué. On retrouve donc la même temporalité que pour le rétrécissement perceptif observé pour les visages d'autres types ou les phonèmes non-natifs, c'est-à-dire qu'après 6 mois, les stimuli non-natifs ne sont plus correctement traités. Lewkowicz, Sowinski, & Place (2008) ont confirmé ces résultats en montrant de plus que les nourrissons de plus 8 mois restent toujours capables de différencier les deux sons et les deux vidéos mais ne sont plus capables de faire l'association. Ce n'est donc pas un déficit pour traiter une information uni-sensorielle qui est à l'origine des résultats observés mais bien un déficit pour traiter l'information de manière multisensorielle, mettant pour la première fois en évidence un rétrécissement perceptif multisensoriel. Dans

l'hypothèse d'un développement en U, des nourrissons plus âgés, de 12 et 18 mois, ont aussi été testés, mais n'ont pas montré de capacité d'association entre vidéo et son. En s'appuyant sur un protocole similaire, Pons et al. (2009) ont testé des nourrissons sur l'intégration audiovisuelle avec des visages humains mais pour des contrastes phonémiques non-natifs. Pour ce faire, ils ont testé des nourrissons anglais et espagnols de 6 et 11 mois avec les phonèmes et les visèmes [ba] et [va]. Ce contraste phonémique n'existe pas en espagnol mais est présent en anglais. Les nourrissons de 6 mois, quelle que soit leur langue maternelle, sont capables de différencier ces deux phonèmes. En revanche, les nourrissons espagnols de 11 mois n'en sont plus capables. Dans cette étude, la vidéo d'une locutrice en train de prononcer [va] était présentée à côté d'une vidéo de la même locutrice en train de prononcer [ba], et un audio jouant soit [ba] soit [va] était joué de manière synchrone avec l'ouverture et la fermeture des lèvres. Les résultats montrent que les nourrissons anglais, quel que soit leur âge, regardent plus longtemps la vidéo qui correspond au son joué. En revanche, seuls les nourrissons espagnols de 6 mois regardent plus le visage qui correspond au son diffusé.

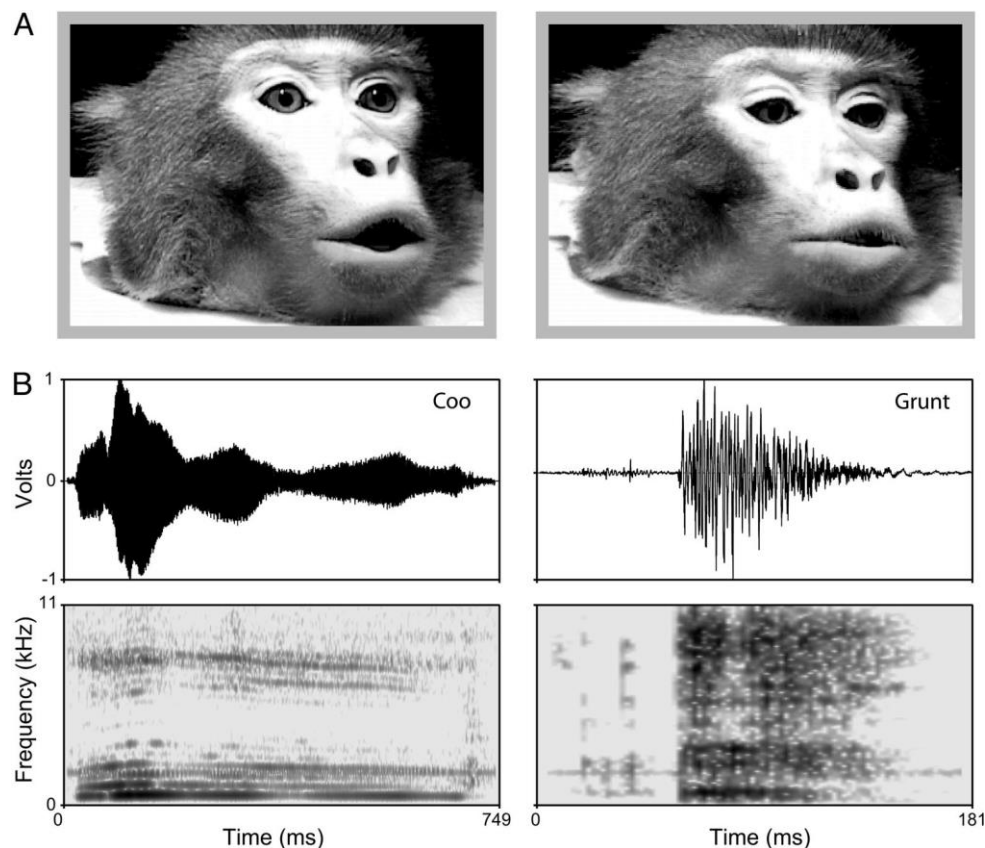


Figure 12. Exemple de stimuli vidéo et audio utilisés dans le test d'intégration multisensorielle de Lewkowicz & Ghazanfar (2006).

A gauche, (A) une capture d'écran de la vidéo d'un singe émettant un « Coo » ainsi que (B haut) l'oscillogramme et (B bas) le spectrogramme correspondant au « Coo ». A droite, (A) une capture d'écran de la vidéo d'un singe émettant un « Grunt » ainsi que (B haut) l'oscillogramme et (B bas) le spectrogramme correspondant au « Grunt ». Durant l'étude, les deux vidéos dont sont tirés les photographies de A sont jouées simultanément, côté à côté, pendant qu'un des deux sons de B est joué en synchronie avec l'ouverture des lèvres. Tiré de Lewkowicz & Ghazanfar (2006).

IV. Implications sur la cognition sociale précoce

Le rétrécissement perceptif n'a cependant pas que des conséquences perceptives ; il va aussi influencer les interactions sociales et la cognition sociale du nourrisson. Plusieurs études se sont intéressées aux relations/interactions qu'ont les nourrissons avec des individus d'un autre type et ou parlant une langue nouvelle.

Une expérience de Kinzler et al. (2007) a montré que des nourrissons de 5 mois regardent plus longtemps le visage d'une personne ayant précédemment parlé leur langue maternelle plutôt que celui d'une personne ayant parlé une langue étrangère. Ces préférences vont à leur tour entraîner des biais et stéréotypes appliqués aux membres exo-groupe. La même étude montre qu'à 10 mois, les nourrissons présentent plus facilement un engagement social avec une personne ayant précédemment parlé leur langue plutôt qu'une langue étrangère. Pour mettre cela en évidence, une locutrice de la langue maternelle du nourrisson et une locutrice d'une langue étrangère sont présentées à tour de rôle via une courte vidéo. Ensuite, les deux locutrices apparaissent côte à côte à l'écran et présentent chacune un jouet (identique) au nourrisson. Puis, les locutrices tendent la main tenant le jouet vers l'écran, celui-ci passant hors-champ ; un expérimentateur caché derrière l'écran fait simultanément apparaître deux jouets, de chaque côté de l'écran, comme s'ils étaient réellement donnés par les locutrices. On observe que les nourrissons vont plus souvent prendre le jouet donné par la locutrice parlant la langue maternelle native, bien que le jouet donné par la locutrice parlant une langue étrangère non-native soit exactement identique. Xiao et al. (2017a) ont récemment évalué la valence que les nourrissons attribuent aux visages de même type et d'autres types. Pour cela, ils ont présenté des séries de six visages statiques entrecoupés de six extraits de musique et ont analysé les temps de fixation des nourrissons. Les visages pouvaient être soit de même type que l'enfant (ici de type asiatique), soit d'autre type (ici de type africain). La musique pouvait être joyeuse ou triste (préalablement jugée par un panel de participants adultes) pour respectivement représenter une valence positive ou négative. Il y avait donc quatre conditions possibles : même

type de visage/musique joyeuse, même type de visage/musique triste, autre type de visage/musique joyeuse, autre type de visage/musique triste. Des nourrissons entre 3 et 12 mois ont passé l'expérience et une seule condition parmi les quatre a été présentée à chaque enfant. Les résultats montrent qu'à partir de 9 mois, les temps de fixation des nourrissons sur les visages sont plus longs dans les conditions « visage de même type/musique joyeuse » et « visage d'autre type/musique triste » que dans les deux autres situations. Il semblerait donc que dès 9 mois, les nourrissons associent les visages de leur type à une valence positive et les visages d'autres types à une valence négative. Cela pourrait être l'indicateur de l'émergence d'un biais racial implicite, pouvant perdurer à l'âge adulte (Baron & Banaji, 2006 ; Dunham et al., 2013). Il est à noter que l'on peut aussi attribuer ces résultats à la familiarité. En effet, les nourrissons sont plus souvent exposés, via les jouets, veilleuses ou dessins animés, à de la musique joyeuse qu'à de la musique triste (82 % des nourrissons avaient déjà été exposés à de la musique joyeuse contre 18 % à de la musique triste, selon les retours des parents de cette étude). Il se pourrait donc que les nourrissons associent les visages qui leur sont familiers avec la musique qui leur est familière et les visages qui leurs sont inhabituels avec la musique qui leur est inhabituelle. Même si l'association entre les visages d'autres types et une valence négative est indirecte, cela ne remet pas en question son existence. Il semble probable que l'attitude de l'entourage du nourrisson soit globalement plutôt positive (de par les expressions faciales : Malatesta & Haviland, 1982 ; ou bien la parole : Kim & Johnson, 2014 ; Trainor et al., 2000) à son égard que plutôt négative. On peut donc supposer que les événements à valence positive provenant de l'entourage direct du nourrisson sont plus courants que ceux à valence négative, renforçant l'association entre visages familiers et événements positifs. Dans une autre étude, Xiao et al. (2017b) se sont intéressés à la « confiance » que pouvaient accorder des nourrissons à des visages de même et d'autres types. Il a été montré que les enfants sont, dès 8 mois, capables de suivre le regard d'un adulte qui prédit des événements futurs (soit l'apparition d'une récompense visuelle dans un coin de l'écran) (Tummeltshammer, et al., 2014 ; Wu & Kirkham, 2010). Ce principe est utilisé dans l'expérience de Xiao et al., les chercheurs ayant présenté sur un écran un visage d'adulte (soit de même type que l'enfant, soit d'un autre type) qui interagissait avec l'enfant puis tournait son regard dans un des quatre coins de l'écran. Une récompense visuelle (un jouet) apparaissait ensuite, 50% du temps dans le coin où le visage avait regardé, 50 % du temps dans un autre coin (l'adulte est donc plus fiable que le niveau de hasard, puisque celui-ci est à 25%). Un léger temps de latence entre le regard de l'adulte et l'apparition de la récompense permettait de voir si l'enfant anticipait l'apparition de la récompense. Les résultats montrent que les enfants de 7 mois suivent le regard de l'adulte uniquement s'il est de même type d'eux. En revanche, si le

regard de l'adulte était fiable à 100 %, les enfants le suivaient quel que soit le type de visage ; s'il était fiable à 25% (pas plus que le niveau de hasard donc), ils ne le suivaient pas de manière significative quel que soit le type de visage. Les nourrissons, dès 7 mois, sont donc plus susceptibles de faire confiance à un adulte de leur type plutôt qu'à un adulte d'un autre type, et ce même si celui-ci n'est pas entièrement fiable. Après le début du rétrécissement perceptif, les nourrissons sont donc plus enclins à traiter plus positivement les individus ressemblants et parlant comme la majorité des individus qu'ils rencontrent que les individus s'écartant de ce schéma. Cela pourrait marquer l'émergence d'un biais racial implicite dès la seconde moitié de la première année de vie, qui pourrait avoir des répercussions jusqu'à l'âge adulte.

V. Moduler le rétrécissement perceptif

Cependant, le rétrécissement perceptif tel que l'on vient de le décrire n'est pas une fatalité. Certaines méthodes expérimentales ou encore la diversité de l'environnement dans lequel le nourrisson évolue se sont montrées être capables de moduler l'impact du rétrécissement perceptif. De plus, si le rétrécissement perceptif est observable dès la seconde moitié de la 1^{ère} année de vie, le système perceptif durant l'enfance reste encore assez plastique pour contrebalancer ces effets. Même chez l'adulte, des améliorations peuvent encore être observées avec un entraînement adapté.

V.1. Par l'expérimentation

V.1.a. *Après l'émergence du rétrécissement perceptif*

L'ORE observé pour les visages peut être réduit par un simple entraînement à reconnaître des visages d'autres types. Ainsi, des participants adultes de type européen entraînés à mémoriser des photos de visages d'un autre type et testés immédiatement après présentent une amélioration des performances comparativement à un groupe n'ayant pas reçu d'entraînement. En revanche, si les participants sont entraînés sur des visages européens, cela n'améliore pas leur reconnaissance de visages européens par rapport à un groupe sans entraînement (Elliott et al., 1973). L'efficacité de ce type d'entraînement a montré une certaine stabilité dans le temps. Des participants adultes testés 5 mois après entraînement (6 sessions de 30-45 minutes sur 2-3 semaines consécutives) gardent les bénéfices de leur entraînement (Goldstein & Chance, 1985). Nous avons précédemment vu que l'ORE entraîne les nourrissons (et par la suite l'enfant et l'adulte) à présenter une préférence pour les individus de leur type. Lebrecht et al. (2009) ont montré que ce biais racial implicite était lui aussi atténué après un entraînement visant à mieux individualiser des visages d'autres types. Des participants adultes de type européens ont en effet

été entraînés à faire la distinction entre différents types de visages afro-américains. Leur biais racial implicite a été mesuré avant et après entraînement par un test d'association implicite, l'« *Affective Lexical Priming Score* » (score d'amorçage lexical affectif). Les résultats montrent qu'avant entraînement, les participants de type européen mettent plus de temps à associer des mots positifs à des visages afro-américain qu'à leur associer des mots négatifs. Après entraînement, les participants sont non seulement plus efficaces pour individualiser des visages afro-américains mais ils montrent aussi une réduction de leur biais racial implicite. Cette réduction de l'ORE et du biais implicite est positivement corrélée par participant. On comprendra alors l'enjeu de réduire au plus tôt l'ORE pour réduire l'impact du biais racial implicite. Xiao et al. (2015) ont testé des enfants chinois de 4 à 5 ans sur une tâche quelque peu similaire. Un visage joyeux ou en colère de type ambigu (morphing 50% entre un visage de type chinois et un visage de type africain) était présenté aux enfants qui devaient le catégoriser soit comme chinois soit comme africain. Avant entraînement, les enfants chinois ont tendance à catégoriser les visages joyeux comme chinois et les visages en colère comme africains. Après un entraînement visant à individualiser des visages de type africain, non seulement l'ORE était amoindri mais de plus le biais racial était moins prononcé, montrant que celui-ci peut être réduit par l'entraînement durant l'enfance.

V.1.b. Avant l'émergence du rétrécissement perceptif

Nous avons vu que l'ORE se développe entre 6 et 9 mois et que l'on peut atténuer l'ORE et l'émergence subséquente du biais racial implicite chez l'enfant et l'adulte. On peut donc se demander ce qui se passerait si des méthodes de ce type étaient appliquées chez le nourrisson. Une première étude de ce type, menée par Pascalis et al. (2005) a montré qu'entre 6 et 9 mois, une exposition hebdomadaire à des visages de singes (via des photographies) pouvait maintenir la capacité des nourrissons à reconnaître ces visages à 9 mois. De la même manière, Heron-Delaney et al. (2011) ont testés les compétences de nourrissons de 9 mois à reconnaître des visages d'autres types (ici des visages chinois) après une exposition régulière à des visages d'autres types (via un livre de photographies, à la maison, plusieurs fois par semaine pendant 3 mois). Bien que les nourrissons de type européen soient habituellement incapables d'individualiser des visages d'autres types à cet âge et selon ce type de protocole, les nourrissons de 9 mois de cette étude ont été capables de différencier de nouveaux visages chinois entre eux. Le groupe contrôle qui a été exposé artificiellement à des visages européens présentent un ORE de manière classique. Spangler et al. (2013) ont fait une étude longitudinale avec des nourrissons passant l'expérience à 3, 6 puis 9 mois afin de mettre en évidence l'effet

d'exposition sur l'ORE. Ils ont pour cela utilisé un protocole avec une habitude durant laquelle ils présentaient un visage d'un autre type. Cette habitude était suivie d'une phase de déshabitude avec un nouveau visage d'un autre type. Les nourrissons de 9 mois de leur étude n'ont pas présenté l'ORE classique, contrairement aux nourrissons du groupe contrôle, qui ne passaient l'expérience qu'une seule fois à 9 mois. Cela vient confirmer l'idée selon laquelle le rétrécissement perceptif est causé par le manque d'exposition aux stimuli non-natifs. A noter toutefois que leur phase d'habitude était particulièrement longue (pouvant aller jusqu'à 18 essais de 40 secondes). Si une simple exposition aux visages d'autres types suffit pour maintenir les capacités de discrimination chez les nourrissons, cette exposition n'est plus suffisante chez les jeunes enfants et il faut bien un entraînement pour noter une amélioration des performances d'individualisation et une baisse du biais racial (Qian et al., 2017, les enfants de l'étude étant en maternelle). D'autres moyens expérimentaux que la présentation de visages d'autres types permettent de restaurer la sensibilité aux stimuli non-natifs. Fair et al. (2012) ont répliqué les études de Pascalis et al. (2002, 2005) mais en allongeant les durées de familiarisation et de test. Ainsi, en passant d'une familiarisation de 20 secondes à une familiarisation de 40 secondes et en passant de tests de 5 secondes à des tests de 10 secondes, les nourrissons de 12 mois s'avèrent toujours capables de discriminer les visages de macaques. Les compétences de discrimination des stimuli non-natifs diminuent donc au cours du développement (puisque plus de temps est nécessaire pour la reconnaissance) mais ne sont pas complètement perdues. Markant et al. (2016) ont montré que la reconnaissance de visages d'autres types pouvait être restaurée chez des nourrissons de 9 mois en augmentant l'attention visuelle des nourrissons vers les visages. Pour ce faire, les auteurs ont créé un biais attentionnel en amont de la tâche principale. Les auteurs ont créé ce biais attentionnel soit vers le côté où le visage allait apparaître lors de la familiarisation (condition congruente), soit du côté opposé (condition incongruente), à l'aide de cible apparaissant sur l'écran. Puis, ils ont présenté des visages d'un autre type durant la phase de familiarisation. Lors d'un test de préférence à la nouveauté consécutif à la familiarisation, seul les nourrissons dont l'attention était dirigée vers les visages durant la familiarisation ont plus regardés les nouveaux visages. Cette étude suggère que diriger l'attention visuelle des nourrissons vers un visage d'autre type peut permettre de les aider à le reconnaître. La reconnaissance de visage d'autres types peut aussi être restaurée chez des nourrissons de plus de 6 mois par l'ajout d'émotion, comme l'ont démontré Quinn et al. (2020). Des visages africains ou asiatiques avec des expressions de colère ou de joie ont été présentés à des nourrissons de type européen lors de familiarisations de 30 secondes puis leur reconnaissance a été testée lors de deux tests de 10 secondes. Les nourrissons ont passé plus de

temps à regarder le nouveau visage, signifiant que le visage de la phase de familiarisation a bien été appris. La plus forte saillance ou bien une meilleure distinguabilité des visages avec émotion par rapport aux visages neutres communément utilisés pourrait permettre d'expliquer ces résultats.

Pour ce qui est du langage, une exposition même à court terme de nourrissons à des langues étrangères leur permet de retrouver leur capacité à distinguer des phonèmes non-natifs. Kuhl et al. (2003) ont exposé des nourrissons américains de 9 mois à du mandarin pendant 12 sessions de 25 minutes réparties sur 4 semaines. Durant ces sessions, un locuteur natif du mandarin lisait des histoires pour enfants et jouait avec le nourrisson. Lors d'un test de discrimination phonétique de contrastes mandarins, les nourrissons américains exposés au mandarin ont présenté des résultats similaires à ceux des nourrissons dont la langue maternelle est le mandarin. En revanche, des nourrissons américains exposés au mandarin mais sans interaction (via des vidéos) ne présentent pas d'amélioration des performances. Toutefois, si une interaction avec un locuteur semble nécessaire lors de l'exposition au langage, les nourrissons peuvent aussi retrouver la sensibilité aux contrastes phonémiques avec un autre type d'engagement qu'une interaction sociale. En effet, Yeung & Werker, (2009), ont démontré que par simple association de phonèmes avec des objets, les nourrissons peuvent à nouveau présenter une sensibilité aux phonèmes non-natifs. Ce rehaussement de la sensibilité n'est en revanche observé que si l'association est constante (toujours le même phonème avec le même objet) et ne l'est pas dès lors que l'association est inconstante.

V.1. Par l'environnement

Les expérimentations en laboratoire montrent que l'exposition durant l'enfance joue un rôle primordial sur le développement du rétrécissement perceptif. Les visages couramment rencontrés dans la vie de tous les jours semblent modeler les capacités perceptives des enfants. On peut donc s'attendre à ce que des enfants grandissant dans un environnement hautement multiculturel, avec des personnes de différents types, parlant différentes langues puissent garder des capacités de discrimination beaucoup plus fines que des enfants grandissant dans des environnements très homogènes.

V.1.a. Par l'exposition à différents types de visages

Les données recueillies dans des milieux hétérogènes (Malpass, 1981 ; Hancock & Rhodes, 2008 ; Feinman & Entwisle, 1976 ; McKone et al., 2007 ; Meissner & Brigham, 2001) ont amenés certains auteurs à proposer l'hypothèse de contact. Celle-ci stipule qu'un manque

d'interaction avec des personnes d'autres types résulte en une diminution des capacités de reconnaissance des personnes de ce type. Globalement, les études portant sur l'hypothèse de contact semblent montrer que la fréquence de l'exposition à des personnes d'autres types conditionne l'ORE ; une exposition passagère est moins efficace qu'une exposition permanente ou que des interactions avec ce type de personnes pour réduire cet effet. En effet, on peut supposer que si des stimuli (ici des visages d'autres types) ne sont pas présents ou ne le sont que très peu dans l'environnement, il n'apparaît pas pertinent de développer des capacités pour les traiter, tout comme il n'apparaît pas nécessaire de distinguer les consonnes rétroflexe [ɖ] et dentale [t] du hindi si elles ne sont pas présentes dans la langue ambiante. Allant dans le sens de cette hypothèse, Carroo (1986) a testé l'ORE pour des personnes de type africain ou afro-américain différant dans leur degré d'exposition à des visages blancs américains (les personnes de type africain étant nigérianes et ayant vécu la plupart de leur vie dans des milieux ne présentant que très peu de personnes blanches). Les résultats montrent que les nigériens ont plus de mal que les afro-américains à reconnaître les visages d'un autre type. De plus, parmi les afro-américains, ceux ayant de meilleures capacités de discrimination des visages blancs sont ceux ayant des amis proches blancs, soulignant l'importance de l'exposition aux visages d'autres types en termes qualitatifs et non uniquement quantitatifs afin de réduire l'ORE. Li et al. (1998) ont montré que les personnes de type européen fan de basketball présentent un ORE moins fort que les non-initiés au basketball. Les joueurs de basketball étant majoritairement de type afro-américain, il apparaît donc pertinent pour les supporters d'identifier facilement les joueurs. Toutefois, les données de la littérature ne sont pas complètement unanimes concernant l'hypothèse de contact. Malpass & Kravitz (1969) ne trouvent aucune corrélation entre l'expérience rapportée des sujets vis-à-vis de personnes d'autres types et leurs capacités effectives de reconnaissance de visages d'autres types. Une étude similaire de Brigham & Barkowitz (1978) ne trouvent notamment aucun lien entre la quantité et la qualité des interactions avec des individus d'autres types et la capacité à reconnaître des photos d'individus d'autres types. De même, Ng et Lindsay (1994) ont testé des étudiants du Canada et de Singapour sur leur capacité à reconnaître des visages d'autres types et ont trouvé que leur exposition à des visages d'un autre type n'expliquait qu'à 0.89% la variance des résultats. Il est à noter de façon intéressante que la plupart des études ne trouvant pas de lien entre l'exposition et l'ORE remontent aux années 1970 voire 1960 tandis que les études plus récentes ont plutôt tendance à mettre en évidence l'existence de ce lien. Il se peut que l'évolution de la société provoque plus de mixité entre les individus de différents types, amenant à des relations plus qualitatives, plus susceptibles de réduire l'ORE (cela apparaît d'autant plus plausible que la

plupart des études portent sur la population américaine, qui a connu le mouvement américain des droits civiques dans les années 60 ayant eu pour conséquence une société moins ségrégationniste). Une méta-analyse par Meissner et Brigham (2001) portant sur 39 recherches différentes et presque 5000 participants met en évidence un faible mais significatif effet de l'exposition, qui expliquerait 2% de l'ORE.

Une raison pouvant expliquer la variabilité des résultats concernant l'influence de l'exposition pourrait être l'âge auquel les individus sont exposés à des individus d'autres types. Une exposition à l'âge adulte pourrait avoir un impact réduit voire nul tandis que la même exposition durant l'enfance pourrait notablement et durablement réduire l'ORE. Feinman & Entwisle (1976) figurent parmi les premiers à avoir étudié l'ORE chez les enfants. Leur étude montre que les enfants fréquentant des écoles avec des enfants de différents types (enfants afro-américains et américains blancs) présentent un ORE moins fort que celui des enfants allant dans des écoles avec des enfants très majoritairement du même type qu'eux. En revanche, cette réduction n'était observée que pour les enfants qui de surcroît vivaient dans un quartier présentant aussi différents types. Plus récemment, Zhou et al. (2019) ont testé des adultes de type asiatique nés dans un milieu majoritairement blanc (au Canada) sur leur reconnaissance de visages blancs, et leurs performances ont été comparées à celles des adultes de type asiatiques ayant immigré dans un pays majoritairement peuplé par des blancs. Les résultats montrent que les adultes de type asiatique nés et ayant grandi dans un pays dont la population est majoritairement blanche ne présentent pas l'ORE pour les visages blancs. Pour les adultes ayant immigré, les auteurs ont trouvé que la force de l'ORE dépend de l'âge d'arrivée dans le nouveau pays ; plus ils ont immigré tôt, plus l'ORE sera faible. De récentes études se sont donc intéressées à des enfants ayant toujours grandi dans des milieux très hétérogènes. Tham et al. (2017) ont comparé l'intensité de l'ORE d'enfants de 5-6 ans ayant grandi dans différents milieux en termes d'homogénéité de type de visages : des enfants britanniques de type européen et grandissant dans des milieux où leur type est majoritaire et des enfants malaisiens de Kuala Lumpur qui sont exposés à différents types de visage (de type malais, chinois et indien). La reconnaissance de visages de type malais, européens, chinois et africains a été testée pour ces deux groupes d'enfants. Alors que les enfants britanniques présentent l'ORE pour les visages d'un autre type que le leur, les enfants malaisiens ne présentent de reconnaissance amoindrie ni pour les visages de type chinois ni pour les visages de type européen. Les auteurs trouvent cependant que la reconnaissance des visages africains par les enfants malaisiens est légèrement moins bonne, mais quand même supérieure à celle des visages africains par les enfants

européens. En revanche, la reconnaissance de visage de même type est meilleure pour les enfants européens que pour les enfants malaisiens. Plusieurs points importants sont à retirer de ces résultats. Premièrement, les enfants malaisiens grandissent dans un environnement où ils sont quotidiennement exposés à des visages de type malais et chinois, mais ils n'ont que peu d'exposition aux visages de type européen (potentiellement un peu via la télévision) et très peu aux visages de type africain. Cependant, les compétences des enfants malaisiens sont meilleures pour tous les types de visages, pas seulement ceux auxquels ils sont exposés. L'exposition à plus d'un type de visage permettrait donc de maintenir des compétences de reconnaissance pour tous les types de visages, même ceux non présents dans l'environnement dans l'enfant. Deuxièmement, les meilleures performances des enfants de type européens pour la reconnaissance de visages de même type qu'eux semblent nous indiquer que le rétrécissement perceptif est bien accompagné d'une amélioration des compétences pour les stimuli natifs. Inversement, un système perceptif plus généraliste est lié à une moins bonne expertise spécifique pour les visages de même type. Tham et al. (2019) se sont aussi intéressés aux nourrissons de type chinois grandissant en Malaisie, montrant que ceux-ci, dès 8 mois, sont capables de reconnaître non seulement des visages féminins chinois mais aussi malais. Leurs capacités pour reconnaître les visages malais est meilleure à 8 mois qu'à 3 mois, suggérant bien un effet de l'expérience. Des recherches sur des enfants adoptés viennent confirmer l'importance de l'âge sur l'effet d'exposition. Sangrigoli et al. (2005) ont montré que des adultes coréens adoptés en France pendant leur jeunesse (ayant été adoptés entre 3 et 9 ans) présentent un traitement des visages similaire à celui des adultes nés et élevés en France (c'est-à-dire qu'ils sont meilleures pour faire la distinction entre des visages de type européen qu'entre des visages de type asiatique). Une autre étude de De Heering et al. (2010) révèle aussi que les enfants adoptés (ayant été adoptés entre 2 et 26 mois et âgés d'entre 6 et 14 ans au moment de l'étude) semblent avoir des capacités de reconnaissance meilleure pour les visages de leur nouvel environnement. En revanche, les enfants de cette étude gardent aussi des capacités de reconnaissance pour les visages de leur type contrairement aux sujets de l'étude de Sangrioli et al. Un critère d'explication possible de ces résultats peut provenir du fait que les participants de l'étude de Sangrioli et al. sont des adultes tandis que ce sont des enfants qui ont été testés par de Heering et al. L'expérience et le développement cérébral ne sont donc pas comparables entre ces deux populations, les enfants pouvant graduellement perdre leurs capacités à reconnaître des visages de leur type avec le temps (par manque d'exposition à ceux-ci et par constante exposition aux visages d'autres types).

V.1.b. Par l'exposition à différentes langues

On pourrait légitimement penser que les nourrissons bilingues font face à un défi difficile, devant apprendre le langage en étant exposés à deux langues, donc avec moins d'exposition totale à chaque langue et avec la difficulté supplémentaire de devoir dissocier les deux flux langagiers (Byers-Heinlein, 2013). C'est pourtant de manière tout à fait naturelle que se fait l'acquisition du langage chez les enfants bilingues qui démontrent pour leurs deux langues les capacités que les monolingues présentent pour une seule langue (Burns et al., 2007 ; Sundara et al., 2008 ; Albareda-Castellot et al., 2011 ; mais voir Bosch & Sebastian-Galles, 2003). Tout comme les nouveau-nés monolingues présentent une préférence pour leur langue maternelle (Moon et al., 1993), les nouveau-nés bilingues présentent une préférence pour leurs deux langues maternelles à la naissance si celles-ci sont de deux classes rythmiques différentes (Byers-Heinlein et al., 2010). La rythmicité est le premier élément de leur langue avec lequel les nourrissons sont familiarisés et qui permettra de la reconnaître (Nazzi et al., 1998), cette information basse fréquence étant probablement perçue *in utero*. Cette préférence montre que les nourrissons bilingues ont déjà, *in utero*, perçus les rythmicités différentes de leurs deux langues. D'un point de vue neuronal, les nourrissons bilingues présentent une trajectoire développementale différente des monolingues lors de l'écoute de phonème de leurs langues, comme le montre Garcia-Sierra et al. (2011) via un paradigme de type *oddball*. Dans ce type de paradigme, un stimulus appelé stimulus « standard », est présenté de nombreuses fois. Un ou des intrus, appelé les stimuli « déviants », sont présentés aléatoirement lors de la présentation du stimulus standard, en nombre bien inférieur. Si le stimulus déviant est distingué (du moins d'un point de vue neuronal), on observe une réponse ERP caractéristique, appelée « *mismatch negativity* ». Plus précisément, les auteurs ont ici choisi un phonème espagnol/anglais (entendu /ta/ en espagnol et /da/ en anglais) comme stimulus standard, présenté 80% du temps. Un phonème déviant espagnol (entendu /da/) était présenté 10% du temps, et un phonème déviant anglais (prononcé /ta/) les 10% restants. Les nourrissons bilingues espagnol-anglais de 6-9 mois ne présente pas de réponse ERP indiquant une discrimination des stimuli lors de l'écoute de ces phonèmes. Les nourrissons monolingues anglais du même âge présentent une réponse neuronale indiquant la discrimination des phonèmes, qu'ils soient natifs ou non (Rivera-Gaxiola et al., 2005). En revanche, à 10-12 mois, la réponse neuronale des nourrissons bilingues indiquent une discrimination des phonèmes, tandis que les nourrissons monolingues du même âge ne présentent plus de réponse indiquant une discrimination pour les phonèmes non-natifs. Une explication est que l'environnement langagier des nourrissons bilingues

présente une plus grande variabilité que celui des monolingues, amenant à une spécialisation moins rapide aux sons de leurs langues. À la vue de ces observations et concernant notre problématique du rétrécissement perceptif, on peut donc se demander si le bilinguisme précoce affecte la perception des sons des langues non-natives, i.e. des langues qui ne sont pas dans l'environnement quotidien des nourrissons. Nous venons de voir que des enfants exposés à plusieurs types de visages gardent des capacités de distinction des visages semblant s'étendre à des types de visages auxquels ils ne sont pourtant pas exposés dans leurs milieux (Tham et al., 2017). Des auteurs avancent en effet que les nourrissons bilingues pourraient présenter un rétrécissement perceptif langagier moins strict que les nourrissons monolingues (Byers-Heinlein, 2013). Petitto et al. (2012) ont utilisé la technique d'imagerie spectroscopique proche infrarouge fonctionnelle (ISPIf) pour observer l'activité neuronale de nourrissons de 4-6 mois et 10-12 mois, monolingues ou bilingues, lors de l'écoute de phonèmes natifs et non-natifs. Leurs résultats montrent une activation du gyrus temporal supérieur gauche (STG) et du gyrus frontal inférieur gauche (IFG) chez les monolingues et les bilingues de 4-6 mois, que ce soit lors de l'écoute de phonèmes natifs ou non-natifs. Le STG est une région habituellement activée chez l'adulte lors de tâches de discrimination de phonèmes tandis que l'IFG est plutôt activé lors de la récupération d'informations sémantiques, syntaxiques et phonologique (Burton & Small, 2006). A 10-12 mois, l'IFG reste activé lors de l'écoute de phonèmes natifs chez les nourrissons monolingues et bilingues mais n'est activé pour les phonèmes non-natifs que chez les nourrissons bilingues. L'étude ne nous permet pas de savoir si cette différence de traitement perdure dans le temps, ou bien s'il s'agit juste d'un délai développemental. De plus, si cela montre que le cerveau des bilingues garde une certaine souplesse quant au traitement des phonèmes non-natifs, cela ne nous renseigne pas sur d'éventuels implications comportementales. Dans le but de vérifier si les nourrissons bilingues peuvent discriminer des phonèmes non-natifs à un âge plus avancé que les nourrissons monolingues, Singh et al. (2017) ont testé des nourrissons monolingues (anglais) et bilingues (anglais-mandarin) sur une tâche de discrimination de phonème natifs et non-natifs. Alors que conformément aux données de la littérature, les nourrissons monolingues de 11.5 mois ne distinguent plus les phonèmes de l'hindi [t] et [t̪], les bilingues du même âge en restent capables.

L'effet du bilinguisme semble aussi impacter la perception visuelle du langage. En effet, Weikum et al. (2007) ont testé la capacité de nourrissons monolingues et bilingues à distinguer des langues différentes uniquement via la modalité visuelle. Pour ce faire, ils ont habitué des nourrissons monolingues anglais et des nourrissons bilingues anglais/français à des vidéos sans

son de locutrices bilingues anglais-français, produisant des phrases chacune dans une langue. Puis, une fois le critère d'habitation atteint, les vidéos silencieuses des mêmes locutrices parlant cette fois la 2^{ème} langue étaient présentées. Les résultats montrent que les nourrissons sont capables de détecter visuellement le changement de langue à 4 et 6 mois. En revanche, et conformément aux données sur le rétrécissement perceptif décrites au chapitre précédent, les nourrissons monolingues n'en sont plus capables à 8 mois. En revanche, les nourrissons bilingues anglais/français restent capables de détecter visuellement le changement de langue à 8 mois. Sebastián-Gallés et al. (2012) ont de plus montré que cette capacité n'est pas limitée aux langues des enfants bilingues ; ainsi, des nourrissons bilingues espagnol/catalan de 8 mois sont capables de distinguer un changement entre français et anglais dans la modalité visuelle seule. Ces études nous montrent que l'exposition à plusieurs langues ne module pas uniquement la perception du langage mais aussi la façon dont l'information audio-visuelle est extraite des visages, ce qui interroge sur l'étendue des liens entre traitement des visages et traitement du langage durant le développement.

Dans cet optique, Singh et al., 2017, en plus de tester l'impact du bilinguisme sur la discrimination de phonèmes non-natifs, se sont intéressés à son influence sur l'émergence du rétrécissement perceptif observé pour les visages, avec l'hypothèse sous-jacente que le rétrécissement perceptif pourrait être affecté de manière générale par l'exposition à une seconde langue. Les résultats montrent cependant que l'ORE que présentent les nourrissons monolingues et bilingues à 11.5 mois est équivalent, la capacité de discrimination de visages étant limitée aux visages de même type. Cependant, Kandel et al. (2016) montrent que des adultes bilingues (exposés durant leur enfance à une deuxième langue) ne présentent pas de moins bonnes capacités de discrimination pour les visages d'autres types que pour les visages de même type. Il semblerait que la façon dont les bilingues traitent les visages, indépendamment de leur type, soit affectée : leur temps de réaction pour traiter les visages sont plus généralement longs que ceux d'adultes monolingues. La façon dont l'exposition à une autre langue durant l'enfance module le rétrécissement perceptif et le traitement des visages reste cependant à éclaircir. Il n'est en effet pas clair si ce qui est observé pour le langage chez les nourrissons bilingues est une différence de traitement persistant à l'âge adulte ou bien simplement un retard dans l'émergence du rétrécissement perceptif. De même, s'il semble que le traitement des visages d'autres types soit toujours affecté chez les nourrissons bilingues, il se peut que la capacité de discrimination de ces visages réapparaisse au cours du développement, lorsque plus d'expérience langagière est acquise. Des études testant des nourrissons bilingues sur d'autres

contrastes phonémiques et à d'autres âges permettraient une vision plus claire de l'impact du bilinguisme sur l'émergence du rétrécissement perceptif.

VI. La nature du rétrécissement perceptif

VI.1. Un phénomène inter-sensoriel

Ces découvertes nous questionnent sur la nature et les origines du rétrécissement perceptif. Si nous venons de voir que le rétrécissement perceptif est observé pour le langage parlé, pour les visages et pour l'intégration-audiovisuelle, les données de la littérature nous montrent que ce phénomène peut être mis en évidence pour d'autres capacités perceptives. Une étude portant sur la langue des signes par Palmer et al. (2012) montre que des nourrissons normo-entendants peuvent discriminer des signes de la langue des signes américaine (ASL) à 4 mois, mais que seuls les nourrissons apprenant l'ASL en sont encore capables à 14 mois. Le rétrécissement perceptif observé pour le langage existe donc quelle que soit la modalité (audio ou visuelle) du langage. De plus, il semblerait que le rétrécissement perceptif ne se limite pas au traitement des visages et du langage. En effet, Hannon & Trehub (2005a, b) ont démontré que la perception de la musique était aussi soumise à ce phénomène. Certains aspects de la musique, notamment le rythme ou les intervalles mélodiques, ont des règles qui dépendent de leurs origines culturelles. Dans le cadre de cette étude, des musiques avec un rythme courant dans la musique occidentale et des musiques avec des rythmes usuels de la musique bulgare ont été présentées à des nourrissons nord-américains de 6 et 12 mois. Une fois familiarisé à un type de rythme, une musique violant la règle rythmique de la familiarisation et une musique conservant cette règle étaient jouées à tour de rôle. Les nourrissons de 6 mois ont présenté une préférence pour la nouveauté, i.e. pour la musique violant la règle rythmique avec laquelle ils avaient été familiarisés, quelle que soit la musique de familiarisation. Les nourrissons de 12 mois ne présentaient en revanche une préférence à la nouveauté que lorsque la musique de familiarisation était occidentale. Les nourrissons de 6 mois sont donc à la fois sensibles aux rythmes occidentaux et à ceux d'autres cultures (ici bulgare) tandis qu'à 12 mois, ils sont uniquement sensibles aux rythmes de leur culture.

On retrouve des signes qui laisse penser que le rétrécissement perceptif pourrait aussi s'observer pour le goût et l'olfaction. Haller et al. (1999) ont évalué chez des participants adultes la préférence pour un goût en fonction de leur exposition à ce goût durant l'enfance. Une partie des participants adultes avait été nourris au lait vanillé durant leur petite enfance tandis que l'autre partie avait été nourris au sein. Les chercheurs ont fait goûter aux participants deux

ketchups différents : un nature et un très légèrement vanillé (normalement à un seuil difficilement identifiable par les participants). La proportion de participants préférant le ketchup vanillé était plus grande au sein des participants ayant été exposé au lait vanillé durant la petite enfance qu'au sein des participants ayant été nourris au sein. Poncelet et al. (2010) ont de plus montré que la réponse neuronale suite à la présentation d'une certaine odeur n'est pas la même en fonction de si les participants ont été précocement exposé ou non à cette odeur. En effet, des français issus de l'immigration algérienne présentent une réponse ERP différente de français d'origine européenne lors d'une exposition à une odeur de menthe, odeur plus présente (ce fait étant confirmé par le questionnaire de l'étude) dans l'environnement des français issus de l'immigration algérienne. L'effet de préférence selon l'exposition précoce à un goût ou une odeur se retrouve en fait très tôt dans le développement. Schaal et al. (2000) ont étudié la réaction de nouveau-nés (par les mouvements de têtes et les mimiques faciales) lors de la présentation d'une odeur d'anis. Une partie des mères avait consommé des aliments anisés durant la grossesse, tandis que l'autre partie les avait évités. Les résultats montrent que les nouveau-nés ayant été exposé *in utero* à l'anéthol (la substance donnant l'odeur caractéristique des aliments anisés) montrent plus de réactions positives lors de la présentation de l'odeur d'anis que les nouveau-nés n'ayant pas été exposé à cette substance *in utero*. Pour l'olfaction et le goût, on retrouve donc des préférences qui nous rappellent ce qui est observé pour les visages ou pour les langues : les types de visages et les langues les plus présentes dans l'environnement du nourrisson sont préférés par celui-ci. Le goût est un sens stimulé de manière très précoce *in utero*, ce qui peut expliquer que l'on observe des préférences très précocement, dès la naissance. Il sera intéressant de voir si, comme pour les langues ou les visages, des capacités plus fines pour distinguer les goûts ou les odeurs présents dans l'enfance se développent par la suite. A notre connaissance, aucune étude à ce jour ne porte sur l'existence possible d'un rétrécissement perceptif pour le toucher, mais il est possible qu'un phénomène similaire puisse se retrouver dans chaque modalité. Ainsi, de nombreuses études montrent que la perception des nourrissons évolue d'un système général et non-spécifique vers un système qui est spécifique à ce que l'enfant voit, entend et peut-être sent dans son environnement. Le rétrécissement perceptif est un phénomène observé à la fois pour le traitement des visages, de la langue orale, de la langue des signes et de la musique, avec une temporalité assez proche, puisque survenant toujours vers la moitié de la première année de vie. On peut donc se demander si le rétrécissement perceptif observé pour chaque modalité est indépendant – différents processus opèrent en parallèle avec une temporalité assez similaire – ou bien si le rétrécissement perceptif est un phénomène global touchant de manière générale les capacités

perceptives des nourrissons durant la première année de vie. Des articles récents sur le rétrécissement perceptif appuient cette deuxième idée. Krasotkina et al. (2018) ont testé des nourrissons de 9 mois sur une tâche de discrimination de visages d'un autre type et sur une tâche de discrimination de phonèmes non-natifs. Les résultats montrent que les scores des nourrissons sont positivement corrélés entre les deux tâches de discrimination. Autrement dit, le rétrécissement perceptif observé pour le traitement des visages est positivement corrélé au rétrécissement perceptif observé pour le langage. Cette observation n'est pas dû au fait que certains nourrissons présentent de meilleures capacités de discrimination ; en effet, le score de discrimination pour les stimuli non-natifs n'est pas corrélé au score de discrimination pour les stimuli natifs. Xiao et al. (2018) ont de la même manière testé les capacités de nourrissons à distinguer des visages de leur propre type et d'un autre type ainsi que des phonèmes natifs et non-natifs. Xiao et al. trouvent eux aussi une corrélation entre le rétrécissement perceptif observé pour le traitement des visages et celui observé pour le traitement du langage. Cependant, la corrélation entre les deux rétrécissements perceptifs observée dans leur étude est négative à 9 mois, et devient positive à 12 mois. Bien que le rétrécissement perceptif soit un phénomène universel, il a été montré que de subtiles différences de temporalités peuvent exister entre les différentes populations étudiées (Kelly et al., 2009). Cela pourrait peut-être expliquer ce décalage, les nourrissons testés par Krasotkina et al. étant allemands et ceux testés par Xiao et al. étant japonais. Xiao et al. expliquent la corrélation négative entre le score de la tâche de discrimination des visages et le score de la tâche de discrimination des phonèmes trouvée à 9 mois par une compétition entre les deux systèmes de traitement (des visages et du langage). Toutefois, la corrélation (négative ou positive) entre les deux rétrécissements perceptifs peut suggérer qu'un processus d'ordre général est responsable de la trajectoire développementale du traitement des visages et du traitement du langage. Enfin, Krasotkina et al. (2021) ont conduit une nouvelle étude sur des nourrissons à 6 et 9 mois en utilisant le même type de tâches de discrimination de phonèmes non-natifs mais ne retrouvent pas cette corrélation entre les scores des nourrissons aux deux tâches. En revanche, l'étude de Krasotkina et al. (2021) est longitudinale et montre que les mêmes nourrissons qui discriminaient les stimuli non-natifs à 6 mois n'en sont plus capables à 9 mois. Cette étude nuance les résultats précédemment obtenus par Krasotkina et al. (2018) et Xiao et al. (2018) et met en valeur l'importance des variations idiosyncratiques dans la temporalité du rétrécissement perceptif. Pascalis et al. (2014) ont émis l'hypothèse que la trajectoire du traitement des visages et du traitement du langage est liée par un même but : permettre aux nourrissons de rapidement communiquer avec leurs partenaires sociaux. Ce but implique le développement des

compétences de communication (compréhension du langage, des émotions etc) mais aussi des compétences d'identification des partenaires sociaux (qui peut passer par le type de langue qu'ils parlent, leur type de visage, ou bien leur culture qui se retrouve dans la musique, l'alimentation, etc.). Cette finalité commune nous amène à supposer qu'un mécanisme général pourrait être responsable des rétrécissements perceptifs.

VI.2. Un processus non spécifique à l'Humain

Notre espèce n'est pas la seule pour laquelle s'adapter rapidement à son groupe social est essentiel. On retrouve des trajectoires développementales proches de celle observée chez l'humain au sein d'autres espèces animales sociales. Nous pouvons par exemple faire un rapprochement entre la formation du groupe social chez les rats et chez les humains. Les rats vivent généralement en groupes d'individus allant de 20 à 200 individus et peuvent faire preuve d'une grande hostilité envers les membres de l'exogroupe alors qu'ils présentent un comportement pro-social entre les membres du groupe (Lore et Flannelly, 1977). Bartal et al. (2014) ont étudié les mécanismes conduisant à l'affiliation à un groupe chez les rats issus de différentes lignées. Les lignées de rats diffèrent par leur apparence physique, mais les rats utilisent probablement aussi des signaux olfactifs pour les distinguer. Bartal et al. ont placé dans une même cage un rat piégé (le piège étant situé au centre de la cage et ne pouvant être ouvert que de l'extérieur) et un rat libre de ses mouvements (rat que nous nommerons « assistant »). Selon la condition, le rat piégé était de la même ou d'une autre lignée que le rat assistant. De plus, les deux rats pouvaient être issus de la même cage ou bien de cages séparées, donc être familier l'un de l'autre ou être des étrangers. Les résultats de cette étude ont montré que lorsque les deux rats sont issus de la même lignée, le rat assistant libère l'autre rat, qu'il soit un compagnon de cage ou non. Lorsque les rats sont issus de lignée différentes, le rat assistant aide l'autre rat uniquement lorsqu'il est issu de la même cage que lui. Cependant, si le rat assistant a été logé pendant un temps suffisant (2 semaines dans le cadre de l'étude) avec des rats de la lignée du rat piégé, le rat assistant libère alors le rat piégé, qu'il soit issu de la même cage ou non que lui. A noter que les rats de l'étude n'ont que quelques semaines ; il n'est pas sûr qu'une exposition à plus tardive à des individus d'une autre lignée permettrait de généraliser l'appartenance à tous les individus de cette lignée. Enfin, si le rat assistant a été élevé uniquement au sein de rats de l'autre lignée et n'a jamais été mis en contact avec des rats de sa propre lignée, il aide les rats de sa lignée d'adoption, qu'ils soient étrangers ou familiers. En revanche, il n'aide pas les rats étrangers de sa propre lignée. Par conséquent, chez le rat, l'affiliation à un groupe social n'est pas innée mais dépendante de l'expérience. L'appartenance

à un groupe social est basée sur les individus présents dans l'environnement du rat pendant son développement. Ce schéma de développement du groupe social est donc assez semblable à ce que nous observons chez l'humain et n'est pas sans rappeler les conséquences du rétrécissement perceptif. Il est réaliste de supposer que, pour les rats comme pour les nourrissons, un mécanisme permet aux individus d'extraire les invariants concernant leurs partenaires sociaux et leur permettent de définir le groupe social.

On pourra également citer le développement du répertoire vocal de certains oiseaux. Au sein d'une même espèce, des populations géographiquement éloignées peuvent présenter des variations dans leurs répertoires vocaux, appelés "dialectes" régionaux, rappelant nos différentes langues. Marler & Tamura (1964) ont étudié l'apprentissage des dialectes par des juvéniles de plusieurs populations de bruant à couronne blanche (white-crowned sparrow, WCS). Marler & Tamura ont isolé des WCS à différents âges et les ont exposés à des dialectes artificiels (créés par les expérimentateurs à partir de chants de WCS). Les résultats montrent qu'une exposition trop tardive ou trop précoce n'influence pas la production vocale du WCS isolé (l'oiseau n'apprend pas le dialecte artificiel). Ils ont découvert qu'il existe une fenêtre de développement pendant laquelle le WCS est capable d'acquérir un dialecte, environ entre 14 et 100 jours après l'éclosion. Cependant, l'exposition aux chants d'autres espèces au cours de cette fenêtre de développement ne modifie pas le chant du WCS. Par conséquent, apprendre le chant de l'espèce semble en partie génétiquement prédéterminé. Nelson (2000) a mis en évidence que les WCS sont capables de différencier les chants conspécifiques des chants d'autres espèces à la naissance. En revanche, apprendre un dialecte est dépendant de l'expérience durant les premiers mois de vie.

Ces études nous indiquent que l'affiliation à un groupe social semble se développer en fonction des individus présent dans l'environnement précoce des individus. Le fait que ce phénomène se produise chez l'homme mais aussi, dans une certaine mesure, chez d'autres espèces animales, nous amène à penser que le mécanisme menant au rétrécissement perceptif est un processus assez généraliste. Plusieurs auteurs que nous allons citer par la suite suggèrent que le mécanisme d'apprentissage statistique pourrait expliquer de manière assez parcimonieuse la formation de catégorie, sur la base de peu d'exemplaires. Ainsi, ce mécanisme pourrait expliquer, au moins en partie, l'émergence du rétrécissement perceptif.

VI.3. L'apprentissage statistique et la mise en place du rétrécissement perceptif

VI.3.a. *L'apprentissage statistique*

L'apprentissage statistique est maintenant établi comme étant un mécanisme crucial d'apprentissage du langage, permettant la segmentation des mots dans un flux de langue continue, seulement par l'exposition, implicitement, et même sans connaissance antérieure de la langue (Safran et al., 1996, 1997 ; Pelucchi, et al. 2009). De plus, il a été montré que l'apprentissage statistique peut également permettre d'extraire les relations qui ne sont pas directement adjacentes (c'est-à-dire des événements récurrents mais séparés par des événements variables), et pourrait donc permettre la détection de régularités grammaticales de la langue (Newport & Aslin, 2004). Cependant, si l'apprentissage statistique est un des mécanismes nécessaires à l'apprentissage du langage, il n'en semble pas spécifique. En effet, on retrouve des capacités d'extractions des régularités pour des stimuli audio qui ne sont pas des mots (Safran et al., 1999), chez d'autres espèces animales que l'humain (chez les primates : Heuser et al., 2001 ; Newport et al., 2004 / chez le rat : Toro et Trobalón, 2005), ou bien encore pour des stimuli visuels (Kirkham et al., 2002 ; Denison et al., 2013). De plus, Shafto et al. (2012) ont d'ailleurs montré une relation linéaire entre la capacité d'apprendre une séquence visuelle à 8.5 mois et une compréhension de vocabulaire 5 mois plus tard. De même, Ellis et al. (2013) ont montré que les performances sur une tâche de prédiction visuelle à 6 mois est positivement corrélée à leur performance sur une tâche de vocabulaire à 22 mois mais pas à 12 ou 16 mois. Par conséquent, la capacité d'apprentissage des dépendances statistiques visuelles est corrélée à la capacité d'apprentissage des dépendances statistiques dans le domaine du langage. Ces résultats suggèrent que l'apprentissage statistique est plutôt un mécanisme général, non spécifique, qui permet l'extraction de patterns récurrents dans l'environnement. Une partie de ces études suggèrent que la capacité d'extraire les régularités statistiques de l'environnement est déjà en partie présente chez les nourrissons dès 2 mois (Kirkham et al., 2002) mais que la capacité à généraliser semble plus évidente après 6 mois (Denison et al., 2013 ; Safran et al., 1996 ; Pelucchi, et al. 2009). Cette trajectoire développementale rappelle ce que nous observons pour le rétrécissement perceptif. L'apprentissage statistique est donc un mécanisme permettant d'extraire implicitement les invariants de l'environnement. Par une simple exposition, ce mécanisme permet aux individus de définir ce qui est probable d'arriver et ce qui ne l'est pas. Les individus vont ensuite former des attentes concernant les nouveaux événements, sur la base de ces probabilités.

VI.3.b. L'apprentissage statistique et son influence sur le rétrécissement perceptif pour les langues

Si les premières recherches sur l'apprentissage statistique dans le cadre du langage se sont principalement intéressées à l'acquisition lexicale (soit segmenter la parole afin d'identifier les mots), des recherches suivantes se sont intéressées à l'acquisition phonologique grâce au mécanisme d'apprentissage statistique. Maye et al. (2002) ont voulu tester l'hypothèse que les nourrissons apprennent (de manière implicite) la distribution statistique des sons de leur langue maternelle. Elles ont donc exposé des nourrissons de 6 et 8 mois à un continuum de son (voir Figure 13). Plus précisément, elles ont utilisé les sons « da » et « ta », qui diffèrent principalement par la présence d'un voisement (pour « da ») ou non (pour « ta »). Les auteures ont ensuite créé six sons intermédiaires, avec différents degrés de voisement, pour créer un continuum entre le son « da » et le son « ta ». Ensuite, un groupe de nourrissons a été exposé à une distribution unimodale des sons de ce continuum, c'est-à-dire qu'ils ont majoritairement entendus les sons intermédiaires du continuum (hybride de « da » et de « ta »). Un autre groupe de nourrissons a été exposé à une distribution bimodale de ce même continuum, c'est-à-dire qu'ils ont majoritairement entendus les sons proches des extrêmes de ce continuum (mais pas les deux extrêmes, donc pas les sons les plus prototypique). Après cette exposition, les deux groupes ont été testés sur une tâche de discrimination phonétique. Les résultats montrent que seuls les nourrissons ayant été exposés à la distribution bimodale sont capables de distinguer les deux extrêmes du continuum. Il est à noter que les deux groupes de nourrissons ont été exposés le même nombre de fois à ces deux exemplaires extrêmes durant la familiarisation ; c'est donc bien la répartition générale des exemplaires qui permet aux nourrissons de distinguer ces exemplaires et non une sur-exposition à ces exemplaires en particulier. Les nourrissons suivent ici les probabilités distributionnelles des stimuli ; ce type d'apprentissage statistique est communément appelé apprentissage distributionnel.

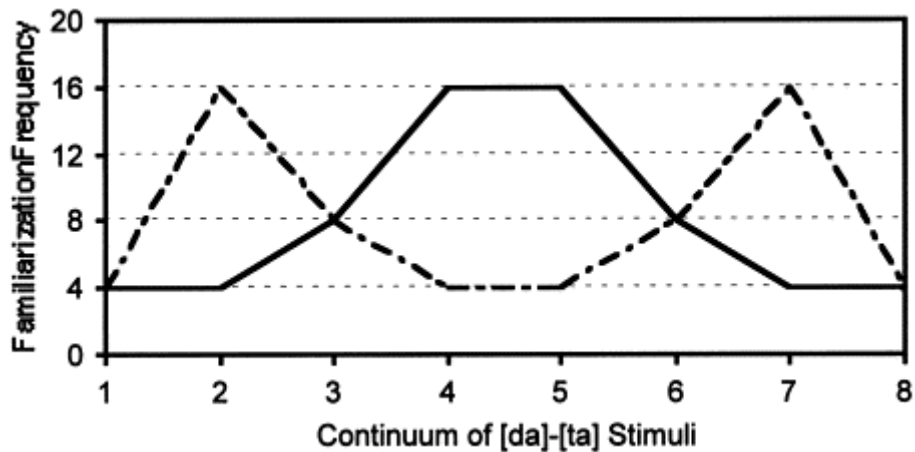


Figure 13. Nombre de présentation des huit stimuli lors de la familiarisation en fonction des deux conditions de l'étude de Maye et al. (2022).

En ligne continue : pour la distribution unimodale. En ligne discontinue : pour la distribution bimodale. Tiré de Maye et al. (2002).

L'apprentissage distributionnel peut mener au PME (Kuhl et al. 1992). Le PME implique qu'un exemplaire prototypique existe pour chaque représentation phonémique. Chaque exemplaire rencontré, bien que non parfaitement identique au prototype, est rattaché au prototype le plus proche par un effet d'aimant perceptuel. En se référant à la Figure 13, cela signifie que pour la distribution bimodale, deux prototypes existeraient (un « ta » et un « da »). Les exemplaires extrêmes ne sont donc pas proches des mêmes pics et sont perçus comme appartenant à deux catégories. Pour les nourrissons exposés à la distribution unimodale, il n'existe qu'un seul pic central. Les deux exemplaires extrêmes sont l'un comme l'autre rattachés au pic central, et sont donc catégorisés comme appartenant au même son. Conformément aux travaux de Maye et al., Wanrooij et al. (2014) ont mis en évidence que l'exposition à une distribution bimodale d'un contraste non-natif améliore les capacités de discrimination de ce contraste par les nourrissons. Dans cette étude, des nourrissons néerlandais de 2 à 3 mois ont été exposés à une distribution unimodale ou bimodale basée sur les voyelles contrastives en anglais [æ]–[e]. Les nourrissons exposés à la distribution bimodale ont montré des réponses ERP de *mismatch* plus importantes que les nourrissons exposés à une distribution unimodale. L'exposition à une distribution bimodale de phonèmes non-natifs a été testée par Yoshida et al. (2010) sur des nourrissons de 10 mois. Lorsque les nourrissons de l'étude de Yoshida et al. ont été familiarisés à une distribution bimodale de phonème non-natifs durant une durée de temps similaire à celle de l'étude de Maye et al., les nourrissons n'ont pas distingué les deux phonèmes lors de la phase de test. Cependant, lorsque la durée de la familiarisation a été doublée, les nourrissons ont

montré des signes d'apprentissage de la distribution bimodale. Plus les nourrissons grandissent, plus leur expérience avec leur langue maternelle est importante. Par conséquent, une exposition plus importante est nécessaire pour permettre aux nourrissons la distinction de phonèmes non présents dans leur langue. Liu & Krager (2014, 2017) ont confirmé ces résultats en évaluant la capacité des nourrissons hollandais de 5, 11 et 14 mois à discriminer des phonèmes mandarins non natifs après une exposition à des distributions unimodales ou bimodales du contraste mandarin. Les nourrissons de 5 mois ont discriminé les deux phonèmes indépendamment de leur exposition durant la familiarisation. Les nourrissons de 11 mois n'ont été capables de discriminer les phonèmes que s'ils appartenaient au groupe de distribution bimodale. En revanche, les nourrissons de 14 mois ont été incapables de discriminer les phonèmes, quelle que soit leur condition de familiarisation. Il est possible qu'une exposition plus longue à la distribution bimodale du contraste non-natif puisse permettre aux nourrissons de 14 mois de retrouver la capacité de distinction de ces phonèmes, comme c'est le cas dans l'étude Yoshida et al. Par conséquent, les nourrissons deviennent de moins en moins sensibles aux contrastes non-natifs en raison de leur expérience grandissante avec leur propre langue. Il faut une exposition de plus en plus importante pour modifier l'espace de distribution phonémique et permettre l'apparition d'un nouveau pic de distribution. Chez les jeunes enfants et les adultes, d'autres mécanismes que seulement l'apprentissage distributionnelle peuvent être nécessaire pour permettre à nouveau la discrimination de phonèmes non-natifs. Nous avons vu que si une simple exposition aux visages d'un autre type suffit aux nourrissons pour contrebalancer l'ORE (Heron-Delaney et al., 2011), il faut en plus des interactions sociales pour contrebalancer cet effet durant l'enfance et l'âge adulte (Qian et al., 2017 ; Goldstein et Chance, 1985). Ong et al. (2015) montre que l'exposition passive ne semble plus suffire non plus pour permettre la discrimination de phonèmes non-natifs chez des participants plus âgés. Les participants adultes de cette étude ont été exposé à une distribution bimodale ou unimodale de phonèmes non-natifs. L'exposition passive à une distribution bimodale ou unimodale n'a pas changé les performances des deux groupes lors d'une tâche de discrimination des phonèmes non-natifs. En revanche, lorsque pendant la familiarisation, il était demandé aux participants d'écouter activement les extraits et d'identifier certains sons intrus (introduit aléatoirement et en moindre nombre), les performances du groupe exposé à la distribution bimodale ont été meilleures que celles du groupe exposé à la distribution unimodale.

Tout récemment, Schatz et al (2021) ont développé un algorithme pour modéliser l'apprentissage phonémique, sans qu'aucune forme préexistante de catégorie phonémique ne

soit nécessaire. Durant une phase d'entraînement, des signaux des paroles sont donnés en entrée au modèle. Lorsqu'un signal de parole est fourni au modèle, le modèle l'échantillonne toutes les 10 ms (voir Figure 14). Chaque échantillon est réparti dans un espace multi-dimensionnel, selon des propriétés basiques du signal audio (amplitude, fréquence...). Ces propriétés acoustiques sont des propriétés du signal audio qui peuvent être perçues par les nourrissons, presque dès la naissance pour la plupart d'entre elles. En échantillonnant un nombre suffisant de signaux, le modèle crée ainsi un espace de distribution des différents sons de la langue à laquelle il a été exposé, sur de simples critères acoustiques. Cet espace de distribution est une mixture de différentes gaussiennes, avec certaines régions plus denses que d'autres (comparables aux distributions de la Figure 13, avec des pics et zones qui se recoupent, mais dans un espace multi-dimensionnel). Lorsque le modèle est testé sur un nouveau signal de parole, celui-ci est à nouveau échantillonné. La probabilité que chaque échantillon appartienne à l'une des gaussiennes issus de la phase d'entraînement du modèle est calculée. Le modèle peut ensuite rattacher l'échantillon à la gaussienne la plus probable. Sur la simple base de propriétés acoustiques du son, il est donc possible de 1) concevoir un modèle permettant de créer un espace de distribution des sons de la langue ambiante et 2) de rattacher des nouveaux sons entendus aux différents pics de la distribution. Ainsi, changer la sensibilité perceptive (c'est-à-dire augmenter la sensibilité de détection de certaines différences audio) est possible sur la simple base d'un apprentissage distributionnel.

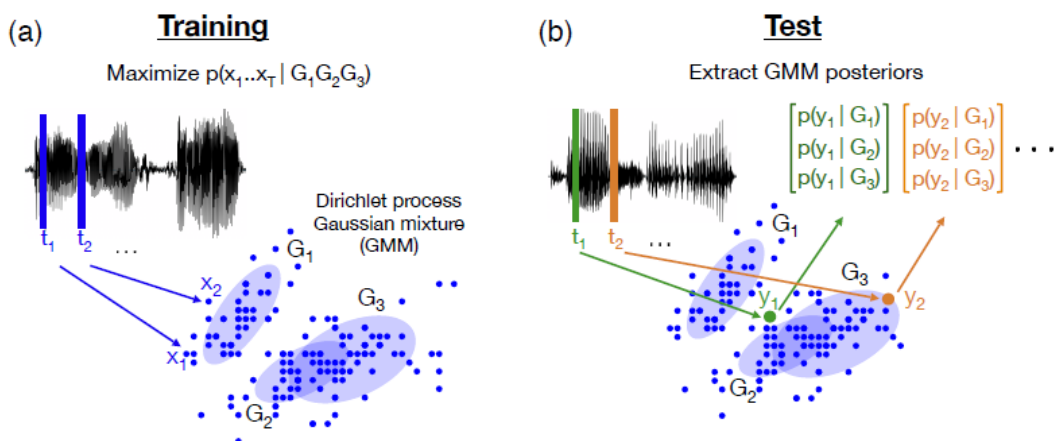


Figure 14. Schéma du fonctionnement du modèle de Schatz et al. (2021) lors de la phase d'entraînement (a) et lors de la phase de test (b).

Lors des deux phases : le signal audio est échantillonné toutes les 10 ms. Chaque échantillon est réparti dans un espace multi-dimensionnelle en fonction de ses propriétés

acoustiques. Le modèle traite l'ensemble des échantillons comme une mixture de gaussienne. Lors de la phase de test uniquement : pour chaque gaussienne, le modèle calcule la probabilité que l'échantillon soit issu de celle-ci. Tiré de Schatz et al. (2021)

Bosseler et al. (2013) ont cherché à savoir si cette stratégie, c'est-à-dire suivre les fréquences de distribution des différents sons, perdure au fil du développement. Dans ce but, ils ont fait écouter des phonèmes à des nourrissons finnois de 6 et 12 mois ainsi qu'à des adultes finnois, tout en enregistrant les oscillations cérébrales *thêtas* via magnétoencéphalographie (MEG). Les oscillations *thêtas* mesurées par MEG sont associées à l'attention et à l'encodage de nouvelles informations (Klimesch, 1999). Deux paires de phonèmes ont été jouées aux participants, une paire de phonèmes natifs (/pa/ et /ta/) et une paire de phonèmes non-natifs (/tchi et /ci/, issu du mandarin) mais cependant distinguée par les adultes et nourrissons finnois. Les phonèmes ont été présentés dans un paradigme de type *oddball*, c'est-à-dire en présentant de manière répétée un premier phonème (le phonème « standard », ici 85% des occurrences) et de manière beaucoup plus ponctuelle un second (le phonème « déviant », 15% des occurrences). Les auteurs se sont intéressés à la différence d'activation des oscillations *thêtas* selon deux conditions : selon si les phonèmes entendus étaient natifs ou non, et selon si les phonèmes entendus étaient standards ou déviants. Pour les nourrissons de 6 mois, les résultats montrent une différence d'activation des oscillations *thêtas* entre les phonèmes standards et les phonèmes déviants, mais pas entre les phonèmes natifs et les phonèmes non-natifs. A l'inverse, pour les nourrissons de 12 mois, les résultats montrent une différence d'activation des oscillations *thêtas* entre les phonèmes natifs et les phonèmes non-natifs, mais pas entre les phonèmes standard et les phonèmes déviants. Enfin, pour les adultes, le pattern de réponse est similaire à celui obtenu chez les nourrissons de 12 mois, à une différence : si pour les 12 mois l'activité *thêta* est plus forte lors de l'écoute des phonèmes natifs, elle est, chez les adultes, plus forte lors de l'écoute des phonèmes non-natifs (voir Figure 15). Les résultats indiquent donc que les nourrissons de 6 mois tournent leur attention vers la fréquence de distribution des différents sons de langues, et pas encore vers leur catégorie, natifs ou non-natifs. On remarque un changement de comportement (implicite) des nourrissons entre 6 et 12 mois : ceux-ci passent de traquer les fréquences de distributions des sons entendus à catégoriser ces sons. Ce mode d'attention semble plus proche de celui retrouvé chez l'adulte. Cette temporalité correspond à la perte de capacité de discrimination des phonèmes non-natifs.

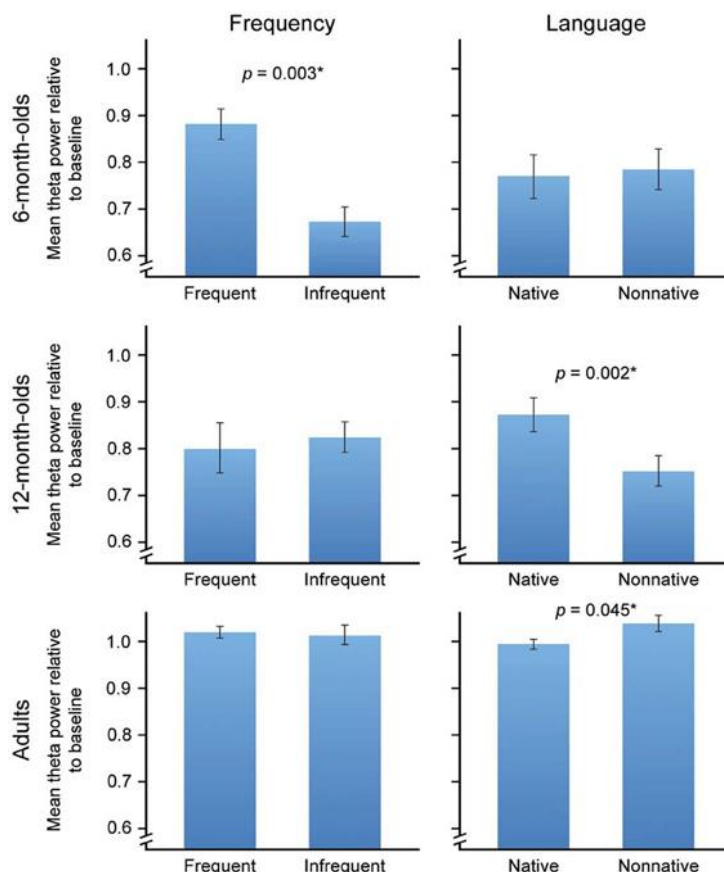


Figure 15. Intensité relative de la réponse *thêta* par conditions de l'étude de Bosseler et al. (2013).

A gauche : pour les phonèmes standards vs déviants. A droite : pour les phonèmes natifs vs non natifs. De haut en bas : pour les nourrissons de 6 mois, pour les nourrissons de 12 mois et pour les adultes. Tiré de Bosseler et al. (2013).

Ces études nous indiquent que l'apprentissage distributionnel permet aux nouveau-nés de répartir les sons de langues dans un espace multi-dimensionnel en fonction de leurs propriétés acoustiques. Lorsque le nombre d'exemplaires rencontrés est suffisant, les nourrissons commencent alors à catégoriser les sons. L'apprentissage distributionnel ne semble pas directement responsable de la création des représentations phonémiques mais permet l'émergence de celles-ci.

VI.3.c. *L'apprentissage statistique et son influence sur le rétrécissement perceptif pour les visages*

Le mécanisme d'apprentissage distributionnelle observée pour le langage est aussi observable pour les visages. Altvater-Mackensen et al. (2016) ont utilisé un protocole semblable à celui de Maye et al. (2002) en utilisant des photographies de visages de deux femmes de type européen.

Les auteurs ont créé des *morphings* entre les deux photographies, en intégrant par tranches de 5% la photographie de la deuxième femme dans la photographie de la première femme, créant ainsi 21 photographies. Les auteurs ont ensuite familiarisé des nourrissons de type européen de 6.5 mois à une distribution unimodale ou bimodale des visages composites (voir Figure 16). Plus précisément, l'intégralité des visages a été montrés aux nourrissons des deux groupes mais dans des fréquences différentes : les nourrissons du groupe exposé à la distribution unimodale ont majoritairement vu le visage issu du *morphing* central 50%/50% (composé à 50% de la photographie de la femme 1 et à 50% de la photographie de la femme 2) tandis que les nourrissons du groupe exposé à la distribution bimodale ont majoritairement vu les visages issus des *morphings* 25%/75% et 75%/25%. De manière importante, les nourrissons des deux groupes ont vu les visages originels le même nombre de fois durant la familiarisation. Ensuite, les auteurs ont enregistré les ERP de ces nourrissons lors de la présentation des visages originels. Les essais de tests alternaient entre des essais *match* (le même visage était présenté deux fois de suite) et des essais *mismatch* (les deux visages étaient présentés de manière consécutive). Les résultats montrent que les ERP des nourrissons familiarisés à la distribution bimodale présentent un profil différent selon les essais *match* et *mismatch*, ce qui n'est pas le cas des nourrissons familiarisés à la distribution unimodale. Rien n'indique que les nourrissons ne perçoivent pas la différence entre les deux visages. Cependant, l'exposition à une distribution unimodale semble les amener à catégoriser de la même manière les deux visages. L'apprentissage distributionnelle peut donc également être impliqué dans le développement du traitement des visages.

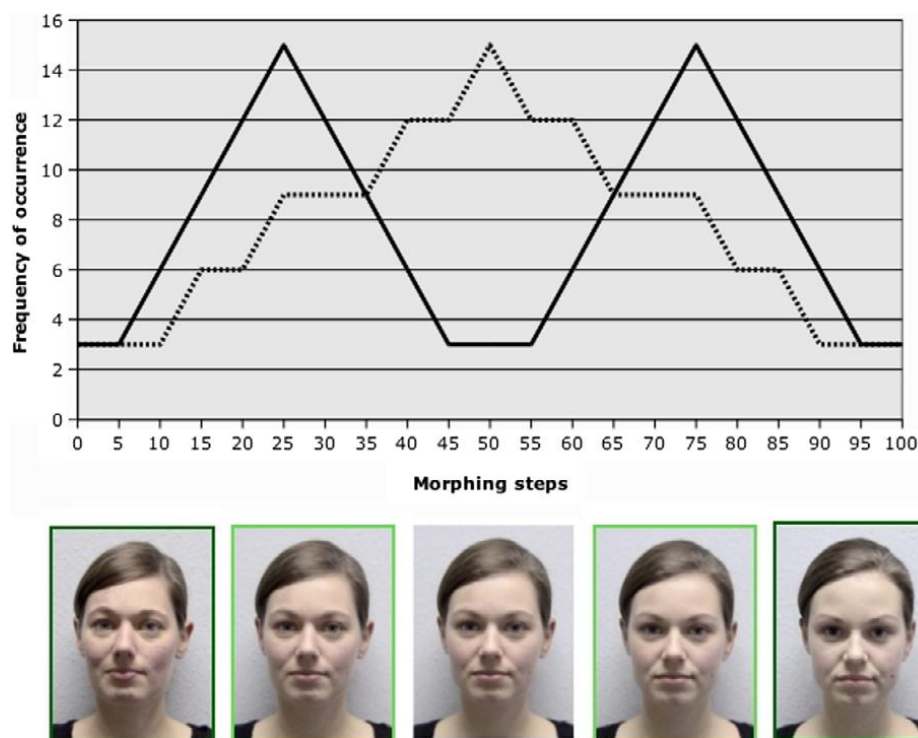


Figure 16. Nombre de présentation des 21 différents stimuli lors de la familiarisation en fonction des deux conditions de l'étude de Altvater-Mackensen et al. (2022).

En ligne continue : pour la distribution bimodale. En pointillé : pour la distribution unimodale. L'axe des x représente le degré de *morphing* entre les photos des deux modèles (0 est la photo originelle du modèle 1 ; 100 est la photo originelle du modèle 2). Les photos originelles, les morphing 25%/75%, 50%/50% et 75%/25 % sont représentés en dessous du graphique. Tiré de Altvater-Mackensen et al. (2016).

Afin d'étudier le rétrécissement perceptif à la lumière de l'apprentissage distributionnel, Krasotkina et al. (2021) ont tout récemment utilisé le même type de protocole que Altvater-Mackensen et al. (2016) en utilisant cette fois-ci des visages d'un autre type que celui des nourrissons (voir Figure 17). Après une familiarisation à une distribution unimodale ou bimodale selon la condition, les nourrissons ont passés une phase de test durant laquelle les deux visages originels étaient présentés à tour de rôle. Deux différences sont à noter par rapport au protocole de Altvater-Mackensen et al. : Krasotkina et al. ont testés des nourrissons de 12 mois et se sont intéressés aux temps de fixations plutôt qu'aux ERP. Les résultats montrent que les nourrissons exposés à une distribution unimodale des visages *morphés* ne présentent pas de temps de regard significativement différent durant la phase de test ; ils ne semblent pas capables de discriminer les deux visages. En revanche, les nourrissons exposés à une distribution bimodale regardent différemment les visages lors des tests de *mismatch* (lorsque les deux

visages présentés à la suite ne sont pas les mêmes). Ainsi, l'exposition à une distribution bimodale semble restaurer la capacité des nourrissons à distinguer des visages d'un autre type. Plus généralement, cette étude indique que la reconnaissance des visages d'un autre type est modulée par la fréquence d'exposition à ceux-ci, du moins à 12 mois.

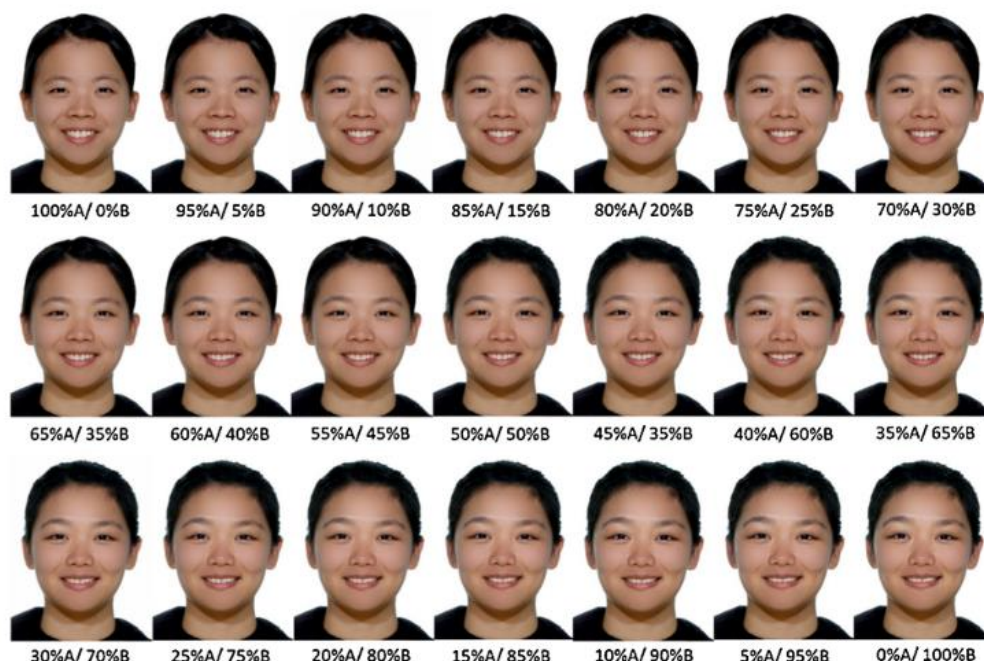


Figure 17. Stimuli utilisés dans l'étude de Krasotkina et al. (2021) pour les différents degrés de morphing. Tiré de Krasotkina et al. (2021).

VI.4. Attentes et apprentissage

Ces études indiquent que la simple exposition aux régularités distributionnelles des sons de langue et des différents visages module les capacités perceptives du nourrisson. Des recherches récentes montrent que les régularités statistiques de l'environnement conduisent les nourrissons à avoir des attentes quant aux événements futurs. Emberson et al. (2015) ont évalué la réponse de nourrissons de 6 mois lors de la présentation de stimuli audio, dans le cadre d'une étude en ISPIf. Lors d'une familiarisation, le stimulus audio était présenté de manière conjointe avec un stimulus visuel pour un groupe test de nourrissons tandis que pour un groupe contrôle, le stimulus audio était présenté seul. Lors d'une phase de test, le stimulus audio était présenté seul (donc violant les attentes de la familiarisation pour le groupe test) ou conjointement avec le stimulus visuel. Les résultats ont montré que la réponse cérébrale des nourrissons face au stimulus de test n'est pas la même selon la condition de familiarisation. Dès 6 mois, la réponse cérébrale des nourrissons est donc déjà en partie dépendante de leurs attentes. De même,

Kouider et al. (2015) ont étudiés la réponse ERP de nourrissons lors d'évènements respectant ou ne respectant pas les attentes des nourrissons. Les auteurs ont familiarisé des nourrissons de 12 mois à la présentation conjointe de sons (deux types de sons : un son de jouet et un son informatique) et de photographies (deux types de photographies : de fleurs ou de visages). Lors de la familiarisation, un son n'était associé qu'à un type de photo (par exemple le son de jouet était associé aux fleurs et le son informatique aux visages). Ensuite, lors de la phase de test, des présentations de stimuli respectant cette règle d'association (dans 75% des cas) ou non. Les résultats mettent en évidence une réponse ERP différentes selon si les associations sont congruentes ou non. Plus précisément, la réponse initiale est plus forte lors des associations congruentes, et la réponse tardive est plus forte lors des associations incongruentes. Les auteurs interprètent la réponse ERP comme une preuve que l'attention sélectives des nourrissons est meilleure pour les évènements attendus. Les résultats de ces études semblent confirmés par les données comportementales. Benitez et al. (2017) ont testés des enfants de 27 mois environ sur une tâche d'apprentissage de paires « nom-objet ». Dans une condition, les objets apparaissaient sur un écran de manière prédictible (ils apparaissaient dans les quatre coins de l'écran, à tour de rôle, dans le sens horaire) tandis que dans une autre condition, ils apparaissaient de manière imprévisible. Les résultats montrent que les enfants ont appris plus efficacement les noms des objets lorsqu'ils étaient nommés dans la condition prédictible que lorsqu'ils étaient nommés dans la condition imprédictible. De manière similaire, Breen et al. (2019) ont familiarisé des enfants de deux ans avec une liste de mots suivant une régularité phonologique (chaque mot commençait avec la même consonne). Ensuite, ces enfants ont été familiarisés à quatre objets, deux d'entre eux labellisés avec des mots suivant la régularité phonologique et deux autres labellisés sans respecter la régularité phonologique. Les résultats montrent que les enfants ont plus facilement mémorisé les objets donc le nom suivait la régularité phonologique lors de la familiarisation. Cependant, les enfants avec des vocabulaires moins développés ont appris les objets indépendamment de s'ils suivaient la régularité phonologique. Ces résultats nous indiquent deux choses. Premièrement, l'apprentissage est facilité lors de situations prédictibles, conformes aux attentes des nourrissons. Deuxièmement, la capacité à former des attentes vis-à-vis d'évènements futurs dépend de la quantité d'expérience par rapport à ce type d'évènements.

Ainsi, durant les premiers mois de vie, les nourrissons vont extraire les régularités concernant les visages, le langage, et leurs associations. L'expérience visuelle et langagière acquise par les nourrissons au cours de cette période provient très majoritairement d'individus de même type

qu'eux et parlant leur langue maternelle. Ce biais d'expérience va les amener à avoir des attentes vis-à-vis de leurs futurs partenaires sociaux. Le respect ou le non-respect de ces attentes lors de nouvelles expositions à des visages, du langage ou les deux conditionnent leurs capacités à traiter, reconnaître ou encore catégoriser ces stimuli, comme nous avons pu le voir dans les différentes études de cette thèse. Le mécanisme d'apprentissage statistique peut donc amener, en partie, aux conséquences du rétrécissement perceptif. Au fil du développement, l'accumulation d'expérience rend l'apprentissage statistique de moins en moins efficace puisque de plus en plus d'expérience devient nécessaire pour former de nouvelles attentes. Ce mécanisme d'apprentissage prend moins de place dans la vie du nourrisson, au profit d'autres mécanismes d'apprentissages plus prosociaux (Conboy et al., 2008 ; Ong et al., 2015 ; Yeung & Werker, 2009 ; Kuhl et al., 2003).

VI.5. Les mécanismes neuronaux à l'origine du rétrécissement perceptif

D'un point de vue neuronal, plusieurs hypothèses ont été proposées l'émergence du rétrécissement perceptif. Une première possibilité est que le rétrécissement perceptif résulte de l'élimination des connections synaptiques peu utilisées (Scott et al., 2007 ; Maurer et al., 2020). Durant les premiers mois de vie des primates, il y a en effet une très forte synaptogenèse (Bourgeois et al., 1994 ; Rakic, 1995). Cette période de surproduction de synapses est suivie par un élagage synaptique, les synapses recevant des *inputs* réguliers étant conservées tandis que les autres sont peu à peu éliminées (Huttenlocher, 1979 ; Huttenlocher et al., 1982). Le système devient ainsi plus adapté aux stimulations récurrentes de l'environnement. Les données expérimentales de Maurer et al. (2020) concernant les enfants synesthètes pourraient aller dans le sens de cette hypothèse. Les synesthètes sont des personnes associant un stimulus d'une modalité avec un autre stimulus, souvent d'une autre modalité : un exemple serait l'attribution de couleurs aux lettres de l'alphabet. Une hypothèse populaire est que la synesthésie résulte d'un défaut d'élagage synaptique (Hubbard et al., 2011 ; Banissy et al., 2012 ; Weiss & Fink, 2009 ; Zamn et al., 2013). Si effectivement le rétrécissement perceptif est la conséquence de l'élagage synaptique durant la petite enfance, les synesthètes devraient présenter un rétrécissement perceptif réduit. C'est effectivement ce que les données expérimentales de Maurer et al. (2020) semblent confirmer. Les synesthètes gardent en effet une sensibilité aux stimuli non-natifs (phonèmes ou visages) supérieure à celle des non-synesthètes. Cependant, le défaut d'élagage synaptique reste une hypothèse. De plus, bien que l'élagage synaptique soit effectivement observé pour certaines régions du cerveau, ce pattern de développement n'est pas uniforme (Quartz & Sejnowski, 1997). Il semblerait même que de manière globale, le nombre

de synapses chez le primate ne diminue pas avant la puberté (Purves et al., 1996). Une autre hypothèse possible est donc que le rétrécissement perceptif résulte de la consolidation des synapses fortement utilisées (Lewkowicz & Ghazanfar, 2009) ; les synapses non utilisées voient leurs poids relatifs diminuer, sans nécessairement que celles-ci soient élaguées. Cette hypothèse est confortée par la temporalité de la croissance synaptique et de l'élagage synaptique, la première ayant principalement lieu durant les 2-3 premières années de vie, le deuxième surtout après cette période et jusqu'à la fin de la vie. Le rétrécissement perceptif s'observant à partir de 9 mois environ, la diminution des capacités observée à cet âge est donc plus susceptible de résulter principalement de la formation et de la consolidation des synapses que de la disparition de celles-ci. Toutefois, les deux hypothèses ne sont pas exclusives, le rétrécissement perceptif pouvant à la fois être la conséquence de la consolidation et de l'élagage. Vogel et al. (2012) proposent à la vue de leurs résultats, que les aires cérébrales en charge de traiter certaines informations évoluent au cours du développement de l'individu. Elles trouvent en effet que les PE de nourrissons en réponses à des visages de même et d'autres types que le leur proviennent d'une région antérieure (enregistrements dans les régions frontales et centrales du scalp, possiblement associés à des activations du cortex préfrontal et cingulaire antérieur) à 5 mois et proviennent de régions plus postérieures (régions occipitales et temporales) à 9 mois. Il y aurait donc un changement quant aux circuits neuronaux traitant les informations relatives aux visages (dans le cadre de cette étude, des visages avec différents types d'émotions et associés à des vocalisations) concordant avec le rétrécissement perceptif. Ainsi, l'ensemble de ces travaux, en lien avec trois hypothèses différentes, montre que les mécanismes à l'origine du rétrécissement perceptif et la nature même du rétrécissement perceptif sont encore assez méconnus.

On note cependant des différences entre les mécanismes sous-tendant le rétrécissement perceptif lors du développement du langage et de celui du traitement des visages. Si l'exposition au langage ne se fait pas durant la première année, des retards irrécupérables se créent (Grimshaw et al., 1998 ; Friedmann & Rusou, 2015). Pour les visages en revanche, la période critique semble se prolonger tant qu'une expérience suffisante n'a pas été accumulée. On peut le voir notamment grâce aux expériences menées chez le singe qui montrent que des macaques non exposés à des visages pendant deux ans peuvent toujours apprendre à distinguer des visages, qu'ils soient de leur espèce ou humains (Sugita, 2008). Une fois que l'expérience avec un type de visage a eu lieu, l'apprentissage d'une autre catégorie devient beaucoup plus compliqué. Le fait que la période critique n'a pas une temporalité bien définie mais reste ouverte tant que l'expérience n'a pas eu lieu est congruent avec ce qui est observé sur le développement de la

vision chez les animaux élevés dans le noir total (Cynader, 1983 ; Mower et al., 1981 ; Timney et al., 1978). Ces observations nous laissent penser que les circuits neuronaux responsables du développement du langage et du traitement des visages mûrissent à des vitesses différentes. Il faut cependant être prudent quant à ces conclusions, les études citées précédemment portant sur le modèle animal. De plus, la temporalité du rétrécissement perceptif que l'on observe pour les visages et de celui que l'on observe pour le langage semble relativement similaire entre la majorité des études (pour le langage : Werker & Tees, 1984 ; Tsushima et al., 1994 ; Tsao et al., 2006 ; pour les visages : Pascalis et al., 2002 ; Kelly et al., 2007, 2009 ; Hayden et al., 2007).

Chapitre 3 : Intégration multisensorielle

Quand on évoque la compréhension du langage oral (non signé), on la rattache évidemment en premier lieu à la modalité auditive. Or, bien que nécessaire et parfois suffisante, l'ouïe n'est pas le seul sens impliqué dans la perception du langage, la vue jouant aussi un rôle essentiel, et parfois même le toucher bien que de manière moins documentée (Gick & Derrick, 2009). On pourra prendre le cas des malentendants, ceux-ci pouvant, grâce à la lecture labiale et gestuelle avoir une compréhension de la parole correcte tout en ayant une audition défailante (Bernstein et al., 2000 ; Erber, 1974 ; Rouger et al., 2007 ; Strelnikov et al., 2009). On utilise d'ailleurs l'expression « lire sur les lèvres », soulignant le caractère audio-visuel de la parole. Bien que beaucoup plus inconsciente chez les personnes disposant d'une audition dans la norme, la lecture labiale est pourtant constamment utilisée dans nos interactions verbales face à face. Une illusion célèbre montrant que l'on « écoute » avec les oreilles mais aussi avec les yeux est l'effet McGurk. Cet effet survient lorsque l'information sonore et l'information visuelle ne sont pas congruentes mais sont compatibles et donc fusionnables. Concrètement, il a été mis en évidence pour la première fois avec la vidéo d'un locuteur prononçant le son /ga/ (la bouche étant ouverte dès la consonne) mais dont la bande audio avait été remplacée par le son /ba/, en synchronie avec l'ouverture des lèvres. Aussi surprenant que cela puisse paraître, une fusion des deux sons fut perçue, soit le son /da/, compatible avec les deux informations à la fois : lèvres ouvertes dès la consonne et spectre acoustique ressemblant à celui d'un /da/ (McGurk & MacDonald, 1976). Ce phénomène sera expliqué plus en détail **Chapitre 6** de cette thèse. En fait, s'il est courant de présenter des stimuli unimodaux (soit des images, soit des sons) dans les études développementales, les informations que nous avons à traiter dans la vie courante sont généralement de nature multimodale. Le présent chapitre traite donc de l'intégration multimodale, en particulier l'intégration audio-visuelle de la parole, et son développement et son rôle chez le nourrisson. Dans un premier temps, nous examinerons en quoi la modalité visuelle contribue à la compréhension de la parole. Nous nous intéresserons ensuite au développement des capacités d'intégration multimodales et particulièrement audio-visuelles chez le nourrisson. Nous verrons que ce développement n'est pas toujours linéaire, et que les capacités d'intégration audio-visuelles semblent soumises, comme les capacités auditives et visuelles, au rétrécissement perceptif évoqué précédemment.

I. Apport de la modalité visuelle sur la compréhension de la parole

Les premières études sur l'intégration audio-visuelle cherchaient généralement à évaluer l'apport de la modalité visuelle sur la compréhension de la parole. Les travaux pionniers de Sumbly & Pollack (1954) visaient à évaluer l'apport de la vision lorsque le signal audio est dégradé. De manière pratique, ils ont présenté des mots, soit uniquement avec de l'audio, soit en audio-visuel. Ils avaient ajouté du bruit à ces mots, à différents niveaux de rapport de signal sur bruit (RSB). Les résultats ont montré que plus le signal était dégradé (plus il y avait de bruit), plus la différence de compréhension entre la condition audio seul et la condition audio-visuelle était importante. Ces résultats furent confirmés par d'autres études montrant que la compréhension de la parole dans un environnement bruité est améliorée par la présence du signal visuel (Benoît, et al., 1994 ; Binnie, et al., 1974 ; Erber, 1969). Voir des mouvements de lèvres cohérents avec le discours prononcé augmente donc l'intelligibilité de celui-ci. Toutefois, cette augmentation de l'intelligibilité est réduite lorsque les mouvements des lèvres sont remplacés par de simples barres horizontales se déplaçant de haut en bas en synchronie avec le discours (Bernstein et al., 2004 ; Schwartz et al., 2004). Cela impliquerait que ce n'est pas simplement une cohérence de la temporalité entre indices visuels et audio qui améliore l'intelligibilité mais que la forme de ces mouvements est elle-même porteuse d'information. En effet, s'il peut paraître évident que les lèvres jouent un rôle majeur – puisqu'elles contribuent à elles seules à environ 2/3 de l'information visuelle extraite du visage (Benoît et al., 1996) – elles ne se sont pas l'unique source d'information. La mâchoire (Goff et al., 1995), la langue si elle est rendue visible (Badin et al., 2010) ou encore les dents (Summerfield, 1992) sont des vecteurs d'informations. Aussi, il semble probable que les mouvements des régions faciales, même distantes de la région de la bouche, se fassent en concordance avec ceux des lèvres ; on trouve en effet qu'avoir accès au visage entier favorise la perception visuelle de la parole par rapport à une présentation des lèvres seules (Thomas & Jordan, 2004). Il a été montré de plus que les mouvements de la tête varient en fonction de la prosodie et peuvent eux aussi faciliter la compréhension du discours (Munhall et al., 2004). Toujours dans un environnement bruité, Schwartz et al., 2004, ont montré que la perception de voyelles qui ne semblent pas distinguables visuellement (par exemple [y] et [u] qui sont toutes des voyelles arrondies) est meilleure en condition audio-visuelle qu'en condition audio seule. Selon les auteurs, les informations visuelles contribuent donc à améliorer l'extraction d'informations acoustiques. Bien que les ajouts de bruit sur le signal audio soient ici artificiels, du bruit dû à l'environnement est presque systématiquement présent lorsque nous percevons la parole. On peut se demander

si l'apport visuel n'est que conjoncturel (sert lorsque le signal audio est dégradé) ou bien améliore toujours l'intelligibilité du discours. Le bruit n'est pas le seul facteur impactant l'intelligibilité à avoir été étudié. Arnold & Hill (2001) ont évalué le rôle des signaux visuels dans la perception d'une seconde langue, d'un accent non-familier ou bien encore d'un discours sémantique complexe. Dans ces trois situations, la présentation du discours en modalité audio-visuelle améliore sa compréhension comparativement à la présentation du discours en modalité audio seule. De la même manière, Navarra & Soto-Faraco, (2007) ont montré que la présentation de contrastes phonémiques catalans à des bilingues espagnol-catalan (de langue maternelle espagnole) se fait correctement lors de la présentation audio-visuelle mais pas lors de la présentation audio (ou visuelle) seule. Il apparaît donc que dans plusieurs situations comprenant une source sonore unique, la compréhension de nos interlocuteurs est renforcée par la possibilité de voir leur visage. On peut toutefois arguer que les trois situations mentionnées ci-dessus (accent non familier, langue secondaire, discours complexe) sont aussi une forme de bruitage du signal. On peut donc se demander si la modalité visuelle apporte à la compréhension de la parole en l'absence de dégradation du signal. Comprendre les émotions de notre interlocuteur est essentiel pour une bonne compréhension du message. Haq et al. (2009) ont montré qu'il est plus facile d'identifier les émotions d'un interlocuteur en condition audio-visuelle qu'en condition audio seule. Un dernier point permettant de mesurer l'impact de la vision sur la compréhension du langage provient des études portant sur les populations aveugles. Ces recherches tendent à montrer un délai dans l'acquisition du langage, les enfants avec une vision normale étant plus précoces pour apprendre les sons correspondant à des différences d'articulation visibles (Miner 1963 ; Ménard et al. 2009). Cependant, ces différences développementales sont faibles et ne semblent pas avoir d'impact à l'âge adulte, Muchnik et al. (1991) suggérant même, à l'inverse, que les non-voyants peuvent présenter de meilleures capacités dans la compréhension de la parole.

II. Intégration multisensorielle chez le nourrisson

II.1. Redondance inter-sensorielle

Certains éléments de la parole peuvent donc être extraits à la fois par la modalité audio et par la modalité visuelle. Le résultat est une possible redondance des informations extraites. Pour certains auteurs, cette redondance inter-sensorielle des informations est essentielle pour l'apprentissage ("The inter-sensory redundancy hypothesis", voir Bahrack & Lickliter, 2000, 2012 ; Shams & Seitz, 2008), le monde dans lequel nous évoluons étant par nature multimodal (Lewkowicz & Ghazanfar, 2009). Toutefois, pour pouvoir tirer avantage de cette redondance,

il faut être capable d'associer différents indices perceptuels au même stimulus. S'il est raisonnable de penser que ces types d'association (par exemple quel animal produit quel son) ne sont pas spontanément présentes chez le nouveau-né, la capacité à former des associations entre deux événements de deux modalités différentes semble opérationnelle dès la naissance. En effet, Slater et al. (1999) a montré que des nouveau-nés pouvaient associer un objet visuel et un son de manière totalement arbitraire. Via une familiarisation, les chercheurs ont présenté aux nouveau-nés une ligne verticale rouge ou une ligne diagonale verte en association avec une voix soit masculine, soit féminine, prononçant deux mots différents. Durant la phase de test, les nouveau-nés ont plus regardé les associations audio-visuelles non présentées lors de la familiarisation. Cette capacité d'associer des événements provenant de deux modalités différentes permettrait de faciliter l'apprentissage. Concrètement, Sai (2005) (ou Burnham, 1993, pour une étude similaire chez des nourrissons de 1 mois), a démontré que le visage de la mère est appris très rapidement après la naissance mais uniquement si la mère parle à son nouveau-né. *In utero*, le fœtus a déjà une expérience de la voix de la mère. L'hypothèse la plus plausible est donc que le nouveau-né associe ce visage à la voix qui lui est déjà familière, ce qui lui permet d'apprendre ce visage plus rapidement.

Nous voyons donc que la plupart des stimulations qui s'exercent sur nous sont de nature multimodale et que les nourrissons ont très tôt des capacités permettant de tirer avantage de cette multimodalité. Cependant, les études emblématiques testant le rétrécissement perceptif chez le nourrisson font souvent usage de stimuli uni-sensoriels, comme des photos (Pascalis, 2002 ; Kelly et al., 2007, 2009 ; Bar-Haim et al., 2006) ou des sons présentés sans support visuel (Eimas, 1971 ; Werker & Tees, 1984). Si ces études montrent bien une différence de traitement entre les stimuli natifs et non-natifs, elles nous renseignent moins sur leurs implications dans un contexte plus écologique. Les visages dynamiques entraînent généralement plus d'attention et de meilleure capacité de discrimination de la part des nourrissons que les visages statiques (Wilcox & Clayton, 1968 ; Knight & Johnston, 1997 ; Otsuka et al., 2009 ; Spencer et al., 2006). De même, les visages dynamiques auxquels des vocalisations sont ajoutées entraînent encore plus d'attention que les visages dynamiques silencieux (Coutrot & Guyader, 2014). Il est donc naturellement envisageable que des protocoles établis avec des stimuli multisensoriels permettent de meilleures performances de mémorisation et d'identification de la part des nourrissons, à condition que ceux-ci soient capables de traiter des stimuli multimodaux.

II.2. Le cas de l'intégration audio-visuelle

Bien que dans une mesure moindre que les adultes, les nourrissons semblent dès les premiers mois de la vie sensibles à la cohérence multimodale entre l'information visuelle et auditive. En effet, lorsque l'on présente deux vidéos côte à côte d'un même locuteur prononçant différentes voyelles (par exemple /i/ et /a/) et qu'un audio correspondant à une seule des deux voyelles est joué, les nourrissons, dès 2 mois, regardent plus longtemps le visage prononçant la voyelle correspondant à l'audio présenté (Kuhl & Meltzoff, 1982 ; Patterson & Werker, 1999 ; 2003 ; voir le **Chapitre 5** de cette thèse pour plus de détails). Aussi, dès 4 mois, les nourrissons sont sensibles à des effets de type McGurk (Burnham & Dodd 2004 ; Desjardins & Werker 2004 ; Rosenblum et al. 1997 ; voir le **Chapitre 6** de cette thèse pour plus de détails). Toujours à cet âge, même en l'absence de synchronie, les nourrissons sont capables d'associer des signaux audio et visuels (Lewkowicz, 2003, 2010). A 3 mois, les nourrissons regardent plus longtemps une vidéo d'un locuteur synchrone avec un extrait audio de parole continue qu'une vidéo d'un locuteur asynchrone avec l'extrait audio (Pickens et al., 1994). Cependant, les auteurs observent un développement en « U », trouvant la capacité à faire la correspondance entre un extrait audio et un extrait vidéo synchrone à 3 et 7 mois mais pas à 5 mois. A 12 mois, cette capacité est même résistante à une présentation séquentielle, toutefois uniquement pour la langue maternelle de l'enfant et dans un style de parole adressé à l'enfant (*infant directed speech* ; Kubicek et al., 2014). A partir de 6 mois environ, les nourrissons font correspondre la voix et le visage d'un même genre et ils sont, à partir de 7 mois, capables de faire correspondre une émotion visuelle et langagière (Patterson & Werker, 2002 ; Walker-Andrews et al., 1991 ; Walker-Andrews, 1986). Lewkowicz & Hansen-Tift (2012) montrent que la trajectoire du regard des nourrissons vers des visages qui parlent évolue au cours de la première année de vie, les fixations visuelles passant des yeux à la bouche (voir Figure 18). Ce changement de l'attention sélective correspond à un moment où les nourrissons commencent à babiller, ce qui pourrait donc favoriser le développement des capacités d'intégration audio-visuelle.

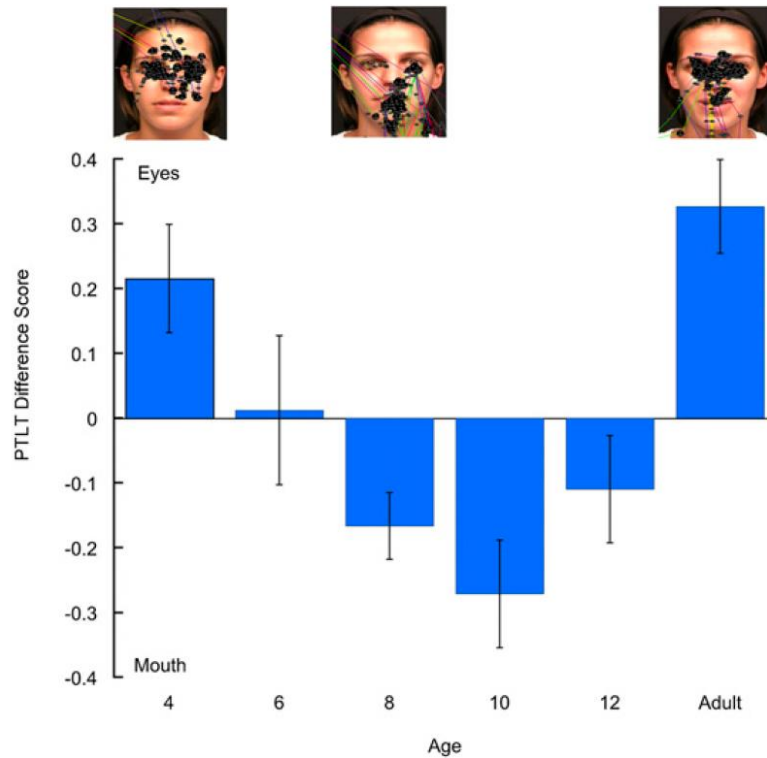


Figure 18. Différence de temps de regard proportionnel vers les yeux et la bouche au cours de la 1ère année de vie et à l’âge adulte. Tiré de Lewkowicz & Hansen-Tift (2012).

Si effectivement la multimodalité des stimuli entraîne de meilleures performances de mémorisation et de reconnaissance comparativement à des stimuli unimodaux et que les nourrissons sont, comme nous venons de le voir, très tôt capables d’appréhender cette multimodalité, alors il est possible que les capacités des nourrissons pour traiter des stimuli non-natifs soient en fait meilleures que ce qui a été conclu précédemment, notamment lorsqu’ils sont présentés de manière multimodale. Dans cette optique, Minar & Lewkowicz (2017) ont présenté des stimuli audio-visuels de visages d’autres types à des nourrissons. Pour cela, ils ont habitué des nourrissons de type européen à une vidéo d’un visage de femme asiatique prononçant le son [a], puis ils ont testé sa reconnaissance en présentant la même vidéo et la vidéo d’une autre femme asiatique prononçant le même son. Les nourrissons les plus âgés (10-12 mois) ont plus regardé le nouveau visage, tout comme les nourrissons les plus jeunes (4-6 mois), indiquant une reconnaissance du visage familier et un intérêt plus grand pour la nouveauté. Face à un visage dynamique et un stimulus multisensoriel, les nourrissons ne présentent donc pas l’ORE classiquement observé avec des stimuli unimodaux (avec simplement une photo de visage). Cependant, il n’est pas toujours clair dans les études combien de temps les nourrissons ont passé à regarder le visage durant l’habituation. Dans l’étude de Minar & Lewkowicz, des essais d’habituation sont présentés jusqu’à ce que la moyenne du temps de regard sur les trois derniers

essais soit de moins de 50% de celle des trois premiers essais, chaque essai d'habituation pouvant durer 60 secondes, ou bien jusqu'à ce que les nourrissons détournent l'attention pendant plus de 1 seconde. Or le simple fait de familiariser plus longtemps les nourrissons aux visages d'autres types peut permettre la reconnaissance, comme l'ont montré Fair et al. (2012) en utilisant des visages de macaques. En passant simplement d'une familiarisation de 20 secondes à 40 secondes ainsi que des tests de 5 secondes à des tests de 10 secondes, des nourrissons de 12 mois ont été capables de reconnaître des visages de macaques durant la phase de test, visages qu'ils ne sont normalement plus en mesure d'identifier à cet âge. On peut donc se poser la question de savoir si les capacités de reconnaissance des nourrissons ne seraient pas tout simplement dues à des expositions plus longues aux stimuli et non à la multimodalité de ces derniers. Minar & Lewkowicz ont cependant conduit une deuxième expérience dans laquelle les mêmes vidéos sont présentées cette fois sans le son. Dans ce cas, les nourrissons de 10-12 mois ont été incapables de reconnaître le visage d'un autre type. Une analyse complémentaire semble de plus montrer que le temps de regard des nourrissons n'est significativement pas plus faible durant la phase d'habituation de cette expérience (les vidéos sans son) que durant l'habituation de la première expérience (les vidéos avec sons), contrairement à ce que l'on pourrait attendre (soit une attention accrue lors de visages parlants, Coutrot & Guyader, 2014). De manière surprenante, les nourrissons de 4-6 mois ne sont plus non plus capables, dans cette deuxième expérience, de reconnaître les visages d'autres types. De surcroît, les nourrissons de 4-6 mois ne sont pas non plus capables de reconnaître les visages de leur propre type lorsqu'ils sont présentés en vidéo et sans son, mais le sont à 10-12 mois. Puisque les temps de regard durant l'habituation ne diffèrent pas entre ces deux expériences, ce n'est *a priori* pas un manque d'attention qui conduit à ne plus montrer de reconnaissance, même pour les visages de même type. L'argument avancé par les auteurs est que la vidéo d'une personne filmée en train de produire un son mais sans le son crée une légère incongruence pour les nourrissons, qui s'attendent une conséquence sonore des mouvements orofaciaux. Une partie des ressources cognitives est donc dédiée à essayer de résoudre cette incongruence, ce qui, en contrepartie, réduit la capacité à traiter les informations faciales permettant la reconnaissance. Les visages plus familiers de même type restent cependant correctement traités par les nourrissons de 10-12 mois, plus expérimentés, malgré la légère incongruence. Cette idée d'incongruence est étudiée dans une troisième et dernière expérience dans l'étude de Minar & Lewkowicz, où le son [a] est remplacé par un bruit (« boing », qui n'est donc plus un son de parole humaine). Cette fois-ci, si les nourrissons de 4-6 mois sont capables de reconnaître les visages de même type et d'autres types associés à ce bruit, ils n'en sont capables que pour les

visages de même type à 10-12 mois. Il se peut donc que les nourrissons les plus jeunes soient moins stricts concernant quels types de son un visage peut produire. Chez les nourrissons plus âgés, associer un visage humain à un son non humain provoque une légère incongruence qui impacte la reconnaissance uniquement pour des visages non familiers. Il est cependant difficile de conclure sur l'effet d'incongruence à la vue de la différence de résultat entre l'expérience deux (sans le son) et trois (avec un bruit). De plus, la reconnaissance par les nourrissons de 4-6 mois des visages associés à un bruit non humain est surprenante. En effet, les nourrissons de 5 mois font *a priori* déjà l'association entre visages humains et vocalisations humaines mais ne font pas l'association entre visage humain et des vocalisations d'autres origines (cri de canard ou de singe) (Vouloumanos et al., 2009). Comme nous l'avons vu au chapitre précédent, Xiao et al. (2017a) ont aussi étudié les effets de congruence entre stimuli visuels et audio, en associant des visages de même type ou d'autres types à de la musique joyeuse ou triste. Bien qu'ici les présentations des visages et des musiques étaient séquentielles (et consistaient donc en une alternance de stimuli unimodaux visuels et unimodaux auditifs plutôt que des stimuli multimodaux), les auteurs ont trouvé un temps de regard plus long dans le cas des associations congruentes qu'incongruentes. Il est donc probable que la congruence ou l'incongruence des associations multimodales est un rôle primordial dans le traitement des stimuli présentés, bien que ce rôle soit à préciser. De plus, si l'on voit qu'une partie des capacités d'intégration audio-visuelle s'améliore au cours du développement, ce n'est pas le cas de toutes. Nous avons vu précédemment que le rétrécissement perceptif impacte vers 6-9 mois la reconnaissance des visages ou la distinction de phonèmes qui ne sont pas ou peu présents dans l'environnement du nourrisson et qu'il semble en être de même pour les capacités d'intégration audio-visuelle (Lewkowicz & Ghazanfar, 2006 ; Lewkowicz et al., 2008 ; Pons et al., 2009).

Chapitre 4 : Méthodologie de l'expérimentation sur le nourrisson

La psychologie développementale présente de nombreuses contraintes spécifiques à l'étude des nourrissons compliquant la collecte de données et l'élaboration de protocoles. Un premier défi est l'absence de langage, rendant impossible d'expliquer la consigne d'une tâche à un nourrisson et obligeant les expérimentateurs à trouver des moyens détournés pour évaluer une capacité. Les nourrissons ont de plus une attention assez courte, et contrairement aux adultes, il est difficile, voire impossible, de les inciter à faire quelque chose qu'ils ne souhaitent pas faire. Cela résulte en un fort taux de passations incomplètes, ne permettant pas un recueil de données exploitables. Un autre défi lié à l'étude des nourrissons est la très grande variabilité individuelle. La psychologie développementale s'intéresse justement à une période de développement où les changements comportementaux et neurologiques sont relativement rapides, les nourrissons testés peuvent donc agir très différemment à peu d'intervalle. De plus, bien qu'il existe des ordres de grandeur concernant les étapes du développement, tous les nourrissons ne présentent pas les mêmes compétences aux mêmes moments. Par exemple, des bébés vont prononcer leurs premiers mots à 12 mois, tandis qu'il faudra attendre le 16^{ème} mois pour d'autres. D'autres compétences ou maturations neurologiques sont plus difficilement perceptibles que celle prise en exemple et pourraient pourtant expliquer des différences de résultats lors de tests expérimentaux. Il est donc nécessaire de trouver des stratégies pour tester les nourrissons de façon appropriée, ce qui passe notamment par des expériences courtes et stimulantes pour maintenir leur attention. Le panel de comportement du nourrisson étant assez limité, il faut s'appuyer sur ce qu'il sait faire, soit entre autres regarder, entendre, sucer, sentir, toucher. Dans une première partie de chapitre, nous décrirons les paradigmes visuels que nous utilisons dans les études de cette thèse. Dans une seconde partie, nous expliquerons le déroulement des expérimentations au BabyLab de Grenoble, du recrutement de participants à l'analyse des données.

I. Un traitement spécialisé

I.1. Le regard préférentiel

Les défis rencontrés lorsque l'on souhaite étudier les capacités cognitives des nourrissons ne sont pas sans rappeler ceux rencontrés par les comportementalistes animaliers. Ce n'est donc pas surprenant que les techniques de test du nourrisson empruntent beaucoup à l'éthologie. L'activité oculomotrice est très certainement l'un des moyens les plus simples pour étudier le développement du nourrisson (Fantz, 1958b). A l'origine popularisé par Fantz dans les années

1950 pour l'étude du comportement animal (Fantz, 1956 ; 1958a ; 1958c), la technique de regard préférentiel est toujours très utilisée de nos jours. L'idée consiste à présenter deux stimuli côte à côte, et d'examiner le temps de regard du nourrisson vers ces stimuli. Généralement, le nourrisson est installé sur les genoux d'un parent ou bien dans un fauteuil adapté. S'il y a une préférence visuelle, soit un temps de regard plus long vers un stimulus plutôt que vers un autre, on peut en conclure que le nourrisson perçoit une différence entre les deux stimuli. Afin de s'assurer que la préférence ne résulte pas d'une configuration spatiale particulière, le côté de présentation des deux stimuli est changé entre les essais. Cependant, l'interprétation des résultats est valable uniquement si le nourrisson a bien regardé les deux stimuli présentés ; il arrive que des nourrissons, particulièrement les plus jeunes, ne regardent qu'un seul côté tout au long de l'expérience, auquel cas il n'est pas possible de conclure. La vision ainsi que les capacités attentives et motrices du nourrisson font que cette technique est adaptée pour les nourrissons à partir de 3 mois.

Cependant, il est à noter que si une préférence visuelle est nécessairement une preuve de différenciation des stimuli, une différenciation des stimuli n'induit pas nécessairement une préférence ; c'est donc là l'une des limites de cette méthode. En effet, un nourrisson peut par exemple différencier un carré vert d'un carré bleu sans pour autant préférer l'une des deux couleurs. Il fixera donc potentiellement les deux stimuli de manière équivalente et l'interprétation des résultats restera très limitée.

I.2. Habituation/déshabituaton

Pour pallier ce problème, il est possible d'utiliser le paradigme d'habituation/déshabituaton. L'habituation est une forme d'apprentissage. Elle consiste en la diminution graduelle et relativement prolongée de l'intensité ou de la fréquence d'apparition d'une réponse suite à la présentation répétée ou prolongée du stimulus l'ayant déclenchée (Thompson, 2009). Le même principe va être utilisé pour le paradigme expérimental d'habituation/déshabituaton (Colombo & Mitchell, 2009). Dans le cadre de ce paradigme, on présente tout d'abord un stimulus au nourrisson, jusqu'à ce que son intérêt diminue, suite à son exposition répétée. C'est la phase d'habituation. Celle-ci dure tant que l'intérêt du nourrisson ne franchit pas un seuil, arbitraire, fixé par l'expérimentateur. Ce seuil peut par exemple être atteint lorsque la moyenne du temps de regard des trois derniers essais est inférieure à 50% de la moyenne du temps de regard lors des trois premiers essais, ou bien ce seuil peut être atteint lorsque la moyenne du temps de regard des deux derniers essais est inférieure à la moyenne du temps de regard des deux essais

les plus longs. Une fois ce seuil d'habituation franchi, vient la phase de déshabituation, lors de laquelle un nouveau stimulus est présenté. Celui-ci peut être présenté côte à côte avec le stimulus d'habituation, ou bien seul. La capacité de distinction de ce stimulus par rapport au premier est inférée si le temps de fixation du nourrisson remonte lorsque le nouveau stimulus est présenté seul, ou bien si une préférence existe lorsque le stimulus d'habituation et le nouveau stimulus sont présentés conjointement (la distinction peut être inférée que la préférence soit vis-à-vis du nouveau ou de l'ancien stimulus, une préférence quelle qu'elle soit indiquant une perception de la différence). Ce paradigme peut être utilisé avec différents type de réponses comportementales, comme par exemple le taux de succion ou le rythme cardiaque, mais la plus employée reste probablement le temps de regard.

Dans la suite de cette thèse, nous parlerons de phase de familiarisation. Nous définissons la phase de familiarisation comme une phase d'habituation, avec pour différence que le seuil arbitraire n'est cette fois plus dépendant du nourrisson mais est fixé à la même durée pour tous les nourrissons (par exemple, la phase de familiarisation termine après 30 secondes de présentation de stimulus). L'avantage est ici d'avoir un contrôle sur la durée de l'expérience empêchant des passations particulièrement longues. Un seuil minimum pourra être fixé quant au temps de regard durant la phase de familiarisation pour s'assurer que les nourrissons soient familiarisés. Dans le cadre de la familiarisation, la phase de réaction à la nouveauté sera enclenchée à la fin du seuil temporel arbitraire fixé en amont, que le bébé regarde encore ou non le stimulus de familiarisation. La suite de l'expérience sera identique à ce que nous avons décrit précédemment, c'est-à-dire que le stimulus de familiarisation sera présenté à côté du nouveau stimulus. Si une préférence pour un côté ressort, on conclura donc à la distinction des deux stimuli. Si le temps de regard est similaire pour les deux stimuli, cela ne voudra pas nécessairement dire que les nourrissons ne sont pas sensibles à leurs différences : pour des stimuli complexes, des temps de familiarisation plus longs peuvent en effet être requis (Hunter et al., 1983).

II. L'expérimentation au Babylab de Grenoble

II.1. Recrutement des participants

Le recrutement prend le plus communément place à la maternité. Deux fois par semaines, un membre de l'équipe se rend sur place et renseigne les tout nouveaux parents sur l'existence du BabyLab. Les coordonnées des parents intéressés sont récupérées, et ils sont recontactés aux alentours des trois mois de leur nourrisson pour un premier rendez-vous. Puis, s'ils sont toujours

intéressés, ils sont recontactés vers les 6, 9 et 12 mois, et éventuellement après, selon les études en place au laboratoire. Le recrutement à la maternité permet d'avoir un échantillon varié, même si des biais peuvent subsister quant à quelles familles se déplacent jusqu'au laboratoire.

II.2. Accueil et recueil de données

Nous recevons tout d'abord les parents dans une salle aménagée pour accueillir des enfants en bas-âge. Lors de cette première phase, nous recueillons le consentement des parents après explications du fonctionnement général du Babylab. Nous n'expliquons les hypothèses de nos études qu'après le recueil des données expérimentales pour que les parents restent autant que possible neutres durant les passations et n'influencent pas leur enfant par leurs propres réactions. Nous en profitons aussi pour administrer un court questionnaire permettant de récolter des informations essentielles pour l'analyse future des données. Ce questionnaire nous permet de mieux connaître le développement de l'enfant ainsi que l'environnement dans lequel il évolue. Parmi les informations demandées, se trouvent des quantifications de l'exposition de l'enfant à différentes langues, de l'exposition à différents types de visages, l'exposition aux visages d'hommes/femmes, ainsi que des données concernant le développement langagier, social et locomoteur de l'enfant. Les parents et l'enfant sont ensuite emmenés dans l'une des deux salles de test.

II.3. Mesures oculométriques avec un appareil EyeLink

Nous emmenons ensuite généralement les parents et l'enfant premièrement dans une salle de test dans laquelle se trouve un oculomètre EyeLink® SR1000. La salle est sans fenêtre, sombre, et éteinte lors des tests. L'enfant est, selon son gabarit, disposé sur les genoux d'un des parents ou dans un siège bébé, en restant en contact direct avec un des parents, positionné juste derrière (voir Figure 19). Des consignes sont données aux parents, leur indiquant qu'ils peuvent rester en contact avec l'enfant pour le rassurer mais qu'il est préférable de ne pas interagir. L'enfant est placé devant l'écran de test (résolution de 1024×768 pixels, taux de rafraîchissement de 100Hz, modèle Iiyama Vision Master Pro 514, 40×30cm) et devant la caméra et l'illuminateur infra-rouge, à une distance permettant une perception nette de la pupille par l'oculomètre. Une paire d'enceinte (CREATIVE Inspire T10) située de part et d'autre de l'écran permet la diffusion des stimuli audio. L'expérimentateur est situé dans la même pièce mais sur le côté et légèrement en arrière du nourrisson, et peut se cacher grâce à un rideau noir. L'expérimentateur place d'abord une petite cible sur le front du nourrisson pour donner un point de repère à la caméra infra-rouge. Un court dessin-animé est diffusé pour permettre de capter l'attention du

nourrisson vers l'écran lors du réglage des paramètres optimaux, puis l'on procède à une calibration en cinq points. Des petites cibles colorées et associées à du bruit permettent cette calibration. Les tests démarrent ensuite. Si le nourrisson change trop de position entre deux tests, une nouvelle calibration peut être effectuée. Le nombre de tests se situe généralement entre deux et cinq, mais dépend de la durée de ceux-ci (une vingtaine de secondes pour le plus court jusqu'à 2-3 minutes pour les plus longs) ainsi que de l'attention du nourrisson.



Figure 19. Dispositif oculométrique au Babylab. Repris de Damon (2015).

II.4. Mesure de temps de regard

La troisième salle du Babylab est une salle d'expérimentation comprenant une caméra classique (voir Figure 20). Les fenêtres de cette salle ont été obstruées pour diminuer la luminosité. Le nourrisson est installé sur les genoux d'un parent, à une distance d'environ 60 cm de l'écran 22 pouces servant à la présentation des stimuli. Des enceintes (Dell A225) sont situées à droite et à gauche de l'écran, certaines études présentant des stimuli sonores. Des paravents se trouvent de chaque côté et permettent aux expérimentateurs et au matériel autre que l'écran d'être caché. De la même manière que pour les tests en oculométrie, des consignes sont données aux parents, leur indiquant d'interagir le moins possible avec l'enfant et de ne pas chercher à recentrer l'attention du nourrisson si celui-ci regarde ailleurs que sur l'écran. En effet, certaines expériences mesurent aussi le taux de désintérêt. S'il est important que le nourrisson regarde l'écran, des maracas sont à disposition des expérimentateurs qui peuvent les agiter pour créer

un bruit juste derrière l'écran et ainsi attirer l'attention des nourrissons sur celui-ci. Deux expérimentateurs sont nécessaires pour les passations : un pour démarrer les expériences depuis l'ordinateur de contrôle caché derrière un paravent et un deuxième pour filmer. L'expérimentateur en charge de lancer les expériences a devant lui un écran lui présentant un retour de ce que voit le nourrisson. Il dispose également d'un retour caméra sur le visage de l'enfant, et d'une minuterie permettant de contrôler la durée de présentation des stimuli selon le critère défini pour chaque expérience, lorsque le défilement n'est pas automatique. Les expériences sont lancées via le logiciel E-prime 2.0 (Psychology Software Tools, Pittsburgh, PA, USA). Le deuxième expérimentateur est situé en face du premier et oriente la caméra vers le visage de l'enfant. Les stimuli sont donc diffusés en aveugle pour cet expérimentateur. Le nourrisson passe entre une et cinq expériences, en fonction de son attention, la durée des expériences allant de 10 secondes à 2 minutes 30.

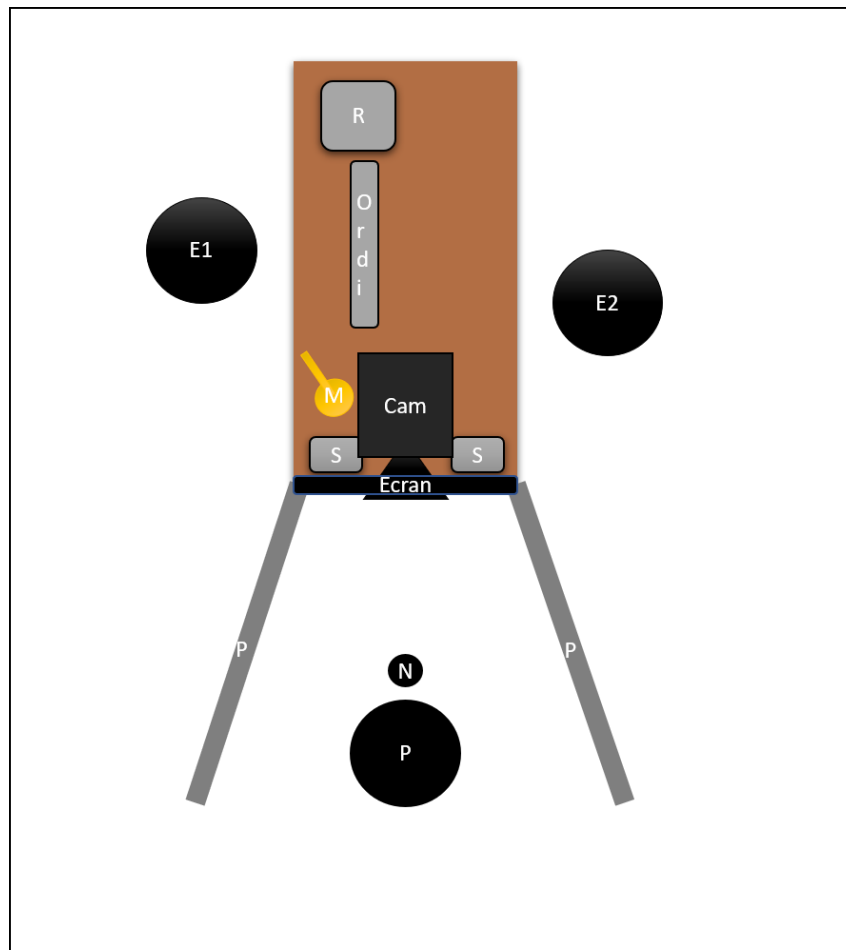


Figure 20. Schéma en vue du dessus de la salle d'expérimentation du Babylab.

E1 = Expérimentateur 1, lance les études ; E2 = Expérimentateur 2, filme le nourrisson ; P = Parent sur lequel le nourrisson est installé ; N = Nourrisson ; Ecran = Ecran sur lequel

apparaissent les études lancées par E1 ; Cam = Caméra de couleur noire, dont l'objectif est incrusté dans un orifice au milieu d'un rideau noir, au-dessus de l'écran ; S = Haut-parleurs gauche et droit ; Ordi = Ordinateur depuis lequel les études sont lancées ; R = Retour caméra permettant à E1 de vérifier le déroulement de l'étude ; M = Maracas permettant si besoin de recentrer l'attention des nourrissons sur l'écran ; P = Panneau occultant

II.5. Codage des vidéos

Les vidéos sont enregistrées grâce à une caméra PANASONIC Full HD basse lumière, adaptée à la pièce de test, peu lumineuse. L'analyse des vidéos se fait via le logiciel VP Video Lab, avec une précision pouvant aller jusqu'à 40ms (soit image par image). Le codage se fait à l'aveugle, c'est-à-dire que le codeur sait de quelle expérience il s'agit (puisque les modalités de codage varient d'une expérience à l'autre) mais ne sait pas quelle est la condition, ni ce que voit le nourrisson. Le travail du codeur est donc d'indiquer sur quelles fenêtres temporelles le nourrisson regarde à droite et à gauche si l'étude implique une présentation simultanée de deux stimuli, ou d'indiquer le temps d'engagement et de désintérêt si la présentation consiste en un seul stimulus. Une sous-partie (20% du total) des vidéos est aléatoirement choisie pour chaque expérience et codée indépendamment par un deuxième codeur, lui aussi aveugle à la condition (O'Connor & Joffe, 2020). La corrélation entre les annotations des deux codeurs est calculée. Un coefficient de corrélation de Pearson donné pour chaque étude permet d'attester de la fiabilité des résultats. Un temps de regard minimal est aussi établi pour chaque expérience, les nourrissons présentant des durées de regard en-dessous de ce seuil étant exclus de l'analyse final. De même, certains nourrissons « bloquent », c'est-à-dire qu'ils ne semblent remarquer qu'un seul stimulus lors de la présentation conjointe de deux stimuli. Les critères et seuil d'exclusion seront précisés pour chaque étude.

PROBLEMATIQUE

Nous avons vu que le développement du traitement du langage et le développement du traitement des visages présentent des similarités. En particulier, les conséquences d'un phénomène appelé rétrécissement perceptif sont observables sur ces deux systèmes de traitement. Cela nous amène à supposer des liens forts entre ces deux systèmes cognitifs durant le développement. Il est donc pertinent de penser que l'expérience rencontrée pour l'un de ces deux systèmes ait des conséquences sur l'autre. Le but de la thèse présentée ici est d'étudier les interactions entre les traitements du langage et des visages durant la première année de vie, durant laquelle le phénomène de rétrécissement perceptif émerge.

Dans un premier volet expérimental composé de deux études, nous nous intéresserons à l'influence du type de visage sur le traitement audio-visuel des visages. La question sous-tendant ces deux études est de savoir si le type d'une personne peut changer la façon dont les nourrissons traitent l'information audio-visuelle. La première étude consistera à tester l'impact du type de visage sur la capacité de correspondance phonémique, capacité présente très tôt dans le développement. La deuxième étude s'intéressera à l'impact du type de visage sur une tâche plus complexe de fusion audio-visuelle via l'effet McGurk.

Dans un deuxième volet aussi composé de deux études, nous nous intéresserons aux associations que les nourrissons forment entre types de visage et types de langues et à leurs conséquences. Dans une troisième étude, nous souhaitons savoir si le rétrécissement perceptif modifie les attentes des nourrissons quant au type de langue associé à un type de visage. Le but de cette étude est aussi méthodologique, puisque nous tenterons d'utiliser un protocole audiovisuel récemment développé et d'examiner si son utilisation peut être généralisée. Dans une quatrième étude, nous nous intéresserons aux conséquences de ces associations sur le traitement des visages et plus particulièrement sur leur reconnaissance.

PARTIE EXPERIMENTALE

Chapitre 5 : Correspondance audio-visuelle de voyelles sur des visages de même type et d'autre type

I. Introduction

Bien que principalement considéré comme une tâche auditive, nous avons vu que la compréhension de la parole est une tâche multimodale. Nous reposons beaucoup sur des indices visuels pour percevoir la parole, en particulier en milieu bruyé (Erber, 1969 ; Ewertzen & Birk-Nielsen, 1971 ; O'Neill, 1954 ; Sumbly & Pollack, 1954). Nous avons aussi vu que progressivement durant la première année de vie, les nourrissons regardent différemment les visages de type familier des visages de types non familiers jusqu'à ne plus bien reconnaître ces derniers (Bar-Haim et al., 2006 ; Kelly et al., 2005, 2007a, 2007b, 2009). A la vue des liens étroits entre traitement des visages et traitement de la parole, il est pertinent de se demander si le rétrécissement perceptif pour les visages impacte les capacités de traitement audio-visuel de la parole. Dans la présente étude, nous souhaitons évaluer l'impact du type de visage sur une tâche d'association audio-visuelle d'un signal de parole. Le développement des compétences d'associations audio-visuelles permet la compréhension multimodale de la parole est peut-être un élément facilitant l'acquisition du langage. Altvater-Mackensen & Grossmann (2014) montrent en effet que les nourrissons de 6 mois présentant une préférence pour un signal de parole audio-visuel congruent plutôt qu'incongruent ont un vocabulaire plus large à 12 mois. Savoir si le type de visage d'une personne impacte les capacités audio-visuelles des nourrissons présente donc un intérêt particulier sur les mécanismes d'acquisition du langage.

I.1. Capacités de distinction des sons de langues

Les nourrissons sont capables dès le premier mois de vie de distinguer des changements de voyelles présentées seules ou après une consonne (Trehub, 1973). Via la méthode de succion non-nutritive, Trehub a en effet montré que des nourrissons âgés d'entre 1 et 4 mois présentent une déshabitude lorsque le son /pi/ est présenté après le son /pa/, lorsque le son /ti/ est présenté après le son /ta/, ou bien encore lorsque la voyelle /u/ est présentée après la voyelle /i/. En revanche, la déshabitude est toujours présente mais plus faible lorsque la voyelle /i/ est présentée après la voyelle /a/. Kuhl (1979, 1983) a de plus démontré que les nourrissons semblent déjà avoir une représentation catégorielle de ces voyelles à 6 mois et répondent uniformément à plusieurs exemplaires différents de la même catégorie (par exemple la voyelle /i/ prononcée par une femme ou un homme) mais répondent différemment à un exemplaire d'une autre catégorie (par exemple la voyelle /i/ et la voyelle /a/ prononcées par une femme).

Marean et al. (1992) ont répliqué ce type de résultats chez des nourrissons aussi jeune que 2 mois, montrant que la plupart d'entre ne réagissent pas à un changement de locuteur sans changement de voyelle.

I.2. Sensibilité à la redondance multisensorielle

En plus d'être capable de distinction phonémique, les nourrissons sont très tôt sensibles à des indices de redondance multisensorielle, c'est-à-dire des événements prenant forme sous différentes modalités sensorielles mais partageant une temporalité commune. Dodd (1979) a montré que des nourrissons âgés entre 10 et 16 semaines portent plus d'attention à des comptines présentées en synchronie avec le mouvement des lèvres du locuteur que lorsque l'audio présentait un retard de 400 ms sur le visuel. Lewkowicz (1986) montre aussi que les nourrissons de 6-8 mois regarde plus un flash lumineux clignotant de manière synchrone avec un son qu'un flash lumineux de même durée mais présenté de manière asynchrone. Les nourrissons font donc des liens entre des événements de natures sensorielles distinctes mais présentant une synchronie. Cependant, Lewkowicz (1996) montre une certaine tolérance à l'asynchronie chez le nourrisson âgé d'entre 2 et 8 mois, ceux-ci détectant une asynchronie à partir de 350 ms pour un son en avance sur une vidéo et à partir de 450 ms pour une vidéo en avance sur un son (alors que ces seuils sont respectivement de 65 ms et 112 ms chez l'adulte). Les études de Pons et al. (2009) et Lewkowicz et al. (2013) (que nous développerons dans la suite de ce chapitre) montre aussi que les nourrissons sont capables de relier deux modalités d'un seul événement multisensoriel présenté de manière non pas synchrone mais séquentiel. C'est donc la congruence de temporalité entre les deux modalités plus que la synchronie même qui permet de relier les modalités.

I.3. La tâche de correspondance audio-visuelle de voyelle

Kuhl & Meltzoff (1982) et MacKain et al. (1983) figurent parmi les premiers à avoir étudié la capacité des nourrissons à faire correspondre les mouvements articulatoires à l'origine d'un son avec ce son. Cette tâche nécessite à la fois que les nourrissons soient capables de distinguer des phonèmes et de lier des modalités sensorielles différentes d'un événement. Pour étudier cette capacité, Kuhl & Meltzoff présentèrent côte à côte deux vidéos d'une même locutrice, d'un côté prononçant le son /i/, de l'autre prononçant le son /a/. Par-dessus, une bande audio jouait soit le son /i/ soit le son /a/, en synchronie avec l'ouverture des lèvres de la locutrice, les deux vidéos étant elles-mêmes synchronisées entre elles. La préférence visuelle pour le visage dont le mouvement des lèvres correspond à la voyelle jouée fut évaluée chez des nourrissons de 4.5 mois. Les résultats montrèrent que les nourrissons passaient 73% du temps à regarder le côté

avec le visage correspondant à la voyelle jouée, signifiant que les nourrissons sont dès 4.5 mois capables de faire correspondre un mouvement articuloire avec sa conséquence auditive. De plus, ils montrèrent que lorsque les voyelles sont des sons purs (sans harmoniques) plutôt que produites par la voix humaine, les nourrissons ne font plus la correspondance entre le son et la forme des lèvres. La conclusion des auteurs fut que la correspondance ne se fait pas uniquement sur des indices temporels mais aussi sur des propriétés spectrales du son produit. De leur côté, MacKain et al. (1983) ont testé la capacité de nourrissons de 5-6 mois à faire correspondre le signal audio de pseudo-mots de type Consonne-Voyelle-Consonne-Voyelle (CVCV) avec la vidéo du geste articuloire correspondant, via un paradigme similaire à celui de Kuhl & Meltzoff. Dans cette étude, les nourrissons n'ont réussi que lorsque la vidéo congruente était présentée sur la droite – le temps de regard était similaire au niveau de chance lorsque la vidéo congruente était sur la gauche. Les auteurs expliquent les résultats de leur étude par une dominance de l'hémisphère gauche dans le traitement du langage. Cependant, les stimuli de la tâche de MacKain et al. (1983) ne sont pas présentés dans un seul hémichamp, rendant l'hypothèse d'une dominance hémisphérique assez faible. Cette hypothèse hémisphérique ne fut d'ailleurs pas vérifiée par les répliques suivantes – à noter cependant que la tâche de Kuhl & Meltzoff (1982) consiste en une seule voyelle tandis que la tâche de McKain et al. (1983) consiste en un stimulus de type CVCV donnant ainsi une tâche potentiellement plus difficile pour les nourrissons. Kuhl & Meltzoff répliquèrent en effet plusieurs fois leurs résultats (1984, 1988, 1991). Tout d'abord, ils répliquèrent leurs résultats obtenus 2 ans plus tôt (Kuhl & Meltzoff, 1984) sur un nouveau groupe de nourrissons de 4.5 mois. Puis, ils montrèrent que les résultats pouvaient être répliqués pour une autre paire de voyelle, en l'occurrence /i/ et /u/ (Kuhl & Meltzoff, 1988). Quelques années plus tard, Kuhl & Meltzoff confirmèrent encore leurs résultats chez le nourrisson mais montrèrent en revanche qu'un simple son tonal suffit chez les sujets adultes pour permettre la correspondance (Kuhl et al., 1991). Par la suite, Walton & Bower (1993) ont répliqué les résultats de Kuhl & Meltzoff via un autre paradigme que la préférence visuelle. Ils ont utilisé une procédure de succion non-nutritive et n'ont présentés qu'un seul visage à la fois. Ce visage restait à l'écran en train d'articuler un son correspondant ou non à la voyelle jouée tant que les nourrissons suçaient la tétine. Les résultats ont montré que les nourrissons de 4 mois présentèrent un taux de succion plus élevé lors de la présentation du stimulus visuel correspondant avec la voyelle jouée, ici soit /i/ soit /u/. Ils ont de plus testé des nourrissons un peu plus âgés, de 6-8 mois, et montré que la correspondance audio-visuelle ne dépend pas de la familiarité avec les sons. En l'occurrence, les nourrissons furent capables de faire correspondre le visage d'une locutrice avec le son /y/ français non familier pour ces

nourrissons (texans). Utilisant les mêmes stimuli que Walton & Bower (1993), Aldridge et al. (1999) ont testé le même protocole de succion non-nutritive quelque peu adaptée sur des nouveau-nés. Les résultats montrent que quelques heures seulement après la naissance, les nouveau-nés regardent plus les vidéos qui correspondent aux voyelles jouées que les vidéos qui ne correspondent pas aux voyelles jouées (pour les paires testées, soit /a-/u/ et /i-/y/). Cela nous indique donc que des capacités de correspondance audio-visuelle sont présentes dès la naissance et ne nécessitent pas d'expériences avec des visages qui parlent. Depuis, Patterson & Werker (1999) ont de nouveau répliqué les résultats de Kuhl & Meltzoff. De la même manière que l'étude originelle, elles ont montré que des nourrissons de 4.5 mois sont capables de faire la correspondance entre une voyelle vu et entendu. Elles ont de plus montré que cet effet est toujours présent lorsque les stimuli visuels sont un peu plus complexes. Alors que seulement l'intérieur du visage de la locutrice était visible dans l'étude de Kuhl et Meltzoff (1982), les cheveux, le coup et le haut du torse des locutrices étaient ici visible. Ils ont aussi démontré que cet effet est robuste au genre en utilisant des locuteurs à la place des locutrices. Patterson & Werker (2003) ont une nouvelle fois répliqués les résultats de cette étude mais cette fois-ci sur un groupe de nourrisson de 2 mois. Là encore, les nourrissons regardèrent plus longtemps la vidéo correspondant à la voyelle jouée, et ce quelle que fût la voyelle, le côté de présentation ou le genre du locuteur. De même, Baier et al. (2007), ont voulu étudier la capacité des nourrissons à faire la correspondance audio-visuelle avec de nouvelles paires de syllabes. Après avoir tout d'abord répliqué les résultats sur le contraste /i-/u/, ils ont montré que des nourrissons de 2-3 mois regardent aussi plus longtemps le visage qui correspond au son entendu lorsque les contrastes sont /a-/u/ ou /i-/wi/. Des auteurs se sont intéressés au lien entre production et perception dans le cadre de cette tâche de correspondance phonème-visème. Tout d'abord, Mugitani et al. (2008) ont utilisé deux sons non langagiers produits par des visages : une consonne roulée bilabiale² (« brrr ») et un sifflement. Ils ont évalué la capacité de nourrissons japonais de 8 mois à faire la correspondance audio-visuelle entre ces sons et les vidéos correspondant à leur production. Si ces sons ne sont pas présents dans la langue japonaise, les nourrissons de 8 mois produisent généralement des trills bilabiales mais pas (ou presque) de sifflements (Mugitani & Kobayashi, 2003, cité par Mugitani et al., 2008). Le protocole utilisé par Mugitani et al. (2008) est légèrement différent de celui de Kuhl & Meltzoff (1982) puisque les vidéos (congruentes ou incongruentes avec le signal audio) sont cette fois-ci présentés séquentiellement et non simultanément. Les résultats montrent que les nourrissons regardent

² Ce type de consonne est peu utilisée dans les langues parlées. On retrouve son usage en Mangbetu (langue parlée dans le nord-est de la République Démocratique du Congo) et dans certaines langues océaniques.

plus la vidéo congruente lorsque le son est un son qu'ils sont capables de produire (soit la trill bilabiale) mais regardent autant la vidéo congruente que la vidéo incongruente lorsque le son n'est pas un son qu'ils sont capables de produire (soit le sifflement). Mugitani et al. (2008) montre ainsi que la capacité de correspondance multimodale du signal de parole semble dépendante de l'expérience de production et semble impactée pour des sons pas ou peu présents dans l'environnement langagiers des nourrissons. Confirmant cette idée, Vilain et al. (2019) ont montré que le babillage des nourrissons a une influence sur leur aptitude à faire de la correspondance audio-visuelle. De manière plus précise, les nourrissons babillant sont plus susceptibles de regarder la vidéo d'une locutrice prononçant /ba/ plutôt que /da/ après avoir entendu des syllabes commençant par la consonne /b/, et inversement. En revanche, les nourrissons ne babillant pas encore regardent indistinctement la vidéo correspondant au même type d'articulation que celle ne correspondant pas. De plus, les nourrissons babillant ne font la correspondance que lorsque le type d'articulation présenté fait partie de leur vocabulaire de babillage (ils ne font par exemple pas la correspondance avec les consonnes /v/ et /z/). Nous pourrions aussi citer Yeung & Werker (2013) qui ont évalué l'impact de la sensori-motricité sur des nourrissons réalisant la tâche. Ils ont pour cela utilisé les voyelles /i/ et /u/ et ont modifiés la forme des lèvres des nourrissons pour qu'elles correspondent ou non avec la forme requise pour produire la voyelle entendue. L'hypothèse principale était qu'un effet facilitateur pourrait résulter de la forme de la bouche congruente avec le son joué ; l'hypothèse alternative était qu'au contraire, un effet d'interférence pourrait en résulter. Les résultats confirment dans un premier temps que les nourrissons de 4.5 mois regardent plus la vidéo de la locutrice qui correspond à la voyelle jouée. Cependant, lorsque la forme des lèvres des nourrissons est incongruente avec la voyelle jouée, les nourrissons passent plus de temps à regarder la vidéo de la locutrice qui ne correspond pas avec la voyelle jouée. Les résultats de Yeung & Werker confirment ainsi le lien entre production et perception mettant ici en évidence un effet d'interférence. Enfin, Altaver-Mackensen & Grossman (2014) ont évalué le lien entre capacité de correspondance multimodale et développement du langage. Pour cela, ils ont utilisé le même type de tâche que Mugitani et al. (2008) avec les syllabes /a/, /e/ et /o/. Les résultats montrent que la préférence à 6 mois pour les signaux audio-visuels congruents est positivement corrélé avec la taille du vocabulaire à 12 mois. Le type de tâche initialement développé par Kuhl & Meltzoff (1982) peut donc être un indicateur du futur développement langagier des nourrissons.

I.4. Un rétrécissement perceptif multisensoriel ?

Bien que peu d'études existent encore à ce sujet, nous avons précédemment vu que le rétrécissement perceptif pour les visages pourrait avoir un impact sur la perception multisensorielle. Nous reciterons ici l'étude de Lewkowicz & Ghazanfar (2006). En utilisant une procédure similaire à la tâche de correspondance audio-visuelle de voyelle, ils ont montré qu'après 9 mois, les nourrissons ne sont plus capables de faire correspondre une vocalisation de singe avec la vidéo d'un singe produisant cette vocalisation. Ils montrent donc que la tâche de correspondance audio-visuelle est affectée par le rétrécissement perceptif pour les visages. Les nourrissons pouvant toujours distinguer les deux types de vocalisations et les deux vidéos (Lewkowicz et al., 2008), la conclusion fut que le rétrécissement perceptif affecte la perception multisensorielle. Grossmann et al. (2012) ont aussi confirmé ces résultats via les PE de nourrissons exposés à ces mêmes stimuli. Les PE des nourrissons de 4 mois diffèrent selon qu'ils regardent la vidéo congruente (correspondant à la vocalisation jouée) ou la vidéo incongruente (ne correspondant pas à la vocalisation jouée). En revanche, les PE des nourrissons de 8 mois ne diffèrent pas lorsque les nourrissons regardent la vidéo de singe congruente ou incongruente avec la vocalisation jouée. Les auteurs ont aussi créé de nouveaux stimuli consistant en des locuteurs humains reproduisant les vocalisations de singes le plus fidèlement possible. De la même manière, les PE des nourrissons ont été enregistrés lors de la situation congruente et incongruente. Cette fois, la différence entre les PE enregistrés en situation congruente et incongruente se maintient à 8 mois, bien que légèrement différente de celle enregistré à 4 mois. Après le rétrécissement perceptif pour les visages, la sensibilité à la congruence multisensorielle semble donc limitée aux visages et aux voix humaines (même si la vocalisation n'est ici pas un son de langue humaine). De plus, Lewkowicz et al. (2010) ont aussi testé cette capacité sur des nouveau-nés et montrent que dès la naissance, les nouveau-nés regardent plus le visage de singe qui correspond à la vocalisation jouée. Avec Aldridge et al. (1999), ils fournissent une preuve de plus que la capacité à lier deux modalités d'un même événement est présente dès la naissance ; de plus, ils montrent que cette capacité de se limite pas aux stimuli natifs.

Pons et al. (2009) ont développé un autre type de protocole pour évaluer la capacité des nourrissons à faire correspondre un phonème à son visème. Pour cela (Figure 21), ils ont présenté simultanément les vidéos silencieuses d'une même locutrice des deux côtés d'un écran en train de prononcer des phonèmes différents (/ba/ et /va/). Après une présentation de chaque côté, un écran noir remplace les vidéos. Durant 45 secondes, un phonème correspondant à l'une

des vidéos est joué de manière répétitive. Puis, les deux vidéos silencieuses sont de nouveau présentées simultanément de chaque côté de l'écran pendant 45 secondes. La présentation de l'audio suivi des vidéos silencieuses est répétée une fois pour contrebalancer le côté de présentation de la vidéo correspondant à l'audio. Le temps de regard vers les vidéos après familiarisation à l'audio était ensuite comparé au temps de regard vers les vidéos avant familiarisation. Le contraste phonémique testé dans leur étude, soit entre /ba/ et /va/, n'est pas pertinent en espagnol. Ils ont donc évalué le temps de regard vers la vidéo de la locutrice qui correspond avec le phonème précédemment joué de nourrissons grandissant dans des environnements monolingues espagnols ou anglais de 6 et 11 mois. Les résultats montrent que les nourrissons de 6 mois, quelle que soit la langue présente dans leur environnement, regarde plus la vidéo de la locutrice correspondant au son précédemment joué. En revanche, les nourrissons de 11 mois ne regarde la vidéo qui correspond au phonème précédemment joué que si le contraste phonémique est présent dans leur langue maternelle. Ici donc, les nourrissons grandissant exposés à la langue espagnole regarde autant la vidéo incongruente que la vidéo congruente.

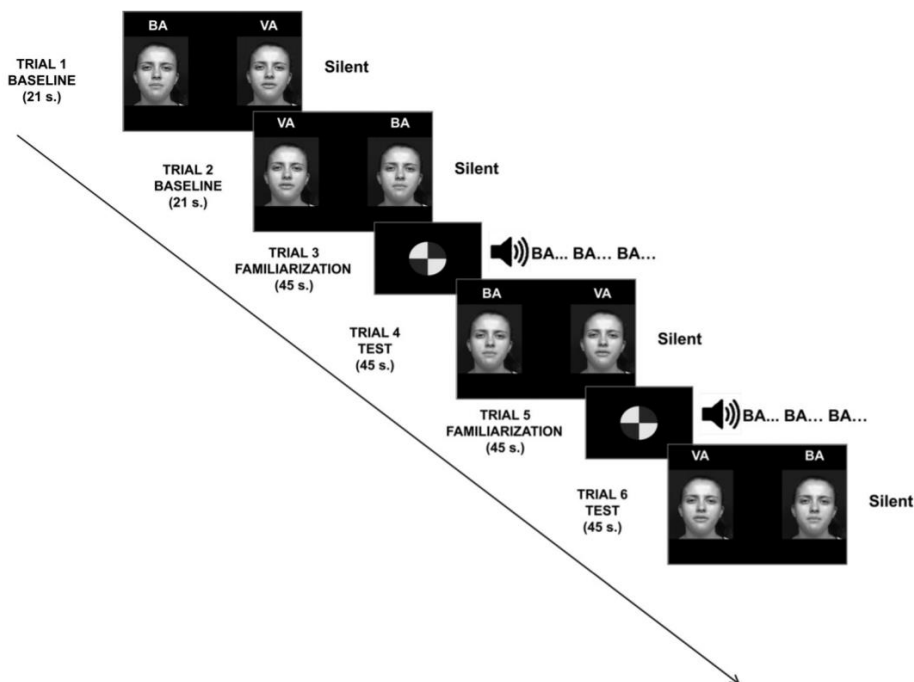


Figure 21. Protocole mis en place par Pons et al. (2009) pour tester la capacité des nourrissons à faire correspondre des phonèmes non-natifs avec des visèmes non-natifs. Tiré de Pons et al. (2009).

Lewkowicz & Pons (2013) ont par la suite évalué la capacité de nourrissons de 6-8 mois et de 10-12 mois à faire correspondre leur langue maternelle ou une langue étrangère à la vidéo de la

locutrice correspondante. Ils ont utilisé un protocole similaire à celui de Pons et al. (2009). La familiarisation à l'audio était cette fois-ci une histoire en langue maternelle ou étrangère selon la condition et durait 20 secondes. Suivait ensuite 30 secondes de présentation des vidéos silencieuses des locutrices, l'une des vidéos correspondant à l'audio précédemment joué (du moins pour les 20 premières secondes), l'autre non (puisque correspondant à l'audio dans l'autre langue). La présentation d'un audio de 20 seconds suivis des vidéos des locutrices de 30 secondes était répété deux fois. Le temps de regard vers les vidéos après familiarisation à l'audio était ensuite comparé au temps de regard vers les vidéos durant la baseline (les vidéos silencieuses avant présentation d'un audio). Lewkowicz et Pons ont trouvé qu'à 6-8 mois, la phase de familiarisation à une langue ne modifiait pas le temps de regard des nourrissons durant la phase de test. En revanche, à 10-12, les nourrissons regardent plus la vidéo de la locutrice qui ne correspond pas à l'audio auquel ils ont été familiarisés mais uniquement lorsque l'audio est dans leur langue maternelle. Cela nous indique que lorsqu'exposé à leur langue maternelle, les nourrissons de 10-12 ont probablement réussi à faire correspondre l'audio entendu et la vidéo de la locutrice et ont plus regardé la vidéo ne correspondant à l'audio par intérêt pour la nouveauté. Le fait que la capacité à faire la correspondance se limite à la langue maternelle montre que le rétrécissement perceptif pour le langage semble impacter cette capacité. Cependant, Kubicek et al. (2014) trouvent des résultats mettant en avant une trajectoire développementale différente en utilisant un protocole assez similaire. Dans leur étude, ils ont testé la capacité de nourrissons de 4.5, 6 et 12 mois à faire correspondre des vidéos de locutrices parlant leur langue maternelle ou une langue étrangère avec l'audio correspondant, présenté de manière synchrone ou séquentielle. Lorsque présentés de manière séquentielle, une histoire en langue native ou étrangère était jouée pendant 30 secondes puis durant la phase de test, les vidéos correspondantes étaient jouées côte à côte, silencieusement. La différence avec l'étude de Lewkowicz & Pons (2013) est une familiarisation à l'audio plus longue (30 secondes au lieu de 20 secondes). Les résultats montrent que les nourrissons regardent plus longtemps la vidéo correspondant à l'audio précédemment joué à 4.5 mois, quelle que soit la langue. En revanche, à 6 mois, ils regardent plus la vidéo correspondante durant la phase de test uniquement lorsque l'audio était en langue native. Les auteurs mettent ici en évidence un phénomène de rétrécissement perceptif multisensoriel dès 6 mois, la correspondance audio-visuelle d'un discours dans une langue non-native ne se faisant pas à cet âge-là. On observe donc ici une différence de temporalité avec l'étude de Lewkowicz & Pons qui n'observe pas de correspondance, quelle que soit la langue à 6-8 mois. Dans la deuxième partie de l'expérience, Kubicek et al. (2014) ont cette fois-ci présenté le stimulus audio de manière synchrone avec les

vidéos. Les auteurs observent que les nourrissons de 6 mois regardent à nouveau plus le visage qui correspond, quelle que soit la langue. Les indices temporels facilitent donc probablement la tâche des nourrissons leur permettant ainsi de faire la correspondance audio-visuelle même pour une langue non-native. A 12 mois cependant, les nourrissons ont un comportement différent et regardent plus la vidéo qui correspond à l'audio uniquement lorsque la langue de la familiarisation est non-native. L'hypothèse avancée par les auteurs est que les nourrissons de 12 mois ont suffisamment d'expérience avec leur langue pour ne pas dépendre des indices de temporalité. Des études montrent en effet que les nourrissons regardent plus la bouche d'un locuteur lorsqu'ils sont exposés à une langue étrangère (Lewkowicz & Hansen-Tift, 2012 ; Kubicek et al., 2013). Les nourrissons pourraient donc bénéficier des indices de temporalité lorsqu'ils sont exposés à une langue étrangère mais pas lorsqu'ils sont exposés à leur langue. Là encore, on observe une divergence avec les résultats de Lewkowicz & Pons qui trouvent que les nourrissons de 6-8 mois ne sont pas capables de faire la correspondance audio-visuelle quelle que soit la langue mais le sont à 10-12 lorsque l'audio est en langue maternelle (mais qui se traduit dans cette étude par plus de regard vers la locutrice native). Des différences méthodologiques sont à noter entre l'étude de Kubicek et al. et celle de Lewkowicz et al. : la durée de la familiarisation à l'audio n'est pas la même (30 secondes pour Kubicek et al. vs 20 secondes pour Lewkowicz et al.), les groupes d'âges ne sont pas exactement les mêmes (6 mois pour Kubicek et al. vs 6-8 mois pour Lewkowicz et al.), les langues testées ne sont pas les mêmes (allemand / français pour Kubicek et al. vs anglais / espagnol pour Lewkowicz et al.) et pour la deuxième partie de l'étude, le protocole n'est pas le même (présentation synchrone pour Kubicek et al. vs présentation séquentielle pour Lewkowicz et al.). Ces facteurs peuvent peut-être expliquer ces divergences de résultats. Toutefois, ces deux études se rejoignent sur la même conclusion : le traitement multisensoriel de la parole semble affectée par le rétrécissement perceptif pour le langage. Aussi, il peut sembler surprenant que les nourrissons réussissent à faire correspondre un audio et une vidéo présentée de manière séquentielle. Bien qu'une tolérance à l'asynchronie existe chez les nourrissons, celle-ci est de l'ordre de quelque centaine de millisecondes (Lewkowicz, 1996). L'argument avancé par Kubicek et al. (2014) est que les nourrissons se servent de l'information prosodique pour faire la correspondance entre l'audio et le visuel. Cette information est la première à laquelle les nourrissons accèdent concernant leur langue, une sensibilité à la prosodie ayant été trouvée in utero (Gervain & Mehler, 2010 ; Gervain, 2018). Dès la naissance, les nouveau-nés peuvent utiliser les indices prosodiques pour différencier des langues (Nazzi et al, 1998). Il est donc possible que cette expérience très

précoce et constante avec la prosodie de leur langue amène les auteurs à observer un rétrécissement perceptif aussi précoce.

I.5. Objectif et hypothèses de l'étude 1

Dans la présente étude, nous avons voulu évaluer l'impact du type de visage sur les capacités d'association audio-visuelle de voyelles des nourrissons. Nous avons pour cela utilisé une tâche simple et de nombreuses fois utilisée dans la littérature. Nous avons donc choisi la tâche de correspondance audio-visuelle de voyelle via un paradigme de préférence visuelle. Nous avons voulu évaluer cet impact avant et après l'émergence du rétrécissement perceptif à proprement parlé, c'est-à-dire le moment à partir duquel les visages d'autres types commencent à ne plus bien être reconnus et soient catégorisés comme une seule catégorie extra-groupe. Nous avons donc testé des nourrissons de 3 mois et des nourrissons de 9 mois avec une paire de voyelle déjà utilisée par le passé et pour laquelle les nourrissons regardent plus la vidéo de la locutrice correspondant à la voyelle jouée (Kuhl & Meltzoff, 1988 ; Yeung & Werker, 2013). Nous avons testé les capacités de correspondance audio-visuelle avec les voyelles /i/ et /u/, paire déjà utilisée dans les études mentionnées ci-dessus. Pour s'assurer de la fiabilité de notre étude, nous avons tout d'abord voulu tester les nourrissons sur la tâche de correspondance audio-visuelle avec des visages de même type qu'eux. Les précédentes études utilisant ce même paradigme expérimental montrent qu'à partir de 2 mois (Patterson & Werker, 2003) et jusqu'à 4.5 mois (Kuhl & Meltzoff, 1982, 1984 ; Patterson & Werker, 1999), les nourrissons regardent plus longtemps la vidéo de la locutrice qui correspond avec la voyelle jouée. Une étude utilisant toutefois un autre paradigme (Walton & Bower, 1993) nous montre que les nourrissons de 6-8 mois préfèrent toujours regarder la vidéo de la locutrice correspondant à la voyelle jouée. Bien qu'il soit probable que les nourrissons plus âgés soient toujours capables de faire cette correspondance audio-visuelle, aucune étude n'a à notre connaissance étudié cette tâche avec un paradigme de préférence visuelle sur des nourrissons de plus de 4.5 mois. Nous testerons ensuite la même tâche et les mêmes groupes d'âges (3 et 9 mois) avec des vidéos de locutrices d'un type de visage non familier pour les nourrissons. Nous avons choisi d'utiliser des vidéos de locutrices de type indien. Les personnes d'origine indienne sont relativement peu fréquentes à Grenoble, nous permettant d'avoir plus de possibilités de tester des nourrissons non exposés à ce type de visage. Nous avons donc différentes hypothèses selon le groupe d'âge et le type de visage utilisé :

- Tout d'abord, nous nous attendons à ce que les nourrissons de 3 mois regardent plus la vidéo de la locutrice qui correspond à la voyelle jouée lorsque la locutrice est du même

type de visage que les nourrissons, conformément à la littérature. Si cette hypothèse n'est pas vérifiée, cela remettrait en cause notre protocole et il serait difficile voire impossible pour nous de conclure sur les autres conditions.

- Pour les nourrissons de 9 mois et toujours pour les vidéos présentant des locutrices de même type, on s'attend potentiellement à ce que les nourrissons regardent plus le visage de la locutrice qui correspond à la voyelle jouée, comme à 3 mois. Cependant, Mercure et al. (2018) ont montré qu'à partir de 6.6 mois, les nourrissons présentent plus d'attention vers la bouche des visages présentant une incongruence audio-visuelle. L'étude de Lewkowicz & Pons (2012) montre aussi que des nourrissons de 10-12 mois regardent la vidéo d'une locutrice qui ne correspond pas un audio précédemment joué. Ces résultats nous indiquent que l'on peut s'attendre à ce que les nourrissons de 9 mois soient toujours en mesure de faire la correspondance audio-visuelle mais que cela se manifeste cette fois-ci par plus de regard vers la vidéo de la locutrice qui ne correspond pas la voyelle jouée plutôt que vers la vidéo de la locutrice qui correspond à la voyelle jouée. Toutefois, aucun résultat n'existant pour ce groupe d'âge dans la littérature, il est possible que cette tâche soit difficilement compatible avec ce groupe d'âge.
- On peut s'attendre à ce que les nourrissons de 3 mois testés avec un visage d'autre type n'aient pas un comportement différent des nourrissons de 3 mois testés avec un visage de même type. En effet, le rétrécissement perceptif n'est pas encore établi. Cependant, bien que la reconnaissance de visage d'autres types soit toujours possible, il existe déjà une préférence pour les visages de même type chez les nourrissons de cet âge (Kelly et al., 2005 ; Bar-Haim et al., 2006 ; Kelly, et al., 2007 b). Il est possible que le traitement des visages de même et d'autres types soit déjà suffisamment différent pour impacter la tâche de correspondance audio-visuelle.
- Si le rétrécissement perceptif impacte les capacités d'association audio-visuelle pour la parole, on peut s'attendre à ce que les nourrissons de 9 mois testés avec un visage d'autre type aient un temps de regard vers la vidéo qui correspond à la voyelle jouée et vers la vidéo qui ne correspond pas à la voyelle jouée assimilable au hasard. Cependant, les voyelles utilisées ici restent des sons connus des nourrissons et quotidiennement entendus contrairement aux vocalisations de singes utilisées par Lewkowicz & Ghazanfar (2008). De plus, bien que traité différemment des visages de même type, les visages d'autres types restent des visages. Aussi, contrairement aux études de Lewkowicz & Pons (2012) et Kubicek et al. (2014), les stimuli audios sont de simples voyelles et non pas un discours complet. Il est donc possible que la tâche soit toujours

réalisable par les nourrissons de 9 mois pour des visages d'autres type. Toutefois, les interprétations dépendront aussi du résultat obtenu sur le groupe de nourrisson de 9 mois testé avec des visages de même type.

II. Matériel et méthodes

II.1. Participants

95 nourrissons ont été testés, 60 dans l'étude avec les visages de même type et 35 dans l'étude avec les visages d'autre type. Nous avons retiré de l'analyse finale les nourrissons qui ne sont que minoritairement exposés à des visages de type européens de la condition avec visages de même type ($N = 3$). De même, nous avons retiré les nourrissons qui ne sont minoritairement exposés à des visages de type européens de la condition avec les visages d'autre type ($N = 1$). Conformément au protocole de Patterson & Werker (1999, 2003), nous avons aussi retiré de l'analyse les nourrissons qui ne regardent pas les deux stimuli lorsque présenté simultanément durant la familiarisation ($N = 4$). Les critères d'exclusion de Patterson & Werker (1999, 2003) comprennent aussi les nourrissons qui regardent moins d'une minute lors de la phase de test et les nourrissons qui ne regardent qu'un seul côté lors de la phase de test mais aucun des nourrissons de notre étude ne remplissait l'un de ces critères. Nous avons retiré des nourrissons supplémentaires à cause de problèmes technique lors de l'enregistrement vidéo ($N = 2$).

85 nourrissons ont donc été inclus dans l'analyse finale. Pour l'étude avec les visages de même type : 26 nourrissons de 3 mois (M âge = 113 jours, SD âge = 7.1 jours, 9 filles) et 25 nourrissons de 9 mois (M âge = 279 jours, SD âge = 8.0 jours, 10 filles). Pour l'étude avec les visages d'autre type : 26 nourrissons de 3 mois (M âge = 111 jours, SD âge = 6.6 jours, 12 filles) et 8 nourrissons de 9 mois (M âge = 279 jours, SD âge = 3.7 jours, 6 filles).

II.2. Stimuli

Pour créer les stimuli de tests, nous avons enregistré quatre locutrices françaises. Deux étaient de type européen, deux de type indien. Une seule session d'enregistrement a eu lieu pour toutes les locutrices pour s'assurer d'une homogénéité entre les enregistrements. Nous avons enregistré les locutrices en train de prononcer plusieurs occurrences du son /u/ et plusieurs occurrences du son /i/, devant un fond bleu.

Nous avons ensuite soigneusement sélectionné une occurrence visuelle de /i/ et une occurrence visuelle de /u/ pour chaque locutrice. Afin que la durée d'ouverture de la bouche soit consistante entre les différentes occurrences et les différentes locutrices tout en gardant un rendu fluide, nous avons doublé certaines images des vidéos pour certaines occurrences. Ainsi, la durée entre

l'ignition de l'ouverture de la bouche et sa fermeture était constante entre toutes les occurrences, d'une durée de 1.92 secondes. Nous avons aussi laissé des images avec un visage neutre, bouche fermée, avant (14 images ; 0.56 secondes) et après (13 images ; 0.52 secondes) l'ouverture de la bouche. Ainsi, toutes nos vidéos avaient une durée de 3 secondes, conformément aux stimuli de Patterson & Werker (1999 ; 2003), et l'ouverture et la fermeture étaient synchronisées entre toutes les vidéos.

Une cinquième locutrice a été enregistrée en chambre anéchoïque afin d'obtenir des enregistrements audios pour doubler les vidéos. La locutrice a produit plusieurs occurrences de /i/ et de /u/ en se calant sur la temporalité des vidéos. Une occurrence de /i/ et une occurrence de /u/ ont été utilisées pour doubler toutes les vidéos. La durée de l'audio était plus courte que la durée d'ouverture de la bouche afin de respecter au mieux la dynamique naturelle de prononciation des voyelles (M /i/ vidéo = 1.92 s, M /u/ vidéo = 1.92 s, M /i/ audio = 0.44 s, M /u/ audio = 0.52 s). Avoir des enregistrements provenant d'une cinquième locutrice ne figurant pas sur les vidéos permet d'éviter toute temporalité trop « parfaite » pour un stimulus.

Les stimuli étaient donc composés de deux stimuli visuels par locutrices (deux locutrices de type européen et deux locutrices de types indien, avec une occurrence de /i/ et une occurrence de /u/ par locutrice) et de deux stimuli audio (une occurrence de /i/ et une occurrence de /u/) étant synchrones avec chaque stimulus visuel.

II.3. Procédure

Un nourrisson ne voyait qu'une seule et unique locutrice durant toute l'étude, la locutrice était contrebalancée entre les nourrissons. La procédure commençait par une phase de familiarisation (voir Figure 22). Lors de celle-ci, un stimulus visuel – la vidéo de 3 secondes d'une locutrice prononçant soit /i/ soit /u/ – était présenté d'un côté de l'écran, trois fois consécutives (pour une durée de 9 secondes) et sans bande son. Le fond était noir et le côté sans visage était donc entièrement noir. Ensuite, le second stimulus visuel était présenté de l'autre côté de l'écran, de la même manière. Puis, les deux stimuli visuels apparaissaient simultanément, du même côté qu'ils étaient précédemment apparus, toujours sans son, pour trois occurrences chacun et simultanées (pour à nouveau une durée de 9 secondes). Cette phase était immédiatement suivie d'un écran noir, pour une durée de 3 secondes. Après ces 30 premières secondes, la phase de test commençait. Lors de la phase de test, les deux stimuli visuels apparaissaient simultanément, du même côté que durant la familiarisation, pour 40 répétitions (soit une durée de 2 minutes). Cependant, à la différence de la phase de familiarisation, un stimulus audio était cette fois-ci joué, de manière synchrone avec les occurrences visuelles. Le stimulus audio était

contrebalancé entre les nourrissons mais un nourrisson n'entendait toujours qu'un seul stimulus audio au cours de l'étude. De même, l'ordre de présentation des stimuli visuels et leurs positions étaient contrebalancés entre les nourrissons.

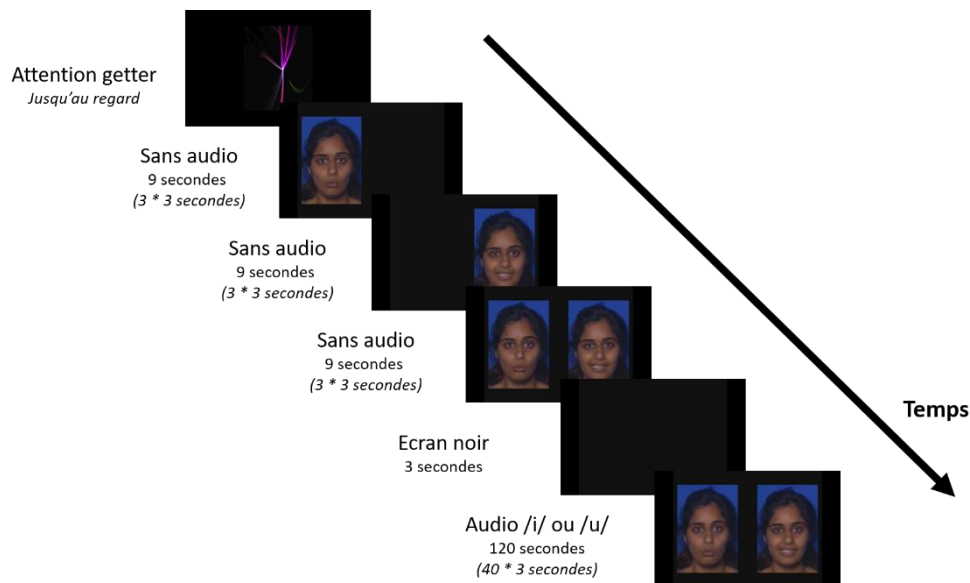


Figure 22. Protocole expérimental utilisé dans l'étude 1.

Le nourrisson était devant l'écran sur les genoux d'un parent. L'étude commençait par un *attention getter* (stimulus visuel permettant de capter l'attention). L'expérimentateur attendait que le nourrisson regarde l'écran avant de lancer la suite du protocole. Le temps de trois répétitions, la vidéo silencieuse d'une locutrice prononçant le son /i/ ou /u/ (selon contrebalancement) apparaissait à droite ou à gauche de l'écran (selon contrebalancement). Le temps de trois nouvelles répétitions, une vidéo silencieuse de la même locutrice apparaissait de l'autre côté de l'écran prononçant l'autre voyelle. Puis, les deux vidéos apparaissaient simultanément du même côté que précédemment, pour trois répétitions chacune. Les deux vidéos laissaient ensuite place à un écran noir pendant 3 secondes. Enfin, les deux vidéos réapparaissaient simultanément, du même côté que précédemment, pour 40 répétitions. Cette fois-ci, un audio, /i/ ou /u/ (selon contrebalancement) était joué en synchronie avec les vidéos.

II.4. Codage et analyse

L'ensemble des vidéos ont été codés à l'aveugle par deux différents codeurs avec une précision de 40 ms. Pour la condition avec les visages de même type, 20 % des vidéos ont été codés par un troisième codeur. Le coefficient de corrélation de Pearson entre les résultats obtenus par les deux codeurs principaux et le troisième codeur est de .998, indiquant un accord presque total.

La récolte de données étant encore en cours pour le groupe de nourrissons de 9 mois dans la condition avec les visages d'autre type (N=8), l'analyse de cette condition sera très succincte.

III. Résultats

III.1. Phase de familiarisation

Un test de wilcoxon par type de visage ne montre pas de différence chez les nourrissons de 3 mois entre le temps de regard vers le /i/ et le temps de regard vers le /u/ lorsque les deux stimuli visuels sont présentés simultanément durant la phase de familiarisation (p pour les visages de même type = .784 ; p pour les visages d'autre type = .276). Il n'y a donc pas de préférence significative pour l'un ou l'autre de ces stimuli lorsque présentés sans son. Une différence significative de temps de regard durant la phase de test devrait donc résulter de l'association entre l'audio et le visuel.

Pour les nourrissons de 9 mois, on observe une différence significative pour ce temps de regard dans la condition avec visage de même type ($p = .027$). Plus précisément, les nourrissons de 9 mois dans la condition visage de même type passent plus de temps à regarder le visage prononçant le son /u/ ($M = 4256$ ms) que le visage prononçant le son /i/ ($M = 3600$ ms). Cette préférence pour le visage prononçant /u/ peut être soit dû à une préférence plus importante de la forme de bouche nécessaire pour prononcer /u/ par rapport à la forme de bouche nécessaire pour prononcer /i/, soit de la préférence d'une vidéo par rapport à l'autre. Le fait que le temps de regard vers le /u/ soit plus long pour les deux locutrices diminue le risque que cette préférence soit dû au fait qu'une vidéo attirerait plus le regard par rapport à une autre. Il est donc possible qu'à 9 mois, les nourrissons préfèrent la forme de bouche nécessaire pour prononcer le son /u/ par rapport à la forme de bouche nécessaire pour prononcer le son /i/. Si cette préférence pour le visage prononçant /u/ est au moins en partie indépendante de l'audio joué, il est possible qu'elle masque une éventuelle capacité de correspondance audio-visuelle des nourrissons de 9 mois lors de la phase de test.

III.2. Phase de test

Les nourrissons de 3 mois passent en moyenne 90.39% du temps à regarder la phase de test et les nourrissons de 9 mois en moyenne 84.32%. Un test de wilcoxon par type de visage et par groupe d'âge ne montre pas d'effet significatif dû à la locutrice sur le temps de regard durant la phase de test (toutes les p -values > .05). Nous nous sommes ensuite intéressés uniquement au temps passé regarder les visages et nous avons calculé la proportion de temps passé à regarder le visage qui correspond et le visage qui ne correspond pas à la voyelle jouée. Pour cela, nous

avons simplement divisé le temps passé à regarder chaque visage avec le temps total passé à regarder un visage. Dans la suite de cette section résultat, lorsque nous exprimerons un pourcentage de temps de regard, celui-ci représentera donc le temps de regard proportionnel vers le visage. Nous avons ensuite comparé cette proportion de temps, exprimé en pourcentage, au niveau de chance de 50%.

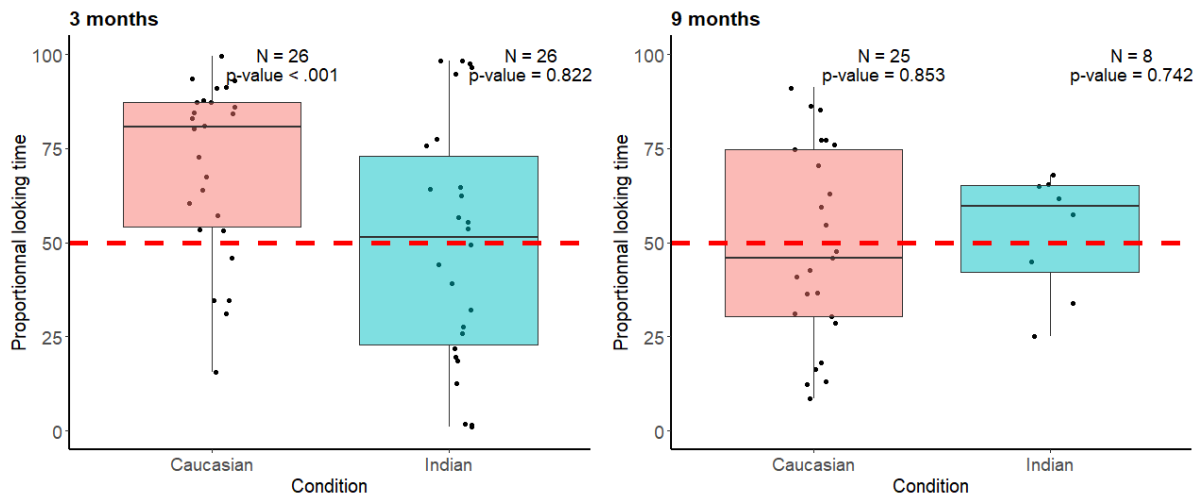


Figure 23. Pourcentage de temps passé à regarder le visage qui correspond à la voyelle jouée par groupe d'âge et par type de visage.

Ce pourcentage est relatif au temps total passé à regarder les stimuli, c'est-à-dire au temps total passé à regarder les visages. Les points noirs représentent les scores individuels ; la dispersion horizontale des points ne représente rien et est uniquement présente à titre de clarté. La ligne horizontale rouge représente le niveau de chance (autant de temps passé à regarder le visage qui correspond que de temps de regard passé à regarder le visage qui ne correspond pas). A gauche, résultats pour les nourrissons de 3 mois. A droite, résultats pour les nourrissons de 9 mois. En rouge clair, résultats pour la condition avec les visages de même type. En bleu clair, résultats pour la condition avec les visages d'un autre type.

III.3. Résultats pour les nourrissons de 3 mois

Tout d'abord, nous avons conduit une ANOVA comprenant quatre facteurs à deux niveaux (pour le sexe de l'enfant, la voyelle jouée, le côté de la vidéo qui correspond à la voyelle et le type de la locutrice) sur le pourcentage de temps passé à regarder le visage qui correspond. L'ANOVA révèle un effet principal du type de la locutrice ($F(1,42) = 4.936, p = 0.032$) mais aucun autre effet principal ni d'effet d'interaction (toutes les p-values > 0.05).

III.3.a. Pour le groupe exposé aux visages de même type

Les nourrissons de 3 mois, dans la condition avec des visages de leur type, regardent en moyenne 69.97% ($SD = 23.03$) du temps le visage qui correspond à la voyelle jouée, ce qui est significativement différent du niveau de chance (test de wilcoxon : $p < .001$; *taille d'effet* = .650 ; voir Figure 23 gauche). Sur les 26 nourrissons testés, 21 nourrissons ont plus regardé le visage congruent que le visage incongruent ce qui est différent de la distribution attendue par le hasard (test binomial, $p = .002$). Il n'y a pas de différence significative de pourcentage de temps passé à regarder le visage congruent en fonction de la locutrice (test de wilcoxon : $p = .799$). Le temps passé à regarder le visage congruent est significativement différent du niveau de chance quel que soit la locutrice (test de wilcoxon : pour la locutrice 1, $N = 15$, $p = .010$; pour la locutrice 2, $N = 11$, $p = .04$). Il n'y a pas non plus de différence significative de pourcentage de temps passé à regarder le visage congruent en fonction de la voyelle jouée (test de wilcoxon : $p = 1$; voir Figure 24 gauche). Le pourcentage de temps passé à regarder le visage congruent est significativement différent du niveau de chance pour la voyelle /u/ (test de wilcoxon : $N = 14$, $p = .002$) et est marginalement significatif pour la voyelle /i/ (test de wilcoxon : $N = 12$, $p = .052$). Nous avons ensuite analysé séparément la 1^{ère} et la 2^{ème} minute du test. Le pourcentage de temps passé à regarder le visage congruent est significativement supérieur au niveau de chance à la 1^{ère} (M PTLT = 67.18, test de wilcoxon : $p = .006$) et à la 2^{ème} minute (M PTLT = 74.62, test de wilcoxon : $p < .001$). La hausse entre la 1^{ère} et la 2^{ème} minute n'est pas significative (test de wilcoxon : $p = .237$). Lorsque découpé en tranche de 30 secondes, le pourcentage de temps passé à regarder le visage congruent n'est pas significativement différent du niveau de chance durant les 30 premières secondes (test de wilcoxon : $p = .071$) mais l'est durant toutes les autres tranches de 30 secondes (tests de wilcoxon : toutes les p -values $< .05$; voir Figure 25 gauche).

III.3.b. Pour le groupe exposé aux visages d'un autre type

Les nourrissons de 3 mois, dans la condition avec des visages d'autre type, regardent en moyenne 49.60% ($SD = 32.18$) du temps le visage congruent, ce qui n'est pas significativement différent du niveau de chance (test de wilcoxon : $p = .822$; *taille d'effet* = .047 ; voir Figure 23 gauche). De plus, sur les 26 nourrissons testés, il y a autant de nourrissons (soit 13) qui ont plus regardé le visage congruent que de nourrissons qui ont plus regardé le visage incongruent ce qui n'est donc pas significativement différent de la distribution attendue par le hasard (test binomial, $p = 1$). Il n'y a pas de différence significative de temps de regard vers le visage congruent en fonction de la locutrice (test de wilcoxon : $p = .16$). Il n'y a pas non plus de

différence significative de pourcentage de temps passé à regarder le visage congruent en fonction de la voyelle jouée (test de wilcoxon : $p = .297$; voir Figure 24 droite). Nous avons ensuite analysé séparément la 1^{ère} et la 2^{ème} minute du test. Le pourcentage de temps passé à regarder le visage congruent n'est pas significativement différent du niveau de chance, ni à la 1^{ère} (M PTLT = 45.74, test de wilcoxon : $p = .501$) ni à la 2^{ème} minute (M PTLT = 54.37, test de wilcoxon : $p = .542$). La hausse entre la 1^{ère} et la 2^{ème} minute n'est pas significative (test de wilcoxon : $p = .360$). Lorsque découpé en tranche de 30 secondes, le pourcentage de temps passé à regarder le visage congruent n'est significativement différent du niveau de chance dans aucune tranche de 30 secondes (tests de wilcoxon : toutes les p-values < .05 ; voir Figure 25 droite).

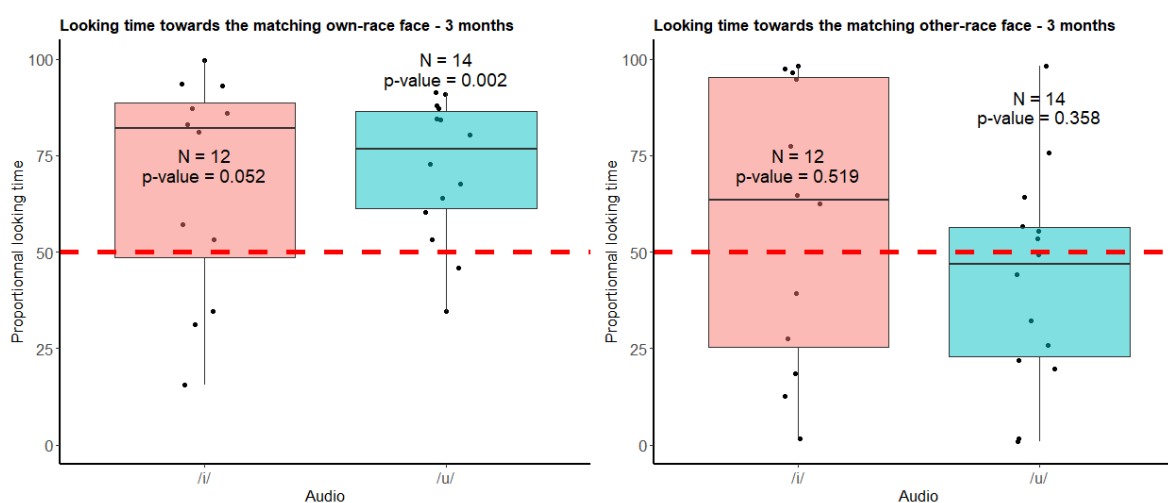


Figure 24. Pourcentage de temps passé à regarder le visage correspondant à la voyelle jouée pour les nourrissons de 3 mois par type de visage et par voyelle.

Ce pourcentage est relatif au temps total passé à regarder les stimuli, c'est-à-dire au temps total passé à regarder les visages. A gauche, les données pour la condition avec visage de même type ; à droite, avec visage d'autre type. En rouge clair, les données lorsque le son /i/ est joué ; en bleu clair, les données lorsque le son /u/ est joué. Les points noirs représentent les scores individuels ; la dispersion horizontale des points ne représente rien et est uniquement présente à titre de clarté. La ligne horizontale rouge représente le niveau de chance (autant de temps passé à regarder le visage congruent que de temps de regard passé à regarder le visage incongruent).

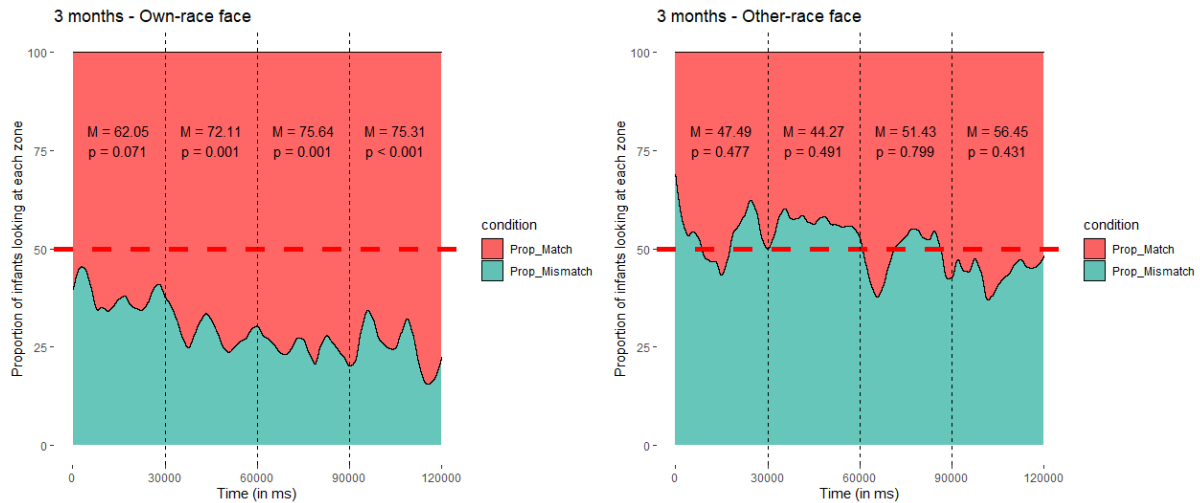


Figure 25. Evolution du pourcentage de temps passé à regarder chaque visage, pour les nourrissons de 3 mois.

Ce pourcentage est relatif au temps total passé à regarder les stimuli, c'est-à-dire au temps total passé à regarder les visages. A gauche, les données pour la condition avec visage de même type ; à droite, avec visage d'autre type. En rouge clair, la proportion de temps passé à regarder le visage congruent ; en bleu clair, la proportion de temps passé à regarder le visage incongruent. La ligne horizontale rouge représente le niveau de chance (autant de temps passé à regarder le visage congruent que de temps de regard passé à regarder le visage incongruent).

III.4. Résultats pour les nourrissons de 9 mois

III.4.a. Pour le groupe exposé aux visages de même type

Les nourrissons de 9 mois, dans la condition avec des visages de leur type, regardent en moyenne 48.91% ($SD = 25.88$) du temps le visage qui correspond à la voyelle jouée, ce qui n'est pas significativement différent du niveau de chance (test de wilcoxon : $p = .853$; *taille d'effet* = .040 ; voir Figure 23 gauche). De plus, sur les 25 nourrissons testés, 11 nourrissons ont plus regardé le visage congruent que le visage incongruent, ce qui n'est pas différent de la distribution attendue par le hasard (test binomial, $p = .69$). Il n'y a pas de différence significative de temps de regard vers le visage congruent en fonction de la locutrice (test de wilcoxon : $p = .81$). Il n'y a pas non plus de différence significative de pourcentage de temps passé à regarder le visage congruent en fonction de la voyelle jouée (test de wilcoxon : $p = .11$; voir Figure 26 gauche). On remarque en revanche mais de manière non significative, que la vidéo de la locutrice prononçant la voyelle /u/ est systématiquement la plus regardée ($M = 58.70$; test de wilcoxon : $p = .096$), qu'elle corresponde avec la voyelle jouée ($M = 57.21$) ou non ($M = 60.32$)

et quelle que soit la locutrice (M pour la locutrice 1 = 60.30 ; M pour la locutrice 2 = 53.61). On retrouve donc la même préférence pour la vidéo de la locutrice produisant le son /u/ durant le test que durant la familiarisation. Nous avons ensuite analysé séparément la 1^{ère} et la 2^{ème} minute du test. Le pourcentage de temps passé à regarder le visage congruent n'est pas significativement différent du niveau de chance, ni à la 1^{ère} (M PTLT = 47.55, test de wilcoxon : $p = .691$) ni à la 2^{ème} minute (M PTLT = 50.13, test de wilcoxon : $p = .979$). La hausse entre la 1^{ère} et la 2^{ème} minute n'est pas significative (test de wilcoxon : $p = .863$). Lorsque découpé en tranche de 30 secondes, le pourcentage de temps passé à regarder le visage congruent n'est significativement différent du niveau de chance dans aucune tranche de 30 secondes (tests de wilcoxon : toutes les p -values < .05 ; voir Figure 27 gauche).

III.4.b. Pour le groupe exposé aux visages d'un autre type

Les nourrissons de 9 mois, dans la condition avec des visages d'autre type, regardent en moyenne 52.68% ($SD = 16.14$) du temps le visage congruent, ce qui n'est pas significativement différent du niveau de chance (test de wilcoxon : $p = .742$; *taille d'effet* = .149 ; Figure 23 droite). De plus, sur les 8 nourrissons testés, 5 nourrissons ont plus regardé le visage congruent que le visage incongruent ce qui n'est différent de la distribution attendue par le hasard (test binomial, $p = .73$). Compte tenu de la faible taille d'échantillon (la récolte de données étant toujours en cours), nous avons choisi de ne pas inclure les analyses complémentaires (sur les effets de locutrices, de voyelles et concernant l'évolution temporelle ; mais voir Figure 26 droite et Figure 27 droite pour les résultats préliminaires).

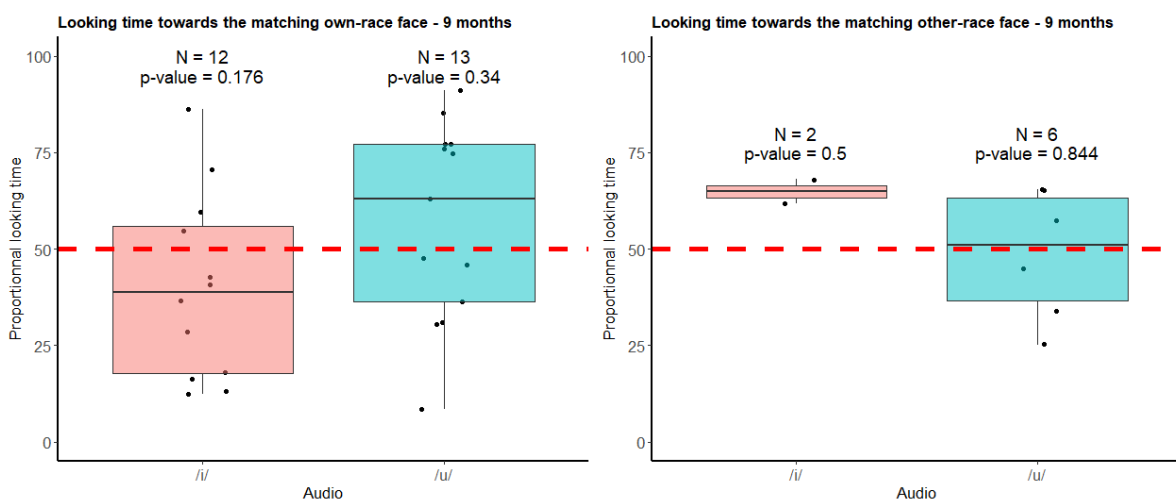


Figure 26. Pourcentage de temps passé à regarder le visage correspondant à la voyelle jouée pour les nourrissons de 9 mois par type de visage et par son joué.

Ce pourcentage est relatif au temps total passé à regarder les stimuli, c'est-à-dire au temps total passé à regarder les visages. A gauche, les données pour la condition avec visage de même type ; à droite, avec visage d'autre type. En rouge clair, les données lorsque le son /i/ est joué ; en bleu clair, les données lorsque le son /u/ est joué. Les points noirs représentent les scores individuels ; la dispersion horizontale des points ne représente rien et est uniquement présente à titre de clarté. La ligne horizontale rouge représente le niveau de chance (autant de temps passé à regarder le visage congruent que de temps de regard passé à regarder le visage incongruent).

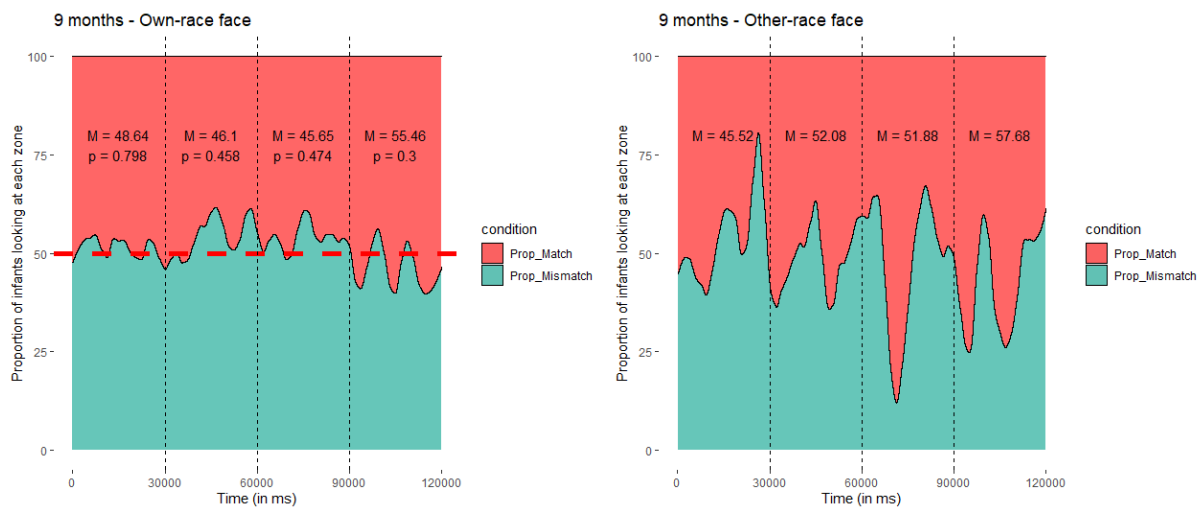


Figure 27. Evolution du pourcentage de temps passé à regarder chaque visage, pour les nourrissons de 9 mois.

Ce pourcentage est relatif au temps total passé à regarder les stimuli, c'est-à-dire au temps total passé à regarder les visages. A gauche, les données pour la condition avec visage de même type ; à droite, avec visage d'autre type. En rouge clair, la proportion de temps passé à regarder le visage congruent ; en bleu clair, la proportion de temps passé à regarder le visage incongruent. La ligne horizontale rouge représente le niveau de chance (autant de temps passé à regarder le visage congruent que de temps de regard passé à regarder le visage incongruent).

IV. Discussion

Nous avons évalué l'impact du type de visage sur une tâche de correspondance audio-visuelle de voyelle chez les nourrissons. Cette tâche a déjà été répliquée plusieurs fois via le même type de protocole et les résultats sont consistants (Kuhl & Meltzoff, 1982, 1984, 1988 ; Patterson & Werker 1999, 2003 ; Yeung & Werker 2013). Les nourrissons sont, entre 2 et 4.5 mois, capables de faire correspondre la vidéo d'un locuteur prononçant une voyelle avec sa conséquence audio,

notamment pour les voyelles /i/ et /u/ présentement testées. Cette tâche n'a jusqu'alors qu'été testé avec des visages de même type que ceux de nourrissons.

IV.1. Un impact du type de visage précoce sur la capacité de correspondance phonémique

Tout d'abord, nous répliquons une fois de plus les résultats obtenus précédemment pour ce type de tâche. Les résultats de notre étude montrent qu'à 3 mois, les nourrissons sont capables de faire correspondre une voyelle présentée dans la modalité audio et dans la modalité visuelle lorsque le visage présenté est un visage de leur type. Cela permet de confirmer la validité de notre étude et donc de nos résultats dans les autres conditions.

A 3 mois dans la condition avec un visage d'autre type, nous observons que les nourrissons regardent autant le visage congruent que le visage incongruent, un type de visage nouveau semble empêcher les participants de faire un matching entre le son entendu et le mouvement audio-visuel. Observer une différence comportementale en fonction du type de visage dès 3 mois peut sembler surprenant puisque les effets typiques du rétrécissement perceptif pour les visages s'observent plus généralement à partir de 9 mois (Kelly et al., 2005, 2009). Cependant, des préférences pour les visages de même type sont présentes dès 3 mois (Bar-Haim et al., 2006 ; Kelly, et al., 2005, 2007b). Bien que ce phénomène soit généralement mis en évidence vers 9 mois, Sangrioli & De Schonen (2004) ont montré que la capacité à reconnaître des visages d'autres types pouvait être impacté dès 3 mois. De plus, Kubicek et al. (2014) ont étudié l'impact du rétrécissement perceptif multi-sensoriel et ont mis en évidence des différences de traitement entre stimuli de parole natifs et non-natifs dès 6 mois soit 4 mois plus tôt que mis en évidence par Lewkowicz & Pons (2013), seulement en utilisant des temps de familiarisation légèrement plus long. Selon le type de stimulus et le protocole utilisé, il est donc possible de trouver des marqueurs du rétrécissement perceptif plus précocement qu'usuellement trouvés. Xiao et al. (2017b) ont montré que dès 9 mois, les nourrissons associent les visages de leurs types à une musique joyeuse et les visages d'autres types à une musique triste. Cela peut être interpréter en termes de familiarité, les musiques joyeuses étant plus familières dans l'environnement d'un nourrisson que les musiques tristes. C'est une illustration qu'une exposition relativement occasionnelle, à des sons de musique, est suffisante pour créer des attentes chez les nourrissons à partir de 9 mois. On peut donc s'attendre à ce qu'une exposition plus régulière puisse être à l'origine d'attentes plus précoces. Dès la naissance, les nourrissons sont très fréquemment exposés aux sons de leur langues. Il est alors relativement sûr de supposer que les liens entre les types de visages et les sons de langues se fassent plus rapidement

qu'entre les types de visage et la musique. Il est donc envisageable que le type de visage ait une influence sur les capacités d'association audio-visuelles de la parole dès 3 mois. Les visages qui parlent sont un élément essentiel de l'environnement des nourrissons ; cette expérience se fait presque uniquement avec des visages de même type (Rennels & Davis, 2008 ; Sugden, Mohamed-Ali, & Moulson, 2014). Les nourrissons ont peut-être dès 3 mois des attentes concernant la correspondance audio-visuelle uniquement pour les visages présents dans leur environnement. Nous avons uniquement intégré à notre étude les nourrissons pour lesquels aucune personne de type indien n'est présente dans l'environnement. Si des visages de type indien ont toutefois été vu par des nourrissons de notre étude, ces rencontres sont tout au plus anecdotiques et très probablement dans un contexte où la personne est masquée. Il est donc très probable que les visages de type indien de notre étude soient les premiers visages de type indien non masqués vu par les nourrissons de notre étude. Il est donc probable que ces visages non familiers nécessitent un temps d'exploration plus long que les visages familiers, temps durant lequel l'information audio-visuelle ne sera pas nécessairement prise en compte. On observe en effet que pour les nourrissons de 3 mois dans la condition avec les visages de même type, le temps de regard vers le visage qui correspond à la voyelle jouée n'est pas significativement différent du niveau de chance durant les 30 premières secondes, mais le devient dans les tranches de 30 secondes suivantes. Pareillement, Patterson & Werker (2003) trouve que le temps de regard vers le visage qui correspond à la voyelle jouée n'est pas significativement différent du niveau de chance durant la 1^{ère} minute chez les nourrissons de 2 mois, mais le devient dans la 2nd minute. Il semble donc que les nourrissons prennent un temps pour explorer le visage avant de traiter l'information audio-visuelle. De manière non significative mais qui mérite d'être mentionné, on observe une dynamique pour le temps de regard des nourrissons dans la condition avec des visages d'autre type qui pourrait confirmer cette hypothèse. En effet, le temps de regard vers le visage qui correspond à la voyelle jouée est dans cette condition le plus grand lors de la dernière tranche de 30 secondes. De la même manière qu'un temps de familiarisation plus long permet de restaurer la reconnaissance de visage de singes chez des nourrissons de 12 mois (Fair et al., 2012), il se peut qu'avec plus de temps de présentation, les nourrissons de 3 mois finissent par montrer une préférence pour le visage d'autre type qui correspond à la voyelle jouée. On peut également faire l'hypothèse qu'ils n'existent peut-être pas d'attentes de synchronie et de correspondance multisensorielle pour des visages d'autres types, jamais rencontrés auparavant.

IV.2. Une tâche peu adaptée pour les nourrissons de 9 mois ?

De manière surprenante, nous ne trouvons pas les mêmes résultats pour les nourrissons de 9 mois. En effet, ils ne regardent pas plus le visage qui correspond à la voyelle jouée lorsque ce visage est de même type qu'eux. Une hypothèse alternative était que les nourrissons de cet âge regarderaient plus le visage qui ne correspond pas avec la voyelle jouée, des éléments de la littérature indiquant que les nourrissons de 9 mois sont attirés par l'incongruence (Mercure et al., 2018). Cette hypothèse n'est pas vérifiée non plus, les nourrissons ne passant pas significativement plus de temps à regarder un visage plutôt que l'autre. Une autre possibilité aurait pu être que les nourrissons regardent le visage congruent pour une locutrice et le visage incongruent pour l'autre, masquant ainsi tout effet. Nous ne trouvons cependant pas d'effet d'interaction dû à la locutrice. En revanche, on observe une préférence des nourrissons pour la vidéo de la locutrice prononçant /u/, significative durant la phase de familiarisation silencieuse. Bien que celle-ci ne soit pas significative durant le test, elle est constante dans toutes les sous-conditions. Cette préférence pour le /u/ fut déjà mise en évidence, bien que de manière non significative, par Baier et al. (2007). En effet, ils trouvèrent que lorsque le /u/ et le /i/ sont présentés simultanément dans la modalité visuelle, le /u/ est plus regardé. Est-ce que cette préférence pour le visuel /u/ comparativement au visuel /i/ suffit à expliquer l'absence de différence significative et pourquoi les nourrissons ont cette préférence sont deux questions sans réponses franches. Il est à noter que la préférence que nous trouvons chez les nourrissons de 9 mois est plus forte que celle trouvée chez les nourrissons de 3 mois par Baier et al. (Baier et al., 2007 : $t(14) = 0.7418$, $p = 0.4705$; la présente étude : $t(23) = 2.40$, $p = 0.020$). Il est imaginable que cette préférence s'accroisse et interfère avec la préférence pour la congruence. Nous ne retrouvons cependant pas de différence de temps de regard significative à 3 mois lors de la familiarisation pour le /i/ ou pour le /u/ ($t(19) = -0.288$, $p = 0.775$).

Quant à répondre au pourquoi, il est possible que les protrusions soient des mouvements de lèvres plus courants chez les nourrissons de 9 mois que les mouvements d'étirements (notamment lors de la tétée). Nous avons montré en introduction de ce chapitre que Yeung & Werker (2013) ont mis en évidence des liens entre la perception de la congruence audio-visuelle et la forme des lèvres des nourrissons. En effet leurs résultats montrent que l'expérience motrice (dans le cas de leur étude, concomitante avec la tâche) affecte la préférence pour la congruence. De manière plus générale, des études chez le nourrisson et chez l'adulte montrent que l'écoute de parole ou de sons recrute des aires liées à la production vocale (Imada et al., 2006 ; Wilson, et al., 2004 ; Yuen et al., 2010). Il se peut donc qu'une expérience motrice plus fréquente avec

une forme de lèvres affecte la préférence pour une vidéo plutôt qu'une autre. Mugitani et al. (2008) montrent que dans leur étude, les nourrissons ont globalement plus regardé le stimulus correspondant à un mouvement de bouche qu'ils arrivent à produire plutôt que le stimulus correspondant à un mouvement de bouche qu'ils ne produisent pas. Un biais du même ordre pourrait avoir lieu dans notre étude, bien que dans une moindre mesure. En effet, Mugitani et al. (2008) observent que les nourrissons passent significativement plus de temps à regarder la vidéo qui correspond au son joué dans le cas où ce son correspond à une forme de bouche produite par l'enfant, ce qui n'est pas notre cas. Altvater-Mackensen et al. (2015) montrent dans leur étude que la préférence des nourrissons pour les vidéos congruentes dépend des voyelles étudiées. Les auteurs ont testé les paires de voyelles suivantes : /a/ vs. /o/ et /a/ vs. /e/. Pour la paire de voyelle /a/ vs. /e/, les auteurs ont trouvé une influence des capacités de production des nourrissons sur leur préférence pour les essais congruents ou incongruents. Plus précisément, les nourrissons qui ne produisent pas de sons similaires à ceux produits par les locutrices ont plus tendance à regarder les vidéos incongruentes tandis que ceux qui produisent des sons similaires ont plus tendance à regarder les vidéos congruentes. Il n'est donc pas impossible qu'un changement de préférence dépendant de l'expérience des nourrissons soit aussi à prendre en compte. Ces résultats sont en accord avec DePaolis et al. (2011, 2013) qui montre que les préférences des nourrissons (âgés d'entre 9 et 12 mois) pour certains sons sont dépendantes de leurs productions vocales. De manière générale, Altvater-Mackensen et al. (2015) ont trouvé que les nourrissons détectent la congruence pour la paire de voyelle /a/ vs. /o/ mais pas pour la paire de voyelle /a/ vs. /e/. Les auteurs avancent que la protusion des lèvres pour prononcer /o/ est peut-être à l'origine de cette différence. La forme des lèvres serait ici plus saillante et plus facilement identifiable que la forme des lèvres à l'origine du son /e/ ou /a/ par exemple. Cela expliquerait pourquoi les résultats ne sont pas homogènes pour toutes les paires de voyelles. Cette différence ne serait pas ici liée aux capacités de production des nourrissons puisque cette différence est indépendante des sons produits par les nourrissons (d'après un questionnaire rempli par les parents). Cependant, bien que la plupart des nourrissons de l'étude ne produisent pas le son /o/, ils produisent probablement des formes de lèvres similaires dans d'autres contextes (lors de la tétée par exemple). Kubicek et al. (2014) ont aussi observé que les protrusions semblent être un mouvement de lèvres attrayants pour les nourrissons. Des nourrissons monolingues allemands exposés à des vidéos silencieuses de locutrices parlant français ou allemand ont préférés regarder la vidéo de la locutrice française à 4.5 et 6 mois. Les auteurs expliquent cette préférence par l'usage plus fréquent de protrusion de lèvres dans le français que dans l'allemand, protrusions qui attirent le regard des nourrissons (potentiellement

par leurs ressemblances aux mouvements de lèvres produits lors de bisous). Enfin, il est possible que l'incongruence soit plus facilement détectée selon la modalité des voyelles. Par exemple, peut-être que la voyelle audio /u/ sur la voyelle visuelle /i/ produit une incongruence plus facilement détectable que la voyelle audio /i/ sur la voyelle visuelle /u/. Rien n'étaye cette possibilité mais celle-ci n'est pas complètement excluable non plus.

Aussi, la tâche de correspondance phonémique est peut-être trop « simple » pour des nourrissons de cet âge qui se désintéressent de celle-ci. Il pourrait être pertinent d'évaluer la tâche de correspondance phonémique avec des syllabes de type CV plutôt que des voyelles isolées, ou bien des courtes séquences de syllabes. Ces stimuli plus complexes pourraient provoquer plus d'intérêt pour la congruence (ou l'incongruence) audio-visuel chez des nourrissons plus âgés.

Enfin, concernant les nourrissons ayant passé l'étude de correspondance audio-visuelle dans la condition avec des visages d'autre type, le faible effectif (N = 8) ne nous permet pour l'instant pas de porter de conclusions. Il est à noter que pour le moment, les nourrissons de 9 mois ne semblent pas regarder plus un visage que l'autre lorsque les visages sont d'un autre type. Nos résultats précédents nous indiquent que :

- les nourrissons de 3 mois dans la condition avec des visages d'autre type ne regardent pas de manière significativement différente du niveau de chance le visage congruent ou incongruent
- les nourrissons de 9 mois dans la condition avec des visages de même type ne regardent pas de manière significativement différente du niveau de chance le visage congruent ou incongruent

Il serait donc surprenant que les nourrissons de 9 mois dans la condition avec des visages d'autres types présentent une préférence pour un des deux visages. Il sera intéressant de voir si l'on retrouve une préférence pour le /u/ dans cette condition aussi.

Chapitre 6 : Effet McGurk sur des visages de même type et d'autre type

I. Introduction

Dans le chapitre précédent, nous nous sommes intéressés à l'impact du rétrécissement perceptif sur une tâche de correspondance audio-visuelle de voyelle. Dès 3 mois, avons mis en évidence un impact du type de visage sur une tâche simple qui peut être un indicateur du développement langagier futur de l'enfant (Altaver-Mackensen & Grossman, 2014 ; Altaver-Mackensen et al., 2015). Dans le cadre du présent chapitre, nous nous sommes intéressés à une tâche un peu plus complexe, notamment dans sa mise place, l'effet McGurk (McGurk & MacDonald, 1976).

I.1. L'effet McGurk

L'effet McGurk est une illusion perceptive créée de manière artificielle qui se produit lors de discordance entre les informations auditives et visuelles durant la perception de la parole. L'intégration audiovisuelle est alors perturbée et se traduit par la perception d'un stimulus autre que le stimulus audio joué. Ce phénomène sert souvent à mettre en évidence le caractère multimodal de la perception de la parole, puisqu'il montre que les deux canaux (audio et visuel) sont responsables de notre perception de la parole. Un exemple pour obtenir cet effet est un locuteur qui prononce le son /ka/ mais est doublé avec le son /pa/ de manière synchrone avec l'ouverture des lèvres. Dans ce cas précis d'effet McGurk, le son généralement perçu par les personnes regardant la vidéo sera /ta/, bien que la perception dépende du récepteur (Green et al., 1991). En effet, cette illusion perceptive présente une forte variabilité inter-individuelle (Benoit et al., 2010 ; Tremblay et al., 2007). Certains individus rapportent même ne jamais percevoir l'illusion (Basu Mallick et al., 2015). Une méta-analyse de 27 études (pour un total de 1044 participants) rapporte qu'environ 14.6% des participants ne sont pas sensibles à l'effet McGurk (Alsuis & Munhall, 2018). En revanche, la variabilité intra-individuelle est faible, la perception de l'effet par un individu étant stable avec le temps (Basu Mallick et al., 2015 trouvent un $r = .91$ pour un test un an après le premier test ; Strand et al., 2014 trouve un $r = .77$).

I.1.a. Paramètres jouant sur la sensibilité à l'effet

La sensibilité ou non des sujets à l'effet peut dépendre de plusieurs paramètres, à la fois dû au sujet et au stimulus, jouant sur la perception de l'effet McGurk. Feng et al. (2019) ont étudié la perception de l'effet McGurk chez des jumeaux (mono- et dizygotiques) et montrent que le

facteur génétique ne jouerait qu'un très faible effet (expliquant 5.5% de la variance). Inversement, une étude sur des musiciens de haut niveau montrent que ceux-ci ne présentent que très peu voire pas l'effet (Proverbio et al., 2016). Si la personne exposée à un stimulus McGurk joue un rôle dans l'ampleur des résultats, la personne produisant l'effet en est aussi responsable. Basu Mallick et al. (2015) ont trouvé que le report d'effet McGurk pouvait varier de 17 % à 58 % selon le locuteur employé. La nature du stimulus employé est aussi en grande partie responsable du résultat observé. Si de nombreuses paires audio-visuelles ont été utilisées pour tester le McGurk, les paires de types Consonne – Voyelle ou Voyelle – Consonne - Voyelle semblent les plus employées bien que l'effet puisse s'observer pour des voyelles seules ou pour des mots (Summerfield & McGrath, 1984 ; Dekle et al., 1992). La voyelle utilisée dans les combinaisons peut aussi influencer l'effet, Green et al. (1988) ayant montré que la voyelle /i/ induit plus d'effet McGurk que la voyelle /a/ qui elle-même en induit plus que la voyelle /u/ (lorsque combiné en V/g- voyelle/ A/b-voyelle/). Un autre facteur jouant sur l'intensité du McGurk provient de l'instruction donnée aux participants. Selon les études, les instructions demandent aux participants de reporter ce qu'ils entendent, ou ce que le locuteur dit, via des réponses ouvertes ou parmi un choix multiple (Hillock-Dunn et al., 2016 ; White et al., 2014 ; Green et al., 1991 ; Jordan & Thomas, 2001). Bien que la différence puisse sembler minime, l'attention sélective portée à une modalité plutôt qu'à une autre peut influencer l'intégration audio-visuelle (Buchan & Munhall, 2011). Aussi, le simple fait de proposer une réponse libre plutôt qu'une réponse au choix parmi un choix multiple engendre moins de report d'effet McGurk (Colin et al., 2005). Enfin, la définition même de l'effet McGurk n'est pas la même selon tous les auteurs. La définition stricte et originelle du terme (McGurk & MacDonald, 1976) implique que le stimulus doit entraîner la perception d'un troisième percept différent du percept audio ou visuel présenté. Ce troisième percept peut résulter d'une fusion entre le percept audio et visuel, comme c'est le cas lors d'un effet McGurk avec un stimulus V/ka/A/pa/, généralement perçu /ta/ ; il peut aussi résulter d'une combinaison, comme c'est le cas avec un stimulus V/ba/A/ga/, généralement perçu /bga/. Certains auteurs ont une vision plus souple du terme et considère comme preuve de l'effet McGurk toute perception rapportée différant de l'audio joué.

De manière intéressante, il n'est pas nécessaire que le sexe de la personne produisant le son soit le même que celui de la personne sur la vidéo (Green et al., 1991). On peut donc se demander quels éléments sont nécessaires ou non pour produire l'effet McGurk. Jordan et al. (2000) ont présenté à des participants adultes l'effet McGurk produit par un visage obscurci, comprenant 28 points lumineux placés principalement au niveau des muscles articulateurs. Les résultats

montrent que certains participants sont toujours sensibles à l'effet McGurk produit par ce type de dispositif, bien que moins nombreux que dans la condition classique où le visage est entièrement visible. L'effet McGurk peut donc être perçu même lorsque produit par un dispositif n'étant pas considéré comme un visage par les participants. Hietanen et al. (2001) ont utilisé des *scrambled faces* pour tester l'importance de la configuration du visage dans la perception de l'effet McGurk. La perception de l'effet McGurk par les participants ne fut pas affectée par les perturbations de la configuration puisque les sujets furent sensibles au McGurk indépendamment de la configuration des éléments du visage. Dans le même ordre d'idée, des études se sont intéressées à l'effet d'inversion sur la perception du McGurk. Nous avons précédemment vu que l'effet d'inversion perturbe grandement le traitement des visages, en particulier les traitements dit configuraux. Jordan & Bevan, 1997, ont trouvé que la perception de l'effet McGurk ne fut pas impactée par une inversion totale (rotation de 180°) mais fut en revanche amoindrie lors de présentation dans des orientations intermédiaires, entre 0 et 180°. Les mouvements des lèvres étant principalement des mouvements verticaux, une inversion totale résulte probablement en moins de perturbation du signal articulatoire qu'une orientation intermédiaire. Bertelson et al. (1994) ou encore Massaro & Cohen (1996) semblent confirmer une légère diminution de l'effet McGurk lorsque présenté via des visages inversés. Rosenblum et al. (2000) ont testé l'impact de l'effet Thatcher sur la perception de l'effet McGurk. L'effet McGurk est réduit, mais seulement lorsque le visage était présenté en configuration canonique. Ces différentes études nous indiquent qu'il n'est pas encore complètement clair à quel point les perturbations configurales perturbent l'intégration audio-visuelle, et ne nous indiquent pas si l'intégration audio-visuelle dans le cadre de l'effet McGurk dépend de sa production par un visage.

Un des paramètres jouant sur l'amplitude de l'effet McGurk provient de la population testée. En effet, regarder le visage d'un interlocuteur est un comportement lié à la culture de chaque pays. Par exemple, les japonais regardent peu le visage d'un interlocuteur, ceci étant considéré impoli (Nomura, 1984 ; cité par Sekiyama, 1997). L'effet McGurk étant conditionné par le regard sur le visage, l'amplitude de cet effet est généralement plus faible dans la population japonaise (Sekiyama, 1994) que dans la population européenne ou américaine (Dekle et al., 1992 ; Green et al., 1991 ; Massaro & Cohen, 1983 ; Rosenblum & Saldana, 1992). Puisqu'un environnement bruyant résulte en plus de regard vers la bouche pour faciliter la compréhension de la parole (Sumbly & Pollack, 1954 ; Erber, 1975 ; Peelle & Sommers, 2015), les études cherchant à produire l'effet McGurk, en particulier au sein de populations asiatiques, ajoutent

souvent du bruit à leur stimuli (Alm et al., 2009 ; Fixmer & Hawkins, 1998 ; Hardison, 1996 ; Jordan & Sergeant, 1998 ; Sekiyama & Tohkura, 1991 ; Ujiie et al., 2015 ; une possibilité est de diminuer l'importance de la composante audio du stimulus, tout simplement en baissant le son, voir Colin et al., 2002). De manière plus précise concernant les patterns de regard, il est à noter que les japonais accordent généralement moins d'importance à l'information articulatoire visuelle (notamment en contexte ambigu) et portent proportionnellement moins d'attention visuelle sur la bouche que sur les yeux (Sekiyama & Tohkura, 1991, 1993 ; Shinozaki et al., 2016). Si les adultes anglais portent plus d'attention à la bouche dès les premiers mouvements des lèvres (avant la production vocale), ce n'est en revanche pas le cas des adultes japonais qui regardent plus les yeux et le nez (Hisanaga et al., 2016).

1.1.b. L'effet McGurk chez le nourrisson

Bien que pouvant présenter des challenges, mesurer l'effet McGurk chez l'adulte est assez simple puisque l'on peut explicitement demander au participant ce qu'il comprend. Chez le nourrisson, il faudra en revanche passer par une mesure indirecte pour pouvoir déduire la sensibilité ou non à l'effet McGurk. Nous venons de voir que dès 2 mois, des nourrissons anglophones sont capables faire correspondre un mouvement de lèvres à sa conséquence audio et ont donc une perception audio-visuelle de la parole (Kuhl and Meltzoff, 1982, 1984 ; Patterson and Werker, 1999, 2002 ; chapitre précédent de cette thèse). Rosenblum et al. (1997) ont testé des nourrissons de 5 mois sur un effet d'intégration visuelle proche de l'effet McGurk. Dans cette étude, les chercheurs ont familiarisé des nourrissons anglophones à un stimulus visuel de /va/ avec un audio /va/ puis ont testé la réaction des nourrissons lorsque l'audio changeait, soit en /ba/ soit en /da/. Chez l'adulte, un stimulus audio-visuel V/va/A/ba/ est perçu dans la grande majorité des cas (98%) comme le son /va/ tandis que le même visuel avec un autre audio, soit V/va/A/da/, est lui perçu comme /da/. Après la familiarisation, les nourrissons présentaient un regain d'intérêt dans ce deuxième cas (V/va/A/da/) mais pas dans la première situation (V/va/A/ba/) présentant donc un pattern similaire à celui des adultes. De la même manière, Desjardins & Werker, 2004, montrèrent que les nourrissons d'un peu plus de 4 mois présentent déjà des signes de sensibilité à l'effet McGurk en utilisant le percept V/vi/A/bi/ ou V/bi/A/vi/ (tous les deux perçus comme /vi/ chez les adultes). Cependant, les résultats furent inconsistants entre le genre des nourrissons. Si les études de Rosenblum et al. (1997) et Desjardins & Werker (2004) impliquent une influence du visuel sur la perception audio, elles ne testent en revanche pas un effet McGurk dans le sens originel et strict du terme (McGurk & MacDonald, 1976 ; Tiippana, 2014). En effet, il s'agit dans les deux cas d'une prédominance

d'une modalité sur l'autre (soit le visuel prime sur l'audio, soit l'audio prime sur le visuel) mais il ne s'agit ni de fusion des deux signaux ni d'une combinaison. L'effet McGurk résultant en une fusion ou une combinaison des deux percepts est cependant sujet à variations, les adultes ne percevant pas tous le même stimulus final. Ce type d'effet est donc plus dur à mettre en évidence chez des nourrissons d'un point de vue méthodologique. Burnham & Dodd, 1996 testèrent l'effet McGurk chez des nourrissons âgés de 4 mois avec un stimulus résultant en une fusion chez les adultes, soit V/ga/A/ba/ (résultant généralement en la perception de /da/ ou /ta/). Via un score de fusion, les chercheurs interprétèrent que les nourrissons habitués à un McGurk V/ga/A/ba/ perçurent /da/, et donc que les nourrissons sont dès 4 mois sensibles à l'effet McGurk. Certains éléments méthodologiques limitent cependant quelque peu l'interprétation. Celle-ci se base en effet sur deux scores, un score audio et un score de fusion représentant respectivement la sensibilité soit au percept audio pur soit au percept issu de la fusion. Le score de fusion pour le groupe habitué au McGurk et le groupe habitué à un stimulus audio-visuel congruent (V/ba/A/ba/) sont relativement proches (85% pour le groupe test et 75% pour le groupe contrôle). De plus, ce score résulte de la différence de temps de regard entre la phase de test et la phase d'habituation, que cette différence soit positive ou négative. Burnham et Dodd ont donc interprété une hausse ou une baisse de temps de regard de la même manière. Concrètement, cela veut dire qu'après habituation à l'effet McGurk, un nourrisson qui regarde plus le percept audio /da/ ou un nourrisson qui regarde moins le percept audio /da/ contribuent tous les deux à augmenter le score de fusion et à conclure que les nourrissons sont sensibles au McGurk. L'absence d'hypothèse quant à la direction de cette différence limite les interprétations. Cependant, Burnham & Dodd, 2004, testèrent à nouveau des nourrissons sur un effet McGurk de type fusion mais cette fois-ci en s'intéressant au temps de fixation. Durant une phase d'habituation, une personne réelle produisait l'effet McGurk en se synchronisant sur l'audio joué. Puis, durant une phase de test, différents audio furent joués, la personne de l'habituation étant toujours présente mais immobile. Ils confirmèrent leurs résultats de 1996, montrant que les nourrissons de 4 mois $\frac{1}{2}$ présentent moins d'intérêt pour le son /ba/ que pour le son /da/ ou /ta/ lorsque habitués à un McGurk V/ga/A/ba/. Walker-Andrew & Gibson, 1986 (cité par Burnham & Dodd, 2004), ayant montré qu'un changement de modalité entre l'habituation et les tests produit un intérêt non pas pour la nouveauté mais pour la familiarité, la hausse d'intérêt des nourrissons de l'étude de Burnham & Dodd fut interprétée comme une préférence pour la familiarité. Cependant, s'il y a bien changement de stimuli entre la phase d'habituation et la phase de test, on pourra argumenter qu'il n'y a pas changement de modalité, la phase de test présentant simultanément un stimulus visuel et un stimulus audio. Cela rend

plus difficilement interprétable les temps de fixation vers /da/ et /ta/ en tant que préférence pour la familiarité. Il ressort de toutes ces études que bien que les nourrissons puissent présenter une sensibilité aux deux modalités dès 4-5 mois, il n'est pas clair dans quelles mesures ils sont sensibles à l'effet McGurk à cet âge. Le fait que de nombreux paramètres peuvent influencer l'impact de l'effet McGurk, que les nourrissons ne présentent pas nécessairement la même sensibilité en fonction de leur genre et la difficulté à mettre en place des protocoles et des mesures reflétant cette sensibilité chez le nourrisson rendent les interprétations difficiles.

Nous avons mentionné que des différences culturelles ont un impact sur la sensibilité à l'effet McGurk. Ces différences ont des conséquences observables dès l'enfance. En effet, entre 6 et 8 ans, l'effet McGurk devient de plus en plus fort chez les enfants monolingues anglais tandis qu'il reste relativement constant chez les enfants japonais (Sekiyama & Burnham, 2008). Les enfants japonais ont généralement des réponses biaisées vers l'audio aux stimuli de McGurk. On peut trouver des différences encore plus tôt dans le développement. Morisawa et al. (2013 ; cité par Sekiyama et al., 2021) ont montrés qu'un shift d'attention visuelle des yeux vers la bouche se produit vers les 12-13 mois chez les nourrissons japonais. Ce même shift attentionnel est observé dès 4-5 mois chez les nourrissons monolingues anglais ou français (Lewkowicz & Hansen-Tift, 2012 ; Morin-Lessard et al., 2019). Le développement des capacités audio-visuelles des nourrissons japonais ne présente donc pas la même trajectoire que celle des nourrissons anglais. Si les nourrissons anglophones ou francophones sont dès 2-3 mois capables de faire correspondre une voyelle vue et entendue (Kuhl & Meltzoff, 1982 ; Patterson & Werker, 1999, 2003 ; résultats du chapitre précédent), les nourrissons japonais ne peuvent le faire que vers 8 mois pour la voyelle /a/ et vers 11 mois pour la voyelle /i/ (Mugitani et al., 2006 ; Mugitani et al., 2004 ; cité par Sekiyama et al., (2021). Tout récemment, Sekiyama et al. (2021) ont évalué l'attention visuelle porté à la bouche et aux yeux d'une locutrice par des nourrissons et enfants japonais, en contexte bruité et normal. Tout d'abord, ils trouvent que progressivement plus d'attention est porté vers la bouche à partir de 10 mois, ce qui est légèrement plus précoce que Morisawa et al. (2013) mais toujours plus tardif que pour les nourrissons anglaise ou français. L'attention porté vers les yeux est accrue à l'âge adulte comparativement à la période entre 10 mois et 3 ans. De plus, ils trouvent que l'ajout de bruit provoque une hausse de l'attention visuelle vers la bouche seulement à 3 ans et pas avant. Cette hausse de l'attention visuelle lorsque le signal audio est bruité ne se retrouve pas chez les jeunes adultes de l'étude, contrairement à ce que laisse penser les résultats de l'étude de Sekiyama & Tohkura, 1991. Cependant, la nature des tâches de ces deux études est très différente : alors

qu'il suffit d'écouter une phrase dans l'étude de Morisawa et al. (2013), il faut reporter ce qui est entendu face à un effet McGurk dans l'étude de Sekiyama & Tohkura. Il est donc tout à fait explicable que la première tâche ne nécessite pas plus d'attention vers la bouche contrairement à la deuxième.

I.2. Objectif et hypothèses de l'étude 2

Ujiie et al. (2020) ont tout récemment étudié l'impact du type de visage dans la perception de l'effet McGurk chez des nourrissons japonais âgés de 8-9 mois. Pour cela, ils ont utilisé un stimulus de McGurk prononcé par une femme soit de type asiatique soit de type européen. Le McGurk était de type « pa – ta – ka » : la locutrice prononçait /ka/ mais l'audio était un enregistrement du son /pa/, ce qui résulte généralement chez le sujet adulte en la perception du percept fusionné /ta/. Une première partie de l'étude était conduite en ISPif. Cette étude montre une activation bilatérale des régions temporales lors d'un McGurk et l'activation de la région temporale gauche uniquement lors d'un stimuli audiovisuel congruent ; ces activations ne sont présentes lorsque le visage est de même type que les nourrissons mais pas lorsqu'il est d'autre type. Dans une deuxième partie, les chercheurs ont voulu voir si la différence de réponse cérébrale s'exprimait par une différence comportementale. Ils ont donc familiarisé les nourrissons au stimulus de McGurk puis ont présenté le percept audio /pa/ (entendu si les nourrissons ne sont pas sensibles au McGurk) et /ta/ (entendu si les nourrissons sont sensibles au McGurk), associé à des images de légumes (la modalité visuelle ne pouvant donc plus influencer sur la perception de l'audio). Selon le principe de préférence vers la nouveauté, le percept audio induisant une plus haute attention durant la phase de test est le percept audio qui n'est pas entendu durant la familiarisation. Si les nourrissons sont sensibles au McGurk, il devrait donc présenter une attention accrue pour le percept audio /pa/ durant la phase de test. C'est bien ce que Ujiie et al. (2020) retrouvent chez les nourrissons japonais de 8-9 mois mais uniquement lorsque le McGurk est produit par un visage de même type qu'eux (voir Figure 28). Lorsque le visage est d'autre type, les nourrissons prêtent autant d'attention aux deux stimuli durant la phase de test. Leur étude met donc en évidence une sensibilité neuronale et comportementale au McGurk chez des nourrissons japonais de 8-9 mois lorsque le McGurk est produit par un visage japonais mais pas lorsqu'il est produit par un visage de type européen.

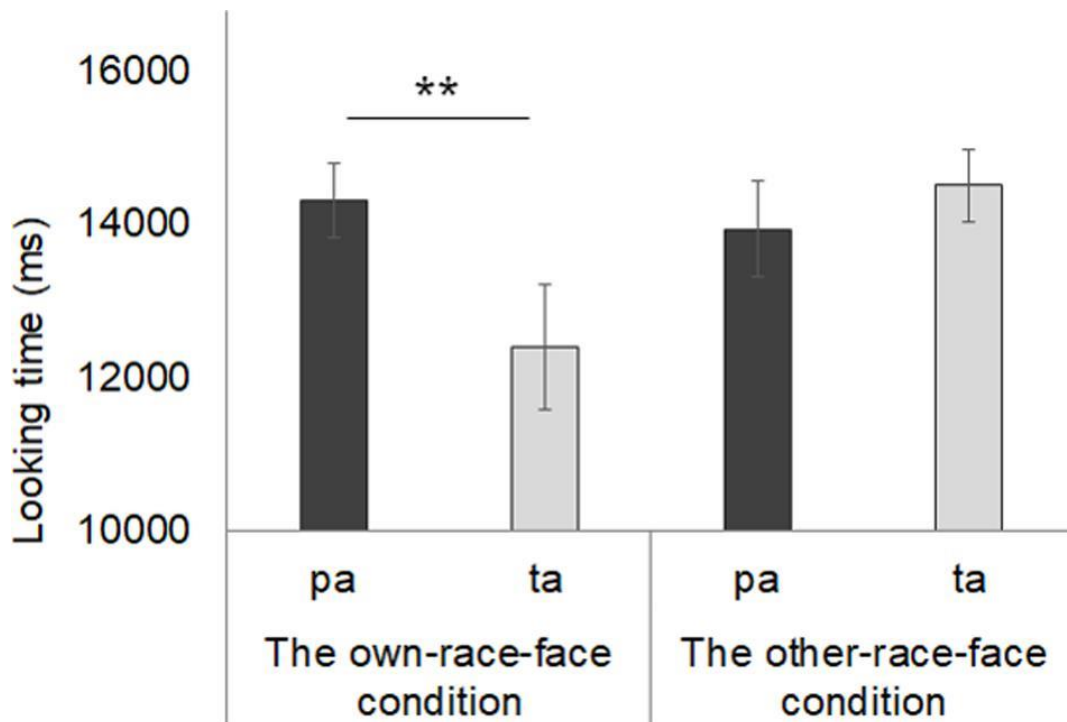


Figure 28. Temps de regard des nourrissons de 8-9 mois dans l'étude de Ujiie et al. (2020) vers le stimulus issu de la fusion (/ta/) et le stimulus audio (/pa/) durant la phase de test, après familiarisation à l'effet McGurk.

A gauche, pour un effet McGurk produit par un visage de même type que les nourrissons. A droite, pour un effet McGurk produit par un visage d'autre type que celui des nourrissons. Tiré de Ujiie et al. (2020).

Dans le cadre de la présente étude, nous avons testé la sensibilité de nourrissons à un effet McGurk produit par des visages de même et d'autre type, en nous basant sur le protocole de Ujiie et al. (2020). Nous avons décidé de tester des nourrissons avant l'émergence du rétrécissement perceptif, à 6 mois, et après son émergence, à 9 et 12 mois. Nous présenterons donc soit un effet McGurk produit par un visage de type européen soit un effet McGurk produit par un visage de type asiatique. Nous avons vu que l'expérience semble jouer un rôle sur le développement de la perception multimodale, comme le suggère les études mentionnées au chapitre précédent. Le rétrécissement perceptif, qui s'observe à la fois pour les visages et pour le langage, et pour de la correspondance audio-visuelle, pourrait donc aussi s'observer lors d'intégration audio-visuel. Si le rétrécissement perceptif impacte aussi l'intégration audio-visuelle, nous nous attendons à observer des différences de sensibilité pour les nourrissons exposés à l'effet McGurk en fonction du type de visage qui le produit. Nous avons vu que la sensibilité à l'effet McGurk dépend de vers où s'oriente le regard sur le visage produisant l'effet. Nous avons aussi vu que les populations japonaises et européennes présentes des

différences dans leurs façons de regarder un visage et que cela présente des conséquences sur le développement des capacités audio-visuelles de traitement de la parole. Nous pouvons donc aussi nous attendre à des différences d'ordre culturelle.

Nous avons choisi un effet McGurk produisant une fusion chez l'adulte, pour s'assurer qu'un phénomène d'intégration audio-visuelle à bien lieu. Nous avons couplé le visuel /ka/ avec le percept audio /pa/, résultant chez l'adulte en la perception de /ta/ (bien qu'une minorité de personnes entendent autre chose, comme /sa/ ou /da/).

Nous pouvons postuler plusieurs hypothèses :

- Si nous nous référons aux résultats comportementaux de Ujiie et al. (2020), nous nous attendons à ce que les nourrissons de 9 et 12 mois soient sensibles à l'effet McGurk mais uniquement pour les visages de leur type. Le rétrécissement perceptif impactant l'intégration audio-visuelle, les nourrissons de 9 et 12 mois pourraient en effet ne plus être capable d'intégration audio-visuelle pour les visages qui ne sont pas de leurs types.
- Pour les nourrissons de 6 mois, il se peut que la tâche soit un peu trop complexe pour permettre la fusion quel que soit le type de visage. Cependant, nous avons vu que les nourrissons français présentent plus tôt que les nourrissons japonais une attention accrue vers la bouche et sont capable de résoudre une tâche de matching audio-visuelle beaucoup plus précocement. Il est donc possible que dès 6 mois, les nourrissons de notre étude soient sensibles à l'effet McGurk.

S'offre à nous à nouveau deux hypothèses :

- Puisque le rétrécissement perceptif n'est pas encore très observable à 6 mois, les nourrissons de 6 mois pourraient être sensibles à l'effet McGurk quel que soit le type de visage.
- Cependant, nous venons de mettre en évidence dans le chapitre précédent que les capacités de matching audio-visuelle des nourrissons de 3 mois sont limités aux visages de leur type. A la vue de ces résultats, on peut s'attendre à ce que la sensibilité à l'effet McGurk ne soit présente que pour les visages de leur type.

II. Matériel et méthode

II.1. Participants

253 nourrissons ont été testés : 131 pour l'étude avec le visage de même type et 122 dans l'étude avec le visage d'autre type. Pour l'étude avec le visage d'autre type, nous avons retirés de l'analyse principale les nourrissons grandissant dans un environnement bilingue ou fréquemment exposé à des visages asiatiques ($n = 28$). Pour l'étude avec le visage de même type, nous avons retirés de l'analyse les nourrissons qui ne sont que minoritairement exposés à des visages de type européens ($n = 1$). Nous avons retiré des deux études les nourrissons pour lesquels des problèmes techniques sont survenus durant l'enregistrement ou le codage des vidéos ($n = 14$). Nous avons de plus retiré de l'analyse les nourrissons regardant plus lors des trois derniers essais de familiarisation que lors des trois premiers essais ($n = 35$; voir la procédure pour plus de détails sur les essais de familiarisation). En effet, ces nourrissons ne présentent probablement pas une familiarisation suffisamment forte pour exprimer un regain d'intérêt face à la nouveauté durant la phase de test. Enfin, nous avons retiré de l'analyse les nourrissons passant moins de 50% du temps de la familiarisation à regarder l'écran ($n = 2$). Même si sensibles au McGurk, ces nourrissons seraient plus habitués au percept audio pure (qu'ils perçoivent lorsqu'ils ne regardent pas l'écran) qu'au percept issu de la fusion (qu'ils perçoivent lorsqu'ils regardent l'écran). Leur préférence à la nouveauté lors de la phase de test ne seraient donc pas indicatives de leur sensibilité à l'effet McGurk.

176 nourrissons ont été inclus dans l'analyse finale. Pour l'étude avec le McGurk produit par le visage de même type : 26 6 mois (M âge = 192 jours, SD âge = 5.6 jours, 12 filles), 30 9 mois (M âge = 281.5 jours, SD âge = 8.5 jours, 17 filles) et 37 12 mois (M âge = 373 jours, SD âge = 9.8 jours, 18 filles). Pour l'étude avec le McGurk produit par le visage d'autre type : 26 6 mois (M âge = 190 jours, SD âge = 9.95 jours, 13 filles), 24 9 mois (M âge = 279 jours, SD âge = 8.4 jours, 11 filles) et 33 12 mois (M âge = 373 jours, SD âge = 6.2 jours, 14 filles).

II.2. Stimuli

Le stimulus de familiarisation consistait en une vidéo d'une locutrice européenne ou asiatique. Les locutrices furent enregistrées devant un fond bleu en train de prononcer les sons /pa/, /ta/ et /ka/. Une occurrence visuelle du /ka/ ainsi qu'une occurrence audio du /pa/ ont été soigneusement sélectionné afin de créer un stimulus audio-visuel provoquant l'effet McGurk. Le stimulus audio /pa/ a donc été synchronisé sur le stimulus visuel /ka/, conduisant à percevoir /ta/ (ou parfois /da/ ou /sa/). Du bruit rose a été rajouté pour renforcer l'effet McGurk

(potentiellement plus faible chez les nourrissons) et pour que notre protocole soit similaire à celui de Ujiie & Yamaguchi, 2020. Cette vidéo était d'une durée de 2.7 secondes. Les stimuli de tests consistaient en trois photos de légumes/fruits (une tomate, un oignon et une courgette ronde).

II.3. Procédure

Les nourrissons étaient installés sur les genoux d'un des parents, à approximativement 60 cm d'un écran de 22 pouces. Un *attention getter* permettait de centrer l'attention des nourrissons sur l'écran avant que la familiarisation ne commence. La familiarisation consistait en six blocs de six présentations de la vidéo de la locutrice. Toutes les six occurrences, un *attention getter* était joué pendant 2.7 secondes pour attirer à nouveau l'attention. La vidéo de familiarisation était donc, en tout et pour tout, jouée 36 fois. L'attention des nourrissons n'était pas redirigée vers l'écran lorsque que ces derniers ne le regardaient plus, puisque nous souhaitions observer une courbe d'habituation.

La phase de test était constituée de deux blocs de test : un bloc pour le son /ta/ (le son perçu en cas de sensibilité au McGurk) et un bloc pour le son /pa/ (le son pur, perçu en cas de non-sensibilité au McGurk). Lors de chaque bloc de test, les images de légumes étaient présentées à la suite, pendant 2.7 secondes, puis une deuxième fois à la suite, sans interruption (pour une séquence de type tomate – oignon – courgette -tomate – oignon -courgette). En même temps que chaque apparition de légume, l'audio /pa/ ou l'audio /ta/ était joué. A la fin du premier bloc de test, un *attention getter* était joué afin de recentrer l'attention du nourrisson sur l'écran pour le début du deuxième bloc de test. Le deuxième bloc de test était identique au premier excepté pour l'audio joué. L'ordre des blocs de test était contrebalancé entre les nourrissons.

Les vidéos étaient ensuite analysées à une précision de 40ms par un codeur aveugle à la condition expérimentale. Un deuxième codeur aveugle aux conditions expérimentales a codé 15 % des vidéos. Le taux d'accord entre les codeurs montre un taux de fiabilité élevé (Pearson $r = .95$).

III. Résultats

III.1. Phase de familiarisation

Nous nous sommes d'abord intéressés au temps de regard durant la familiarisation. Un test de Wilcoxon nous indique que les nourrissons regardent significativement plus durant les trois premières phases de familiarisations que durant les trois dernières (pour la familiarisation 1 à

3 : $M = 50414$ ms, $SD = 4717$ ms ; pour la familiarisation 4 à 6 : $M = 42089$ ms, $SD = 8548$ ms ; $p < .001$; à noter que c'est aussi le cas sans le critère d'inclusion ; voir Figure 29 gauche).

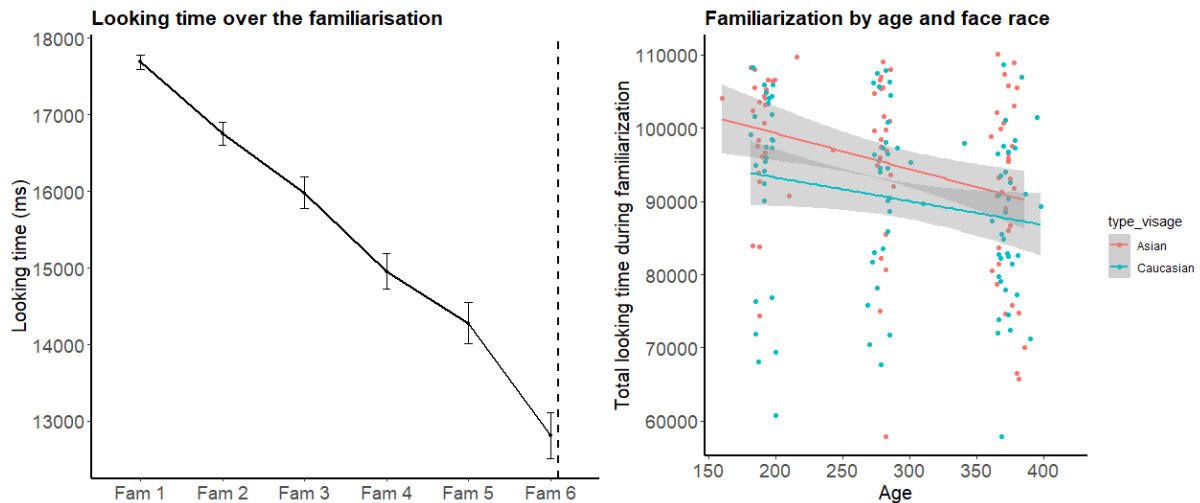


Figure 29. Temps de regard des nourrissons durant la phase de familiarisation, en fonction des essais, de l'âge et des conditions.

A gauche, temps de regard durant chaque phase de familiarisation, tout groupe d'âge et toute condition confondue. A droite, temps de regard durant l'intégralité de la phase de familiarisation par groupe d'âge et type de visage (en bleu, visage de même type ; en rouge, visage d'autre type).

Une analyse des rangs transformés (Aligned Rank Transform, ART soit une ANOVA pour données non paramétriques et avec des variables factorielles ; voir Wobbrock et al. 2011) 3 (age : 6, 9 et 12 mois) * 2 (type de visage : asiatique ou européen) nous indique un effet de l'âge ($F(2,170) = 8.31, p < .001$) et du type de visage ($F(1,170) = 8.26, p = .005$) sur l'attention visuelle que porte les nourrissons durant la phase de familiarisation. Un test de comparaison post-hoc nous indique que globalement, les nourrissons passent moins de temps à regarder la familiarisation en grandissant même si cette différence n'est significative qu'entre 6 et 12 mois ($t = 4.017, p < .001$; M temps de regard des nourrissons de 6 mois = 96310 ms ; SD temps de regard des nourrissons de 6 mois = 11759 ms ; M temps de regard des nourrissons de 12 mois = 88865 ms ; SD temps de regard des nourrissons de 12 mois = 11796 ms) et la regardent plus lorsque le visage est d'autre type ($t = 2.873, p = .005$; M temps de regard vers le visage asiatique = 95015 ms ; SD temps de regard vers le visage asiatique = 11745 ms ; M temps de regard vers le visage européen = 90262 ms ; SD temps de regard vers le visage européen = 12049 ms ; voir Figure 29 droite). Il n'y a en revanche pas d'effet d'interaction ($F(2,170) = .27, p = .763$).

III.2. Phase de test

III.2.a. Modèle ART

Nous avons d'abord calculé le temps de regard proportionnel vers les stimuli de tests. Le temps de regard proportionnel vers un stimulus de test est simplement calculé en divisant le temps de regard vers celui-ci par le temps total de regard durant les deux phases de test et ce pour chaque nourrisson. La somme du temps de regard proportionnel vers /pa/ et du temps de regard proportionnel vers /ta/ est donc égal à 100% pour chaque nourrisson. Il n'est donc pas informatif d'inclure les deux valeurs dans l'ART puisque les deux informations sont redondantes. Utiliser le temps de regard proportionnel de regard présente l'avantage de s'affranchir de l'effet de l'âge lors de son inclusion dans un modèle. Nous avons en effet vu que globalement, l'attention visuelle est plus faible pour les nourrissons de 12 mois que pour les nourrissons de 6 mois. Cela a donc un impact sur les temps de regard absolu mais pas sur les temps de regard relatif. Un modèle portant sur les temps de regard proportionnel nous permet donc de nous assurer que si un effet principal de l'âge est significatif, il est dû à un changement dans la sensibilité au McGurk et non à une simple différence attentionnelle au cours du développement. Nous avons donc effectué une ART 3 (âge : 6, 9 et 12 mois) * 2 (type de visage : asiatique ou européen) * 2 (ordre : /pa/ pouvant soit être le 1^{er} stimulus de test soit le 2nd) sur le temps de regard proportionnel vers le stimulus test /pa/ (un modèle vers le temps de regard proportionnel /ta/ donne donc strictement les mêmes résultats). Nous ne trouvons pas d'effet principal de l'âge ($F(2,164) = .08, p = .927$) ni du type de visage ($F(1, 164) = .89, p = .347$) mais un effet principal de l'ordre ($F(1, 164) = 128.07, p < .001$). Plus précisément, les nourrissons regardent plus le premier test que le deuxième ($t = 4.105, p < .001$; M temps de regard proportionnel vers le premier test = 84.33 % ; SD temps de regard proportionnel vers le premier test = 14.74 % ; M temps de regard proportionnel vers le second test = 70.63 % ; SD temps de regard proportionnel vers le second test = 17.00 %). Nous trouvons un effet de l'interaction entre l'âge et le type de visage ($F(2,164) = 3.52, p = .032$) ainsi qu'entre l'ordre et l'âge ($F(2,164) = 3.91, p = .022$). Ce dernier effet est probablement imputable à une baisse attentionnelle plus forte des nourrissons les plus âgés au deuxième test et n'est donc que peu intéressant pour notre analyse. En revanche, l'effet d'interaction entre l'âge et le type de visage semble nous indiquer que la sensibilité au McGurk n'évolue pas de la même manière au cours du développement selon le type de visage. Aucun test post-hoc n'est cependant significatif.

III.2.b. Analyse comparative avec Ujiie et al. (2020)

Conformément à l'analyse menée par Ujiie et al. (2020), nous avons exploré pour chaque catégorie d'âge et type de visage les différences en termes de temps de regard. Nous avons donc conduit des t tests (avec une correction de Bonferroni) pour chaque sous-groupe sur le temps de regard vers chaque stimulus durant la phase de test (puisque nous conduisons un t test par sous-groupe, il n'est plus nécessaire d'utiliser les temps de regard proportionnel et nous utiliserons désormais les temps de regard en millisecondes). Aucune différence ne s'est avérée significative. Nous avons vu grâce à notre modèle que le 2^{ème} test est beaucoup moins regardé que le 1^{er} et ce indépendamment du stimulus. Le deuxième test survient après la phase de familiarisation qui est déjà relativement longue pour des nourrissons (36 répétitions du stimulus de familiarisation pour une durée totale d'environ 1min40) et après une première phase de test d'environ 17 secondes. De plus, alors que le stimulus de familiarisation est la vidéo d'un visage, les stimuli de test sont des photos de légumes. Il est donc normal que l'attention chute rapidement lors de cette phase de test. Cette chute rapide pourrait cependant masquer un effet observable au 1^{er} test.

III.2.c. Analyse sur le 1^{er} test avec temps de regard pondérés

Nous avons donc décidé de conduire une 2nd analyse uniquement sur le 1^{er} test de l'étude. En revanche, ce choix nous amène à diviser nos effectifs par deux, chaque nourrisson ne contribuant dorénavant qu'à un groupe test : le groupe exposé à /ta/ ou le groupe exposé à /pa/ lors de ce premier test. De plus, bien que la familiarisation soit strictement identique au sein de chaque catégorie de visage, l'attention est très fluctuante d'un nourrisson à l'autre. Pour représenter plus fidèlement l'attention des nourrissons durant la phase de test, nous avons pondéré le temps de regard durant la phase de test par rapport au temps de regard durant la phase de familiarisation par type de visage et par groupe d'âge, pour chaque nourrisson. A titre d'exemple, un nourrisson qui regarderait 90% du temps la familiarisation et regarderait le test 80% du temps ne présente pas de regain d'intérêt. En revanche, un nourrisson qui regarderait la familiarisation 60% du temps et qui regarderait le test 70% du temps présente un regain d'intérêt. Sans correction par rapport au niveau attentionnel, le temps de regard du premier nourrisson est plus grand lors du test que le temps de regard du second nourrisson, ce qui ne reflète pas fidèlement leurs regain ou perte d'intérêts durant le test. Pondérer le temps de regard lors de la phase de test permet donc de s'affranchir de ce problème. Nous avons donc tout d'abord calculé le temps moyen de regard durant la familiarisation par sous-groupe d'âge et de type de visage. Ensuite, nous avons divisé cette moyenne par le temps moyen de regard de

chaque nourrisson afin de créer un ratio de pondération. Enfin, pour chaque nourrisson, le temps de regard durant le premier test est multiplié par ce ratio.

La seule différence significative que nous observons indique les nourrissons de 12 mois regardent plus la phase de test avec le son /pa/ que la phase de test avec le son /ta/ lorsque ceux-ci ont été familiarisés avec un visage d'autre type (le seuil de significativité n'est cependant pas atteint avec une correction de Bonferroni ; voir Figure 30 et Tableau 1). Compte tenu de l'effectif divisé par deux, il est normal que notre puissance statistique soit moindre sur ce type d'analyse. Les quelques lignes suivantes sont donc purement descriptives et non statistiques. Il semblerait que les nourrissons familiarisés avec un visage d'autre type regardent indifféremment la phase de test à 6 mois, que le stimulus audio soit /pa/ ou /ta/. En revanche, toujours pour les nourrissons familiarisés avec un visage d'autre type, il semblerait qu'à partir de 9 mois, les nourrissons ayant pour stimulus de test le son /pa/ passent plus de temps à regarder la phase de test que les nourrissons ayant pour stimulus de test le son /ta/. Cette tendance pourrait suggérer que la sensibilité à l'effet McGurk présenté avec un visage d'autre type émerge vers 9 mois. Pour les nourrissons familiarisés avec un visage de même type, on observe à 6 mois que les nourrissons ayant pour stimulus de test le son /pa/ passent plus de temps à regarder la phase de test que les nourrissons ayant pour stimulus de test le son /ta/. Cette différence s'atténue nettement à 9 et 12 mois. Cette tendance pourrait suggérer qu'une sensibilité au McGurk est présente dès 6 mois.

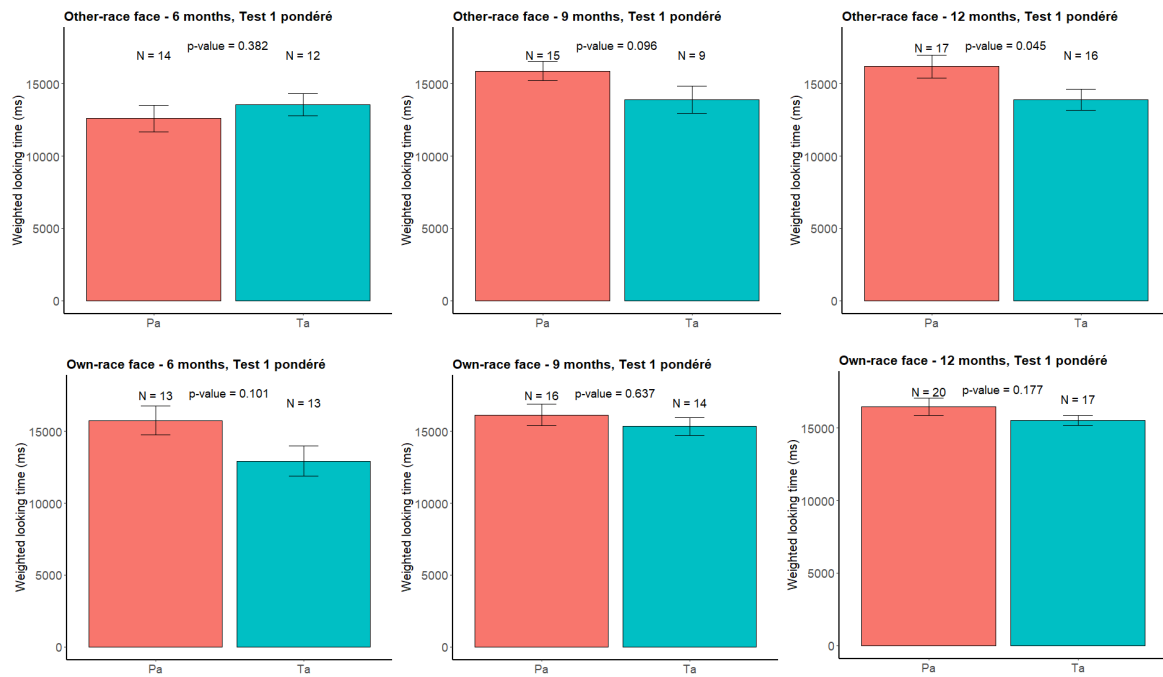


Figure 30. Temps de regard pondérés des nourrissons durant la première phase de test. Respectivement de gauche à droite : pour les nourrissons de 6, 9 et 12 mois. Respectivement en haut et en bas : pour la condition avec visage d'autre type et de même type. En rouge, le temps de regard des nourrissons ayant pour premier stimulus de test le son /pa/. En bleu, le temps de regard des nourrissons ayant pour premier stimulus de test le son /ta/.

Tableau 1. Moyenne du temps de regard pondéré vers le stimulus /pa/ durant la phase de test pour les nourrissons de 6, 9 et 12 mois, pour la condition avec visage de même type et pour la condition avec visages d'autres types. Les p-values sont issues de test de Wilcoxon comparant le temps de regard durant la présentation du stimulus /pa/ contre le temps de regard durant la présentation du stimulus /ta/.

Age group (months)	Own-race (Caucasian)				Other-race (Asian)			
	N(N/pa/)	M(SD)	p	Taille d'effet (Z/N—√).	N(N/pa/)	M(SD)	p	Taille d'effet
6	26(13)	15757(3583)	.101	.327	26(14)	12591(3418)	.382	.177
9	28(16)	16135(3002)	.637	.091	24(15)	15886(2622)	.096	.347
12	37(20)	16466(2609)	.177	.225	33(17)	16195(3282)	.045	.351

III.2.d. *Analyse des taux de déshabituat*

Une autre manière généralement efficace de refléter le regain ou non d'intérêt durant la phase de test est de comparer le temps de regard durant la première phase de test au temps de regard durant la dernière phase de familiarisation (voir Figure 31). Il n'est plus nécessaire ici d'utiliser les temps de regards pondérés par l'attention durant la familiarisation. En effet, pour chaque sous-groupe d'âge et de type de visage, la comparaison du test est faite par rapport à la dernière familiarisation et reflète donc déjà le niveau attentionnel durant celle-ci.

Cette analyse révèle les mêmes tendances qu'observées pour les temps de regard pondérés sur le test 1, mais de manière beaucoup plus marquée (Voir Tableau 2). Pour les nourrissons exposés au McGurk avec un visage d'autre type :

- A 6 mois, on n'observe aucun regain d'intérêt en phase de test quel que soit l'audio joué.
- A 9 mois, on observe un regain d'intérêt significatif pour la phase de test présentant l'audio /pa/ ($p = .008$; *taille d'effet* = .616) mais pas pour la phase de test présentant l'audio /ta/.
- A 12 mois, on observe un regain d'intérêt significatif quelque que soit l'audio de la phase de test (pour /pa/ : $p < .001$; *taille d'effet* = .809 ; pour /ta/ : $p = .013$; *taille d'effet* = .556).

Pour les nourrissons exposés au McGurk avec un visage de même type :

- A 6 mois, il y a un regain d'intérêt significatif pour la phase de test avec l'audio /pa/ ($p = .016$; *taille d'effet* = .591) mais pas pour la phase de test avec l'audio /ta/.
- A 9 mois, on observe un regain d'intérêt significatif quelque que soit l'audio de la phase de test (pour /pa/ : $p < .001$; *taille d'effet* = .879 ; pour /ta/ : $p = .004$; *taille d'effet* = .679).
- A 12 mois, on observe un regain d'intérêt significatif quelque que soit l'audio de la phase de test (pour /pa/ : $p < .001$; *taille d'effet* = .860 ; pour /ta/ : $p < .001$; *taille d'effet* = .867).

Nous observons donc que pour les nourrissons ayant été familiarisé à l'effet McGurk avec un visage d'autre type, une préférence pour le son /pa/ et donc potentiellement la sensibilité à l'effet McGurk émerge environ à partir de 9 mois. Pour les nourrissons ayant été familiarisé à l'effet McGurk avec un visage de même type qu'eux, la préférence pour le /pa/ et donc potentiellement la sensibilité à l'effet McGurk est présente dès 6 mois (voir Figure 32). Le regain d'intérêt observable à 9 et 12 mois quel que soit l'audio de la phase de test est discuté par la suite.

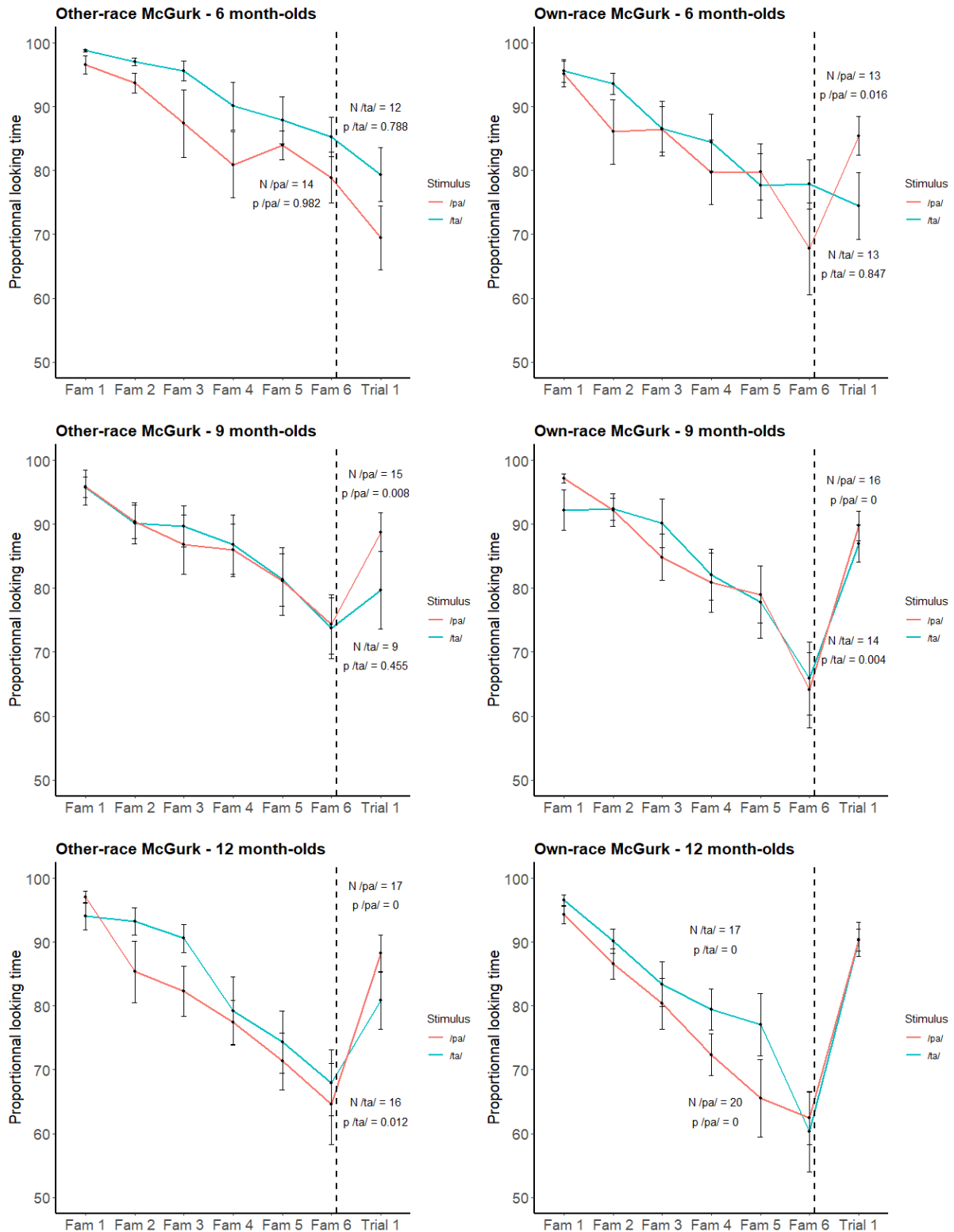


Figure 31. Temps de regard durant chaque phase de familiarisation et durant la première phase de test (après la ligne verticale de pointillée).

Respectivement de haut en bas : pour les nourrissons de 6, 9 et 12 mois. Respectivement à gauche et à droite : pour la condition avec visage d'autre type et de même type. En bleu, le temps de regard des nourrissons ayant pour premier stimulus de test le son /ta/. En

rouge, le temps de regard des nourrissons ayant pour premier stimulus de test le son /pa/. La partie à gauche des pointillés est identique pour les nourrissons quelle que soit la couleur de la courbe. Les barres d'erreurs représentent l'écart-type. Les tests exécutés sont des tests de Wilcoxon avec comme hypothèse H1 un regain d'intérêt lors de la phase de test.

Tableau 2. Moyenne du temps de regard proportionnel vers les stimuli /pa/ et /ta/ durant la phase de test pour les nourrissons de 6, 9 et 12 mois, pour la condition avec visage de même type et pour la condition avec visages d'autres types. Les p-values sont issues de test de Wilcoxon comparant le temps de regard durant la présentation du premier stimulus de test contre le temps de regard durant la présentation de la dernière phase de familiarisation.

Age group (months)		Own-race (Caucasian)				Other-race (Asian)			
		N	M(SD)	p	Taille d'effet ($Z/N-\sqrt{}$)	N	M(SD)	p	Taille d'effet
6	Pa	13	85.44(11.11)	.016	.591	14	69.46(18.60)	.982	.545
	Ta	13	74.43(18.88)	.847	.281	12	79.35(14.58)	.788	.226
9	Pa	16	89.74(9.08)	<.0001	.879	15	88.78(11.63)	.008	.616
	Ta	14	86.98(10.94)	.004	.679	9	79.65(18.15)	.455	.059
12	Pa	20	90.41(11.96)	<.00001	.860	17	88.24(11.84)	<.001	.809
	Ta	17	90.29(7.06)	<.0001	.867	16	80.83(17.88)	.013	.556

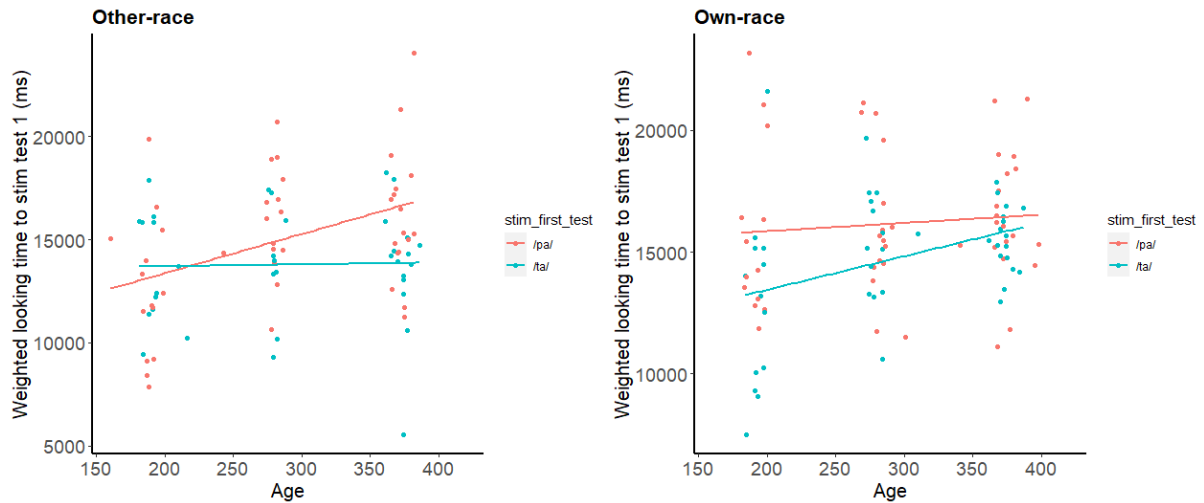


Figure 32. Temps de regard pondéré durant la première phase de test par type de visage (à gauche, d'autre type ; à droite, de même type) en fonction de l'âge des nourrissons et du stimulus de test (en rouge, /pa/ ; en bleu, /ta/).

IV. Discussion

Nous avons déjà souligné dans l'introduction la difficulté voire l'impossibilité à mettre en place un protocole mesurant l'effet McGurk de manière non ambiguë chez des nourrissons. Le nôtre n'échappe pas à la règle et plusieurs points limitent donc nos interprétations.

Tout d'abord, notre absence de résultats dans l'analyse principale reprenant les mêmes procédés d'analyse que Ujiie et al. (2020) ne nous permet pas de faire une comparaison directe de nos résultats avec les leurs. Nos hypothèses étaient ici qu'une sensibilité à l'effet McGurk de la part des nourrissons aurait dû se traduire par plus de temps de regard vers la phase de test présentant l'audio /pa/ comparativement à la phase de test présentant l'audio /ta/ ; à contrario, une absence de sensibilité à l'effet McGurk de la part des nourrissons aurait dû se traduire par plus de temps de regard vers la phase de test présentant l'audio /ta/ comparativement à la phase de test présentant l'audio /pa/. Cependant, nous ne trouvons pas de différence entre les temps de regard entre les deux phases de tests, quel que soit l'âge des nourrissons ou le type de visage par lequel le McGurk a été présenté. Nous ne pouvons donc ni conclure à une sensibilité ni à une absence de sensibilité envers le McGurk.

IV.1. L'effet McGurk présenté avec un visage de même type

L'analyse portant sur la différence de temps de regard entre la dernière phase de familiarisation et la première phase de test nous permet d'interpréter en partie nos résultats. Concernant le McGurk présenté avec un visage de même type, on observe dès 6 mois un regain d'intérêt

lorsque le son de la phase de test est le son /pa/ mais pas lorsque le son est /ta/. On peut donc considérer que les nourrissons sont sensibles à l'effet McGurk dès 6 mois lorsque présenté via un visage de même type qu'eux. En revanche, on ne trouve aucun regain d'intérêt pour les nourrissons ayant été exposé au McGurk via un visage asiatique. Les résultats de Ujiie et al. (2020) montrent que les nourrissons sont sensibles à l'effet McGurk présenté avec des visages de même type qu'eux mais ne montrent pas de différence de temps de regard en fonction du son joué lorsque le McGurk est présenté avec un visage d'autre type et ce à 8-9 mois. Nous avons vu que le traitement audio-visuel se fait plus tardivement par les nourrissons japonais que par les nourrissons européens (Mugitani et al., 2006 ; Mugitani, et al., 2004 ; cité par Sekiyama et al., 2021), ce qui peut expliquer l'émergence plus tardive de l'effet McGurk chez les nourrissons japonais que chez les nourrissons européens.

Pour les nourrissons de 9 et 12 mois exposé à l'effet McGurk produit par un visage de même type, nous trouvons un regain d'intérêt significatif à la fois chez les nourrissons exposés à la phase de test comprenant le son /pa/ et chez les nourrissons exposés à la phase de test comprenant le son /ta/. La conclusion est donc moins directe. Le changement de stimuli entre la phase de familiarisation et la phase de test peut poser problème. Il nous fallait évidemment retirer toute influence visuelle sur la perception sonore durant la phase de test pour mesurer la réaction face aux percepts audio « purs ». Nous avons donc opté pour les mêmes stimuli que Ujiie et al (2020), soit des images de légumes, objets visuels n'influençant pas la perception sonore. Cependant, ce changement visuel entre la phase de familiarisation et la phase de test entraîne très probablement un regain d'intérêt, indépendamment du stimulus audio joué. Le regain d'intérêt pour le changement audio, soit le regain d'intérêt que nous souhaitons mesurer, se retrouve donc en partie masqué par le regain d'intérêt pour le changement visuel. Lewkowicz (1988a, b) a montré que les nourrissons soumis à des stimuli audio-visuels présentent des dominances de modalité, c'est-à-dire qu'une modalité peut-être préférentiellement traitée. A 6 mois, Lewkowicz nous indique que les nourrissons présentent une dominance de la modalité audio lorsque confronté à des stimuli audio-visuels. Cela peut expliquer pourquoi les nourrissons de 6 mois de notre étude ne présentent pas un regain d'intérêt systématique lors du changement visuel entre la phase de familiarisation et la phase de test. En revanche, cette dominance semble moins évidente à 10 mois. Cela pourrait expliquer pourquoi on obtient un regain d'intérêt chez les nourrissons de 9 mois, encore plus fort chez les 12 mois, indépendamment de l'audio joué. Toutefois, la significativité ainsi que la taille d'effet du regain d'intérêt sont plus importantes chez les nourrissons exposés à la phase de test présentant le son

/pa/ que chez les nourrissons exposés à la phase de test présentant le son /ta/, et ce à 9 et à 12 mois. On peut donc envisager que ce « surplus » d'intérêt chez les nourrissons exposés à la phase de test avec l'audio /pa/ est dû à la différence des percepts audio entre la phase de test et la phase de familiarisation. Il est en effet probable que les nourrissons, sensibles à l'effet McGurk présenté avec un visage de même type à l'âge de 6 mois, soient toujours sensibles à cet effet à l'âge de 9 et 12 mois. Une hypothèse alternative est cependant envisageable à la vue des résultats du chapitre précédent, montrant qu'il n'est pas clair que les nourrissons présente à 9 mois la capacité de correspondance audio-visuelle qu'ils présentent à 3 mois.

IV.2. L'effet McGurk présenté avec un visage d'autre type

Pour les nourrissons exposés à l'effet McGurk produit par un visage d'autre type, on observe à 9 mois un regain d'intérêt significatif pour les nourrissons exposés au son /pa/ lors de la phase de test et pas de regain d'intérêt pour les nourrissons exposés au son /ta/ (ou plus précisément un regain d'intérêt non significatif). On a donc le même pattern de réponse que pour les nourrissons de 6 mois exposés au McGurk avec des visages de même type. Ces résultats pourraient indiquer que l'émergence de la sensibilité à l'effet McGurk est dépendante du type de visage et retardée lorsque présenté via un visage d'autre type. Nous trouvons une sensibilité à l'effet McGurk présenté via des visages d'autres types chez des nourrissons de 9 mois, alors que Ujiie et al. (2020) ne trouvent pas d'effet dans cette même condition chez des nourrissons japonais de 8-9 mois. Là encore, le traitement audio-visuel plus tardif des nourrissons japonais peut expliquer cette absence chez les nourrissons japonais de 8-9 mois. Il est cependant probable qu'une sensibilité à l'effet McGurk, indépendante du type de visage, émerge un peu plus tard chez les nourrissons japonais.

Enfin, pour les nourrissons de 12 exposés à l'effet McGurk produit par un visage d'autre type, on observe un regain d'intérêt significatif à la fois chez les nourrissons exposés à la phase de test comprenant le son /pa/ et chez les nourrissons exposés à la phase de test comprenant le son /ta/. Cependant, la significativité et la taille d'effet du regain d'intérêt sont plus importantes chez les nourrissons exposés à la phase de test présentant le son /pa/ que chez les nourrissons exposés à la phase de test présentant le son /ta/. On retrouve donc un pattern similaire à celui des nourrissons de 9 mois exposé à l'effet McGurk via des visages de même type. On peut donc, de la même manière, conclure que les nourrissons sont toujours sensibles à l'effet McGurk présenté via un visage d'autre type à l'âge de 12 mois, mais que cette sensibilité est en partie masqué par un regain d'intérêt lié au visuel.

IV.3. Limites du protocole

Nous avons mentionné le fait que le changement de stimulus visuel entre la phase de familiarisation et la phase de test contribue à masquer notre effet. Un deuxième facteur limitant notre interprétation est l'exposition des nourrissons de notre étude aux deux stimuli audios avant la phase de test. En effet, nous observons une habituation au fil des essais de familiarisation, avec une baisse d'intérêt. Cette baisse d'intérêt est nécessaire pour permettre une déshabitude lors de la phase de test. Cependant, elle entraîne un effet pervers. Un nourrisson qui serait sensible à l'effet McGurk entendrait le son /ta/ lorsque qu'il regarde la vidéo de familiarisation. A mesure que son intérêt baisse, celui-ci passera de plus en plus de temps à regarder hors-écran, et sera donc exposé à l'audio seul. Ce nourrisson entendra donc /pa/ lorsqu'il détournera la tête de l'écran. Un nourrisson, même sensible à l'effet McGurk, peut donc potentiellement être habitué au percept résultant de la fusion, /ta/, à et au percept audio seul, /pa/. Nous avons cependant pris le soin de retirer de l'analyse les nourrissons ayant passé plus de temps à regarder en dehors de l'écran que l'écran. L'habituation au percept audio est donc être moindre que l'habituation au percept fusionné. Toutefois, cela peut résulter en moins de regain d'intérêt pour le son « nouveau » qui n'est finalement pas entièrement nouveau pour le nourrisson. Nous pouvons donc avoir un deuxième effet de masquage, cette fois provoquant une baisse d'intérêt pour le percept audio pur.

De plus, nous avons vu en introduction de ce chapitre que la perception de l'effet McGurk par les adultes n'est pas uniforme (Green et al., 1991). Cela veut dire que le percept fusionné ne sera pas le même en fonction du sujet. Concernant notre McGurk, la plupart des adultes entendent /ta/ mais un nombre non négligeable entend autre chose. Nous n'avons pas de raisons de penser qu'il en soit différemment pour les nourrissons, Burnham & Dodd (2004) suggérant d'ailleurs que la plupart des nourrissons de leur étude entendent un percept différent de celui entendu par les adultes. Pour une partie des nourrissons de notre étude, le percept audio /ta/ entendu lors de la phase de test peut être nouveau et ce même s'ils sont sensibles à l'effet McGurk. Ces nourrissons présenteraient alors un regain d'intérêt quel que soit l'audio, voire potentiellement plus pour le son /ta/ que pour le son /pa/ pour la raison évoquée précédemment impliquant une légère habituation au son /pa/ lors de regards hors-écran. Ainsi, des nourrissons pourtant sensibles au McGurk contribuerait à un plus fort regain d'intérêt pour le percept fusionné /ta/ que pour le percept audio pur qui devrait être nouveau, /pa/. Enfin, il est à noter que chez les adultes, même chez un sujet réceptif, l'effet McGurk n'est pas observé à chaque

répétition de l'illusion (Brancazio, 2004 ; Brancazio & Miller, 2005 ; Massaro & Cohen, 1983). Ce dernier point peut donc encore diminuer le regain d'intérêt que l'on observe.

En prenant en compte ces diverses limitations, nos résultats pourraient indiquer que les nourrissons français sont sensibles à l'effet McGurk dès 6 mois pour des visages de même type qu'eux et dès 9 mois pour des visages d'autres types. À la vue de la littérature sur le rétrécissement perceptif, ces résultats sont surprenants. Il est établi qu'à partir de 9 mois, les visages d'autres types ne sont plus traités de la même manière que les visages de même type au point de ne plus être reconnus en situation de test (Kelly et al., 2007b, 2009). L'hypothèse la plus évidente aurait été que le type de visage par lequel est présenté le McGurk n'aurait une influence sur cet effet qu'après le rétrécissement perceptif, soit à partir de 9 mois. On observe cependant une émergence de l'effet McGurk après ce cap des 9 mois pour les visages d'autres types. Cependant, si la reconnaissance des visages d'autres type semble affectée à partir de 9 mois, des différences de traitements entre visage de même type et d'autre type sont déjà présent avant (Bar-Haim et al., 2006 ; Kelly, et al., 2005, 2007b ; Sangrioli & De Schonen, 2004). Nous avons montré dans le chapitre précédent que dès 3 mois le type de visage affecte une tâche de matching phonémique. De la même manière que les nourrissons de 3 mois échouent à la tâche de matching phonémique avec des visages d'autres types alors qu'ils y arrivent avec des visages de même type, les nourrissons de 6 mois ne sont pas sensibles à l'effet McGurk présenté avec un visage d'autre type alors qu'ils le sont lorsque présenté avec un visage de même type. Tüttenberg & Wiese (2021) a très récemment démontré que chez les adultes, reconnaître des visages d'autres types demande plus d'efforts cognitifs que reconnaître des visages de même type. Les nourrissons de moins de 9 mois, qui sont encore capables de reconnaître des visages d'autres types, présentent cependant déjà une préférence pour les visages de leurs types (Kelly et al., 2005, 2007a). Le scanning de ces visages se différencie aussi de celui des visages de même type à partir de 3 mois (Wheeler et al., 2011 ; Liu et al., 2011 ; Xiao et al., 2013). Le traitement des visages d'autres types est donc différent de celui des visages de même type dès 3 mois et est probablement plus demandant en termes de ressources cognitifs que celui des visages de même type. On peut donc imaginer que cette sur-demande en ressources cognitifs pour traiter le visage amène les nourrissons à ne pas traiter la tâche audio-visuelle. A 9 mois cependant, il est possible que les nourrissons traitent le visage à un niveau subordonnée (« c'est un visage d'autre type »), d'une manière rapide et plus économe en ressources cognitives. Ainsi, la tâche audio-visuelle peut être effectuée.

IV.4. La tâche de McGurk comme proxy de l'intégration audio-visuelle

De manière plus générale, des questionnements quant à la pertinence d'utiliser l'effet McGurk comme un proxy des capacités d'intégration audio-visuelle peuvent se poser. En effet, la sensibilité à cet effet est souvent utilisée comme une évidence de l'intégration audio-visuelle de la parole. Pourtant, il ne s'agit là que d'un cas très particulier d'intégration visuelle ; il n'est pas dit que le traitement audio-visuel de stimuli qui sont congruents se fasse de la même manière. Grant & Seitz (1998) ont testé des adultes souffrant de perte auditive sur des tâches de McGurk et de perceptions de phrases complètes dans un environnement sonore bruité. Si les auteurs ont trouvé que les participants bénéficient d'indices visuels pour traiter le signal audio dans un environnement bruité, ils ne trouvent en revanche pas de corrélation entre les résultats des participants à la tâche de McGurk et à la tâche de compréhension de phrases. Plus récemment, Van Engen et al. (2017) ne trouvent qu'une très faible corrélation entre la compréhension du discours en environnement bruitée et la sensibilité à l'effet McGurk. De même, Brown et al. (2018) trouvent que la sensibilité à l'effet McGurk et la performance en lecture labiale ne sont que faiblement corrélées ($r = .29$). Autrement dit, la sensibilité à l'effet McGurk représente bien la capacité à percevoir le signal de parole de manière multimodale mais potentiellement seulement dans le cadre de l'effet McGurk.

Chapitre 7 : Association langage/visage

I. Introduction

I.1. Associations voix/visages

A partir de 6 mois, il semblerait que les nourrissons commencent à associer un visage avec une voix en fonction de son sexe (Walker-Andrews et al., 1991). Pour tester cela, les auteurs ont utilisé un protocole similaire à celui de Kuhl & Meltzoff (1982) : un visage masculin et un visage féminin sont présentés côte à côte en train de raconter une histoire, pendant qu'un audio de l'histoire est joué, soit avec une voix masculine soit avec une voix féminine. Les nourrissons regardent plus le visage dont le sexe correspond au sexe perçu dans l'audio. Cette capacité à associer une voix à un visage du sexe correspondant ne semble pas présente avant 6 mois (Patterson & Werker, 2002 ; Walker-Andrews et al., 1991). Ceci nous indique que les nourrissons ont à partir de 6 mois des attentes concernant la voix d'un individu en fonction de son sexe. De plus, le fait que ces associations ne sont formées qu'à partir de 6 mois suggère que ces attentes se forment avec l'expérience. Le fait de voir quasi systématiquement des visages masculins produire des voix masculines et des visages féminins produire des voix féminines amène les nourrissons à avoir de fortes attentes vis-à-vis du type de voix produit par un visage. De la même manière, Bahrick et al. (1998) ont présenté des vidéos de visages d'enfants et d'adultes à des nourrissons de 4 et 7 mois, tout en présentant une bande audio d'une comptine récitée par un enfant ou un adulte. Les résultats montrent que les nourrissons de 7 mois regardent globalement plus la vidéo correspondant à la voix entendue. Les nourrissons de 4 mois de l'étude présentent une tendance à faire la correspondance audio-visuelle entre la voix et la vidéo, mais de manière moins forte et moins immédiate que les nourrissons de 7 mois. Vouloumanos et al. (2009) montrent aussi que les nourrissons de 5 mois associent les sons de langues humaines (même de langues inconnues par les nourrissons) avec des visages humains et des cris de macaques avec des visages de singes. Il est peu probable que les nourrissons de cet âge présentent une grosse expérience avec les visages de singes et les cris de singes. Il est donc possible que l'association entre les vocalisations de singes et les visages de singes se fasse par élimination : lorsque les nourrissons entendent du langage (même étranger), ils s'attendent à ce qu'il soit produit par un visage humain ; si la vocalisation n'est pas rattachable à un langage humain, il est envisageable que les nourrissons la rattachent à un visage non humain. Cependant, cette même étude montre que les nourrissons de 5 mois ne regardent pas plus un visage de canard lorsque des vocalisations de canard sont jouées que lorsque du langage

est joué. Un élément avancé par les auteurs pour expliquer cette différence est que les visages de canards ne sont potentiellement pas considérés comme des visages par les nourrissons. Les visages de canards diffèrent évidemment plus des visages humains que les visages de macaques (voir Figure 33). Les visages de singes sont d'ailleurs encore distingués correctement à cet âge par les nourrissons, bien qu'ils n'en soient plus capables après 9 mois (Pascalis et al., 2002). Enfin, les auteurs ont aussi couplé des photos de visages humains et de singes avec des sons humains mais non langagiers (des rires). Les nourrissons n'ont pas plus regardé les visages humains lorsqu'ils étaient associés à ces sons que lorsqu'ils étaient associés aux vocalisations de singes. Une possibilité est que les signaux non langagiers (ici le rire) présentent moins d'éléments distinctifs que le langage qui permettent de les catégoriser comme une production vocale humaine. Dans le même ordre d'idée, Uttley et al. (2013) ont étudié les associations que font les nourrissons entre type de visage et type de langue. Pour cela, ils ont présenté à des nourrissons de 6 mois des visages de même type (de type européen) ou d'un autre type qu'eux (de type asiatique et plus précisément des visages chinois), un à la fois. A chaque fois qu'un visage était présenté, un audio constitué de petites phrases adressées au nourrisson en langue maternelle (en anglais) ou étrangère (en mandarin) était joué. Dans une première étude, la congruence était une variable inter-sujet, c'est-à-dire qu'un groupe de nourrissons ne voyaient que des essais congruents (des visages de même type avec langue maternelle ou bien des visages d'autre type avec une langue étrangère) tandis qu'un autre groupe de nourrissons ne voyaient que des essais non congruents. Les résultats montrent que les nourrissons du groupe congruent regardent plus longtemps les visages que les nourrissons du groupe incongruent. Encore une fois, il est probable que l'expérience des nourrissons leur ait appris à associer les visages de leur type et leur langue maternelle. En revanche, il est peu probable que les nourrissons aient des attentes très définies concernant les visages d'autres types ou les langues étrangères, n'étant que très occasionnellement exposés à ces stimuli (les nourrissons de l'étude étant de plus issus de familles monolingues). Il est donc plus qu'improbable que les nourrissons de 6 mois aient associé les visages chinois spécifiquement au mandarin ; l'interprétation plus rationnelle est que les nourrissons associent le type de visage non familier à la langue non familière. Dans une seconde étude, la variable inter-sujet était le type de visage. Un groupe de nourrissons ne voyaient que des visages de même type (mais avec à la fois des essais congruents et incongruents) tandis qu'un autre groupe de nourrissons ne voyaient que des visages d'un autre type. Les résultats sont ici quelque peu différents de la première étude. Si l'on retrouve à nouveau que les nourrissons regardent plus longtemps les visages d'un autre type lorsqu'ils sont associés à une langue étrangère que lorsqu'ils sont associés à leur langue maternelle, le temps

de regard des nourrissons vers les visages de même type n'est en revanche pas différent en fonction de la langue avec laquelle ces visages sont associés. Il n'est donc pas complètement clair dans quelle mesure ces associations sont stables à cet âge. Une troisième étude des auteurs apporte un élément supplémentaire permettant une interprétation plus précise des résultats. Dans cette étude, les auteurs ont associé les visages (du même type et d'un autre type) à du langage joué à l'envers, ne ressemblant donc plus à du langage mais contenant les mêmes informations de durées, volumes et fréquences. Les résultats montrent ici que les nourrissons ne regardent pas plus un type de visage que l'autre lorsque ce type de stimulus audio est joué. Autrement dit, les nourrissons n'associent pas tous les types de sons inconnus ou non fréquents aux visages d'un autre type ; ils associent uniquement les sons indétectables comme du langage mais non connu ou non familier. Ces résultats vont dans le même sens que ceux trouvés par Vouloumanos et al. (2009) montrant que les nourrissons n'associent pas les sons humains mais non langagier avec les visages humains. Ces résultats sont aussi en accord avec les autres résultats de Vouloumanos et al. (2009) montrant que les nourrissons n'associent pas les vocalisations de canards avec les visages de canard, probablement trop différent des visages humains pour être identifiés en tant que visage produisant des vocalisations. Ces recherches montrent que l'expérience semble faire émerger des attentes chez les nourrissons quant au type de vocalisation que doit produire un visage, ici en fonction de son sexe, de son espèce, de son âge et de son type. D'un point de vue expérimental, cela se traduit par des temps de regard plus long vers les visages lorsque les vocalisations sont « congruentes » (c'est-à-dire conformes aux attentes ou n'allant pas à l'encontre de celles-ci).



Figure 33. Exemple de stimuli utilisé par Vouloumanos et al. (2009).

A la vue des stimuli, il apparaît plausible que les visages de canards ne soient pas catégorisés comme des visages par les nourrissons. Cela expliquerait l'incapacité des

nourrissons à relier ces visages aux vocalisations animales. Tiré de Vouloumanos et al. (2009).

I.2. Protocole d'association intermodale

Récemment, Xiao et al. (2016) ont développé un nouveau protocole d'association intermodale (voir Figure 34). Cette étude a déjà été mentionné dans l'introduction mais il est important d'en rappeler ici les principes, puisque nous utilisons le même protocole pour notre étude. Les auteurs de cette étude ont voulu évaluer la valence émotionnelle (positive ou négative) que les nourrissons associent aux visages selon leur type. Pour caractériser cette valence, les auteurs ont utilisé de la musique, qui pouvait être joyeuse ou triste (caractérisé principalement par le mode, majeur ou mineur, et le tempo, généralement plus rapide pour les musiques joyeuses). Ils ont présenté de manière séquentielle les visages et la musique. Plus précisément, une photo de visage apparaissait pendant 10 secondes puis disparaissait. Pendant à nouveau 10 secondes, un *attention getter* étant présent à l'écran et de la musique était joué. Cette alternance de visage et de musique était répétée jusqu'à ce que six visages et six extraits de musiques soient présentés. Les visages et les extraits de musique étaient à chaque fois nouveaux afin de conserver l'attention des nourrissons au fil de l'expérience. Cependant, les nourrissons étaient assignés à un seul type de visage et à un seul type de musique. Les visages étaient donc soit tous de même type que le nourrisson soit tous d'un autre type. Les extraits musicaux étaient soit tous joyeux soit tous tristes. Quatre combinaisons de stimuli étaient alors possibles : des visages de même type entrecoupés de musique triste, des visages de même type entrecoupés de musique joyeuse, des visages d'un autre type entrecoupés de musique triste, des visages d'un autre type entrecoupés de musique joyeuse. L'hypothèse des auteurs était que les nourrissons associeraient plus facilement les visages de leur type à une valence positive et les visages d'autres types à une valence négative, de manière de plus en plus marquée au fil du développement. Par conséquent, les combinaisons visage de même type / musique joyeuse et visage d'un autre type / musique triste furent définies comme congruentes tandis que les combinaisons visage de même type / musique triste et visage d'un autre type / musique joyeuse furent définies comme incongruentes. Les résultats confirment l'hypothèse des auteurs. En effet, les nourrissons passent plus de temps à regarder les visages dans les conditions congruentes qu'incongruentes à mesure qu'ils grandissent. Si cette distinction n'est pas significative à 3 mois, elle le devient à 9 mois, soit après le rétrécissement perceptif. Avec l'âge, les nourrissons deviennent donc plus susceptibles d'associer des émotions positives avec des visages de leur type et des émotions négatives avec des visages d'un autre type. Une possibilité

alternative est de considérer les résultats en termes de familiarité : la musique joyeuse est plus familière pour les nourrissons que la musique triste (82% des nourrissons de l'étude ont une expérience fréquente de la musique joyeuse tandis que seulement 18 % des nourrissons ont une expérience, qui plus est seulement occasionnelle, de la musique triste). Par conséquent, il est probable que les nourrissons aient associé les stimuli familiers entre eux et les stimuli non familiers entre eux.

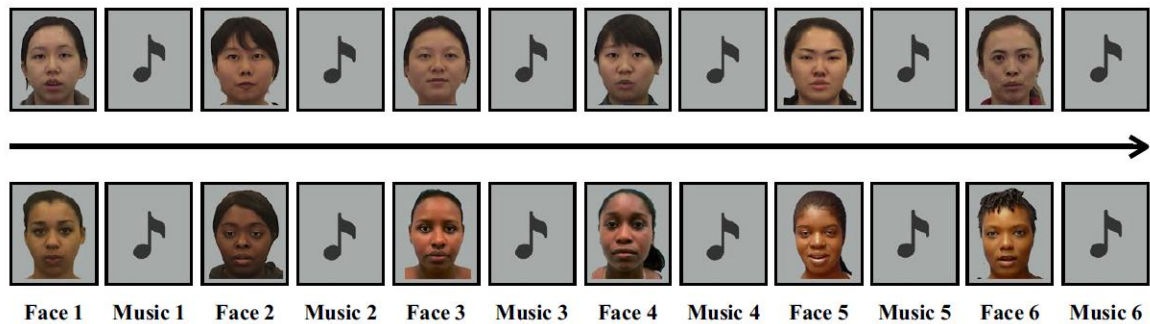


Figure 34. Protocole mise en place par Xiao et al. (2016). Tiré de Xiao et al. (2016).

I.3. Objectif et hypothèses de l'étude 3

Le but de cette étude est double. Le premier objectif est de permettre de mieux comprendre l'évolution des associations que forment les nourrissons entre type de visages et type de langues. Nous associerons donc des visages de même type que les nourrissons (ici de type européen) et d'un autre type (de type asiatique) avec la langue maternelle des nourrissons (le français) et une langue étrangère (l'allemand). Pour mieux comprendre l'évolution de ces associations, nous testerons des nourrissons de 3, 6, 9 et 12 mois, soit avant et après le rétrécissement perceptif. Le deuxième objectif est méthodologique. Nous souhaitons appliquer le protocole mis en place par Xiao et al. (2016) pour des associations entre langage et visage. Cette étude est faite en collaboration avec une équipe allemande de l'université de Giessen. Les nourrissons de 3 mois seront testés par l'équipe allemande tandis que les autres groupes d'âges seront testés au laboratoire de Grenoble. Pour les visages d'autres types, nous utiliserons des visages de type asiatique, relativement peu présents, que ce soit à Giessen ou à Grenoble. Pour la langue étrangère, nous utiliserons l'allemand à Grenoble tandis que le français sera utilisé à Giessen. Les deux langues utilisées étant les mêmes, cela nous permettra de comparer les résultats de manière fiable, sans effet de distance linguistique.

Nos hypothèses sont les suivantes :

- A 3 mois, les nourrissons n'ont peut-être pas encore des attentes assez fortes concernant les associations voix/visages pour induire une différence de temps de regard. Cependant,

nos résultats de l'étude 1 (Chapitre 5) montrent que dès 3 mois, le type de visage d'une personne impacte la manière de traiter le signal audio-visuel. Il est donc possible que d'autres différences dans la manière de traiter les stimuli non natifs des stimuli natifs soient présentes à cet âge.

- A 6 mois, les attentes des nourrissons concernant les associations entre type de visage et type de langue semblent émerger mais ne sont pas clairement définies (Uttley et al., 2013). En nous basant sur les résultats d'Uttley et al., nous nous attendons à ce que les nourrissons regardent plus les visages lors des associations congruentes pour les stimuli non natifs (lorsque les visages d'autres types sont associés à une langue étrangère), mais pas nécessairement pour les stimuli natifs.
- A 9 et 12 mois, nous nous attendons à ce que les nourrissons regardent différemment les visages lors des associations congruentes et lors des associations incongruentes. Plus précisément, nous nous attendons à ce que les nourrissons regardent plus les visages lors des associations congruentes que lors des associations incongruentes, conformément aux résultats trouvés par Xiao et al. (2016). Nous nous attendons potentiellement à un effet plus fort à 12 mois qu'à 9 mois.

II. Matériels et méthodes

II.1. Participants

118 nourrissons ont été inclus dans l'analyse finale répartis de la manière suivante selon les groupes d'âges et les conditions :

Tableau 3. Répartition des nourrissons par groupes d'âges et par condition (type de visage et type de langue). Entre parenthèse : la moyenne d'âge et l'écart type de l'âge par sous-groupe.

	Visages de même type		Visages d'autre type	
	Langue maternelle	Langue étrangère	Langue maternelle	Langue étrangère
6 mois	11 (M = 190 ; SD = 7.9)	14 (M = 191 ; SD = 7.9)	10 (M = 190 ; SD = 7)	8 (M = 194 ; SD = 10.5)
9 mois	9 (M = 280 ; SD = 9.5)	10 (M = 279 ; SD = 7.4)	5 (M = 277 ; SD = 4.2)	6 (M = 286 ; SD = 11)

12 mois	18 (M = 376 ; SD = 10.2)	11 (M = 376 ; SD = 8.2)	6 (M = 376 ; SD = 14)	10 (M = 378 ; SD = 7.6)
---------	-----------------------------	----------------------------	--------------------------	----------------------------

Aucun nourrisson bilingue ou exposé à des visages d'autres types n'est inclus dans les conditions non-natives.

II.2. Stimuli

Pour les stimuli audios, nous avons enregistré deux locutrices françaises et deux locutrices allemandes récitant un texte adapté du petit chaperon rouge. Les locutrices avaient pour consignes de réciter le texte comme si elles s'adressaient à des enfants. Le texte fut spécialement adapté pour que des séquences de 10 secondes sans phrases interrompues puissent être extraites. Ainsi, six séquences de 10 secondes ont été composées à partir de ces enregistrements. Le volume fut normalisé entre toutes les séquences.

Pour les stimuli visuels, six photographies en couleur de visages de femmes de type européen et six photographies de visages de femmes de type asiatique furent utilisées. Les visages des modèles étaient cadrés à partir du bas du cou., et les cheveux étaient visibles. L'expression des modèles était neutre. Le fond de chaque photographie était noir.

II.3. Procédure

Les nourrissons étaient installés soit sur les genoux d'un des parents, soit dans un siège adapté pour nourrisson. Les nourrissons étaient assignés à l'une des quatre conditions de manière semi-aléatoires (les nourrissons bilingues ou exposés à des visages d'autres types étaient assignés aux conditions natives) : visages de même type associés à la langue maternelle, visages de même type associé à une langue étrangère, visages d'un autre type associés à la langue maternelle, visages d'un autre type associé à une langue étrangère. Pour tous les groupes, l'étude commençait par un visage apparaissant au centre de l'écran (dimension sur l'écran 25 * 19 cm ; voir Figure 35). Celui-ci restait à l'écran pendant 10 secondes, puis disparaissait au profit d'un *attention getter* (de type formes géométriques lumineuses). Cet *attention getter* restait à l'écran pendant 10 secondes pendant lesquels un stimulus audio était joué. Cette alternance entre visage et stimulus langagier continuait jusqu'à ce que les six différents visages et les six différentes portions de l'histoire du petit chaperon rouges soient présentés. L'ordre de présentation des différents visages et des différents extraits vocaux était aléatoire. La locutrice (deux locutrices par type de langue) était contrebalancée entre les nourrissons.

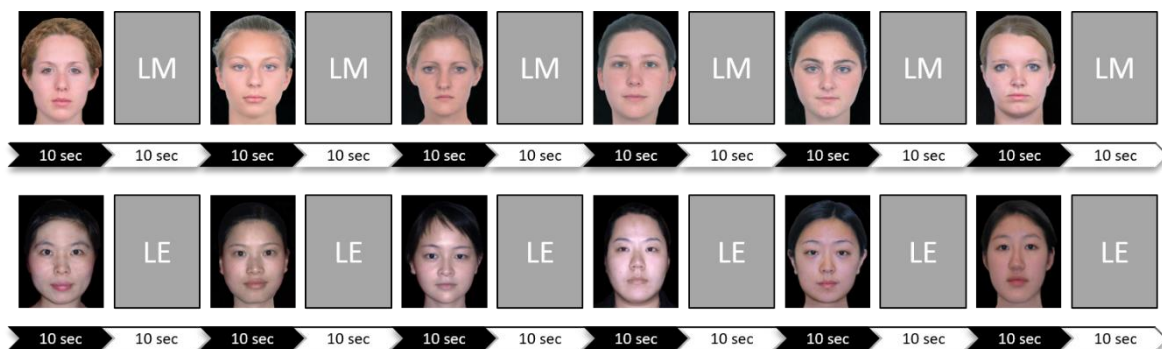


Figure 35. Schéma de la procédure de l'étude 3.

En haut : pour la condition avec visages de même type et langue maternelle. En bas : pour la condition avec visages d'autre type et langue étrangère. LM = langue maternelle ; LE = langue étrangère.

II.4. Analyse

Le temps total de fixation a été comptabilisé pour chaque visage vu par chaque nourrisson (six essais par nourrissons). Le premier visage présenté se fait avant association avec une langue. Le temps de regard sur le premier essai et le temps de regard sur les essais 2 à 5 ont donc été analysés séparément. Les essais présentant un temps de regard nul ne sont pas pris en compte dans les analyses (31 essais sur 708). Les temps de regard sur les essais 2 à 5 ont été cumulés pour chaque nourrisson puis divisé par le nombre d'essais (avec un temps de regard non nul) pour donner par nourrisson un temps moyen de regard vers le visage, après association avec une langue. Les temps de regard sont exprimés en secondes (un essai durant 10 secondes).

III. Résultats

Avant toute chose, nous nous sommes intéressés à l'évolution du temps de regard au fil des essais et selon l'âge, indépendamment des conditions. Une ANOVA comprenant le groupe d'âge (6, 9 ou 12 mois) ainsi que l'essai (de 1 à 6) montre un effet principal des deux facteurs mais pas d'effet d'interaction (effet du groupe d'âge : $F(2,659) = 5.952, p = .003, Eta^2 = .017$; effet de l'essai : $F(5,659) = 9.111, p < .001, Eta^2 = .063$). Concernant l'effet de l'âge, une analyse descriptive ainsi qu'un test de Kruskal-Wallis montrent que le temps de regard vers les visages diminue avec l'âge ($H(2) = 9.855, p = .007$). Des tests post-hoc de Wilcoxon montrent cependant que cette baisse n'est significative qu'entre les nourrissons de 6 mois et les nourrissons de 12 mois ($p = .007$). Concernant l'effet de l'essai, une analyse descriptive et un test de Kruskal-Wallis montre que le temps de regard vers les visages baisse au fil des essais ($H(5) = 43.061, p < .001$). Des post-hoc de Wilcoxon montrent que cette baisse n'est

significative qu'entre le 1^{er} essai et les 5 autres (pour le temps de regard vers l'essai 1 contre les autres essais : toutes les p -values $> .01$; pour les autres comparaisons : $p > .05$; p -values ajustées avec des corrections de Bonferroni).

III.1. Temps de regard avant association

Nous nous sommes ensuite intéressés au temps de regard vers le visage lors du premier essai, avant association avec une langue. Cette analyse porte sur le premier essai de 117 nourrissons, un nourrisson ayant un temps de regard nul et ayant donc été retiré. Parmi les 117 nourrissons, 73 ont passé la condition avec un visage de même type (M temps de regard = 7.45, SD = 2.16) et 44 avec un visage d'autre type (M temps de regard = 6.97, SD = 2.36). Nous avons d'abord conduit une ANOVA avec comme facteur le type de visage et l'âge, puisque l'âge était un facteur significatif lorsque tous les essais étaient pris en compte. Le modèle n'indique aucun effet significatif. Un test de Wilcoxon confirme cette absence de différence significative concernant le temps de regard vers le premier visage présenté en fonction de son type (test de Wilcoxon : $p = .326$; voir Figure 36).

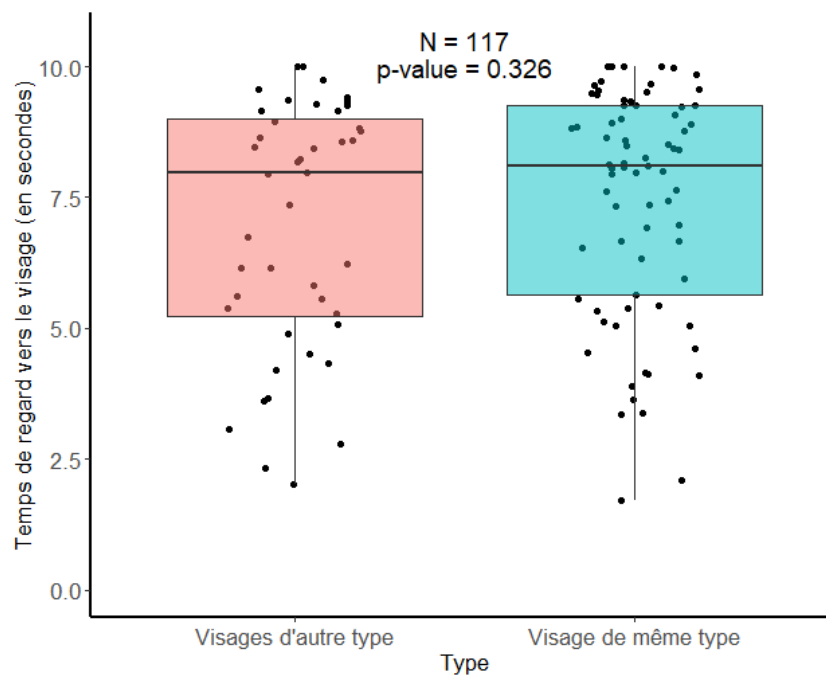


Figure 36. Temps de regard vers les visages durant le premier essai (avant association avec une langue) en fonction de son type. La p -value correspond à un test de Wilcoxon.

III.2. Temps de regard après association

Nous avons ensuite comparé le temps de regard vers les visages après associations avec une langue. Nous avons d'abord conduit une ANOVA sur le temps de regard en fonction du type

de visage (de même ou d'autre type), du type de langue (maternelle ou étrangère) et de l'âge (6, 9 ou 12 mois). Aucun effet ne ressort de manière significative (toutes les p-values > .05).

Notre hypothèse principale étant que le temps de regard vers les visages change en fonction du type de langue associé après le rétrécissement perceptif, nous avons tout de même conduit une analyse de l'évolution du temps de regard en fonction de l'âge, par condition (voir Figure 37). Nous avons fait une régression linéaire du temps de regard moyen par essai en fonction de l'âge, pour chaque condition. Conformément à l'étude Xiao et al. (2016), nous avons calculé le coefficient de corrélation de Pearson pour chaque régression linéaire. On observe une absence de corrélation entre le temps de regard vers les visages et l'âge quelle que soit la condition (toutes les p-values > .05).

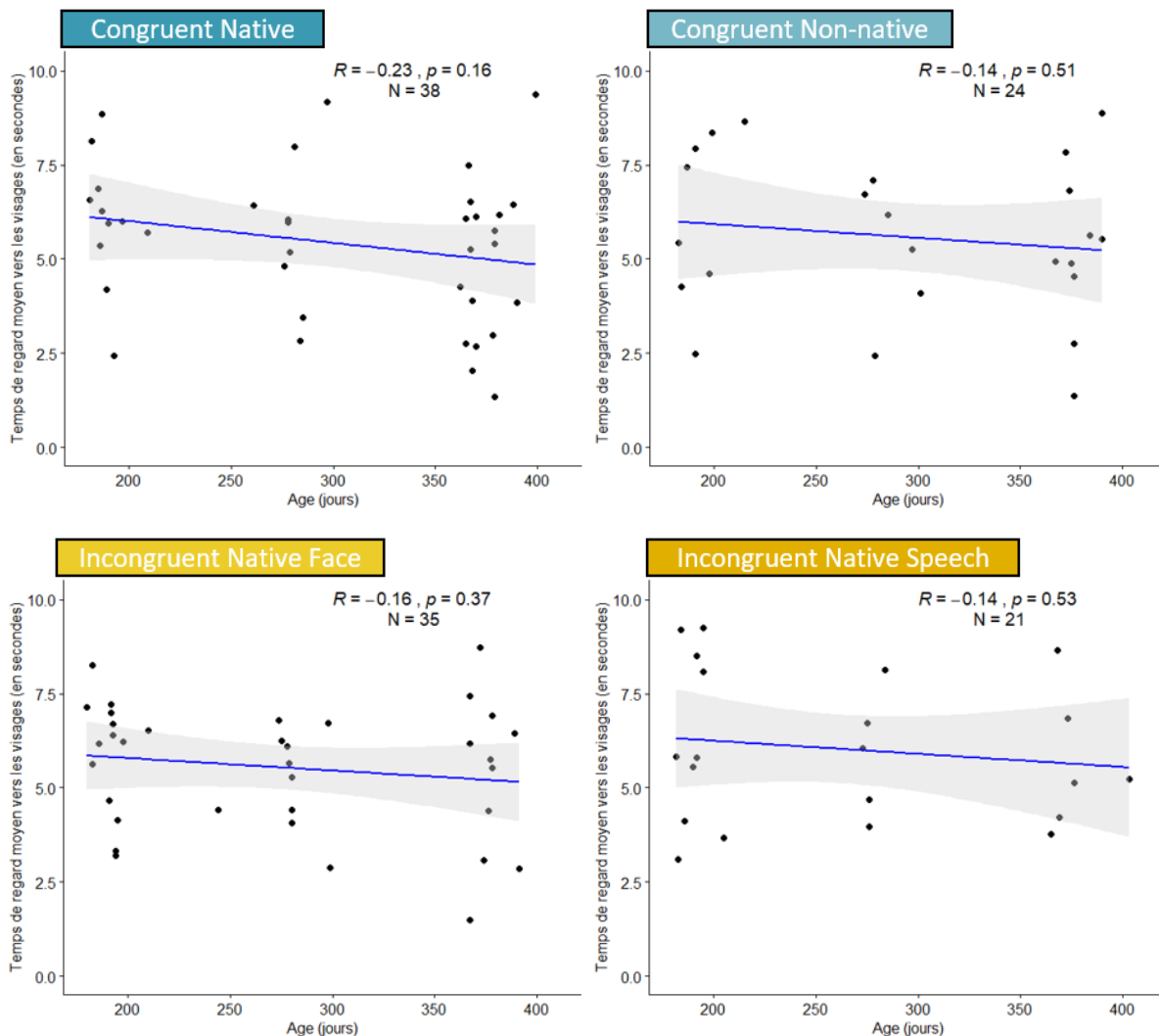


Figure 37. Evolution du temps de regard moyen par essai (en secondes) vers les visages en fonction de l'âge (en jours), pour chaque condition.

En haut à gauche : pour la condition avec visages de même type et langue maternelle. En haut à droite : pour la condition avec visages d'autre type et langue étrangère. En bas à gauche : pour la condition avec visages de même type et langue maternelle. En bas à droite : pour la condition avec visages d'autre type et langue maternelle. La *p-value* et le *R* correspondent à une corrélation de Pearson.

Nous nous sommes aussi intéressés au temps de regard vers les visages par groupe d'âge (voir Figure 38). L'équipe allemande avec laquelle nous collaborons sur cette étude ont de cette manière mis en évidence des différences des temps de regard au sein des nourrissons de 3 mois en fonction de l'association présentée. Ils trouvent en effet que les nourrissons regardent plus les visages dans les situations congruentes que dans la situation incongruente lorsque les visages de même type sont associés à une langue étrangère. Via le même type d'analyse, nous ne trouvons cependant pas de différence significative pour les autres groupes d'âges.

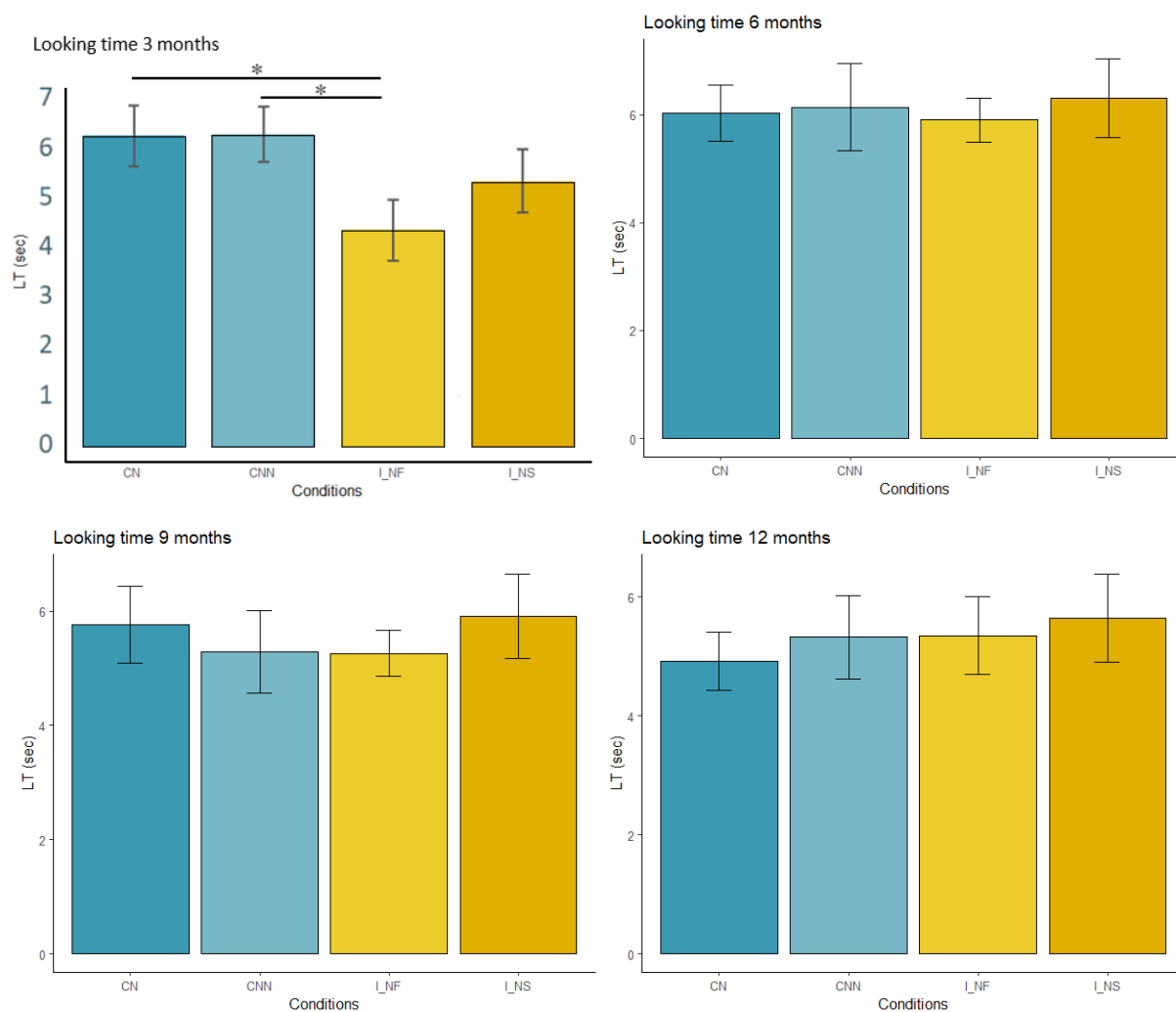
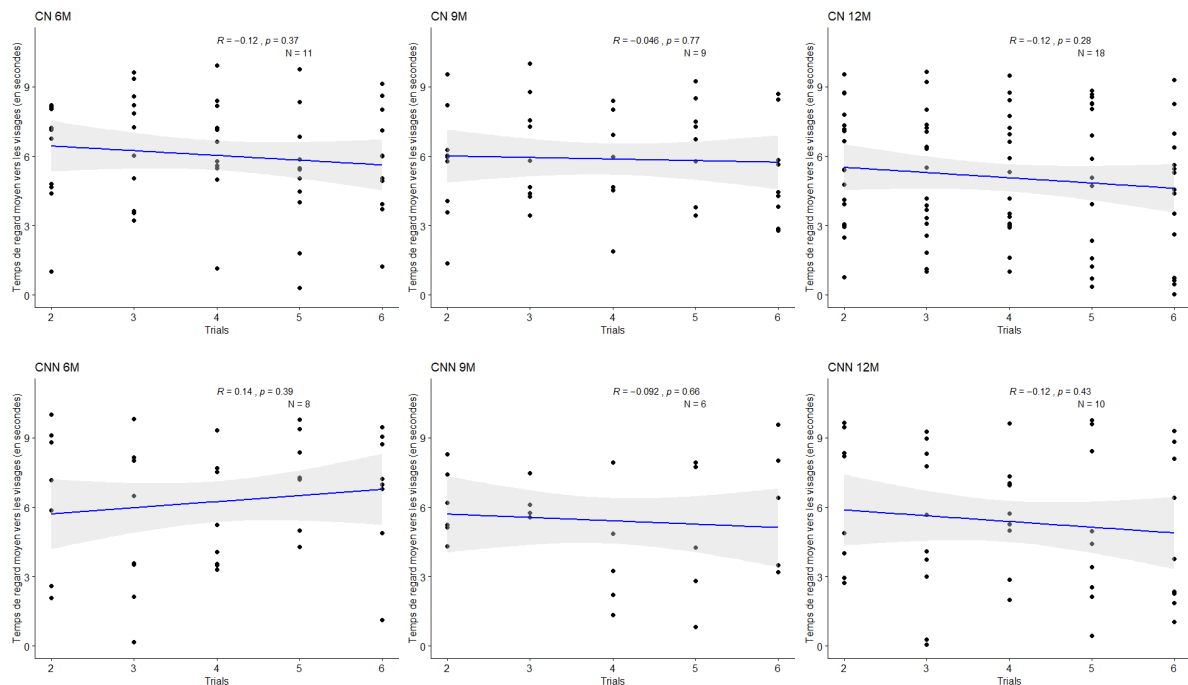


Figure 38. Temps de regard moyen (en secondes) vers les visages par âge et par condition.

En haut à gauche = pour les nourrissons de 3 mois (tiré des résultats de Krasotkina et al.) ; en haut à droite = pour les nourrissons de 6 mois ; en bas à gauche = pour les nourrissons de 9 mois ; en bas à droite = pour les nourrissons de 12 mois. CN = congruent natif (visage de même type et langue maternelle) ; CNN = congruent non-natif (visage d'autre type et langue étrangère) ; I_NF = incongruent visage natif (visage de même type et langue étrangère) ; I_NS = incongruent langue native (visage d'autre type et langue maternelle).

Enfin, nous avons effectué des régressions linéaires sur les temps de regard par essai pour chaque groupe d'âge et sous-condition (voir Figure 39). Cette analyse permet de voir si l'intérêt au fil des essais baisse (ou augmente) différemment en fonction des sous-condition. Nous ne trouvons cependant pas de différence significative pour cette analyse.



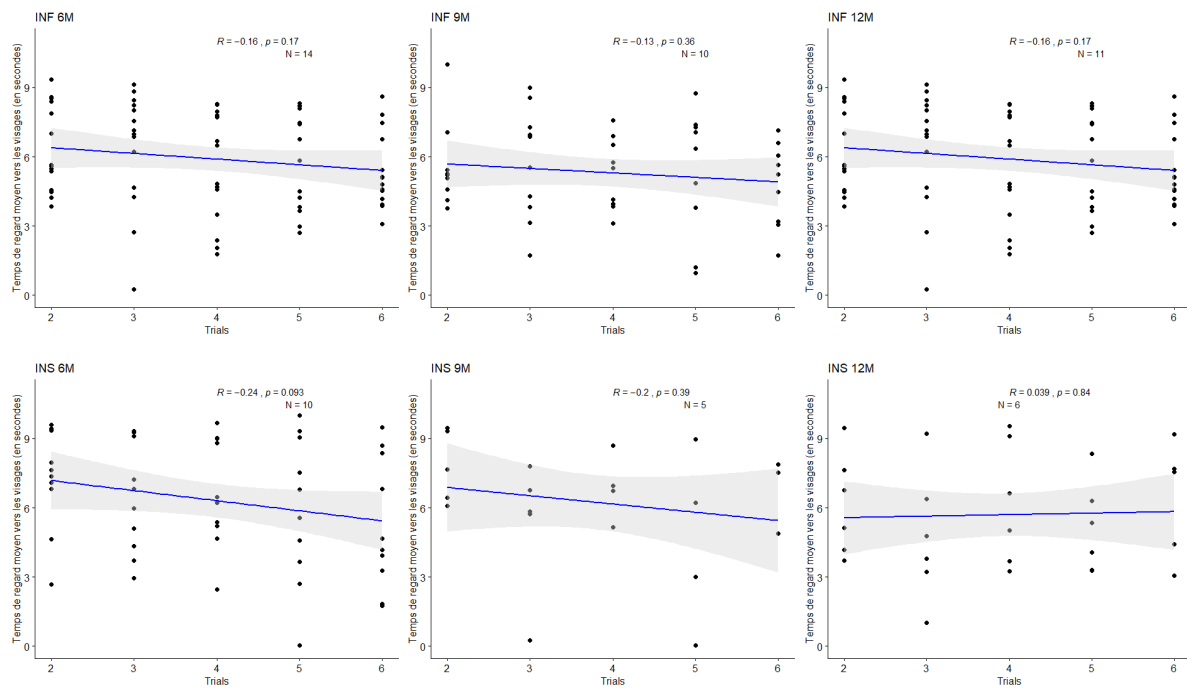


Figure 39. Temps de regard (en secondes) vers le visage en fonction de l'essai, par groupe d'âge et par condition.

A gauche, pour les nourrissons de 6 mois ; au milieu, pour les nourrissons de 9 mois ; à droite, pour les nourrissons de 12 mois. De haut en bas, la condition avec visages de même type et langue maternelle, la condition avec visages d'autre type et langue étrangère, la condition avec visages de même type et langue étrangère, la condition avec visages d'autre type et langue maternelle.

IV. Discussion

Le but de cette étude était de mieux caractériser les associations que les nourrissons forment entre les différents types de visages et les différents types de langues au cours de la première année de vie. Pour cela, nous avons utilisé un nouveau protocole d'association intermodale mis en place par Xiao et al. (2016). Nous nous sommes intéressés au temps de regard vers les visages en fonction de leur type et de la langue qui leur était associée. Notre hypothèse principale était que les associations congruentes seraient regardées différemment des associations incongruentes, particulièrement après le rétrécissement perceptif. Plus précisément, nous nous attendions à une hausse du temps de regard vers les visages durant la présentation d'associations congruentes. Nos résultats ne permettent cependant pas de mettre en évidence une différence dans l'évolution des temps de regard vers les visages au fil du développement en fonction de l'association voix/visage durant le rétrécissement perceptif.

IV.1. Baisse générale du temps de regard

De manière globale, nous observons une baisse du temps de regard à mesure que les nourrissons grandissent. Cette différence n'est significative qu'entre 6 et 12 mois. Nous retrouvons ici la même différence que nous observons pour notre étude sur l'effet McGurk (Chapitre 6), ou nous trouvons aussi que les nourrissons de 6 mois regardent significativement plus la phase de familiarisation que les nourrissons de 12 mois. De même, pour l'étude sur la correspondance audio-visuelle de voyelle (Chapitre 5), nous trouvons que les nourrissons de 3 mois passent plus de temps à regarder les stimuli que les nourrissons de 9 mois. Lorsqu'un seul stimulus visuel est présent à l'écran, les nourrissons plus jeunes semblent moins détourner le regard que les nourrissons les plus âgés. Cela peut être dû à une moindre mobilité ou encore une attention moins exigeante. En revanche, cette légère baisse d'attention ne diffère pas de manière significative en fonction de l'association voix/visage présentée.

IV.2. Nature de l'association langage/visage

Une possibilité est que le type d'association mise en place dans le protocole de Xiao et al. (2016) fonctionne pour la musique mais pas pour le type de langue. Le protocole mis en place par Xiao consiste originellement à associer de manière séquentielle de la musique à des visages. Dans le cadre de notre étude, la musique est remplacée par du langage. Le type d'association résultant de l'étude de Xiao et de la nôtre n'est potentiellement pas équivalent. Dans l'étude de Xiao, les nourrissons n'infèrent (à priori) pas que la musique est produite par les visages ; les nourrissons associent la valence (positive/négative) ou la familiarité de la musique à celle des visages. En revanche, il est possible qu'aucune valence ne soit extrait des stimuli langagiers. Aussi, il est à noter que lorsque le signal de parole est présenté de manière asynchrone avec la vidéo (Lewkowicz, 1996) et même de manière séquentielle avec une vidéo (Pons et al., 2009 ; Lewkowicz & Pons, 2013), les nourrissons semblent relier les deux événements en se basant entre autres sur des indices temporels. De plus, notre étude 4 (Clerc et al., 2021) ainsi que l'étude d'Hillairet de Boisferon et al. (2021) montrent que les nourrissons peuvent associer du langage parlé avec un visage même lorsque le visage n'est pas présenté via une vidéo mais simplement via une photo. En revanche, il n'est pas dit que les nourrissons puissent relier de la parole présentée de manière séquentielle avec une photo. Dans la présente étude, le langage n'est jamais présenté de manière conjointe avec les visages ; durant les stimuli de langage, une cible mouvante est présente à l'écran. Il se peut donc que les nourrissons ne soient pas en mesure de faire l'association. L'expérience serait alors pour eux une suite d'extraits langagiers et une suite de visages, indépendantes l'une de l'autre.

IV.3. Des associations plus précoces ?

Toutefois, les résultats concernant les nourrissons de 3 mois montrent que ceux-ci regardent plus les visages lors des associations congruentes que lors des associations incongruentes et plus spécifiquement lorsque c'est la langue qui est non-native. Cela laisse entendre que les nourrissons arrivent, dans une certaine mesure, à former des associations entre un type de langue et un type de visage via ce protocole. Observer une différence comportementale en fonction de l'association langue/visage dès 3 mois peut sembler surprenant. Cependant, nous avons vu avec l'étude 1 que dès 3 mois, les nourrissons traitent différemment l'information audio-visuelle en fonction du type de visage qui la produit. De plus, l'association entre un visage de même type que le nourrisson et sa langue maternelle fait partie des associations les plus présentes dans l'environnement du nourrisson. On observe des différences de temps de regard concernant les associations entre type de visage et musique vers 9 mois (Xiao et al., 2016). Il est possible qu'un type d'association beaucoup plus présent dans l'environnement du nourrisson puisse entraîner des modifications comportementales de manière plus précoce. Les résultats de Uttley et al. (2013) montrent que les nourrissons semblent former des associations langues/visages dès 6 mois, ce qui est donc plus tôt que les 9 mois trouvé pour l'association musique/visage. Cela vient corroborer l'idée que l'omniprésence des visages et du langage dans l'environnement des nourrissons entraîne des associations plus précoces ; à noter toutefois que les protocoles sont assez différents entre l'étude de Uttley et celle de Xiao et la nôtre. Nous ne trouvons pas de différence significative de temps de regard vers les visages en fonction de l'association langue/visage après 3 mois. Notre hypothèse que le temps de regard vers les visages dans les associations congruentes se renforceraient avec l'âge n'est donc pas vérifié. Toutefois, les résultats sont à prendre avec précaution. En effet, la collecte de donnée est toujours en cours, car le nombre de sujet par condition n'est pas atteint. Nous avons encore plusieurs conditions pour certains groupes d'âge ou nous avons moins de 10 nourrissons. Les interprétations seront donc à reconsidérer après la collecte complète des données.

IV.4. Catégorisation de la langue

On peut aussi se demander si les nourrissons de l'étude arrivent clairement à identifier la langue qui leur est présentée. La prosodie est un des premiers éléments par lequel les nouveau-nés parviennent à identifier leur langue (Mehler et al., 1988 ; Nazzi et al., 1998, 2000). La prosodie est une information contenue principalement dans les basses fréquences, Nazzi et al. (1998) montrant que les nourrissons arrivent toujours à identifier les éléments rythmiques du langage lorsqu'un filtre passe-bas est appliqué. In utero, ce sont les basses fréquences qui sont le mieux

transmises au fœtus, au contraire des informations contenues dans les hautes fréquences (Bench, 1968 ; Busnel et al., 1992). Ainsi, les fœtus sont déjà familiarisés à la prosodie de leur langue (Gervain & Mehler, 2010 ; Gervain, 2018). L'expérience précoce des nourrissons avec la prosodie de leur langue leur permet alors de l'identifier dès la naissance. Cela peut expliquer pourquoi les nourrissons de 3 mois arrivent à catégoriser les extraits langagiers comme appartenant à leur langue ou à une langue étrangère (le français et l'allemand n'appartenant pas à la même classe rythmique). En grandissant, les nourrissons vont tirer d'autres informations de leur langue, notamment des informations phonémiques (Bosch & Sebastián-Gallés 2001, Ramon-Casas et al. 2009). L'acquisition lexicale commencerait dès la deuxième moitié de la première année de vie (Swingley, 2009), l'attention des nourrissons pouvant donc basculer de la rythmicité globale de la langue à la segmentation et l'identification de mots. Si leur attention est détournée des informations rythmiques globales, il n'est pas sûr que les nourrissons puissent identifier clairement si un extrait appartient à leur langue maternelle ou à une langue étrangère.

De plus, la plupart des études mettant en évidence les capacités de discrimination de langues des nourrissons se font généralement via des protocoles intra-sujets, présentant au même nourrisson deux langues différentes (Mehler et al., 1988 ; Nazzi et al., 1998, 2000 ; Bosch & Sebastián-Gallés, 1997). Via notre protocole, un nourrisson n'a pas d'élément de comparaison puisqu'il n'est exposé qu'à une seule langue. Catégoriser un extrait court comme appartenant à leur langue ou à une autre langue est donc peut-être particulièrement difficile pour eux. La plupart des études en oculométrie montrent que les visages de même type et d'autres types ne sont pas regardés significativement différemment en termes de temps de regard lorsque présentés seuls (Liu et al., 2010 ; Wheeler et al., 2011 ; Xiao et al., 2013). Si l'association type de langue/type de visage n'est pas réellement faite par les nourrissons de notre étude, il est normal de ne pas observer de différence significative de temps de regard sur la simple base du type de visage. De manière intéressante, des données non publiées de l'équipe allemande sur une autre étude consolident cette interprétation. Leur étude montre en effet que des nourrissons de 5 à 13 mois ne présentent pas de comportements différents dans une tâche de suivi de regard en fonction de si la personne à l'écran parle la langue maternelle du nourrisson ou une langue étrangère. Les extraits de langues sont là aussi très courts (quelques secondes) et ne suffisent peut-être pas aux nourrissons pour catégoriser la langue correctement à ces âges-là. En revanche, les nourrissons plus âgés, entre 15-18 mois, font une distinction entre les locuteurs parlant la langue maternelle et ceux parlant une langue étrangère, suivant plus couramment le regard de la personne lorsqu'elle parle leur langue maternelle. Suivant cette idée, il est plus

probable qu'une catégorisation de la langue se fasse vers la fin de notre étude qu'au début. Nous avons donc effectué le même type d'analyse que précédemment mais portant uniquement sur le temps de regard lors du dernier essai, là où la catégorisation du langage et donc une potentielle association type de langue/type de visage est le plus susceptible d'émerger. Voici ce que nous observons :

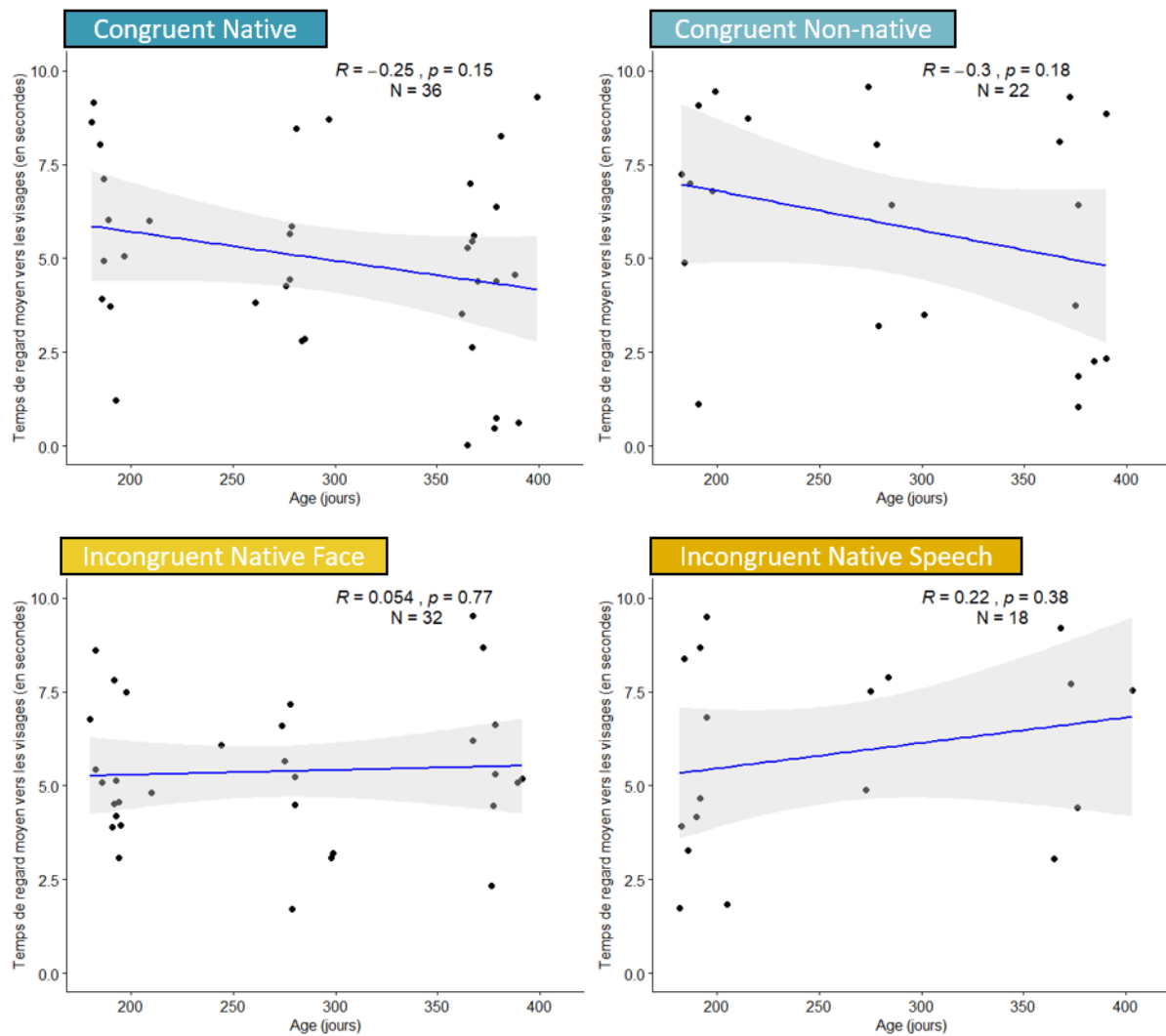


Figure 40. Evolution du temps de regard (en secondes) pour le dernier essai vers le visage en fonction de l'âge (en jours), pour chaque condition.

En haut à gauche : pour la condition avec visages de même type et langue maternelle. En haut à droite : pour la condition avec visages d'autre type et langue étrangère. En bas à gauche : pour la condition avec visages de même type et langue maternelle. En bas à droite : pour la condition avec visages d'autre type et langue maternelle. La p-value et le R correspondent à une corrélation de Pearson.

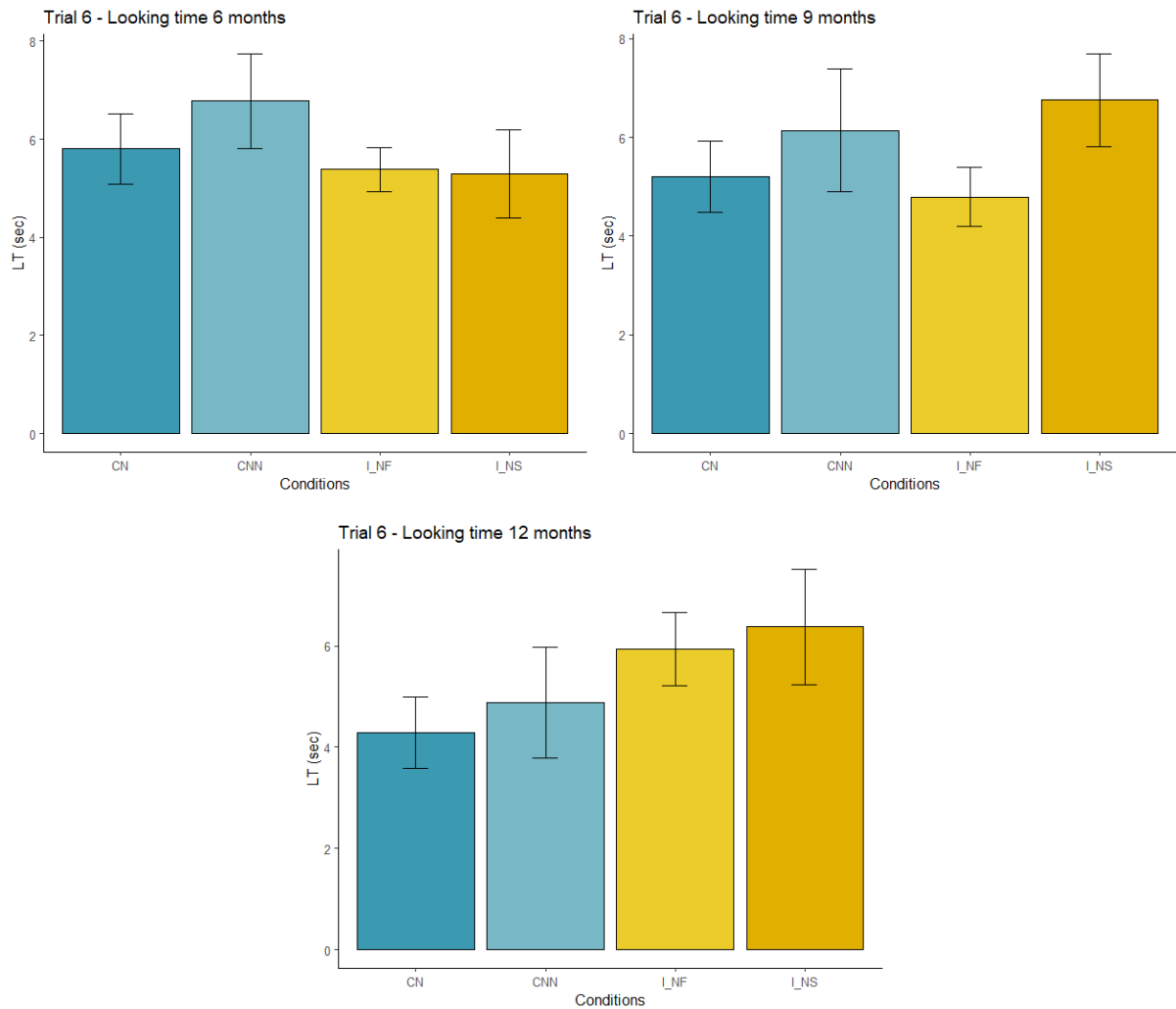


Figure 41. Temps de regard (en secondes) lors du dernier essai vers le visage par âge et par condition.

En haut à gauche = pour les nourrissons de 6 mois ; en haut à droite = pour les nourrissons de 9 mois ; en bas au milieu = pour les nourrissons de 12 mois. CN = congruent natif (visage de même type et langue maternelle) ; CNN = congruent non-natif (visage d'autre type et langue étrangère) ; I_NF = incongruent visage natif (visage de même type et langue étrangère) ; I_NS = incongruent langue native (visage d'autre type et langue maternelle).

Tout d'abord, il faut tenir compte du fait que bien que le nombre de nourrissons par groupe reste le même, l'analyse porte ici sur un seul essai par nourrisson, diminuant ainsi la puissance statistique. Aucun résultat n'atteint ici le seuil de significativité. Ceci étant dit, le pattern observé sur le dernier essai est différent de ce que l'on observe sur l'ensemble des essais. Plus précisément, il apparaît que les associations congruentes deviennent moins regardées à mesure que les nourrissons grandissent, ce qui n'est pas le cas des association incongruentes. Cette trajectoire développementale pourrait se rapprocher de celle trouvé par Mercure et al. (2018)

ou bien encore Lewkowicz & Pons (2012). En effet, leurs études montrent que des nourrissons d'environ 7 mois et de 10-12 mois (respectivement) montrent plus d'intérêt pour des associations voix/visage incongruentes plutôt que congruentes. Bien évidemment, les interprétations des résultats de notre étude sont pour le moment très incertaines mais méritent d'être réévalué lorsque plus de nourrissons seront inclus dans l'étude.

Chapitre 8 : Influence de la langue sur la reconnaissance de visages d'autre type

I. Introduction

Nous avons vu en introduction de cette thèse que le phénomène appelé rétrécissement perceptif est observable à la fois pour le développement du traitement des visages et pour le développement du traitement du langage. Pour le traitement des visages, nous avons vu que le rétrécissement perceptif a pour conséquence l'émergence de l'ORE, impliquant que les visages de personnes de type non majoritaire dans l'environnement du nourrisson sont plus difficilement reconnus, et ce dès 9 mois. Les études de Kelly et al. (2007a, 2009) illustrent bien ce phénomène ; si un visage d'autre type est présenté durant une phase habituation (durant généralement entre 40 et 60 secondes), il ne sera plus reconnu à partir de 9 mois lorsque présenté côte à côte avec un nouveau visage d'autre type lors d'un test de reconnaissance de 5 secondes. Cependant, nous avons aussi vu que dans d'autres contextes expérimentaux, la reconnaissance de visage d'autre types est toujours possible après 9 mois. Le simple fait d'augmenter les temps de familiarisation et de tests, d'ajouter des émotions ou de renforcer l'attention visuelle avant la présentation du visage permet de restaurer la reconnaissance des visages d'autre types (ou même d'une autre espèce) chez les nourrissons (Fair et al., 2012 ; Quinn et al., 2020 ; Markant et al., 2016). Minar & Lewkowicz (2017) ont aussi montré qu'une phase de familiarisation multisensorielle, consistant en un visage associé à un son, permettait à des nourrissons de 10-12 mois de reconnaître des visages d'autre type, ce qu'une présentation unimodale ne permettait pas. Sachant donc que la reconnaissance de visage d'autre type peut être restaurée après 9 mois et à la vue des liens étroits entre traitement des visages et traitement du langage, le but de notre étude était d'étudier l'impact du langage sur la reconnaissance de visage d'autre type, après l'émergence du rétrécissement perceptif.

Uttley et al (2013) ont apportés des éléments indiquant que les nourrissons tendent à former des associations entre type de visage et type de langue. Les résultats montrent que les nourrissons semblent dès 6 mois avoir une représentation multimodale des individus incluant type de visage et langue parlée.

I.1. L'effet de familiarité du langage

Des recherches récentes portant sur l'effet de familiarité du langage (Language Familiarity Effect dans la littérature anglosaxonne, *LFE*) mettent aussi en évidence un lien entre traitement du visage et du langage. Le LFE est un effet robuste montrant que la voix d'un locuteur est plus

facilement identifiée si celui-ci parle la/une langue maternelle du participant (Goggin et al., 1991 ; Johnson et al., 2018 ; Perrachione et al., 2011 ; Perrachione & Wong, 2007) ; cet effet est robuste aussi chez le nourrisson (Fecher & Johnson, 2018 ; Johnson et al., 2011). La langue qu'une personne parle ne semble pas seulement impacter la reconnaissance de sa voix, elle impacte aussi la reconnaissance physique. Les adultes présentent en effet plus de facilités à reconnaître des personnages animés (représentés par des formes géométriques) ou même de vraies personnes (de même type qu'eux) quand ceux-ci sont associés à leur langue maternelle (Bregman & Creel, 2014 ; Kadam et al., 2016 ; Baus et al., 2017). Puisque le LFE impacte la reconnaissance de visage chez les adultes, Hillairet de Boisferon et al. (2021) ont étudié si une influence similaire existe chez le nourrisson. Pour ce faire, les auteurs ont familiarisé pendant 30 secondes des nourrissons de 9 et 12 mois à un visage féminin de même type qu'eux (de type européen) associé à leur langue maternelle (le français) ou à une langue étrangère (l'allemand). Puis, la reconnaissance du visage de la familiarisation était testée via la présentation simultanée du visage de la familiarisation et d'un nouveau visage féminin de même type. Lors de ce type de tâche, les nourrissons présentent un intérêt pour la nouveauté ; la reconnaissance du visage de la familiarisation était inférée lorsque plus de temps était passé à regarder le nouveau visage. La reconnaissance d'un visage de même type, sans association avec du langage, est une tâche que les nourrissons de cet âge réussissent. Ici, les auteurs ont trouvé que les nourrissons ne reconnaissent plus le visage de la familiarisation lorsque celui-ci a été associé à une autre langue que la leur durant la familiarisation. Dès 9 mois, les nourrissons semblent déjà avoir des attentes sur les associations visages/langages qu'ils rencontrent.

I.2. Objectif et hypothèses de l'étude 4

La littérature nous indique donc d'une part que le rétrécissement perceptif peut être atténué et d'autre part que la langue que parle un individu peut impacter sa reconnaissance. Le but de notre recherche est donc de savoir si la familiarité d'une langue (maternelle ou étrangère) peut moduler l'ORE chez les nourrissons. Les études sur le LFE chez l'adulte indiquent que la langue maternelle semble avoir un effet bénéfique sur la reconnaissance de visage de même type ; de ce fait, présenter des visages d'autres types avec la langue maternelle du nourrisson pourrait améliorer la reconnaissance de ces visages. En lien avec cette idée, la langue que parle les individus peut agir comme un critère de classification sociale, les adultes et les enfants traitant différemment les personnes parlant une langue étrangère des personnes parlant leur(s) langue(s) maternelle(s) (Lieberman et al., 2017, 2018). L'accent est aussi un élément du langage qui impacte la perception que les gens ont du locuteur (Ryan et al., 1977 ; Elwell et al., 1984 ; Kalin

& Rayko, 1978). Dès 9 ans, les enfants ont déjà des préjugés quant à la fiabilité, l'intelligence ou la gentillesse d'individus en fonction de leur accents (Kinzler et al., 2009 ; Kinzler & DeJesus, 2013). Pour ce qui est de la différence de traitement en fonction de la langue parlée, les nourrissons, dès 10 mois, présente un engagement social plus facile avec les personnes parlant leur langue maternelle (Kinzler et al., 2007 ; Marno et al., 2016). Cependant, une autre possibilité est que l'association entre les visages d'autres types et la langue maternelle soit inattendue pour les nourrissons, comme le suggère les résultats de Uttley et al (2013). Très majoritairement, l'expérience que les nourrissons ont avec leur langue maternelle est associée à des visages de leur type. Ainsi, par inférence statistique, les attentes du nourrisson sont que lorsque celui-ci entend sa langue maternelle, il s'attend à ce qu'elle soit produite par un visage de son type. En accord avec cette idée, l'association entre leur langue maternelle et un visage d'autre type serait en désaccord avec les attentes des nourrissons et pourrait impacter leur capacité à traiter et donc reconnaître le visage. En revanche, l'association entre une langue étrangère et un visage d'autre type ne devrait pas avoir la même conséquence. Les nourrissons n'étant en effet que très peu exposés aux visages d'autres types que les leurs ainsi qu'à des langues étrangères, ils n'ont que peu voire pas d'attentes à leurs égards. Ainsi, l'association entre langue étrangère et visage d'autre type pourrait ne pas avoir d'effet, et simplement résulter sur le classique ORE observée pour les visages d'autres types. Une dernière possibilité est que la familiarisation multisensorielle aide les nourrissons à mémoriser le visage, comme le suggère l'étude de Minar & Lewkowicz (2017). L'association entre un flux audio et un stimulus visuel pourrait résulter en un stimulus multisensoriel plus attractif, améliorant l'attention des nourrissons durant la phase de familiarisation. En revanche, pour que cette dernière hypothèse soit possible, il ne faut pas que l'association proposée soit incongruente avec les attentes des nourrissons ; c'est donc lors de l'association avec une langue que cette hypothèse pourrait permettre de rétablir la reconnaissance des visages d'autres types.

Bien que beaucoup d'études se soient intéressées à la reconnaissance de visages d'autres types par les nourrissons, il n'y a pas d'études (à notre connaissance) qui se soient intéressées à l'impact du type de langue sur la reconnaissance de ces visages. Cette question est pertinente car les visages que nous tâchons de reconnaître dans la vie de tous les jours sont généralement associés à du langage (pour les nourrissons comme pour les adultes). Nous avons ainsi testé des nourrissons après le rétrécissement perceptif afin de voir si celui-ci pouvait être modulé. Des nourrissons de 9 et 12 mois ont été évalués sur une tâche de reconnaissance de visage d'autre après familiarisation, soit en association avec leur langue maternelle soit en association avec

une langue étrangère. Avoir deux groupes d'âge nous permet d'éventuellement observer des effets de maturation et/ou d'expérience. Nous avons choisi le français et l'allemand, ces deux langues appartenant à des classes rythmiques différentes que les nouveau-nés sont déjà capables de différencier (Mehler et al., 1988 ; Nazzi et al., 1998 ; Ramus et al., 1999). Nous avons utilisé des visages asiatiques comme visages d'autres types, l'ORE ayant été montré plusieurs fois chez les nourrissons européens testés avec des visages asiatiques (Anzures et al., 2012 ; Hayden et al., 2007 ; Kelly et al., 2007a).

II. Matériel et méthode

II.1. Participants

80 nourrissons ont participé à l'étude, 40 9 mois (M age = 279 days, SD age = 7.2 days, 21 filles, 19 garçons) et 40 12 mois (M age = 377 days, SD age = 14.4 days, 15 filles, 25 garçons). Les nourrissons étaient des enfants nés à terme, recrutés à la maternité de Grenoble. Comme nos stimuli de visages d'un autre type consistaient en des visages asiatiques, nous avons exclu les nourrissons non majoritairement exposés à des visages européens et ceux exposés à des visages asiatiques, ces nourrissons étant susceptibles de présenter un ORE réduit pour les visages asiatiques. De plus, tous les nourrissons grandissaient dans un environnement parlant français. Pour chaque groupe d'âge, la moitié des nourrissons (soit 20) ont passé la condition de familiarisation avec la langue maternelle et l'autre moitié l'a passé avec la langue étrangère. Des nourrissons supplémentaires ont dû être retirés de l'étude suite à une incapacité à compléter l'étude.

II.2. Stimuli

Les stimuli de familiarisation consistaient en une photo associée à une bande audio. Les deux photos utilisées étaient celles de deux femmes asiatiques présentant une expression neutre (voir Figure 42). Le visage utilisé lors de la familiarisation était contrebalancé entre les nourrissons. Le visage non utilisé pour la familiarisation était ensuite utilisé comme nouveau visage durant la phase de test. La bande audio était une courte histoire de 30 secondes pour enfant (tirée de Boucle d'or), racontée soit en français soit en allemand à un niveau de conversation ambiante (65 dB \pm 5 dB). Deux locutrices bilingues furent enregistrées pour cela, et la locutrice était contrebalancé entre les nourrissons.

Durant les tests de reconnaissances, les stimuli utilisés étaient la photo de la familiarisation et la photo de l'autre femme asiatique non utilisée lors de la familiarisation. Les deux photos occupaient le même espace sur l'écran (20.6 cm * 18 cm) et étaient présentés côte à côte sur

l'écran devant un fond noir, séparé par un intervalle de 12 cm. Aucun audio n'était joué durant la phase de test.



Figure 42. Stimuli utilisés. Les stimuli étaient en couleurs et présentés sur un écran 22 pouces.

II.3. Procédure

Les nourrissons étaient assis sur les genoux de leurs parents à une distance approximative de 60 cm de l'écran. L'étude était divisée en une étape de familiarisation et une étape de test. Au début de la familiarisation, un attention-getter était joué au centre de l'écran. La photo de familiarisation était ensuite présentée au milieu de l'écran pendant 30 secondes, durant lesquelles l'audio était joué. La familiarisation était suivie d'un attention-getter puis de deux tests de 5 secondes entrecoupés d'un autre *attention getter*. Durant les tests, la photo de la familiarisation et la nouvelle photo étaient disposées côte à côte sur l'écran. Chaque test durait 5 secondes à partir du premier regard vers l'écran. La position droite/gauche des photos était contrebalancé entre les deux tests.

La direction du regard du nourrisson fut ensuite analysée grâce à un logiciel permettant un codage image par image. Le codeur était aveugle à la condition passée et à quelle photo se

trouvait de quel côté. Un second codeur permis d'établir un taux de confiance de très haute fiabilité (r de Pearson = .98) sur 20% des vidéos (O'Connor & Joffe, 2020).

III. Résultats

III.1. Tests de familiarisation

Nous nous sommes d'abord intéressés au temps de regard durant la familiarisation. Pour cela, nous avons réalisé une ANOVA 2 * 2 (âge : 9 mois vs 12 mois / Familiarité du langage : natif vs non-natif). L'analyse n'a révélé aucun effet significatif, ni de l'âge ($F < 1$) ni du type de langue ($F < 1$) ni de l'interaction ($F(1,76) = 1.048, p = .309$).

Des t-tests consécutifs ont confirmé l'absence de différence significative ($p > .05$) concernant le temps de regard durant la familiarisation entre les deux groupes d'âges ($M = 24.89s, SD = 4.55$ pour les 9 mois et $M = 24.79s, SD = 4.17$ pour les 12 mois) et entre les deux conditions de langue ($M = 24.49s, SD = 4.47$ pour la langue étrangère et $M = 25.18s, SD = 4.23$ pour la langue maternelle).

III.2. Préférence lors du test de reconnaissance

La proportion de temps passé à regarder le nouveau visage est calculé en divisant le temps que passe les nourrissons à regarder le nouveau stimulus par le temps total qu'ils passent à regarder les deux visages (le nouveau et le familier) ; cette proportion sert d'indicateur de préférence pour la nouveauté.

Dans un premier temps, nous avons réalisé une ANOVA 2 * 2 (âge : 9 mois vs 12 mois / Familiarité du langage : natif vs non-natif) sur ces scores de préférence pour la nouveauté. Les résultats indiquent un effet proche de la significativité concernant le type de langue ($F(1,76) = 3.482, p = .066$) mais d'effet de l'âge ($F(1,76) = 1.208, p = 0.275$) ni d'effet d'interaction entre les deux facteurs ($F < 1$).

Nous avons ensuite réalisé des t-test pour comparer les scores de préférence pour la nouveauté au niveau de hasard de 50% (c'est-à-dire autant de temps passé à regarder le nouveau visage et le visage de la familiarisation ; voir Figure 43). Lorsque familiarisé au visage associé avec une langue étrangère, le score de préférence pour la nouveauté des 9 mois ainsi que celui des 12 mois est significativement au-dessus du niveau du hasard (pour les 9 mois : $M = 54.85, SD = 8.16, t(19) = 2.66, p = 0.016, d$ de Cohen = 0.84; pour les 12 mois: $M = 57.14, SD = 8.11, t(19) = 3.94, p < .001, d$ de Cohen = 1.25). Cependant, lorsque familiarisé au visage associé à leur langue maternelle ni les 9 mois ni les 12 mois ne présentent une préférence pour la nouveauté

significative (pour les 9 mois : $M = 51.13$, $SD = 8.40$, $t(19) < 1$, $p = .555$; pour les 12 mois : $M = 53.29$, $SD = 11.23$, $t(19) = 1.31$, $p = .205$).

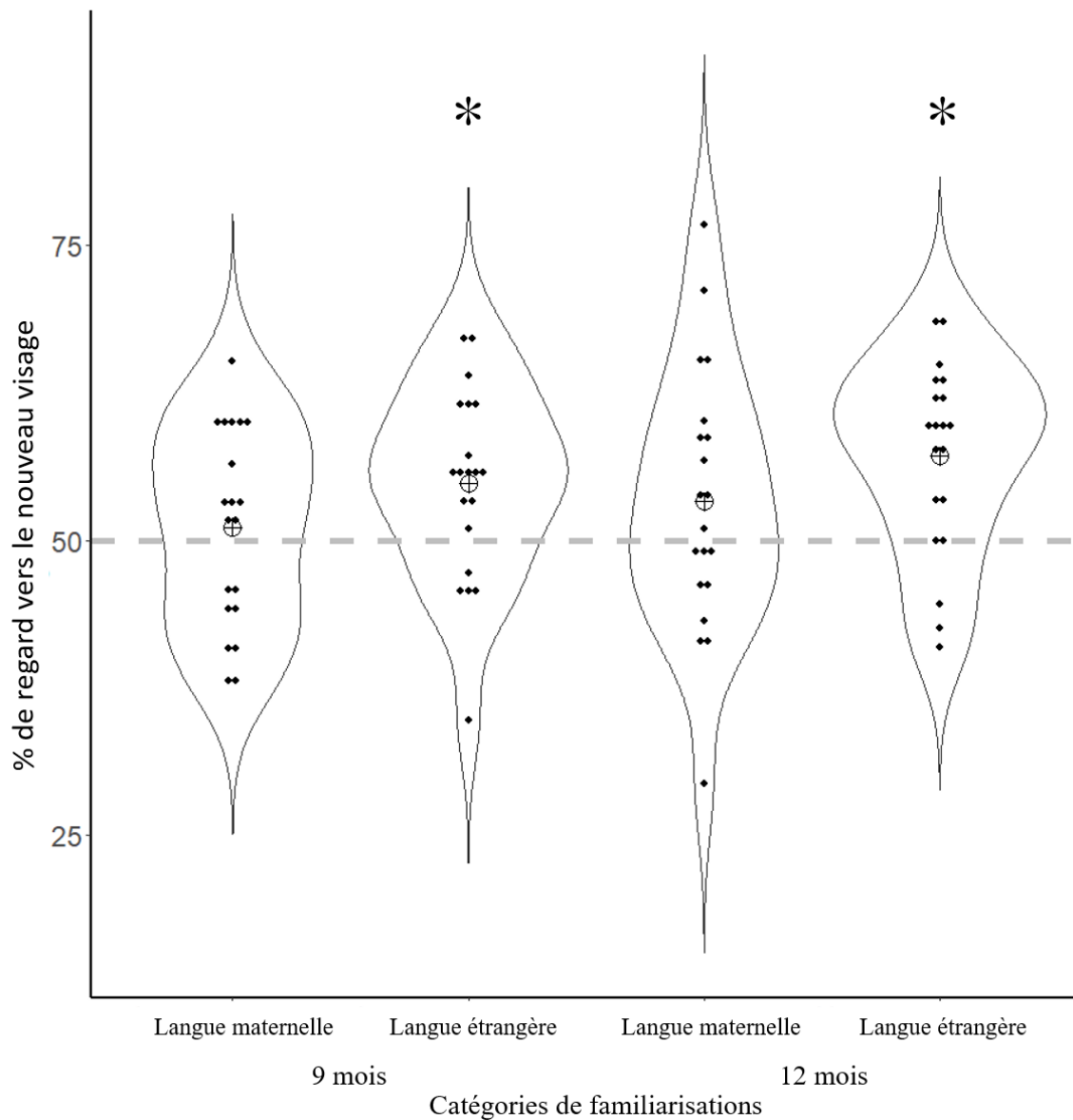


Figure 43. Score de préférence à la nouveauté en fonction de la familiarité de la langue et de l'âge.

Les croix cerclées représentent la Moyenne pour chaque groupe d'âge. Le trait en pointillé représente le niveau de hasard (50 %). Les scores qui diffèrent significativement du niveau de hasard sont marqués d'un astérisque. Le nombre de participant dans chaque condition est de 20.

IV. Discussion

Cette étude visait à évaluer si la familiarité d'un langage pouvait impacter la reconnaissance de visage d'autres types par des nourrissons de 9 et 12 mois. Après une familiarisation de 30 secondes, la reconnaissance pour les visages d'autres types a été restaurée uniquement lorsque les visages d'un autre type furent associés à une langue étrangère mais pas lorsqu'ils furent associés à la langue maternelle des nourrissons.

Il est établi dans la littérature que les visages d'autres types présentés sans langage ne sont pas correctement reconnus par les nourrissons après le rétrécissement perceptif (Kelly et al., 2007a, 2009). Dans une étude précédente, Hillairet de Boisferon et al. (2021) ont trouvé que les visages de même type que le nourrisson ne sont plus reconnus lorsqu'ils sont associés à une langue étrangère au nourrisson, mais ils sont toujours reconnus lorsque les visages de même type sont associés à la langue maternelle du nourrisson. Les résultats de la présente étude sont en contradiction avec le LFE puisque la langue maternelle n'améliore ici pas la reconnaissance des visages d'autres types.

IV.1. La langue et le visage en tant que catégoriseurs sociaux

Le langage fait plus que transmettre des informations référentielles. En effet, c'est aussi un moyen de tirer des informations concernant notre interlocuteur, ce qui façonnent notre attitude vis-à-vis de celui-ci (Cargile et al., 1994 ; Giles & Billings, 2004). L'exposition au langage étant plus précoce que l'exposition aux visages (le système auditif du fœtus devient fonctionnel vers 25 semaines tandis que l'exposition aux visages commence à la naissance), Kinzler & Spelke (2011) ont avancé l'idée que le langage joue un rôle plus crucial pour la formation de catégorie social que les visages. Les nourrissons pourraient donc ainsi déterminer en fonction de la langue que parle un individu si celui-ci appartient à leur groupe social ou non. Bernstein et al. (2007) a montré que pour les adultes, présenter quelqu'un comme un membre extra- ou intra-groupe (par exemple en le présentant comme appartenant à la même université ou non que le participant) est suffisant pour impacter la reconnaissance de cet individu par le participant lors d'un test de reconnaissance. Les individus présentés comme individus intra-groupe étaient mieux reconnus que les individus présentés comme extra-groupe. Sachant que la reconnaissance des individus intra-groupe est meilleure que la reconnaissance des individus extra-groupe, on peut s'attendre à ce que les individus parlant la même langue que nous soient identifiés comme intra-groupe et soient donc mieux identifiés que les individus parlant une autre langue. C'est en effet ce que suggèrent Baus et al. (2017) pour les adultes. Cependant,

nous ne retrouvons pas cet effet dans notre étude. Premièrement, bien que nous ayons testé des nourrissons après le rétrécissement perceptif, de nombreux changements développementaux concernant le traitement des visages et du langage existent entre la première année de vie et l'âge adulte. Deuxièmement, il n'y avait pas de différence perceptive entre les individus à reconnaître et les participants ; dans les deux cas, il s'agissait d'étudiants blancs. Dans notre étude, les individus à reconnaître sont des femmes adultes de type asiatique tandis que les participants sont des nourrissons de type européen. Cela pourrait expliquer pourquoi nous ne retrouvons pas un effet facilitateur de la langue maternelle, puisque le visage présenté n'appartient pas au groupe social du participant. Il existe donc une incongruence entre l'apparence visuelle et la langue que parle les personnes à reconnaître, qui pourrait empêcher les nourrissons de catégoriser ces individus comme intra-groupe.

IV.2. Redondance inter-sensoriel et incongruence

Il semble donc que l'interaction entre le type de visage et le type de langue joue un rôle plus important que le type de visage ou le type de langue seul. Nos résultats ainsi que ceux de Hillairet de Boisferon et al. (2021) suggèrent que lorsqu'un nourrisson voit un visage de type familial, il s'attend à entendre sa langue maternelle et que lorsqu'un nourrisson entend sa langue maternelle, il s'attend à ce qu'elle soit produite par un visage de type familial. Si l'association visage/langue va à l'encontre des attentes du nourrisson, l'apprentissage du visage ne se fait pas correctement comme c'est le cas lorsqu'un visage de leur type est associé à une langue étrangère (Hillairet de Boisferon et al., 2021) ou lorsqu'un visage d'autre type est associé à leur langue maternelle (la présente étude). Le fait que les visages d'autre-types soient reconnus par les nourrissons lorsqu'associés avec une langue étrangère pourrait résulter de l'association multimodale non-incongruente (plutôt que congruente) entre les signaux audios et visuels. Dans la vie de tous les jours, la communication est presque toujours multimodale, reposant principalement sur des indices audios et visuels, produit par le langage et le visage (Chen & Rao, 1998 ; Rosenblum, 2008). Bien que chaque modalité puisse fournir des informations sur l'identité, l'état émotionnel, ou encore l'état attentionnel de nos partenaires sociaux (Kamachi et al., 2003), la redondance inter-sensoriel est supposée améliorer la salience perceptives des stimulus, permettant de ce fait une meilleure prise d'informations et donc un apprentissage facilité (« The inter-sensory redundancy hypothesis », voir Bahrnick et al., 2012 ; Bahrnick & Lickliter, 2000 ; Shams & Seitz, 2008). La redondance perceptive a été mise en évidence dans plusieurs modalités sensorielles. Par exemple, Seigneuric et al. (2010) ont présenté à des participants adultes une photographie sur laquelle apparaissait plusieurs objets, connus pour

avoir des odeurs caractéristiques (par exemple une orange et de la lavande). Les auteurs ont trouvé que lorsqu'il était diffusé une odeur congruente avec un des objets (par exemple une odeur d'orange), l'objet était plus rapidement regardé. De plus, l'objet était légèrement moins longtemps regardé que lorsque l'odeur n'était pas congruente, ce qui pourrait indiquer qu'il était plus rapidement identifié. A noter aussi que l'odeur était présente à un niveau très faible (seul 10 participants sur 48 ont reporté avoir senti une odeur), la congruence n'avait donc pas besoin d'être consciemment perçue pour guider l'exploration visuelle. Ainsi, ajouter du langage à un visage pourrait aider les nourrissons à l'identifier, à condition que l'association bimodale ne soit pas en désaccord avec leurs connaissances et leurs attentes. En accord avec cette possibilité, Minar & Lewkowicz (2017) ont montré qu'ajouter des sons (des voyelles) de manière dynamique avec des visages d'autre type que ceux des nourrissons permettait de restaurer la reconnaissance de ces visages par des nourrissons âgés d'entre 10 et 12 mois. Dans une récente étude, Xiao et al. (2018) ont montré que la congruence entre des stimuli audios et visuels améliore l'attention visuelle chez des nourrissons de 9 mois. Dans leur étude, ils ont associé des visages de même type ou d'autre type à de la musique joyeuse ou triste, dans un protocole similaire à celui de l'étude Chapitre 7. Les auteurs ont trouvé que les nourrissons sont, à partir de 9 mois, plus susceptibles d'associer les visages de leur type avec une valence positive (la musique joyeuse) et les visages d'autres types avec une valence négative (la musique triste). Plus précisément, le temps de regard durant les associations congruentes était plus long que les temps de regard durant les associations incongruentes. Une possibilité est que les nourrissons associent ce qui est familier avec ce qui est familier (la musique joyeuse étant plus courante dans les programmes pour enfants que la musique triste) et ce qui est non-familier avec ce qui est non-familier. Dans notre étude, la redondance multisensorielle peut donc améliorer la salience des visages et aider les nourrissons à accéder à l'identité. En accord avec l'apprentissage statistique et les résultats précédents trouvés par Hillairet de Boisferon et al. (2021), il semble que l'expérience des nourrissons façonnent leurs attentes vis-à-vis des individus et des événements, et dans notre cas, vis-à-vis de quel type de visage parle ou ne parle pas leur langue maternelle (Lany & Gómez, 2008 ; Lew-Williams & Saffran, 2012).

IV.3. Reconsidérer le rétrécissement perceptif pour les visages

Le fait que des nourrissons de plus de 9 mois reconnaissent des visages d'autres types lorsqu'associés avec une langue étrangère soulève des interrogations vis-à-vis du rétrécissement perceptif. Les résultats de notre étude viennent soutenir l'idée que le rétrécissement perceptif n'est pas un état aussi définitif que l'on a pu penser par le passé. En

effet, des études précédentes ont déjà montré qu'un entraînement spécifique durant l'enfance (Heron-Delaney et al., 2011 ; Spangler et al., 2013) ou même à l'âge adulte (Goldstein & Chance, 1985 ; Lebrecht et al., 2009) peut atténuer le rétrécissement perceptif. De récents travaux réalisés en Malaisie, un pays avec une grande hétérogénéité de types de visages, montrent que grandir dans un tel environnement réduit voir même supprime l'ORE (Estudillo et al., 2019 ; Tham et al., 2017, 2018). Enfin, l'exposition à plusieurs langues pourrait aussi influencer le traitement des visages et réduire et impacter le rétrécissement perceptif. Via une étude de suivi de regard, Singh et al. (2019) ont montré que le type d'une personne n'influence pas la façon que les nourrissons bilingues âgés d'entre 18 et 20 mois ont de suivre leur regard, tandis que les nourrissons monolingues du même âge suivent plus le regard des personnes de même type qu'eux. Ceci suggère que l'expérience des nourrissons vis-à-vis des langues influence la manière dont les nourrissons traite les visages de même et d'autres types. Allant dans ce sens, d'autres études chez les adultes montrent que les adultes bilingues, contrairement aux adultes monolingues, présentent un ORE réduit (Kandel et al., 2016 ; Burns et al., 2019).

IV.4. Conclusions

Notre résultat principal est que les visages d'autres types peuvent toujours être reconnus par les nourrissons de 9 et 12 mois lorsque ces visages sont associés à une langue étrangère. Ce résultat nous questionne sur l'ORE durant la petite enfance, car il montre que la présentation multimodale d'un visage d'autre type peut aider les nourrissons à surmonter l'ORE, du moment que les attentes des nourrissons ne soient pas violées. De manière complémentaire, des études précédentes (Hillairet de Boisferon et al, 2021 ; Minar & Lewkowicz, 2017) ont montrées que n'importe quelle association multimodale considéré comme incongruente par les nourrissons impactent la reconnaissance des visages. Il est donc possible qu'à partir de 9 mois, la représentation que les nourrissons ont de leurs partenaires sociaux est multimodale, incluant le type de langue et le type de visage. L'étude de Hillairet de Boisferon et al. (2021) et la nôtre montrent que les nourrissons ont des attentes fortes et non négociables quant à la langue que doit parler un visage de même type qu'eux ; une conséquence de ces attentes est que si une incongruence existe entre le type de langue et le type de visage, l'apprentissage du visage ne se fait pas correctement.

Concernant les individus exo-groupes, les nourrissons n'ont probablement que de très faibles voir pas d'attentes quant à comment ceux-ci doivent parler et à quoi ils doivent ressembler. Quinn et al. (2015) montrent qu'il est possible pour les nourrissons de former des catégories pour des visages non familiers au cours d'une expérience. Dans cette étude, ils ont montré que

les nourrissons de 6 mois forment une catégorie par type de visage présentés (c'est-à-dire une catégorie visages asiatiques, une catégorie visages africains, une catégorie visages européens etc.) tandis qu'à 9 mois, ils ne forment plus qu'une catégorie contenant les visages de même type et une catégorie englobant tous les autres visages, indépendamment de leurs types. Il est donc possible, dans notre étude, que les nourrissons forment une pré-catégorie multimodale exo-groupe incluant les visages d'autres types et les langues étrangères.

Bien que probablement moins écologique que des visages dynamiques, nous avons choisis d'utiliser des photos de visages. Ce choix s'explique par l'étude de Hillairet de Boisferon et al. (2021). Dans leur étude, des visages dynamiques furent tout d'abord utilisés mais le résultat fut que les visages ne furent plus reconnus, même lorsque de même type et parlant la langue maternelle du nourrisson. Un visage qui parle amène probablement les nourrissons à focaliser leur attention sur la bouche plutôt que sur le reste du visage, les amenant ainsi à faire omission d'éléments permettant la reconnaissance du visage. Hillairet de Boisferon et al. (2021) ont donc ensuite utilisé des visages statiques accompagné d'une piste audio, comme dans la présente étude. Cette limitation étant prise en compte, l'association entre les visages statiques et l'audio aide (de quelque manière que ce soit) les nourrissons à encoder les visages au niveau individuel durant la phase de familiarisation, ce qui les aide à surmonter l'ORE durant la phase de test.

La présente étude fournit est une preuve de plus que des liens étroits existent entre le traitement des visages et le traitement du langage lors du développement. En effet, ajouter une langue étrangère à un visage d'autre type semble améliorer les capacités de reconnaissances des nourrissons de 9 et 12 mois. Les résultats de cette étude ainsi que celle de Hillairet de Boisferon et al. (2020) montre le rôle des associations voix/visage dans la reconnaissance des individus.

Chapitre 9 : Discussion générale

I. Récapitulatif de l'objectif de la thèse et des résultats principaux

La majorité des études portant sur le rétrécissement perceptif évalue son impact sur le traitement des visages ou sur le traitement du langage. L'objectif de cette thèse était de mieux comprendre les interactions entre ces deux systèmes cognitifs durant la première année de vie et durant le rétrécissement perceptif. Pour parvenir à cet objectif, nous avons réalisé quatre études portant sur des nourrissons de 3 à 12 mois. Ces études visent à évaluer l'impact du type de visage sur le traitement du langage ou inversement, l'impact du langage sur le traitement du visage. Dans leur globalité, ces études confirment une forte influence mutuelle des deux systèmes cognitifs durant le développement du nourrisson, allant dans le sens d'études récentes (Maurer & Werker, 2014 ; Pascalis et al., 2014, 2017 ; Krasotkina et al., 2018 ; Xiao et al., 2018). Nous mettons en évidence d'autres implications du rétrécissement perceptif, plus précoces que celles généralement mises en évidence (Werker & Tees, 1984 ; Tsushima et al., 1994 ; Pascalis et al., 2002 ; Kelly et al., 2007a, 2009 ; Krasotkina et al. 2018, 2021 ; Tsao et al., 2006). Nous montrons aussi que les conséquences du rétrécissement perceptif ne sont pas définitives, le système perceptif est encore très plastique, comme ont déjà pu le montrer Fair et al. (2012), Markant et al. (2016) ou encore Quinn et al. (2020). Ces études apportent de nouveaux éléments nous permettant de mieux décrire la mise en place du rétrécissement perceptif.

Dans l'étude 1 (Chapitre 5), nous avons mis en évidence un impact du type de visage sur le traitement audio-visuel du signal de parole. Les nourrissons de 3 mois font la correspondance entre un signal visuel de parole et un signal audio de parole uniquement lorsque le visage est de type familier. Cette étude nous amène à penser que des conséquences du rétrécissement perceptif sont observables dans le développement bien plus tôt que ce qui avait été généralement mis en évidence jusqu'à présent : si la reconnaissance de visage d'autres types ne semble impactée qu'à partir de 9 mois, le type de visage peut impacter d'autres capacités cognitives (comme ici la correspondance audiovisuelle phonémique) plus tôt. De plus, les résultats de notre étude sont en accord avec ceux de Kuhl & Meltzoff (1982, 1984, 1988), Patterson & Werker (1999, 2003) et Yeung & Werker (2013) et confirment que les nourrissons sont très tôt sensibles à la congruence intersensorielle, qui est ici une congruence audio-visuelle. Les résultats sur les nourrissons de 9 mois sont pour le moment incomplets et plus complexes à interpréter.

Pour notre deuxième étude (Chapitre 6), nous avons tenté d'évaluer l'impact du type de visage sur l'effet McGurk. Les résultats révèlent des différences de perception chez les nourrissons exposés à cette illusion audio-visuelle, en fonction du type de visage qui produit l'illusion. Les résultats révèlent aussi des différences en fonction de l'origine des nourrissons, impliquant une part d'influence culturelle dans la sensibilité à l'effet McGurk. Les interprétations que l'on peut tirer des résultats ne sont cependant pas claires. Une possibilité est que la sensibilité à l'effet McGurk émerge plus tard pour les visages d'autres types. De premier abord, ces résultats semblent difficilement conciliables avec les résultats de l'étude 1, qui montrent au contraire que le type de visage a un impact dès 3 mois sur les capacités de traitement audio-visuel du signal de parole. Nous reviendrons sur ce point surprenant dans la suite de cette discussion. Aussi, l'effet McGurk est une expérience très spécifique : il s'agit d'une illusion perceptive qui ne survient pas en conditions naturelles, et n'est peut-être pas très représentative du développement général des capacités audio-visuelles des nourrissons.

Pour ces deux études, il serait intéressant de savoir quelles parties du visage sont regardés par les nourrissons. Il serait ainsi possible de vérifier si les nourrissons faisant la correspondance audio-visuelle ou étant sensibles à l'effet McGurk sont les nourrissons qui regardent le plus la bouche. Si c'est le cas, l'évolution des *patterns* de regard au fil de l'expérience pourrait nous indiquer si une durée d'exposition plus longue permettrait aux nourrissons de faire la correspondance ou la fusion audio-visuelle avec les visages d'autres types. En effet, si les nourrissons passent progressivement plus de temps à regarder la bouche que le reste du visage, cela confirmerait le fait qu'ils effectuent la tâche audio-visuelle après une longue exploration du visage.

Dans l'étude 3 (Chapitre 7), nous avons voulu tester un nouveau protocole nous permettant d'évaluer les associations que les nourrissons forment entre type de langue et type de visage. Les résultats nous montrent, là aussi, une influence de l'expérience plus précoce que celle généralement mise en évidence. Dès 3 mois, les nourrissons sont plus susceptibles de regarder les associations n'allant pas à l'encontre de leurs attentes. En revanche, nous ne trouvons pas de trajectoire développementale claire après 3 mois.

Enfin, pour l'étude 4 (Chapitre 8), nous avons montré que le type de langue avec lequel un visage est associé impacte sa reconnaissance. En associant un visage d'autre type avec une langue étrangère nous restaurons la capacité des nourrissons de 9 et 12 mois à le reconnaître. Si dans l'étude 3, nous n'observions pas de différence en termes de temps de regard sur les

associations type de langue/type de visage chez les nourrissons de 9 et 12 mois, ces associations ont en revanche des conséquences sur les capacités cognitives de mémorisation et de reconnaissance des visages.

De manière intéressante, les associations « type de langue/type de visage » les plus regardées par les nourrissons de 3 mois dans l'étude 3 (associations que nous avons définies comme congruentes) sont les associations permettant la reconnaissance des visages par les nourrissons de 9 et 12 mois dans l'étude 4. Cependant, on ne trouve pas cette correspondance entre les deux études à 9 ou 12 mois : dans l'étude 3, les visages sont regardés pendant la même durée lors des associations congruentes ou incongruentes par les nourrissons de 9 et 12 mois. Il semble que les mêmes types d'associations se forment à 3, 9 et 12 mois ; cependant, leur conséquence comportementale est différente. A 3 mois ces associations provoquent plus d'attention de la part des nourrissons, tandis qu'après 9 mois elles conditionnent la reconnaissance des individus. Nous reviendrons sur ce changement comportemental entre 3 et 9 mois dans la suite de cette discussion.

II. Remise en contexte dans le cadre du modèle *face-space*

Nous pouvons interpréter nos résultats dans le contexte du modèle *face-space* de Valentine (1991).

Avant 6 mois, on observe expérimentalement que les nourrissons sont encore en mesure de discriminer et d'individualiser les visages d'un autre type (Kelly, et al., 2007 ; Kelly et al., 2009). Dans un premier temps, l'espace de distribution des visages est initialement vierge, dépourvu de catégories et de prototypes. Chaque expérience avec un visage sera encodée dans l'espace de distribution par le nouveau-né (voir Figure 44 – A). Un même visage sera donc encodé autant de fois qu'il sera vu. Comme pour le modèle d'apprentissage distributionnel des sons de langues de Schatz (2021), nous pouvons représenter les points d'encodage des différents visages comme formant un mélange de gaussiennes. Plus les nourrissons sont jeunes, et donc moins ils ont d'expérience avec les visages, plus l'espace de distribution est sensible aux entrées. Ainsi, chaque nouvelle entrée remodèle l'espace et réévalue la contribution de chaque dimension. Bien que la totalité (ou quasi-totalité) des visages rencontrés durant les premiers mois de vie soient de même type que celui des nourrissons, les nourrissons ne catégorisent pas encore les visages d'un autre type différemment des visages de même type. En revanche, les dimensions pertinentes pour différencier les visages de même type ne sont pas les mêmes que les dimensions pertinentes pour différencier les visages d'autres types. L'espace de distribution

des visages n'est donc pas « adapté » pour les visages d'un autre type, puisqu'il est composé uniquement (ou presque) de visage de même type. Les résultats de notre étude 1 montrent que les nourrissons de 3 mois ne sont pas en mesure de réussir la tâche de correspondance phonémique avec des visages d'un autre type. Nous avons cependant vu que l'évolution des temps de regard peut suggérer qu'une durée d'exposition plus longue permettrait aux nourrissons de réaliser cette tâche. Une hypothèse pourrait être que les nourrissons de 3 mois encodent le visage d'un autre type dans l'espace de distribution des visages, comme ils le feraient pour un visage de leur propre type. Cependant, comme l'espace de distribution et les dimensions de l'espace sont façonnés par les visages de même type, l'encodage d'un visage d'un autre type nécessite plus de temps. De la même manière, les résultats de notre étude sur l'effet McGurk (étude 2) peuvent montrer que les nourrissons ne font pas la fusion audio-visuelle lorsque le stimulus auditif est présenté avec un visage d'un autre type mais font la fusion audio-visuelle lorsque le son est présenté avec un visage de même type que le leur. Les nourrissons, pour qui les visages de même type qu'eux sont familiers, n'ont pas besoin de beaucoup de temps pour encoder le visage dans l'espace de distribution. A nouveau, ce n'est pas le cas pour les visages d'un autre type, qui nécessitent plus de temps d'encodage. Ce temps supplémentaire empêche peut-être ainsi la réalisation de la tâche audio-visuelle dans la durée limitée de l'étude. Cette interprétation reste à prendre avec beaucoup de précautions compte tenu des limites évoquées pour l'étude sur l'effet McGurk (étude 2).

Après 9 mois, les nourrissons ne sont plus capables de discriminer correctement les visages d'un autre type et ne font, de plus, pas de distinction entre les différents autres types (Quinn et al., 2015). L'expérience est, à cet âge, suffisante pour former un nuage de points central dans l'espace de distribution des visages (voir Figure 44 – B). Le modèle de Valentine suppose que ce nuage de points central permet de créer une représentation d'un visage prototypique. Dans ce cadre, nous pouvons expliquer comment les nourrissons de plus de 9 mois vont catégoriser les visages en fonction du rattachement ou non à ce prototype : les visages de même type sont encodés dans le nuage central et sont donc rattachés au prototype. En revanche, les visages d'un autre type sont placés en périphérie de celui-ci. Même si une expérience ponctuelle avec des visages d'autres types existe, elle est dans la plupart des cas trop minime pour former des nuages de points périphériques. Par conséquent, les visages sont catégorisés de manière binaire : ils appartiennent au nuage de points central ou non, sans distinction plus précise pour ceux qui n'y appartiennent pas. Les résultats de notre étude sur l'effet McGurk (étude 2) peuvent sembler montrer que les nourrissons font, à partir de 9 mois, la fusion audio-visuelle même lorsque le

son est présenté avec un visage d'un autre type. L'espace de distribution des visages étant plus stable à cet âge, celui-ci s'adapte peu à une nouvelle entrée et l'encodage est plus rapide. De plus, les nourrissons de plus de 9 mois semblent commencer à se servir de l'expérience accumulée pour catégoriser les visages en fonction de leur appartenance ou non au groupe social (Quinn et al., 2015). Ce basculement (implicite) de stratégie permet peut-être aux nourrissons de traiter plus rapidement le visage d'autre type. Ainsi, la tâche audio-visuelle peut être réalisée dans la durée impartie. De même, cette interprétation est à prendre avec beaucoup de précautions.

Les enfants et les adultes sont capables de catégoriser les individus d'un autre type selon leur type, distinction qu'ils ne faisaient pas à 9 mois ; en revanche, ils présentent toujours de moins bonnes performances pour individualiser les visages d'un autre type (Levin, 1996 ; Dunham et al., 2015 ; Setoh et al., 2019). Au cours de l'enfance, l'expérience avec des visages d'autres types peut devenir suffisante pour former des nuages de points périphériques (voir Figure 44 – C). Par conséquent, la catégorisation peut être possible pour les individus ne se rattachant pas au nuage de points central mais se rattachant à un nuage de points situé en périphérie. Cependant, une même distance entre deux points est plus difficile à distinguer en périphérie qu'au centre de l'espace de distribution, la reconnaissance des individus n'appartenant pas au nuage de points central (i.e. un individu d'un autre type) reste plus difficile. Une question peut se poser pour la catégorisation d'individus métis. Peery & Bodenhausen (2008) et Lewis (2016) ont étudié cette question et les résultats montrent que les individus métis (avec un parent du même type que le participant et un parent d'un autre type) sont catégorisés comme appartenant à l'autre type que celui du participant. Nous considérons que ces résultats sont conformes à la géométrie de l'espace de distribution des visages. Toujours dans le cadre théorique de Valentine (1991), nous pouvons faire l'hypothèse que dans l'espace de distribution des visages, un individu métis ne se situerait ni dans le nuage de points constituant les visages de même type ni dans le nuage de points constituant les visages de l'autre type. Cet individu se situerait dans l'espace entre ces deux nuages de points, à équidistance du périmètre du nuage de points central et du nuage de point en périphérie (voir Figure 45). Le nuage de points central, composé des individus de même type, est plus large que les nuages de points périphériques, composés d'individus d'autres types, puisque l'exposition aux visages de même type est beaucoup plus fréquente que l'exposition aux visages d'un autre type. Ainsi, ce visage métis se trouve plus proche du centre du nuage de points le plus petit, soit le nuage de points composés des visages d'un autre type. Le visage métis est donc rattaché à ce nuage de point et est catégorisé comme

appartenant à un autre type. L'évolution du traitement des visages par les nourrissons peut donc être formalisée dans le cadre du modèle théorique de Valentine (1991).

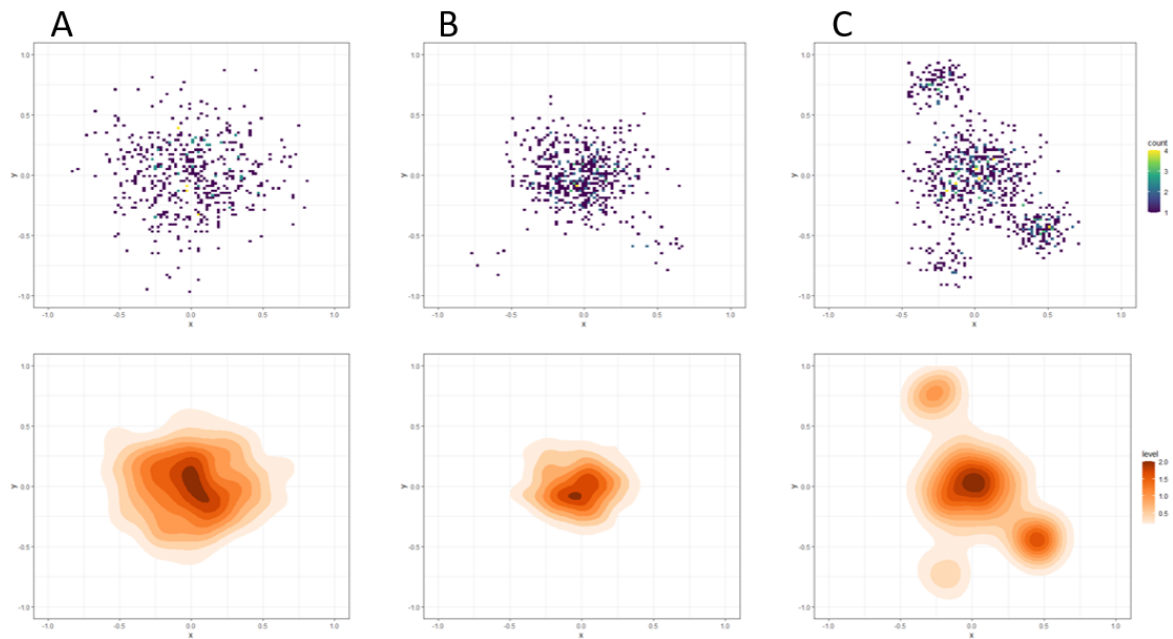


Figure 44. Représentation arbitraire du développement de l'espace de distribution des visages, inspirée des propositions théoriques de Valentine (1991).

L'espace est représenté en deux dimensions. En haut : représentation graphique où chaque point représente un nombre d'occurrence. En bas : représentation graphique en densité. A : avant 6 mois, les dimensions de l'espace sont encore instables ; les visages d'un autre type sont éloignés du centre et donc moins bien reconnus mais ils ne sont pas catégorisés différemment. B : vers 9 mois, l'expérience est suffisante pour former un nuage de points central ; les visages d'un autre type sont catégorisés différemment des visages de même type. C : progressivement durant l'enfance, l'expérience devient suffisante pour former des nuages de points en périphérie ; les visages d'autres types sont catégorisés selon leur type.

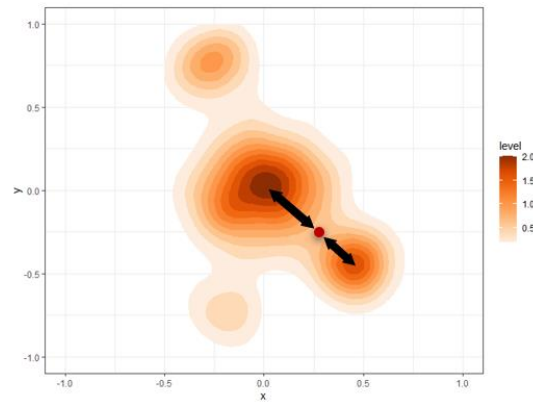


Figure 45. Proposition d’extension du modèle explicatif aux visages métis. Les individus métis, situés à mi-distance entre la périphérie du nuage de points central et d’un nuage de points périphérique, sont plus proches du centre du nuage périphérique (soit du prototype de visage d’un autre type) ; ils sont catégorisés comme appartenant au groupe de visages d’un autre type.

III. Reconsidérer le rétrécissement perceptif

III.1. Des marqueurs plus précoces

Nous pouvons nous questionner sur la définition du rétrécissement perceptif. Dans la littérature scientifique, le terme de rétrécissement perceptif a généralement été associé au fait de ne plus pouvoir correctement distinguer des stimuli non-natifs entre eux. Cette conséquence est observable vers 9 mois, que ce soit pour les visages ou les sons de langues (Kelly et al., 2007a, 2009 ; Werker & Tees, 1984 ; Tsushima et al., 1994 ; Tsao et al., 2006). Cependant, la difficulté à reconnaître des stimuli non-natifs n’est qu’une des manifestations du rétrécissement perceptif. D’autres marqueurs du rétrécissement perceptif sont observables plus précocement. L’étude 1 de cette thèse montre que les capacités de traitement audio-visuel des nourrissons de 3 mois sont déjà affectées par le type de visage de la personne produisant le signal de parole. L’étude 3 indique que les nourrissons de 3 mois sont déjà capables de former des associations entre type de langue et type de visage, fondées sur leur expérience. Ces résultats vont donc dans le sens de plusieurs études mettant en avant des différences de traitement entre les visages de même type et les visages d’un autre type bien avant 9 mois (Kelly, et al. ; 2005, 2007b ; Bar-Haim et al., 2006). Le rétrécissement perceptif étant une conséquence du manque d’exposition à des stimuli non-natifs, c’est un processus graduel qui commencent dès la naissance (voir *in utero* pour le langage). Il semble probable que d’autres marqueurs du rétrécissement perceptif que ceux que nous avons mentionnés soient présents avant 9 mois.

III.2. Des conséquences différentes avec des stimuli multisensoriels

Les études 1 et 2 indiquent que la trajectoire développementale des capacités audio-visuelles n'est pas la même en fonction du type de visage de la personne produisant le signal de parole. Aussi, l'étude 4 et l'étude de Hillairet de Boisferon et al. (2021) nous indiquent que lorsque les stimuli sont bimodaux, les résultats diffèrent de ce qui peut être trouvé lorsque les stimuli sont unimodaux (Kelly et al., 2007, 2009). Sur la base d'indices visuels uniquement, nous avons vu que les nourrissons de 9 mois n'arrivent pas à distinguer deux individus d'un autre type que le leur mais arrivent à les distinguer lorsqu'ils sont de leur propre type. De même, Johnson et al. (2011) ont montré que des nourrissons de 7 mois n'arrivent pas à distinguer deux voix lorsque celles-ci parlent une langue étrangère, tandis qu'ils y arrivent lorsqu'elles parlent leur langue maternelle. Les rencontres avec des individus de même type parlant la langue maternelle sont des événements tellement familiers dans la vie du nourrisson que même des « demi-stimuli », amputés d'une modalité sensorielle (c'est-à-dire un visage de type familier ou une voix parlant la langue familière), sont suffisants pour permettre l'identification. Ce n'est pas le cas pour les stimuli non-natifs, qui ne semblent pas pouvoir être reconnus, sauf à condition de bénéficier de l'intégralité de l'information. L'impact du rétrécissement communément observé par les études scientifiques doit être remis dans le contexte de la vie quotidienne du nourrisson. De manière générale, les stimuli que nous rencontrons sont multimodaux (Lewkowicz & Ghazanfar, 2009). Des auteurs ont avancé que cette multimodalité aide la prise d'informations et donc l'apprentissage (Bahrick & Lickliter, 2000, 2012 ; Shams & Seitz, 2008). Dans la vie courante, les nourrissons ne sont que rarement exposés à un visage statique silencieux ou à un signal de parole sans visage mais sont très couramment exposés à des visages qui parlent. Le rétrécissement perceptif est généralement mis en évidence sur des stimuli expérimentaux qui ne présentent pas autant d'informations que les stimuli rencontrés dans le quotidien des nourrissons. La reconnaissance d'individus est peut-être, en conditions expérimentales, plus difficile qu'elle ne l'est en conditions réelles, du fait de cette réduction de l'information. Nos études soulignent l'intérêt de considérer des stimuli multisensoriels pour qui souhaite évaluer l'impact du rétrécissement perceptif. Toutefois, il faut garder en tête une limite concernant l'étude 4. Les stimuli que nous présentons dans l'étude 4 sont certes bimodaux et permettent d'avoir accès à plus d'informations que des stimuli unimodaux, mais il ne s'agit pas de stimuli très écologiques non plus.

Un élément important pour que la multimodalité des stimuli facilite l'apprentissage est la congruence des signaux sensoriels (Press et al., 2004 ; Lippert et al., 2007). Autrement dit, il

faut que l'information perçue dans une modalité soit en accord avec l'information perçue dans une autre modalité. Par exemple, dans l'étude 1, les nourrissons de 3 mois regardent plus le visage correspondant au son joué que le visage ne lui correspondant pas (lorsque les visages sont du même type qu'eux). Les nourrissons sont, dès la naissance, sensibles à la redondance inter-sensorielle et ils sont rapidement capables d'en tirer profit (Izard et al., 2009 ; Guellaï et al., 2016 ; Leleu et al. 2019 ; voir Guellaï et al., 2019 pour une revue). Nous avons parlé de l'apprentissage statistique en introduction. Nous avons vu comment ce mécanisme permet aux nourrissons de former des attentes quant aux événements futurs sur la base des événements précédemment rencontrés. Par la suite, le traitement des informations ne se fera pas de la même manière selon que les événements sont conformes ou non aux attentes formées par l'individu. Si la redondance inter-sensorielle est conforme aux attentes des nourrissons, alors la prise d'informations sera facilitée. Les résultats récents de l'étude de Hillairet de Boisferon et al. (2021) vont dans ce sens. En effet, les résultats indiquent que la reconnaissance d'individus est facilitée lorsque leur présentation est conforme aux attentes des nourrissons. Les résultats de notre étude sur la reconnaissance de visage d'un autre type en association avec du langage (étude 4) semblent apporter une précision à cette idée. Dans notre étude, les nourrissons de 9 et 12 mois reconnaissent des individus d'un autre type parlant une langue étrangère. Les nourrissons de cet âge ne possèdent pas d'attente sur le type de langue que les individus d'un type inconnu doivent parler car ils n'en ont jamais vu. De même, les nourrissons de cet âge ne possèdent pas d'attente sur le type de visage dont doit provenir une langue inconnue car ils ne l'ont jamais entendue. En revanche, les nourrissons ont des attentes très fortes concernant leur langue maternelle : l'intégralité de leur expérience vis-à-vis de celle-ci (ou presque) provient de visage de leur propre type. De la même manière, ils ont des attentes très fortes concernant leur type de visage. S'il est probablement juste de dire que l'apprentissage est facilité lorsque les attentes des nourrissons sont respectées, il semble aussi important de dire que l'apprentissage est perturbé lorsqu'il y a un non-respect des attentes. Avec des stimuli plus écologiques, nous pouvons imaginer le principe suivant : tous les individus qui parlent sont en fait reconnaissables par les nourrissons. La seule condition pour cette reconnaissance est que ces événements audio-visuels n'aillent pas à l'encontre des attentes des nourrissons. Nous avons testé la reconnaissance de visages en fonction de la langue qui leur est associée. Un moyen de confirmer cette hypothèse serait de tester la reconnaissance de voix en fonction du type de visage qui leur est associé. Selon cette hypothèse, les nourrissons devraient pouvoir identifier une voix parlant leur langue maternelle uniquement lorsqu'elle est apprise avec un visage de

même type que le leur ; à l'inverse, ils devraient pouvoir reconnaître une voix parlant une langue étrangère uniquement lorsqu'elle est apprise avec un visage d'un autre type que le leur.

IV. Le cas des milieux hétérogènes

Une question que nous avons déjà abordée en introduction vient naturellement à la lecture de ces résultats. Qu'en est-il des nourrissons bilingues ou bien des nourrissons grandissant entourés de plusieurs types de visages ? Nous avons vu que c'est sur très peu d'exemplaires (mais présentant un nombre d'occurrences répétées) que se construit la représentation du partenaire social « type ». Cette faculté semble d'autant plus impressionnante lorsque les quelques exemplaires d'apprentissage ne sont pas réguliers ou concordant entre eux, c'est-à-dire pour les nourrissons grandissant dans des environnements hétérogènes en termes de types de visages ou de langues. Il est donc d'autant plus intéressant de se demander comment s'établit la formation du partenaire social « type » chez ces nourrissons. Cependant, il est très compliqué d'établir une trajectoire développementale pour les nourrissons qui ne sont pas monolingues et/ou qui sont exposés à plusieurs types de visages. Le monolinguisme est relativement comparable quels que soit les nourrissons considérés. Globalement, les personnes s'occupant de nourrissons monolingues parlent tous une même et unique langue. L'exposition à une (ou plusieurs) langue étrangère peut exister de façon très ponctuelle, mais ces langues étrangères ne sont pas le moyen de communication utilisé dans l'environnement du nourrisson. De même, l'homogénéité du type de visage dans l'environnement du nourrisson est assez facile à évaluer. Les personnes s'occupant du nourrisson sont toutes de même type, soit du type dominant dans l'environnement. En revanche, lorsque l'exposition au langage ne provient pas d'une langue unique ou bien lorsque plusieurs types de visages sont présents dans l'environnement, on peut presque différencier autant de situations que de nourrissons. On peut recenser plusieurs facteurs de variabilité pour ces nourrissons. Premièrement, le nombre de langues ou le nombre de type de visages présents dans l'environnement. Grandir dans un environnement bilingue n'est peut-être pas équivalent à grandir dans un environnement trilingue quant à la mise en place du partenaire social « type ». Les nourrissons trilingues présentent peut-être une flexibilité plus grande que les nourrissons bilingues sur le type de langue qu'un individu peut parler et pourraient garder une plus grande flexibilité quant à qui peut appartenir à son groupe social. De manière similaire, le développement du traitement des visages n'est potentiellement pas identique selon le nombre de types de visages présents dans l'environnement. Potentiellement, plus il y a de types de visages différents présents dans l'environnement du nourrisson, moins le type d'une personne prend d'importance pour définir l'appartenance au groupe social.

Deuxièmement, la nature des langues ou des types de visages peut être un facteur influant sur le développement. Un nourrisson exposé à des visages de type européen et de type asiatique ne se servira pas nécessairement des mêmes indices visuels pour discriminer les visages qu'un nourrisson entouré de visages de type européen et de type africain. A titre d'exemple, Tham et al. (2017) montrent que les enfants malaisiens grandissant dans un environnement constitué de visages de type malais et de type chinois ne présentent pas l'ORE pour les visages de type européen (auquel ils ne sont pourtant pas exposés fréquemment) mais le présentent pour les visages de type africain. On peut imaginer un effet similaire pour les langues : les nourrissons pourraient catégoriser différemment les individus parlant une langue étrangère en fonction de la proximité de celle-ci avec leurs propres langues. Troisièmement, le ratio des différentes langues ou des différents types de visages peut aussi influencer le développement. Une exposition dans une proportion 50%/50% n'est peut-être pas équivalente à une exposition dans une proportion 25%/75% sur l'émergence des catégories phonémiques ou du visage prototypique. Quatrièmement, les personnes par lesquelles les nourrissons sont exposés à ces langues ou ces types de visages peuvent aussi influencer le développement. Ce quatrième point varie souvent conjointement avec le troisième point. Pour reprendre l'exemple des enfants malaisiens, Tham et al. (2017) montrent de plus que les enfants de leur étude ne présentent pas l'ORE pour les visages féminins mais le présentent pour les visages masculins. Les enfants ont probablement vu plus de visages féminins différents (les visages féminins étant généralement plus présents dans l'environnement des enfants que les visages masculins) permettant des attentes moins strictes pour les visages féminins que pour les visages masculins. Enfin, un point spécifique aux langues est la flexibilité langagière des individus dans l'environnement de l'enfant. Les individus parlant plusieurs langues peuvent parler très majoritairement une langue ou bien passer d'une langue à l'autre indifféremment. Dans le deuxième cas, les nourrissons auront peut-être des attentes moins rigides concernant la langue que doit parler un individu.

V. Conclusion

Les travaux de cette thèse cherchent à améliorer notre compréhension des interactions entre le traitement des visages et du langage durant le développement. A travers quatre études, nous apportons des éléments supplémentaires montrant que ces deux systèmes cognitifs sont étroitement liés durant le développement, et ce, dès 3 mois. Nous montrons que les conséquences développementales du rétrécissement perceptif observé pour le traitement des visages impactent le traitement du signal de parole très tôt dans le développement. Nous avons

aussi mis en évidence que les conséquences du rétrécissement perceptif sur les visages associés à un signal de parole ne sont pas équivalentes à celle observées sur un visage seul.

Dans la continuation de ce cheminement, il semble important de concevoir des études avec des stimuli ressemblant plus à ceux rencontrés dans l'environnement réel du nourrisson. Il apparaît aussi intéressant d'étudier l'impact du rétrécissement perceptif sur d'autres marqueurs que la reconnaissance d'individus ou de phonèmes. Une autre perspective envisageable dans la continuité de cette thèse, serait de suivre les interactions entre traitement du langage et traitement des visages chez les jeunes enfants. Enfin, l'étude des populations bilingues (ou multilingues) et exposées à différents types de visages apporterait de nombreuses précisions sur nos questionnements même si elle reste un défi sur le plan de la mise en œuvre expérimentale.

REFERENCES

- Adams, R. J., Courage, M. L., & Mercer, M. E. (1994). Systematic measurement of human neonatal color vision. *Vision Research*, 34(13), 1691-1701. [https://doi.org/10.1016/0042-6989\(94\)90127-9](https://doi.org/10.1016/0042-6989(94)90127-9)
- Adams, R. J., & Courage, M. L. (1998). Human newborn color vision: Measurement with chromatic stimuli varying in excitation purity. *Journal of Experimental Child Psychology*, 68(1), 22-34. <https://doi.org/10.1006/jecp.1997.2407>
- Adams, R. J., & Courage, M. L. (2002). Using a single test to measure human contrast sensitivity from early childhood to maturity. *Vision research*, 42(9), 1205-1210. [https://doi.org/10.1016/S0042-6989\(02\)00038-X](https://doi.org/10.1016/S0042-6989(02)00038-X)
- Adolph, K. E., Kretch, K. S., & LoBue, V. (2014). Fear of heights in infants? *Current Directions in Psychological Science*, 23, 60–66. <https://doi.org/10.1177/0963721413498895>
- Aharon, I., Etcoff, N., Ariely, D., Chabris, C. F., O'Connor, E., & Breiter, H. C. (2001). Beautiful faces have variable reward value: fMRI and behavioral evidence. *Neuron*, 32(3), 537-551. [https://doi.org/10.1016/S0896-6273\(01\)00491-3](https://doi.org/10.1016/S0896-6273(01)00491-3)
- Albareda-Castellot, B., Pons, F., & Sebastián-Gallés, N. (2011). The acquisition of phonetic categories in bilingual infants: New data from an anticipatory eye movement paradigm. *Developmental science*, 14(2), 395-401. <https://doi.org/10.1111/j.1467-7687.2010.00989.x>
- Aldridge, M. A., Braga, E. S., Walton, G. E., & Bower, T. G. R. (1999). The intermodal representation of speech in newborns. *Developmental Science*, 2(1), 42-46. <https://doi.org/10.1111/1467-7687.00052>
- Alm, M., Behne, D. M., Wang, Y., & Eg, R. (2009). Audio-visual identification of place of articulation and voicing in white and babble noise. *The Journal of the Acoustical Society of America*, 126(1), 377-387. <https://doi.org/10.1121/1.3129508>
- Alsius, A., Paré, M., & Munhall, K. G. (2018). Forty years after hearing lips and seeing voices: the McGurk effect revisited. *Multisensory Research*, 31(1-2), 111-144. <https://doi.org/10.1163/22134808-00002565>
- Altwater-Mackensen, N., & Grossmann, T. (2014). Learning to match auditory and visual speech cues: Social influences on acquisition of phonological categories. *Child Development*, 86(2), 362–378. <https://doi.org/10.1111/cdev.12320>
- Altwater-Mackensen, N., Mani, N., & Grossmann, T. (2015). Audiovisual speech perception in infancy: The influence of vowel identity and infants' productive abilities on sensitivity to (mis) matches between auditory and visual speech cues. *Developmental Psychology*, 52(2), 191. <https://doi.org/10.1037/a0039964>
- Altwater-Mackensen, N., Jessen, S., & Grossmann, T. (2017). Brain responses reveal that infants' face discrimination is guided by statistical learning from distributional information. *Developmental Science*, 20(2). <https://doi.org/10.1111/desc.12393>

- Anzures, G., Wheeler, A., Quinn, P. C., Pascalis, O., Slater, A. M., Heron-Delaney, M., Tanaka, J. W., & Lee, K. (2012). Brief daily exposures to Asian females reverses perceptual narrowing for Asian faces in Caucasian infants. *Journal of Experimental Child Psychology*, *112*(4), 484-495. <https://doi.org/10.1016/j.jecp.2012.04.005>
- Argyle, M. & Cook, M. (1994). Gaze and mutual gaze. *The British Journal of Psychiatry*, *165*(6), 848-850. <https://doi.org/10.1017/S0007125000073980>
- Arnold, P., & Hill, F. (2001). Bisensory augmentation: A speechreading advantage when speech is clearly audible and intact. *British Journal of Psychology*, *92*(2), 339-355. <https://doi.org/10.1348/000712601162220>
- Ashworth III, A. R., Vuong, Q. C., Rossion, B., & Tarr, M. J. (2008). Recognizing rotated faces and Greebles: What properties drive the face inversion effect? *Visual Cognition*, *16*(6), 754-784. <https://doi.org/10.1080/13506280701381741>
- Aslin, R. N., & Salapatek, P. (1975). Saccadic localization of visual targets by the very young human infant. *Perception & Psychophysics*, *17*(3), 293-302. <https://doi.org/10.3758/BF03203214>
- Badin, P., Tarabalka, Y., Elisei, F., & Bailly, G. (2010). Can you 'read' tongue movements? Evaluation of the contribution of tongue display to speech understanding. *Speech Communication*, *52*(6), 493-503. <https://doi.org/10.1016/j.specom.2010.03.002>
- Bahrack, L. E., & Lickliter, R. (2000). Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Developmental psychology*, *36*(2), 190. <https://doi.org/10.1037/0012-1649.36.2.190>
- Bahrack, L. E., & Lickliter, R. (2012). The role of intersensory redundancy in early perceptual, cognitive, and social development. *Multisensory development*, 183-206. <https://doi.org/10.1093/acprof:oso/9780199586059.003.0008>
- Baier, R., Idsardi, W. J., & Lidz, J. (2007). Two-month-olds are sensitive to lip rounding in dynamic and static speech events. In *AVSP*.
- Banks, M. S., & Salapatek, P. (1978). Acuity and contrast sensitivity in 1-, 2-, and 3-month-old human infants. *Investigative Ophthalmology & Visual Science*, *17*(4), 361-365.
- Banks, M. S., & Shannon, E. (1993). Spatial and Chromatic Visual Efficiency in Human Neonates. *Visual perception and cognition in infancy*, 23.
- Banissy, M. J., Stewart, L., Muggleton, N. G., Griffiths, T. D., Walsh, V. Y., Ward, J., & Kanai, R. (2012). Grapheme-color and tone-color synesthesia is associated with structural brain changes in visual regions implicated in color, form, and motion. *Cognitive Neuroscience*, *3*(1), 29-35. <https://doi.org/10.1080/17588928.2011.594499>
- Bardi, L., Regolin, L., & Simion, F. (2011). Biological motion preference in humans at birth: Role of dynamic and configural properties. *Developmental science*, *14*(2), 353-359. <https://doi.org/10.1111/j.1467-7687.2010.00985.x>

- Bar-Haim, Y., Ziv, T., Lamy, D., & Hodes, R. M. (2006). Nature and nurture in own-race face processing. *Psychological Science*, *17*, 159–163. <https://doi.org/10.1111/j.1467-9280.2006.01679.x>
- Baron, A. S., & Banaji, M. R. (2006). The development of implicit attitudes: Evidence of race evaluations from ages 6 and 10 and adulthood. *Psychological science*, *17*(1), 53-58. <https://doi.org/10.1111/j.1467-9280.2005.01664.x>
- Baron-Cohen, S., Campbell, R., Karmiloff-Smith, A., Grant, J., & Walker, J. (1995). Are children with autism blind to the mentalistic significance of the eyes?. *British Journal of Developmental Psychology*, *13*(4), 379-398. <https://doi.org/10.1111/j.2044-835X.1995.tb00687.x>
- Bartal, I. B. A., Rodgers, D. A., Sarria, M. S. B., Decety, J., & Mason, P. (2014). Pro-social behavior in rats is modulated by social experience. *Elife*, *3*. <https://doi.org/10.7554/eLife.01385.001>
- Bartlett, J. C., & Searcy, J. (1993). Inversion and configuration of faces. *Cognitive psychology*, *25*(3), 281-316. <https://doi.org/10.1006/cogp.1993.1007>
- Bartlett, J. C., Hurry, S., & Thorley, W. (1984). Typicality and familiarity of faces. *Memory & Cognition*, *12*(3), 219-228. <https://doi.org/10.3758/BF03197669>
- Basu Mallick, D., F Magnotti, J., & S Beauchamp, M. (2015). Variability and stability in the McGurk effect: contributions of participants, stimuli, time, and response type. *Psychonomic bulletin & review*, *22*(5), 1299-1307. <https://doi.org/10.3758/s13423-015-0817-4>
- Batki, A., Baron-Cohen, S., Wheelwright, S., Connellan, J., & Ahluwalia, J. (2000). Is there an innate gaze module? Evidence from human neonates. *Infant Behavior and Development*, *23*(2), 223-229. [https://doi.org/10.1016/S0163-6383\(01\)00037-6](https://doi.org/10.1016/S0163-6383(01)00037-6)
- Baus, C., Bas, J., Calabria, M., & Costa, A. (2017). Speak My Language and I Will Remember Your Face Better: An ERP Study. *Frontiers in Psychology*, *8*, 709. <https://doi.org/10.3389/fpsyg.2017.00709>
- Bearden, E. D., Wilson, J. D., Zharov, V. P., & Lowery, C. L. (2001, July). Deep penetration of light into biotissue. *Laser-Tissue Interaction XII: Photochemical, Photothermal, and Photomechanical*, 4257, 417-425. <https://doi.org/10.1117/12.434729>
- Benitez, V. L., & Saffran, J. R. (2018). Predictable events enhance word learning in toddlers. *Current Biology*, *28*(17), 2787-2793. <https://doi.org/10.1016/j.cub.2018.06.017>
- Benoît, C., Mohamadi, T., & Kandel, S. (1994). Effects of phonetic context on audio-visual intelligibility of French. *Journal of Speech, Language, and Hearing Research*, *37*(5), 1195-1203. <https://doi.org/10.1044/jshr.3705.1195>
- Benoît, C., Guiard-Marigny, T., Le Goff, B., & Adjoudani, A. (1996). Which components of the face do humans and machines best speechread?. *Speechreading by humans and machines*, *150*, 315-328. https://doi.org/10.1007/978-3-662-13015-5_24
- Benoit, M. M., Rajj, T., Lin, F. H., Jääskeläinen, I. P., & Stufflebeam, S. (2010). Primary and multisensory cortical activity is correlated with audiovisual percepts. *Human brain mapping*, *31*(4), 526-538. <https://doi.org/10.1002/hbm.20884>

- Benson, P. J., & Perrett, D. I. (1991). Perception and recognition of photographic quality facial caricatures: Implications for the recognition of natural images. *European Journal of Cognitive Psychology*, 3(1), 105-135. <https://doi.org/10.1080/09541449108406222>
- Benson, P. J., & Perrett, D. I. (1994). Visual processing of facial distinctiveness. *Perception*, 23(1), 75-93. <https://doi.org/10.1068/p230075>
- Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *Journal of cognitive neuroscience*, 8(6), 551-565. <https://doi.org/10.1162/jocn.1996.8.6.551>
- Bernstein, L. E., Tucker, P. E., & Demorest, M. E. (2000). Speech perception without hearing. *Perception & Psychophysics*, 62(2), 233-252. <https://doi.org/10.3758/bf03205546>
- Bernstein, L. E., Auer Jr, E. T., & Takayanagi, S. (2004). Auditory speech detection in noise enhanced by lipreading. *Speech Communication*, 44(1-4), 5-18. <https://doi.org/10.1016/j.specom.2004.10.011>
- Bertenthal B.I., Proffitt D.R., Cutting J.E. (1984). Infant sensitivity to figural coherence in biomechanical motions. *Journal of experimental Psychology*, 37, 213-230. [https://doi.org/10.1016/0022-0965\(84\)90001-8](https://doi.org/10.1016/0022-0965(84)90001-8)
- Bertenthal B.I., Proffitt D.R., Kramer S.J. (1987) Perception of Biomechanical Motions by Infants: Implementation of Various Processing Constraints. *Journal of experimental Psychology: Human Perception and Performance*, 13, 577-585. <https://doi.org/10.1037/0096-1523.13.4.577>
- Bertelson, P., Vroomen, J., Wiegand, G., & de Gelder, B. (1994). Exploring the relation between McGurk interference and ventriloquism. *Proceedings of the 1994 International Conference on Spoken Language Processing*, 2, 559-562.
- Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of experimental psychology: human perception and performance*, 14(3), 345. <https://doi.org/10.1037/0096-1523.14.3.345>
- Best, C. T., & Avery, R. A. (1999). Left-hemisphere advantage for click consonants is determined by linguistic significance and experience. *Psychological Science*, 10(1), 65-70. <https://doi.org/10.1111/1467-9280.00108>
- Binnie, C. A., Montgomery, A. A., & Jackson, P. L. (1974). Auditory and visual contributions to the perception of consonants. *Journal of speech and hearing research*, 17(4), 619-630. <https://doi.org/10.1044/jshr.1704.619>
- Birch, E. E., Gwiazda, J., Bauer Jr, J. A., Naegele, J., & Held, R. (1983). Visual acuity and its meridional variations in children aged 7-60 months. *Vision research*, 23(10), 1019-1024. [https://doi.org/10.1016/0042-6989\(83\)90012-3](https://doi.org/10.1016/0042-6989(83)90012-3)
- Bosch, L., & Sebastián-Gallés, N. (1997). Native-language recognition abilities in 4-month-old infants from monolingual and bilingual environments. *Cognition*, 65(1), 33-69. [https://doi.org/10.1016/S0010-0277\(97\)00040-1](https://doi.org/10.1016/S0010-0277(97)00040-1)

- Bosch, L., & Sebastián-Gallés, N. (2001). Evidence of early language discrimination abilities in infants from bilingual environments. *Infancy*, 2(1), 29-49. https://doi.org/10.1207/S15327078IN0201_3
- Bosch, L., & Sebastián-Gallés, N. (2003). Simultaneous bilingualism and the perception of a language-specific vowel contrast in the first year of life. *Language and speech*, 46(2-3), 217-243. <https://doi.org/10.1177/00238309030460020801>
- Bothwell, R. K., Brigham, J. C., & Malpass, R. S. (1989). Cross-racial identification. *Personality and Social Psychology Bulletin*, 15(1), 19-25. <https://doi.org/10.1177/0146167289151002>
- Bourgeois, J. P., Goldman-Rakic, P. S., & Rakic, P. (1994). Synaptogenesis in the prefrontal cortex of rhesus monkeys. *Cerebral cortex*, 4(1), 78-96. <https://doi.org/10.1093/cercor/4.1.78>
- Bosseler, A., Taulu, S., Pihko, E., Mäkelä, J., Imada, T., Ahonen, A., & Kuhl, P. (2013). Theta brain rhythms index perceptual narrowing in infant speech perception. *Frontiers in Psychology*, 4, 690. <https://doi.org/10.3389/fpsyg.2013.00690>
- Bowlby, J. (1982). Attachment and loss: retrospect and prospect. *American journal of Orthopsychiatry*, 52(4), 664. <https://doi.org/10.1111/j.1939-0025.1982.tb01456.x>
- Braddick, O., Atkinson, J., & Wattam-Bell, J. (2003). Normal and anomalous development of visual motion processing: motion coherence and ‘dorsal-stream vulnerability’. *Neuropsychologia*, 41(13), 1769-1784. [https://doi.org/10.1016/S0028-3932\(03\)00178-7](https://doi.org/10.1016/S0028-3932(03)00178-7)
- Braddick, O., Birtles, D., Wattam-Bell, J., & Atkinson, J. (2005). Motion-and orientation-specific cortical responses in infancy. *Vision research*, 45(25-26), 3169-3179. <https://doi.org/10.1016/j.visres.2005.07.021>
- Brancazio, L. (2004). Lexical influences in audiovisual speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, 30(3), 445. <https://doi.org/10.1037/0096-1523.30.3.445>
- Brancazio, L., & Miller, J. L. (2005). Use of visual information in speech perception: Evidence for a visual rate effect both with and without a McGurk effect. *Perception & Psychophysics*, 67(5), 759-769. <https://doi.org/10.3758/BF03193531>
- Brants, M., Wagemans, J., & Op de Beeck, H. P. (2011). Activation of fusiform face area by Greebles is related to face similarity but not expertise. *Journal of Cognitive Neuroscience*, 23(12), 3949-3958. https://doi.org/10.1162/jocn_a_00072
- Breen, E., Pomper, R., & Saffran, J. (2019). Phonological learning influences label–object mapping in toddlers. *Journal of Speech, Language, and Hearing Research*, 62(6), 1923-1932. https://doi.org/10.1044/2019_JSLHR-L-18-0131
- Bregman, M. R., & Creel, S. C. (2014). Gradient language dominance affects talker learning. *Cognition*, 130(1), 85-95. <https://doi.org/10.1016/j.cognition.2013.09.010>
- Breiter, H. C., Etcoff, N. L., Whalen, P. J., Kennedy, W. A., Rauch, S. L., Buckner, R. L., ... & Rosen, B. R. (1996). Response and habituation of the human amygdala during visual processing of facial expression. *Neuron*, 17(5), 875-887. [https://doi.org/10.1016/S0896-6273\(00\)80219-6](https://doi.org/10.1016/S0896-6273(00)80219-6)

- Brigham, J. C., & Barkowitz, P. (1978). Do “They all look alike?” The Effect of Race, Sex, Experience, and Attitudes on the Ability to Recognize Faces 1. *Journal of Applied Social Psychology*, 8(4), 306-318. <https://doi.org/10.1111/j.1559-1816.1978.tb00786.x>
- Brigham, J. C. (1986). The influence of race on face recognition. *Aspects of Face Processing*, 170-177. https://doi.org/10.1007/978-94-009-4420-6_17
- Brigham, J. C. (1990). Target person distinctiveness and attractiveness as moderator variables in the confidence-accuracy relationship in eyewitness identifications. *Basic and Applied Social Psychology*, 11(1), 101-115. https://doi.org/10.1207/s15324834basp1101_7
- Brown, V. A., Hedayati, M., Zanger, A., Mayn, S., Ray, L., Dillman-Hasso, N., & Strand, J. F. (2018). What accounts for individual differences in susceptibility to the McGurk effect?. *PloS one*, 13(11), e0207160. <https://doi.org/10.1371/journal.pone.0207160>
- Bruce, V., & Young, A. (1986). Understanding face recognition. *British journal of psychology*, 77(3), 305-327. <https://doi.org/10.1111/j.2044-8295.1986.tb02199.x>
- Bruyer, R., & Courvoisier, P. (1990). Distinctiveness of faces or distinctiveness of persons?. *Perceptual and motor skills*, 71(3), 967-984. <https://doi.org/10.2466/pms.1990.71.3.967>
- Buchan, J. N., & Munhall, K. G. (2011). The influence of selective attention to auditory and visual speech on the integration of audiovisual speech information. *Perception*, 40(10), 1164-1182. <https://doi.org/10.1068/p6939>
- Bullier, J. (2001). Integrated model of visual processing. *Brain research reviews*, 36(2-3), 96-107. [https://doi.org/10.1016/S0165-0173\(01\)00085-6](https://doi.org/10.1016/S0165-0173(01)00085-6)
- Burnham, D. (1993). Visual recognition of mother by young infants: facilitation by speech. *Perception*, 22(10), 1133-1153. <https://doi.org/10.1068/p221133>
- Burnham, D., & Dodd, B. (1996). Auditory-visual speech perception as a direct process: The McGurk effect in infants and across languages. In *Speechreading by humans and machines* (pp. 103-114). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-13015-5_7
- Burnham, D., & Dodd, B. (2004). Auditory–visual speech integration by prelinguistic infants: Perception of an emergent consonant in the McGurk effect. *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology*, 45(4), 204-220. <https://doi.org/10.1002/dev.20032>
- Burns, T. C., Yoshida, K. A., Hill, K., & Werker, J. F. (2007). The development of phonetic representation in bilingual and monolingual infants. *Applied Psycholinguistics*, 28(3), 455-474. <https://doi.org/10.1017/S0142716407070257>
- Burns, E. J., Tree, J., Chan, A. H., & Xu, H. (2019). Bilingualism shapes the other race effect. *Vision research*, 157, 192-201. <https://doi.org/10.1016/j.visres.2018.07.004>
- Burton, M. W., & Small, S. L. (2006). Functional neuroanatomy of segmenting speech and nonspeech. *Cortex*, 42(4), 644-651. [https://doi.org/10.1016/S0010-9452\(08\)70400-3](https://doi.org/10.1016/S0010-9452(08)70400-3)

- Bushnell, I. W. R. (2001). Mother's face recognition in newborn infants: Learning and memory. *Infant and Child Development: An International Journal of Research and Practice*, 10(1-2), 67-74. <https://doi.org/10.1002/icd.248>
- Byers-Heinlein, K., Burns, T. C., & Werker, J. F. (2010). The roots of bilingualism in newborns. *Psychological science*, 21(3), 343-348. <https://doi.org/10.1177/0956797609360758>
- Byers-Heinlein, K. (2013). Parental language mixing: Its measurement and the relation of mixed input to young bilingual children's vocabulary size. *Bilingualism: Language and Cognition*, 16(1), 32-48. <https://doi.org/10.1017/S1366728912000120>
- Calder, A. J., Young, A. W., Benson, P. J., & Perrett, D. I. (1996). Self priming from distinctive and caricatured faces. *British Journal of Psychology*, 87(1), 141-162. <https://doi.org/10.1111/j.2044-8295.1996.tb02581.x>
- Campbell, R., Landis, T. & Regard, M. (1986) Face recognition and lip-reading: A neurological dissociation. *Brain*, 109, 509–521. <https://doi.org/10.1093/brain/109.3.509>
- Carey, S., & Diamond, R. (1977). From piecemeal to configurational representation of faces. *Science*, 195(4275), 312-314. <https://doi.org/10.1126/science.831281>
- Carey, S., & Diamond, R. (1994). Are faces perceived as configurations more by adults than by children?. *Visual cognition*, 1(2-3), 253-274. <https://doi.org/10.1080/13506289408402302>
- Cargile, A. C., Giles, H., Ryan, E. B., & Bradac, J. J. (1994). Language attitudes as a social process: A conceptual model and new directions. *Language & Communication*, 14(3), 211-236. [https://doi.org/10.1016/0271-5309\(94\)90001-9](https://doi.org/10.1016/0271-5309(94)90001-9)
- Carroo, A. W. (1986). Other race recognition: A comparison of Black American and African subjects. *Perceptual and Motor Skills*, 62(1), 135-138. <https://doi.org/10.2466/pms.1986.62.1.135>
- Carroo, A. W. (1987). Recognition of faces as a function of race, attitudes, and reported cross-racial friendships. *Perceptual and Motor Skills*, 64(1), 319-325. <https://doi.org/10.2466/pms.1987.64.1.319>
- Cernoch, J. M., & Porter, R. H. (1985). Recognition of maternal axillary odors by infants. *Child development*, 56(6), 1593-1598. <https://doi.org/10.2307/1130478>
- Chance, J., Goldstein, A. G., & McBride, L. (1975). Differential experience and recognition memory for faces. *The Journal of Social Psychology*, 97(2), 243-253. <https://doi.org/10.1080/00224545.1975.9923344>
- Chance, J. E., & Goldstein, A. G. (1981). Depth of processing in response to own-and other-race faces. *Personality and Social Psychology Bulletin*, 7(3), 475-480. <https://doi.org/10.1177/014616728173017>
- Chen, T., & Rao, R. R. (1998). Audio-visual integration in multimodal communication. *Proceedings of the IEEE*, 86(5), 837-852. <https://doi.org/10.1109/5.664274>

- 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>
- Chiroro, P., & Valentine, T. (1995). An investigation of the contact hypothesis of the own-race bias in face recognition. *The Quarterly Journal of Experimental Psychology Section A*, 48(4), 879-894. <https://doi.org/10.1080/14640749508401421>
- Clerc, O., Fort, M., Schwarzer, G., Krasotkina, A., Vilain, A., Méary, D., ... & Pascalis, O. (2021). Can language modulate perceptual narrowing for faces? Other-race face recognition in infants is modulated by language experience. *International Journal of Behavioral Development*. <https://doi.org/10.1177/01650254211053054>
- Cohen, M. E., & Ross, L. E. (1978). Latency and accuracy characteristics of saccades and corrective saccades in children and adults. *Journal of Experimental Child Psychology*, 26(3), 517-527. [https://doi.org/10.1016/0022-0965\(78\)90130-3](https://doi.org/10.1016/0022-0965(78)90130-3)
- Colin, C., Radeau, M., Deltenre, P., Demolin, D., & Soquet, A. (2002). The role of sound intensity and stop-consonant voicing on McGurk fusions and combinations. *European Journal of Cognitive Psychology*, 14(4), 475-491. <https://doi.org/10.1080/09541440143000203>
- Colin, C., Radeau, M., & Deltenre, P. (2005). Top-down and bottom-up modulation of audiovisual integration in speech. *European Journal of Cognitive Psychology*, 17(4), 541-560. <https://doi.org/10.1080/09541440440000168>
- Colombo, J., & Mitchell, D. W. (2009). Infant visual habituation. *Neurobiology of learning and memory*, 92(2), 225-234. <https://doi.org/10.1016/j.nlm.2008.06.002>
- Conboy, B. T., Sommerville, J. A., & Kuhl, P. K. (2008). Cognitive control factors in speech perception at 11 months. *Developmental Psychology*, 44(5), 1505–1512. <https://doi.org/10.1037/a0012975>
- Courage, M. L., & Adams, R. J. (1990). Visual acuity assessment from birth to three years using the acuity card procedure: cross-sectional and longitudinal samples. *Optometry and vision science: official publication of the American Academy of Optometry*, 67(9), 713-718. <https://doi.org/10.1097/00006324-199009000-00011>
- Coutrot, A., & Guyader, N. (2014). How saliency, faces, and sound influence gaze in dynamic social scenes. *Journal of vision*, 14(8), 5-5. <https://doi.org/10.1167/14.8.5>
- Cynader, M. (1983). Prolonged sensitivity to monocular deprivation in dark-reared cats: effects of age and visual exposure. *Developmental Brain Research*, 8(2-3), 155-164. [https://doi.org/10.1016/0165-3806\(83\)90002-0](https://doi.org/10.1016/0165-3806(83)90002-0)
- Damasio, A. R., Damasio, H., & Van Hoesen, G. W. (1982). Prosopagnosia: anatomic basis and behavioral mechanisms. *Neurology*, 32(4), 331-331. <https://doi.org/10.1212/WNL.32.4.331>
- DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science*, 208(4448), 1174-1176. <https://doi.org/10.1126/science.7375928>
- Dehaene, S., & Cohen, L. (2007). Cultural recycling of cortical maps. *Neuron*, 56(2), 384-398. <https://doi.org/10.1016/j.neuron.2007.10.004>

- Del Giudice, M. (2011). Alone in the dark? Modeling the conditions for visual experience in human fetuses. *Developmental psychobiology*, 53(2), 214-219. <https://doi.org/10.1002/dev.20506>
- De Heering, A., De Liedekerke, C., Deboni, M., & Rossion, B. (2010). The role of experience during childhood in shaping the other-race effect. *Developmental science*, 13(1), 181-187. <https://doi.org/10.1111/j.1467-7687.2009.00876.x>
- Dekle, D. J., Fowler, C. A., & Funnell, M. G. (1992). Audiovisual integration in perception of real words. *Perception & Psychophysics*, 51(4), 355-362. <https://doi.org/10.3758/BF03211629>
- Denison, S., Reed, C., & Xu, F. (2013). The emergence of probabilistic reasoning in very young infants: evidence from 4.5- and 6-month-olds. *Developmental psychology*, 49(2), 243. <https://psycnet.apa.org/doi/10.1037/a0028278>
- DePaolis, R. A., Vihman, M. M., & Keren-Portnoy, T. (2011). Do production patterns influence the processing of speech in prelinguistic infants? *Infant Behavior and Development*, 34(4), 590-601. <https://doi.org/10.1016/j.infbeh.2011.06.005>
- DePaolis, R. A., Vihman, M. M., & Nakai, S. (2013). The influence of babbling patterns on the processing of speech. *Infant Behavior and Development*, 36(4), 642-649. <https://doi.org/10.1016/j.infbeh.2013.06.007>
- De Renzi, E., Perani, D., Carlesimo, G. A., Silveri, M. C., & Fazio, F. (1994). Prosopagnosia can be associated with damage confined to the right hemisphere—an MRI and PET study and a review of the literature. *Neuropsychologia*, 32(8), 893-902. [https://doi.org/10.1016/0028-3932\(94\)90041-8](https://doi.org/10.1016/0028-3932(94)90041-8)
- Desimone, R. (1991). Face-selective cells in the temporal cortex of monkeys. *Journal of cognitive neuroscience*, 3(1), 1-8. <https://doi.org/10.1162/jocn.1991.3.1.1>
- Desjardins, R. N., & Werker, J. F. (2004). Is the integration of heard and seen speech mandatory for infants?. *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology*, 45(4), 187-203. <https://doi.org/10.1002/dev.20033>
- Diamond, R., & Carey, S. (1986). Why faces are and are not special: an effect of expertise. *Journal of experimental psychology: general*, 115(2), 107. <https://doi.org/10.1037/0096-3445.115.2.107>
- DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition?. *Neuron*, 73(3), 415-434. <https://doi.org/10.1016/j.neuron.2012.01.010>
- Dodd, B. (1979). Lip-reading in infants: Attention to speech presented in—and out—of synchrony. *Cognitive Psychology*, 11, 478-484. [https://doi.org/10.1016/0010-0285\(79\)90021-5](https://doi.org/10.1016/0010-0285(79)90021-5)
- Dunham, Y., Chen, E. E., & Banaji, M. R. (2013). Two signatures of implicit intergroup attitudes: Developmental invariance and early enculturation. *Psychological science*, 24(6), 860-868. <https://doi.org/10.1177/0956797612463081>
- Dunham, Y., Stepanova, E. V., Dotsch, R., & Todorov, A. (2015). The development of race-based perceptual categorization: Skin color dominates early category judgments. *Developmental Science*, 18(3), 469-483. <https://doi.org/10.1111/desc.12228>

- Dupierrix, E., de Boisferon, A. H., Méary, D., Lee, K., Quinn, P. C., Di Giorgio, E., ... & Pascalis, O. (2014). Preference for human eyes in human infants. *Journal of Experimental Child Psychology*, 123, 138-146. <https://doi.org/10.1016/j.jecp.2013.12.010>
- Eilers, R. E., Gavin, W. J., & Oller, D. K. (1982). Cross-linguistic perception in infancy: early effects of linguistic experience. *Journal of Child Language*, 9(2), 289-302. <https://doi.org/10.1017/S0305000900004724>
- Ellis, E. M., Gonzalez, M. R., & Deák, G. O. (2014). Visual prediction in infancy: What is the association with later vocabulary?. *Language Learning and Development*, 10(1), 36-50. <https://doi.org/10.1080/15475441.2013.799988>
- Ellis, H. D., & Lewis, M. B. (2001). Capgras delusion: a window on face recognition. *Trends in cognitive sciences*, 5(4), 149-156. [https://doi.org/10.1016/S1364-6613\(00\)01620-X](https://doi.org/10.1016/S1364-6613(00)01620-X)
- Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, 171(3968), 303-306. <https://doi.org/10.1126/science.171.3968.303>
- Elliott, E. S., Wills, E. J., & Goldstein, A. G. (1973). The effects of discrimination training on the recognition of white and oriental faces. *Bulletin of the Psychonomic Society*, 2(2), 71-73. <https://doi.org/10.3758>
- Elwell, C. M., Brown, R. J., & Rutter, D. R. (1984). Effects of accent and visual information on impression formation. *Journal of Language and Social Psychology*, 3(4), 297-299. <https://doi.org/10.1177/0261927X8400300404>
- Emberson, L. L., Richards, J. E., & Aslin, R. N. (2015). Top-down modulation in the infant brain: Learning-induced expectations rapidly affect the sensory cortex at 6 months. *Proceedings of the National Academy of Sciences*, 112(31), 9585-9590. <https://doi.org/10.1073/pnas.1510343112>
- Emery, N. J. (2000). The eyes have it: the neuroethology, function and evolution of social gaze. *Neuroscience & biobehavioral reviews*, 24(6), 581-604. [https://doi.org/10.1016/S0149-7634\(00\)00025-7](https://doi.org/10.1016/S0149-7634(00)00025-7)
- Engel, S. A., Glover, G. H., & Wandell, B. A. (1997). Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cerebral cortex*, 7(2), 181-192. <https://doi.org/10.1093/cercor/7.2.181>
- Erber, N. P. (1969). Interaction of audition and vision in the recognition of oral speech stimuli. *Journal of speech and hearing research*, 12(2), 423-425. <https://doi.org/10.1044/jshr.1202.423>
- Erber, N. P. (1974). Visual perception of speech by deaf children: Recent developments and continuing needs. *Journal of Speech and Hearing Disorders*, 39(2), 178-185. <https://doi.org/10.1044/jshd.3902.178>
- Estudillo, A. J., Lee, J. K. W., Mennie, N., & Burns, E. (2020). No evidence of other-race effect for Chinese faces in Malaysian non-Chinese population. *Applied Cognitive Psychology*, 34(1), 270-276. <https://doi.org/10.1002/acp.3609>

- Eswaran, H., Lowery, C. L., Wilson, J. D., Murphy, P., & Preissl, H. (2004). Functional development of the visual system in human fetus using magnetoencephalography. *Experimental Neurology*, *190*, 52-58. <https://doi.org/10.1016/j.expneurol.2004.04.007>
- Ewertesen, H. W., & Nielsen, H. B. (1971). A comparative analysis of the audiovisual, auditive and visual perception of speech. *Acta oto-laryngologica*, *72*(1-6), 201-205. <https://doi.org/10.3109/00016487109122473>
- Fair, J., Flom, R., Jones, J., & Martin, J. (2012). Perceptual learning: 12-month-olds' discrimination of monkey faces. *Child development*, *83*(6), 1996-2006. <https://doi.org/10.1111/j.1467-8624.2012.01814.x>
- Fox R, Mc Daniels C (1982) The perception of biological motion by human infants. *Science*, *218*, 486-487. <https://doi.org/10.1126/science.7123249>
- Fantz, R. L. (1956). A method for studying early visual development. *Perceptual and Motor Skills*, *6*, 13-15. <https://doi.org/10.2466/pms.1956.6.g.13>
- Fantz, R. L. (1958a). Form preference in newly hatched chicks. *Journal of Comparative and Physiological Psychology*, *50*, 422-430. <https://doi.org/10.1037/h0044973>
- Fantz, R. L. (1958b). Pattern vision in young infants. *The Psychological Record*, *8*, 43-47.
- Fantz, R. L. (1958c). Visual discrimination in a neonate chimpanzee. *Perceptual and Motor Skills*, *8*, 59-66. <https://doi.org/10.2466/pms.1958.8.3.59>
- Farah, M. J. (1996). Is face recognition 'special'? Evidence from neuropsychology. *Behavioural brain research*, *76*(1-2), 181-189. [https://doi.org/10.1016/0166-4328\(95\)00198-0](https://doi.org/10.1016/0166-4328(95)00198-0)
- Farroni, T., Valenza, E., Simion, F., & Umiltà, C. (2000). Configural processing at birth: Evidence for perceptual organisation. *Perception*, *29*(3), 355-372. <https://doi.org/10.1068/p2903rvw>
- Farroni, T., Csibra, G., Simion, F., & Johnson, M. H. (2002). Eye contact detection in humans from birth. *Proceedings of the National academy of sciences*, *99*(14), 9602-9605. <https://doi.org/10.1073/pnas.152159999>
- Fecher, N., & Johnson, E. K. (2018). Effects of language experience and task demands on talker recognition by children and adults. *The Journal of the Acoustical Society of America*, *143*(4), 2409-2418. <https://doi.org/10.1121/1.5032199>
- Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral cortex (New York, NY: 1991)*, *1*(1), 1-47. <https://doi.org/10.1093/cercor/1.1.1-a>
- Feng, G., Zhou, B., Zhou, W., Beauchamp, M. S., & Magnotti, J. F. (2019). A laboratory study of the McGurk effect in 324 Monozygotic and Dizygotic Twins. *Frontiers in neuroscience*, *13*, 1029. <https://doi.org/10.3389/fnins.2019.01029>
- Fischer, B., Biscaldi, M., & Gezeck, S. (1997). On the development of voluntary and reflexive components in human saccade generation. *Brain research*, *754*(1-2), 285-297. [https://doi.org/10.1016/S0006-8993\(97\)00094-2](https://doi.org/10.1016/S0006-8993(97)00094-2)

- Feinman, S., & Entwisle, D. R. (1976). Children's ability to recognize other children's faces. *Child development*, 47(2), 506-510. <https://doi.org/10.2307/1128809>
- Fixmer, E., & Hawkins, S. (1998). The influence of quality of information on the McGurk effect. *International Conference on Auditory-Visual Speech Processing*.
- Freiwald, W., Tsao, D. & Livingstone, M. (2009). A face feature space in the macaque temporal lobe. *Nat Neurosci* 12, 1187–1196. <https://doi.org/10.1038/nn.2363>
- Friede, R. L., & Hu, K. H. (1967). Proximo-distal differences in myelin development in human optic fibers. *Zeitschrift für Zellforschung und mikroskopische Anatomie*, 79(2), 259-264. <https://doi.org/10.1002/cne.901300304>
- Friedmann, N., & Rusou, D. (2015). Critical period for first language: the crucial role of language input during the first year of life. *Current opinion in neurobiology*, 35, 27-34. <https://doi.org/10.1016/j.conb.2015.06.003>
- Fulford, J., Vadeyar, S. H., Dodampahala, S. H., Moore, R. J., Young, P., Baker, P. N., ... & Gowland, P. A. (2003). Fetal brain activity in response to a visual stimulus. *Human brain mapping*, 20(4), 239-245. <https://doi.org/10.1002/hbm.10139>
- Fukushima, J., Hatta, T., & Fukushima, K. (2000). Development of voluntary control of saccadic eye movements: I. Age-related changes in normal children. *Brain and Development*, 22(3), 173-180. [https://doi.org/10.1016/S0387-7604\(00\)00101-7](https://doi.org/10.1016/S0387-7604(00)00101-7)
- 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>
- Garey, L. J., & De Courten, C. (1983). Structural development of the lateral geniculate nucleus and visual cortex in monkey and man. *Behavioural Brain Research*, 10(1), 3-13. [https://doi.org/10.1016/0166-4328\(83\)90145-6](https://doi.org/10.1016/0166-4328(83)90145-6)
- Gauthier, I., & Tarr, M. J. (1997). Becoming a “Greeble” expert: Exploring mechanisms for face recognition. *Vision research*, 37(12), 1673-1682. [https://doi.org/10.1016/S0042-6989\(96\)00286-6](https://doi.org/10.1016/S0042-6989(96)00286-6)
- Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (1999). Activation of the middle fusiform face area increases with expertise in recognizing novel objects. *Nature neuroscience*, 2(6), 568-573. <https://doi.org/10.1038/9224>
- Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000a). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature neuroscience*, 3(2), 191-197. <https://doi.org/10.1038/72140>
- Gauthier, I., Tarr, M. J., Moylan, J., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000b). The fusiform “face area” is part of a network that processes faces at the individual level. *Journal of cognitive neuroscience*, 12(3), 495-504. <https://doi.org/10.1162/089892900562165>

- Gervain, J., & Mehler, J. (2010). Speech perception and language acquisition in the first year of life. *Annual review of psychology*, *61*, 191-218. <https://doi.org/10.1146/annurev.psych.093008.100408>
- Gervain, J. (2018). The role of prenatal experience in language development. *Current Opinion in Behavioral Sciences*, *21*, 62-67. <https://doi.org/10.1016/j.cobeha.2018.02.004>
- Gibson, E. J., & Walk, R. D. (1960). The "visual cliff". *Scientific American*, *202*(4), 64-71. <https://doi.org/10.1038/scientificamerican0460-64>
- Gick, B., & Derrick, D. (2009). Aero-tactile integration in speech perception. *Nature*, *462*(7272), 502-504. <https://doi.org/10.1038/nature08572>
- Giles, H., & Billings, A. C. (2004). Assessing Language Attitudes: Speaker Evaluation Studies. *The Handbook of Applied Linguistics*, 187-209. <https://doi.org/10.1002/9780470757000.ch7>
- Glass, P. (2002). Development of the visual system and implications for early intervention. *Infants & Young Children*, *15*(1), 1-10. <https://doi.org/10.1097/00001163-200207000-00003>
- Goff, B. L., Guiard-Marigny, T., & Benoît, C. (1995). Read my lips... and my jaw! How intelligible are the components of a speaker's face?. *Fourth European Conference on Speech Communication and Technology*.
- Goggin, J. P., Thompson, C. P., Strube, G., & Simental, L. R. (1991). The role of language familiarity in voice identification. *Memory & Cognition*, *19*(5), 448-458. <https://doi.org/10.3758/BF03199567>
- Going, M., & Read, J. D. (1974). Effects of uniqueness, sex of subject, and sex of photograph on facial recognition. *Perceptual and Motor Skills*, *39*(1), 109-110. <https://doi.org/10.2466/pms.1974.39.1.109>
- Goldstein, A. G. (1979a). Race-related variation of facial features: Anthropometric data I. *Bulletin of the Psychonomic Society*, *13*(3), 187-190. <https://doi.org/10.3758/BF03335055>
- Goldstein, A. G. (1979b). Facial feature variation: Anthropometric data II. *Bulletin of the Psychonomic Society*, *13*(3), 191-193. <https://doi.org/10.3758/BF03335056>
- Goldstein, A. G., & Chance, J. E. (1980). Memory for faces and schema theory. *The journal of psychology*, *105*(1), 47-59. <https://doi.org/10.1080/00223980.1980.9915131>
- Goldstein, A. G., & Chance, J. E. (1985). Effects of training on Japanese face recognition: Reduction of the other-race effect. *Bulletin of the Psychonomic Society*, *23*(3), 211-214. <https://doi.org/10.3758/BF03329829>
- 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.
- Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds "L" and "R.". *Neuropsychologia*, *9*(3), 317-323. [https://doi.org/10.1016/0028-3932\(71\)90027-3](https://doi.org/10.1016/0028-3932(71)90027-3)
- Gottfried, J. A., O'Doherty, J., & Dolan, R. J. (2003). Encoding predictive reward value in human amygdala and orbitofrontal cortex. *Science*, *301*(5636), 1104-1107. <https://doi.org/10.1126/science.1087919>

- Grant, K. W., Walden, B. E., & Seitz, P. F. (1998). Auditory-visual speech recognition by hearing-impaired subjects: Consonant recognition, sentence recognition, and auditory-visual integration. *The Journal of the Acoustical Society of America*, 103(5), 2677-2690. <https://doi.org/10.1121/1.422788>
- Green, D. G., Powers, M. K., & Banks, M. S. (1980). Depth of focus, eye size and visual acuity. *Vision research*, 20(10), 827-835. [https://doi.org/10.1016/0042-6989\(80\)90063-2](https://doi.org/10.1016/0042-6989(80)90063-2)
- Green, K. P., Kuhl, P. K., & Meltzoff, A. N. (1988). Factors affecting the integration of auditory and visual information in speech: The effect of vowel environment. *The Journal of the Acoustical Society of America*, 84(S1), S155. <https://doi.org/10.1121/1.2025888>
- Green, K. P., Kuhl, P. K., Meltzoff, A. N., & Stevens, E. B. (1991). Integrating speech information across talkers, gender, and sensory modality: Female faces and male voices in the McGurk effect. *Perception & psychophysics*, 50(6), 524-536. <https://doi.org/10.3758/BF03207536>
- Grimshaw, G. M., Adelstein, A., Bryden, M. P., & MacKinnon, G. E. (1998). First-language acquisition in adolescence: Evidence for a critical period for verbal language development. *Brain and Language*, 63(2), 237-255. <https://doi.org/10.1006/brln.1997.1943>
- Grossmann, T., Missana, M., Friederici, A. D., & Ghazanfar, A. A. (2012). Neural correlates of perceptual narrowing in cross-species face-voice matching. *Developmental Science*, 15(6), 830-839. <https://doi.org/10.1111/j.1467-7687.2012.01179.x>
- Gross, C. G., Rocha-Miranda, C. D., & Bender, D. B. (1972). Visual properties of neurons in inferotemporal cortex of the macaque. *Journal of neurophysiology*, 35(1), 96-111. <https://doi.org/10.1152/jn.1972.35.1.96>
- Guellaï, B., Streri, A., Chopin, A., Rider, D., & Kitamura, C. (2016). Newborns' sensitivity to the visual aspects of infant-directed speech: Evidence from point-line displays of talking faces. *Journal of Experimental Psychology: Human Perception and Performance*, 42(9), 1275. <https://doi/10.1037/xhp0000208>
- Guellaï, B., Callin, A., Bevilacqua, F., Schwarz, D., Pitti, A., Boucenna, S., & Gratier, M. (2019). Sensus communis: Some perspectives on the origins of non-synchronous cross-sensory associations. *Frontiers in Psychology*, 10, 523. <https://doi.org/10.3389/fpsyg.2019.00523>
- ainline, L. (1978). Developmental changes in visual scanning of face and nonface patterns by infants. *Journal of Experimental Child Psychology*, 25(1), 90-115. [https://doi.org/10.1016/0022-0965\(78\)90041-3](https://doi.org/10.1016/0022-0965(78)90041-3)
- Haith, M. M., Bergman, T., & Moore, M. J. (1977). Eye contact and face scanning in early infancy. *Science*, 198(4319), 853-855. <https://doi.org/10.1126/science.918670>
- Halgren, E., Dale, A. M., Sereno, M. I., Tootell, R. B., Marinkovic, K., & Rosen, B. R. (1999). Location of human face-selective cortex with respect to retinotopic areas. *Human brain mapping*, 7(1), 29-37. [https://doi.org/10.1002/\(SICI\)1097-0193\(1999\)7:1<29::AID-HBM3>3.0.CO;2-R](https://doi.org/10.1002/(SICI)1097-0193(1999)7:1<29::AID-HBM3>3.0.CO;2-R)

- Haller, R., Rummel, C., Henneberg, S., Pollmer, U., & Köster, E. P. (1999). The influence of early experience with vanillin on food preference later in life. *Chem. Senses*, 24, 465-467. <https://doi.org/10.1093/chemse/24.4.465>
- Hancock, K. J., & Rhodes, G. (2008). Contact, configural coding and the other-race effect in face recognition. *British Journal of Psychology*, 99(1), 45-56. <https://doi.org/10.1348/000712607X199981>
- Hannon, E. E., & Trehub, S. E. (2005a). Metrical categories in infancy and adulthood. *Psychological science*, 16(1), 48-55. <https://doi.org/10.1111/j.0956-7976.2005.00779.x>
- Hannon, E. E., & Trehub, S. E. (2005b). Tuning in to musical rhythms: Infants learn more readily than adults. *Proceedings of the National Academy of Sciences*, 102(35), 12639-12643. <https://doi.org/10.1073/pnas.0504254102>
- Haq, S., Jackson, P. J., & Edge, J. (2009). Speaker-dependent audio-visual emotion recognition. *AVSP*, 53-58.
- Hardison, D. M. (1996). Bimodal speech perception by native and nonnative speakers of English: factors influencing the McGurk effect. *Language Learning*, 46(1), 3-73. <https://doi.org/10.1111/j.1467-1770.1996.tb00640.x>
- Hasselmo, M. E., Rolls, E. T., & Baylis, G. C. (1989). The role of expression and identity in the face-selective responses of neurons in the temporal visual cortex of the monkey. *Behavioural brain research*, 32(3), 203-218. [https://doi.org/10.1016/S0166-4328\(89\)80054-3](https://doi.org/10.1016/S0166-4328(89)80054-3)
- Haxby, J. V., Grady, C. L., Horwitz, B., Ungerleider, L. G., Mishkin, M., Carson, R. E., Herscovitch P., Schapiro M.B. & Rapoport, S. I. (1991). Dissociation of object and spatial visual processing pathways in human extrastriate cortex. *Proceedings of the National Academy of Sciences*, 88(5), 1621-1625. <https://doi.org/10.1073/pnas.88.5.1621>
- Haxby, J. V., Ungerleider, L. G., Clark, V. P., Schouten, J. L., Hoffman, E. A., & Martin, A. (1999). The effect of face inversion on activity in human neural systems for face and object perception. *Neuron*, 22(1), 189-199. [https://doi.org/10.1016/s0896-6273\(00\)80690-x](https://doi.org/10.1016/s0896-6273(00)80690-x)
- Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in cognitive sciences*, 4(6), 223-233. [https://doi.org/10.1016/s1364-6613\(00\)01482-0](https://doi.org/10.1016/s1364-6613(00)01482-0)
- Hayden, A., Bhatt, R. S., Joseph, J. E., & Tanaka, J. W. (2007). The other-race effect in infancy: Evidence using a morphing technique. *Infancy*, 12(1), 95-104. <https://doi.org/10.1111/j.1532-7078.2007.tb00235.x>
- Hecaen, H., & Angelergues, R. (1962). Agnosia for faces (prosopagnosia). *Archives of neurology*, 7(2), 92-100. <https://doi.org/10.1001/archneur.1962.04210020014002>
- Hendrickson, A. (1992). A morphological comparison of foveal development in man and monkey. *Eye*, 6(2), 136-144. <https://doi.org/10.1038/eye.1992.29>
- Hespos SJ, Baillargeon R (2001) Knowledge about containment events in very young infants. *Cognition* 78: 204–245. [https://doi.org/10.1016/s0010-0277\(00\)00118-9](https://doi.org/10.1016/s0010-0277(00)00118-9)

- Heron-Delaney, M., Anzures, G., Herbert, J. S., Quinn, P. C., Slater, A. M., Tanaka, J. W., ... & Pascalis, O. (2011). Perceptual training prevents the emergence of the other race effect during infancy. *PloS one*, 6(5), e19858. <https://doi.org/10.1371/journal.pone.0019858>
- Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a non-human primate: Statistical learning in cotton-top tamarins. *Cognition*, 78(3), B53-B64. [https://doi.org/10.1016/S0010-0277\(00\)00132-3](https://doi.org/10.1016/S0010-0277(00)00132-3)
- Hietanen, J. K., Manninen, P., Sams, M., & Surakka, V. (2001). Does audiovisual speech perception use information about facial configuration?. *European Journal of Cognitive Psychology*, 13(3), 395-407. <https://doi.org/10.1080/09541440126006>
- Hillairet de Boisferon, A., Kubicek, C., Gervain, J., Schwarzer, G., Loevenbruck, H., Vilain, A., ... & Pascalis, O. (2021). Language familiarity influences own-race face recognition in 9-and 12-month-old infants. *Infancy*, 26(4), 647-659. <https://doi.org/10.1111/infa.12404>
- Hillock-Dunn, A., Grantham, D. W., & Wallace, M. T. (2016). The temporal binding window for audiovisual speech: Children are like little adults. *Neuropsychologia*, 88, 74-82. <https://doi.org/10.1016/j.neuropsychologia.2016.02.017>
- Hisanaga, S., Sekiyama, K., Igasaki, T., & Murayama, N. (2016). Language/culture modulates brain and gaze processes in audiovisual speech perception. *Scientific reports*, 6(1), 1-10. <https://doi.org/10.1038/srep35265>
- Hobson, R. P., Ouston, J., & Lee, A. (1988). What's in a face? The case of autism. *British journal of psychology*, 79(4), 441-453. <https://doi.org/10.1111/j.2044-8295.1988.tb02745.x>
- Hoffman, E. A., & Haxby, J. V. (2000). Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nature neuroscience*, 3(1), 80-84. <https://doi.org/10.1038/71152>
- Hubbard, E. M., Brang, D., & Ramachandran, V. S. (2011). The cross-activation theory at 10. *Journal of neuropsychology*, 5(2), 152-177. <https://doi.org/10.1111/j.1748-6653.2011.02014.x>
- Hunter, M. A., Ames, E. W., & Koopman, R. (1983). Effects of stimulus complexity and familiarization time on infant preferences for novel and familiar stimuli. *Developmental Psychology*, 19(3), 338. <https://doi.org/10.1037/0012-1649.19.3.338>
- Huttenlocher, P. R. (1979). Synaptic density in human frontal cortex—developmental changes and effects of aging. *Brain Res*, 163(2), 195-205. [https://doi.org/10.1016/0006-8993\(79\)90349-4](https://doi.org/10.1016/0006-8993(79)90349-4)
- Huttenlocher, P. R., de Courten, C., Garey, L. J., & Van der Loos, H. (1982). Synaptogenesis in human visual cortex—evidence for synapse elimination during normal development. *Neuroscience letters*, 33(3), 247-252. [https://doi.org/10.1016/0304-3940\(82\)90379-2](https://doi.org/10.1016/0304-3940(82)90379-2)
- 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. <https://doi.org/10.1097/01.wnr.0000223387.51704.89>
- Irving, E. L., Steinbach, M. J., Lillakas, L., Babu, R. J., & Hutchings, N. (2006). Horizontal saccade dynamics across the human life span. *Investigative ophthalmology & visual science*, 47(6), 2478-2484. <https://doi.org/10.1167/iovs.05-1311>

- Ishai, A., Ungerleider, L. G., Martin, A., Schouten, J. L., & Haxby, J. V. (1999). Distributed representation of objects in the human ventral visual pathway. *Proceedings of the National Academy of Sciences*, 96(16), 9379-9384. <https://doi.org/10.1073/pnas.96.16.9379>
- Ishai, A., Pessoa, L., Bickle, P. C., & Ungerleider, L. G. (2004). Repetition suppression of faces is modulated by emotion. *Proceedings of the National Academy of Sciences*, 101(26), 9827-9832. <https://doi.org/10.1073/pnas.0403559101>
- Ishai, A., Schmidt, C. F., & Boesiger, P. (2005). Face perception is mediated by a distributed cortical network. *Brain research bulletin*, 67(1-2), 87-93. <https://doi.org/10.1016/j.brainresbull.2005.05.027>
- Itier, R. J., Van Roon, P., & Alain, C. (2011). Species sensitivity of early face and eye processing. *NeuroImage*, 54(1), 705-713. <https://doi.org/10.1016/j.neuroimage.2010.07.031>
- Izard, V., Sann, C., Spelke, E. S., and Streri, A. (2009). Newborn infants perceive abstract numbers. *Proceedings of the National Academy of Science USA*, 106(25), 10382–10385. <https://doi.org/10.1073/pnas.0812142106>
- Jacques, S. L., Weaver, D. R., & Reppert, S. M. (1987). Penetration of light into the uterus of pregnant mammals. *Photochemistry and photobiology*, 45(5), 637-641. <https://doi.org/10.1111/j.1751-1097.1987.tb07391.x>
- Jacques, C., & Rossion, B. (2006). The speed of individual face categorization. *Psychological science*, 17(6), 485-492. <https://doi.org/10.1111/j.1467-9280.2006.01733.x>
- Johnson, E. K., Westrek, E., Nazzi, T., & Cutler, A. (2011). Infant ability to tell voices apart rests on language experience. *Developmental Science*, 14(5), 1002-1011. <https://doi.org/10.1111/j.1467-7687.2011.01052.x>
- Johnson, E. K., Bruggeman, L., & Cutler, A. (2018). Abstraction and the (Misnamed) Language Familiarity Effect. *Cognitive Science*, 42(2), 633-645. <https://doi.org/10.1111/cogs.12520>
- 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. [https://doi.org/10.1016/0010-0277\(91\)90045-6](https://doi.org/10.1016/0010-0277(91)90045-6)
- Jordan, T. R., & Bevan, K. (1997). Seeing and hearing rotated faces: Influences of facial orientation on visual and audiovisual speech recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 23(2), 388. <https://doi.org/10.1037/0096-1523.23.2.388>
- Jordan, T. R., & Sergeant, P. C. (1998). Effects of facial image size on visual and audiovisual speech recognition, in *Hearing by Eye II: Advances in the Psychology of Speech Reading and Audio-Visual Speech*, pp. 155–176.
- Jordan, T.R., Mccotter, M.V. & Thomas, S.M. (2000) Visual and audiovisual speech perception with color and gray-scale facial images. *Perception & Psychophysics*, 62, 1394–1404. <https://doi.org/10.3758/BF03212141>
- Jordan, T. R., & Thomas, S. M. (2001). Effects of horizontal viewing angle on visual and audiovisual speech recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 27(6), 1386–1403. <https://doi.org/10.1037/0096-1523.27.6.1386>

- Kadam, M. A., Orena, A. J., Theodore, R. M., & Polka, L. (2016). Reading ability influences native and non-native voice recognition, even for unimpaired readers. *The Journal of the Acoustical Society of America*, *139*(1), EL6-EL12. <https://doi.org/10.1121/1.4937488>
- Kalin, R., & Rayko, D. S. (1978). Discrimination in Evaluative Judgments against Foreign-Accented Job Candidates. *Psychological Reports*, *43*, 1203-1209. <https://doi.org/10.2466/pr0.1978.43.3f.1203>
- Kamachi, M., Hill, H., Lander, K., & Vatikiotis-Bateson, E. (2003). 'Putting the Face to the Voice': Matching Identity across Modality. *Current Biology*, *13*(19), 1709-1714. <https://doi.org/10.1016/j.cub.2003.09.005>
- Kandel, S., Burfin, S., Méary, D., Ruiz-Tada, E., Costa, A., & Pascalis, O. (2016). The impact of early bilingualism on face recognition processes. *Frontiers in Psychology*, *7*, 1080. <https://doi.org/10.3389/fpsyg.2016.01080>
- Kanwisher N., Chun M.M., McDermott J., & Ledden P.J. (1996). Functional imaging of human visual recognition. *Cognitive Brain Research*, *5*, 55–67. [https://doi.org/10.1016/S0926-6410\(96\)00041-9](https://doi.org/10.1016/S0926-6410(96)00041-9)
- Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of neuroscience*, *17*(11), 4302-4311. <https://doi.org/10.1523/jneurosci.17-11-04302.1997>
- Kanwisher, N., Tong, F., & Nakayama, K. (1998). The effect of face inversion on the human fusiform face area. *Cognition*, *68*(1), B1-B11. [https://doi.org/10.1016/S0010-0277\(98\)00035-3](https://doi.org/10.1016/S0010-0277(98)00035-3)
- Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Gibson, A., Smith, M., Ge, L., & Pascalis, O. (2005). Three month olds, but not newborn, prefer own race face. *Developmental Science*, *8*, 31–36. <https://doi.org/10.1111/j.1467-7687.2005.0434a.x>.
- Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Ge, L., & Pascalis, O. (2007a). The other-race effect develops during infancy: evidence of perceptual narrowing. *Psychological Science*, *18*, 1084–1089. <https://doi.org/10.1111/j.1467-9280.2007.02029.x>
- Kelly, D. J., Liu, S., Ge, L., Quinn, P. C., Slater, A. M., Liu, Q., & Pascalis, O. (2007b). Cross-race preferences for same-race faces extend beyond the African versus Caucasian contrast in 3-month-old infants. *Infancy*, *11*, 87–95. <https://doi.org/10.1080/15250000709336871>
- Kelly, D. J., Liu, S., Lee, K., Quinn, P. C., Pascalis, O., Slater, A. M., & Ge, L. (2009). Development of the other-race effect during infancy: evidence toward universality? *Journal of Experimental Child Psychology*, *104*, 105–114. <https://doi.org/10.1016/j.jecp.2009.01.006>
- Kim, H. I., & Johnson, S. P. (2014). Detecting 'infant-directedness' in face and voice. *Developmental Science*, *17*(4), 621-627. <https://dx.doi.org/10.1111%2Fdesc.12146>
- Kinzler, K. D., Dupoux, E., & Spelke, E. S. (2007). The native language of social cognition. *Proceedings of the National Academy of Sciences*, *104*(30), 12577-12580. <https://doi.org/10.1073/pnas.0705345104>

- Kinzler, Katherine D., Shutts, K., DeJesus, J., & Spelke, E. S. (2009). Accent Trumps Race in Guiding Children's Social Preferences. *Social Cognition*, 27(4), 623-634. <https://doi.org/10.1521/soco.2009.27.4.623>
- Kinzler, Katherine D., & Spelke, E. S. (2011). Do infants show social preferences for people differing in race? *Cognition*, 119(1), 1-9. <https://doi.org/10.1016/j.cognition.2010.10.019>
- Kinzler, Katherine D., & DeJesus, J. M. (2013). Northern = smart and Southern = nice: The development of accent attitudes in the United States. *Quarterly Journal of Experimental Psychology*, 66(6), 1146-1158. <https://doi.org/10.1080/17470218.2012.731695>
- Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, 83(2), B35-B42. [https://doi.org/10.1016/S0010-0277\(02\)00004-5](https://doi.org/10.1016/S0010-0277(02)00004-5)
- Klaver, P., Marcar, V., & Martin, E. (2011). Neurodevelopment of the visual system in typically developing children. *Progress in Brain Research*, 189, 113-136. <https://doi.org/10.1016/B978-0-444-53884-0.00021-X>
- Klein, C., & Foerster, F. (2001). Development of prosaccade and antisaccade task performance in participants aged 6 to 26 years. *Psychophysiology*, 38(2), 179-189. <https://doi.org/10.1017/S0048577201981399>
- Kleinke, C. L. (1986). Gaze and eye contact: a research review. *Psychological bulletin*, 100(1), 78. <https://doi.org/10.1037/0033-2909.100.1.78>
- Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain research reviews*, 29(2-3), 169-195. [https://doi.org/10.1016/S0165-0173\(98\)00056-3](https://doi.org/10.1016/S0165-0173(98)00056-3)
- Knight, B., & Johnston, A. (1997). The role of movement in face recognition. *Visual cognition*, 4(3), 265-273. <https://doi.org/10.1080/713756764>
- Knoblauch, K., Vital-Durand, F., & Barbur, J. L. (2001). Variation of chromatic sensitivity across the life span. *Vision research*, 41(1), 23-36. [https://doi.org/10.1016/S0042-6989\(00\)00205-4](https://doi.org/10.1016/S0042-6989(00)00205-4)
- Kobayashi, H., & Kohshima, S. (1997). Unique morphology of the human eye. *Nature*, 387(6635), 767-768. <https://doi.org/10.1038/42842>
- Kouider, S., Long, B., Le Stanc, L., Charron, S., Fievet, A. C., Barbosa, L. S., & Gelskov, S. V. (2015). Neural dynamics of prediction and surprise in infants. *Nature communications*, 6(1), 1-8. <https://doi.org/10.1038/ncomms9537>
- Kovarski, C., (2014). Les anomalies de la vision chez l'enfant et l'adolescent, (2nd Edition), *Lavoisier*, p.1017.
- Kranz, F., & Ishai, A. (2006). Face perception is modulated by sexual preference. *Current biology*, 16(1), 63-68. <https://doi.org/10.1016/j.cub.2005.10.070>
- Krasotkina, A., Götz, A., Höhle, B., & Schwarzer, G. (2018). Perceptual narrowing in speech and face recognition: evidence for intra-individual cross-domain relations. *Frontiers in Psychology*, 1711. <https://doi.org/10.3389/fpsyg.2018.01711>

- Krasotkina, A., Götz, A., Höhle, B., & Schwarzer, G. (2021). Perceptual narrowing in face-and speech-perception domains in infancy: A longitudinal approach. *Infant Behavior and Development*, 64, 101607. <https://doi.org/10.1016/j.infbeh.2021.101607>
- Kubicek, C., Hillairet de Boisferon, A., Dupierriex, E., Loevenbruck, H., Gervain, J., et al. (2013). Face-scanning behavior to silently-talking faces in 12-month-old infants: The impact of pre-exposed auditory speech. *International Journal of Behavioral Development*, 37, 106–110. <https://doi.org/10.1177/0165025412473016>
- Kubicek, C., Gervain, J., De Boisferon, A. H., Pascalis, O., Løevenbruck, H., & Schwarzer, G. (2014). The influence of infant-directed speech on 12-month-olds' intersensory perception of fluent speech. *Infant Behavior and Development*, 37(4), 644-651. <https://doi.org/10.1016/j.infbeh.2014.08.010>
- Kuhl, P. K. (1979). Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories. *The Journal of the Acoustical Society of America*, 66(6), 1668-1679. <https://doi.org/10.1121/1.383639>
- Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science*, 218(4577), 1138-1141. <https://doi.org/10.1126/science.7146899>
- Kuhl, P. K. (1983). Perception of auditory equivalence classes for speech in early infancy. *Infant behavior and development*, 6(2-3), 263-285. [https://doi.org/10.1016/S0163-6383\(83\)80036-8](https://doi.org/10.1016/S0163-6383(83)80036-8)
- Kuhl, P. K., & Meltzoff, A. N. (1984). The intermodal representation of speech in infants. *Infant behavior and development*, 7(3), 361-381. [https://doi.org/10.1016/S0163-6383\(84\)80050-8](https://doi.org/10.1016/S0163-6383(84)80050-8)
- Kuhl, P. K., & Meltzoff, A. N. (1988). Speech as an intermodal object of perception. In A. Yonas (Ed.), *Perceptual development in infancy: The Minnesota Symposia on Child Psychology*, 20, 235–266). Hillsdale, NJ: Earlbaum.
- Kuhl, P. K., Williams, K. A., & Meltzoff, A. N. (1991). Cross-modal speech perception in adults and infants using nonspeech auditory stimuli. *Journal of Experimental Psychology: Human perception and performance*, 17(3), 829-840. <https://doi.org/10.1037/0096-1523.17.3.829>
- Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, 255(5044), 606-608. <https://doi.org/10.1126/science.1736364>
- Kuhl, P. K. (1994). Learning and representation in speech and language. *Current opinion in neurobiology*, 4(6), 812-822. [https://doi.org/10.1016/0959-4388\(94\)90128-7](https://doi.org/10.1016/0959-4388(94)90128-7)
- Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences*, 97(22), 11850-11857. <https://doi.org/10.1073/pnas.97.22.11850>
- Kuhl, P. K., Tsao, F. M., & Liu, H. M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences*, 100(15), 9096-9101. <https://doi.org/10.1073/pnas.1532872100>
- Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental science*, 9(2), 13-21. <https://doi.org/10.1111/j.1467-7687.2006.00468.x>

- Lany, J., & Gómez, R. L. (2008). Twelve-Month-Old Infants Benefit From Prior Experience in Statistical Learning. *Psychological Science*, 19(12), 1247-1252. <https://doi.org/10.1111/j.1467-9280.2008.02233.x>
- Lebrecht, S., Pierce, L. J., Tarr, M. J., & Tanaka, J. W. (2009). Perceptual other-race training reduces implicit racial bias. *PloS one*, 4(1), e4215. <https://doi.org/10.1371/journal.pone.0004215>
- Lee, B. B. (1996). Receptive field structure in the primate retina. *Vision research*, 36(5), 631-644. [https://doi.org/10.1016/0042-6989\(95\)00167-0](https://doi.org/10.1016/0042-6989(95)00167-0)
- Leleu, A., Rekow, D., Poncet, F., Schaal, B., Durand, K., Rossion, B., & Baudouin, J. Y. (2019). Maternal odor shapes rapid face categorization in the infant brain. *Developmental Science*, 23(2), e12877. <https://doi.org/10.1111/desc.12877>
- Levin, D. T. (1996). Classifying faces by race: The structure of face categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(6), 1364. <https://psycnet.apa.org/doi/10.1037/0278-7393.22.6.1364>
- Lewis, M. B. (2016). Arguing that Black is White: Racial categorization of mixed-race faces. *Perception*, 45(5), 505-514. <https://doi.org/10.1177%2F0301006615624321>
- Lewis, T. L., & Maurer, D. (1992). The development of the temporal and nasal visual fields during infancy. *Vision Research*, 32(5), 903-911. [https://doi.org/10.1016/0042-6989\(92\)90033-F](https://doi.org/10.1016/0042-6989(92)90033-F)
- Lewis, M. B., & Johnston, R. A. (1997). The Thatcher illusion as a test of configural disruption. *Perception*, 26(2), 225-227. <https://doi.org/10.1068/p260225>
- Lewkowicz, D. J. (1986). Developmental changes in infants' bisensory response to synchronous durations. *Infant Behavior and Development*, 9(3), 335-353. [https://doi.org/10.1016/0163-6383\(86\)90008-1](https://doi.org/10.1016/0163-6383(86)90008-1)
- Lewkowicz, D. J. (1988a). Sensory dominance in infants: I. Six-month-old infants' response to auditory-visual compounds. *Developmental Psychology*, 24(2), 155. <https://doi.org/10.1037/0012-1649.24.2.155>
- Lewkowicz, D. J. (1988b). Sensory dominance in infants: II. Ten-month-old infants' response to auditory-visual compounds. *Developmental Psychology*, 24(2), 172. <https://doi.org/10.1037/0012-1649.24.2.172>
- Lewkowicz, D. J. (1996). Perception of auditory-visual temporal synchrony in human infants. *Journal of Experimental Psychology: Human Perception and Performance*, 22(5), 1094. <https://doi.org/10.1037/0096-1523.22.5.1094>
- Lewkowicz, D. J. (2003). Learning and discrimination of audiovisual events in human infants: The hierarchical relation between intersensory temporal synchrony and rhythmic pattern cues. *Developmental psychology*, 39(5), 795. <https://doi.org/10.1037/0012-1649.39.5.795>
- Lewkowicz, D. J., & Ghazanfar, A. A. (2006). The decline of cross-species intersensory perception in human infants. *Proceedings of the National Academy of Sciences*, 103(17), 6771-6774. <https://doi.org/10.1073/pnas.0602027103>

- Lewkowicz, D. J., Sowinski, R., & Place, S. (2008). The decline of cross-species intersensory perception in human infants: Underlying mechanisms and its developmental persistence. *Brain research*, 1242, 291-302. <https://doi.org/10.1016/j.brainres.2008.03.084>
- Lewkowicz, D. J., & Ghazanfar, A. A. (2009). The emergence of multisensory systems through perceptual narrowing. *Trends in cognitive sciences*, 13(11), 470-478. <https://doi.org/10.1016/j.tics.2009.08.004>
- Lewkowicz, D. J. (2010). Infant perception of audio-visual speech synchrony. *Developmental psychology*, 46(1), 66. <https://doi.org/10.1037/a0015579>
- Lewkowicz, D. J., Leo, I., & Simion, F. (2010). Intersensory perception at birth: newborns match nonhuman primate faces and voices. *Infancy*, 15(1), 46-60. <https://doi.org/10.1111/j.1532-7078.2009.00005.x>
- Lewkowicz, D.J., Hansen-Tift, A.M. (2012). Infants deploy selective attention to the mouth of a talking face when learning speech. *Proceedings of the National Academy of Sciences of the United States of America*, 109, 1431–1436. <https://doi.org/10.1073/pnas.1114783109>
- Lewkowicz, D. J., & Pons, F. (2013). Recognition of amodal language identity emerges in infancy. *International Journal of Behavioral Development*, 37(2), 90-94. <https://doi.org/10.1177/0165025412467582>
- Lew-Williams, C., & Saffran, J. R. (2012). All words are not created equal: Expectations about word length guide infant statistical learning. *Cognition*, 122(2), 241-246. <https://doi.org/10.1016/j.cognition.2011.10.007>
- Li, J. C., Dunning, D., & Malpass, R. S. (1998). Cross-racial identification among European-Americans: Basketball fandom and the contact hypothesis. *Unpublished manuscript, Cornell University, Ithaca, NY*. <https://doi.org/10.1111/j.1467-9280.2006.01679.x>
- Lieberman, Z., Woodward, A. L., & Kinzler, K. D. (2017). Preverbal Infants Infer Third-Party Social Relationships Based on Language. *Cognitive Science*, 41, 622-634. <https://doi.org/10.1111/cogs.12403>
- Lieberman, Z., Howard, L. H., Vasquez, N. M., & Woodward, A. L. (2018). Children's expectations about conventional and moral behaviors of ingroup and outgroup members. *Journal of Experimental Child Psychology*, 165, 7-18. <https://doi.org/10.1016/j.jecp.2017.03.003>
- Light, L. L., Kayra-Stuart, F., & Hollander, S. (1979). Recognition memory for typical and unusual faces. *Journal of Experimental Psychology: Human Learning and Memory*, 5(3), 212. <https://doi.org/10.1037/0278-7393.5.3.212>
- Lindsay, D. S., Jack, P. C., & Christian, M. A. (1991). Other-race face perception. *Journal of applied psychology*, 76(4), 587. <https://doi.org/10.1037/0021-9010.76.4.587>
- Lindsay, R. C. L., & Wells, G. L. (1983). What do we really know about cross-race eyewitness identification. *Evaluating witness evidence: Recent psychological research and new perspectives*, 219-233. <https://doi.org/10.1037/0003-066X.48.5.577>

- Lippert, M., Logothetis, N. K., & Kayser, C. (2007). Improvement of visual contrast detection by a simultaneous sound. *Brain research*, *1173*, 102-109. <https://doi.org/10.1016/j.brainres.2007.07.050>
- Liu, J., Higuchi, M., Marantz, A., & Kanwisher, N. (2000). The selectivity of the occipitotemporal M170 for faces. *Neuroreport*, *11*(2), 337-341. <https://doi.org/10.1097/00001756-200002070-00023>
- Liu, L., & Kager, R. (2014). Perception of tones by infants learning a non-tone language. *Cognition*, *133*(2), 385-394. <https://doi.org/10.1016/j.cognition.2014.06.004>
- Liu, L., & Kager, R. (2017) Statistical learning of speech sounds is most robust during the period of perceptual attunement. *Journal of Experimental Child Psychology*. <https://doi.org/10.1016/j.jecp.2017.05.013>
- Liu, S., Quinn, P. C., Wheeler, A., Xiao, N., Ge, L., & Lee, K. (2011). Similarity and difference in the processing of same-and other-race faces as revealed by eye tracking in 4-to 9-month-olds. *Journal of experimental child psychology*, *108*(1), 180-189. <https://doi.org/10.1016/j.jecp.2010.06.008>
- Liu, S., Xiao, W. S., Xiao, N. G., Quinn, P. C., Zhang, Y., Chen, H., ... Lee, K. (2015). Development of visual preference for own- versus other-race faces in infancy. *Developmental Psychology*, *51*, 500-511. <https://doi.org/10.1037/a0038835>
- Livingstone, M., & Hubel, D. (1988). Segregation of form, color, movement, and depth: anatomy, physiology, and perception. *Science*, *240*(4853), 740-749. <https://doi.org/10.1126/science.3283936>
- Lord, C., Rutter, M., & DiLavore, P. C. (1999). Autism Diagnostic Observation Schedule--Generic. *Dissertation Abstracts International Section A: Humanities and Social Sciences*.
- Lore, R., & Flannelly, K. (1977). Rat societies. *Scientific American*, *236*(5), 106-118.
- Lotmar, W. (1976). A theoretical model for the eye of new-born infants. *Albrecht von Graefes Archiv für klinische und experimentelle Ophthalmologie*, *198*(2), 179-185. <https://doi.org/10.1007/BF00410018>
- MacKain, K., Studdert-Kennedy, M., Spieker, S., & Stern, D. (1983). Infant intermodal speech perception is a left-hemisphere function. *Science*, *219*(4590), 1347-1349. <https://doi.org/10.1126/science.6828865>
- Magoon, E. H., & Robb, R. M. (1981). Development of myelin in human optic nerve and tract: a light and electron microscopic study. *Archives of ophthalmology*, *99*(4), 655-659. <https://doi.org/10.1001/archoph.1981.03930010655011>
- Malatesta, C. Z., & Haviland, J. M. (1982). Learning display rules: The socialization of emotion expression in infancy. *Child development*, *53*(4), 991-1003. <https://doi.org/10.2307/1129139>
- Malpass, R. S., & Kravitz, J. (1969). Recognition for faces of own and other race. *Journal of personality and social psychology*, *13*(4), 330. <https://doi.org/10.1037/h0028434>

- Malpass, R. S. (1981). Effective size and defendant bias in eyewitness identification lineups. *Law and Human behavior*, 5(4), 299. <https://doi.org/10.1007/BF01044945>
- Marean, G. C., Werner, L. A., & Kuhl, P. K. (1992). Vowel categorization by very young infants. *Developmental Psychology*, 28(3), 396. <https://doi.org/10.1037/0012-1649.28.3.396>
- Markant, J., Oakes, L. M., & Amso, D. (2016). Visual selective attention biases contribute to the other-race effect among 9-month-old infants. *Developmental psychobiology*, 58(3), 355-365. <https://doi.org/10.1002/dev.21375>
- Marler, P., & Tamura, M. (1964). Culturally transmitted patterns of vocal behavior in sparrows. *Science*, 146(3650), 1483-1486. <https://doi.org/10.1126/science.146.3650.1483>
- Marno, H., Guellai, B., Vidal, Y., Franzoi, J., Nespor, M., & Mehler, J. (2016). Infants' selectively pay attention to the information they receive from a native speaker of their language. *Frontiers in Psychology*, 7, 1150. <https://doi.org/10.3389/fpsyg.2016.01150>
- Marr, D., & Poggio, T. (1979). A computational theory of human stereo vision. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 204(1156), 301-328. <https://doi.org/10.1098/rspb.1979.0029>
- Massaro, D. W., & Cohen, M. M. (1993). Perceiving asynchronous bimodal speech in consonant-vowel and vowel syllables. *Speech Communication*, 13(1-2), 127-134. [https://doi.org/10.1016/0167-6393\(93\)90064-R](https://doi.org/10.1016/0167-6393(93)90064-R)
- Massaro, D. W., & Cohen, M. M. (1996). Perceiving speech from inverted faces. *Perception & Psychophysics*, 58(7), 1047-1065. <https://doi.org/10.3758/BF03206832>
- 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>
- Maurer, D., Ghouloum, J. K., Gibson, L. C., Watson, M. R., Chen, L. M., Akins, K., ... & Werker, J. F. (2020). Reduced perceptual narrowing in synesthesia. *Proceedings of the National Academy of Sciences*, 117(18), 10089-10096. <https://doi.org/10.1073/pnas.1914668117>
- Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, 82(3), B101-B111. [https://doi.org/10.1016/S0010-0277\(01\)00157-3](https://doi.org/10.1016/S0010-0277(01)00157-3)
- Mayer, D. L., Fulton, A. B., & Cummings, M. F. (1988). Visual fields of infants assessed with a new perimetric technique. *Investigative ophthalmology & visual science*, 29(3), 452-459.
- McCarthy, G., Puce, A., Gore, J. C., & Allison, T. (1997). Face-specific processing in the human fusiform gyrus. *Journal of cognitive neuroscience*, 9(5), 605-610. <https://doi.org/10.1162/jocn.1997.9.5.605>
- McCubbin, J., Murphy, P., Eswaran, H., Preissl, H., Yee, T., Robinson, S. E., & Vrba, J. (2007). Validation of the flash-evoked response from fetal MEG. *Physics in Medicine & Biology*, 52(19), 5803. <https://doi.org/10.1088/0031-9155/52/19/005>

- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264(5588), 746-748. <https://doi.org/10.1038/264746a0>
- McKone, E., Brewer, J. L., MacPherson, S., Rhodes, G., & Hayward, W. G. (2007). Familiar other-race faces show normal holistic processing and are robust to perceptual stress. *Perception*, 36(2), 224-248. <https://doi.org/10.1068/p5499>
- Méary, D., Kitromilides, E., Mazens, K., Graff, C., & Gentaz, E. (2007). Four-day-old human neonates look longer at non-biological motions of a single point-of-light. *PloS one*, 2(1), e186. <https://doi.org/10.1371/journal.pone.0000186>
- 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. [https://doi.org/10.1016/0010-0277\(88\)90035-2](https://doi.org/10.1016/0010-0277(88)90035-2)
- Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law*, 7(1), 3. <https://doi.org/10.1037/1076-8971.7.1.3>
- Ménard, L., Dupont, S., Baum, S. R., & Aubin, J. (2009). Production and perception of French vowels by congenitally blind adults and sighted adults. *The Journal of the Acoustical Society of America*, 126(3), 1406-1414. <https://doi.org/10.1121/1.3158930>
- Mercure, E., Quiroz, I., Goldberg, L., Bowden-Howl, H., Coulson, K., Gliga, T., ... & MacSweeney, M. (2018). Impact of language experience on attention to faces in infancy: Evidence from unimodal and bimodal bilingual infants. *Frontiers in psychology*, 9, 1943. <https://doi.org/10.3389/fpsyg.2018.01943>
- Merigan, W. H., & Maunsell, J. H. (1990). Macaque vision after magnocellular lateral geniculate lesions. *Visual neuroscience*, 5(4), 347-352. <https://doi.org/10.1017/S0952523800000432>
- Merigan, W. H., Byrne, C. E., & Maunsell, J. H. (1991). Does primate motion perception depend on the magnocellular pathway?. *Journal of Neuroscience*, 11(11), 3422-3429. <https://doi.org/10.1523/JNEUROSCI.11-11-03422.1991>
- Minar, N. J., & Lewkowicz, D. J. (2017). Overcoming the other-race effect in infancy with multisensory redundancy: 10–12-month-olds discriminate dynamic other-race faces producing speech. *Developmental science*, 21(4), e12604. <https://doi.org/10.1111/desc.12604>
- Miner, L. E. (1963). A study of the incidence of speech deviations among visually handicapped children. *Journal of Visual Impairment & Blindness*, 57(1), 10-14. <https://doi.org/10.1177/0145482X6305700103>
- 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. <https://doi.org/10.3758/BF03211209>
- Moon, C., Cooper, R. P., & Fifer, W. P. (1993). Two-day-olds prefer their native language. *Infant behavior and development*, 16(4), 495-500. [https://doi.org/10.1016/0163-6383\(93\)80007-U](https://doi.org/10.1016/0163-6383(93)80007-U)

- Morin-Lessard, E., Poulin-Dubois, D., Segalowitz, N., & Byers-Heinlein, K. (2019). Selective attention to the mouth of talking faces in monolinguals and bilinguals aged 5 months to 5 years. *Developmental psychology*, 55(8), 1640. <https://doi.org/10.1037/dev0000750>
- Morisawa, N., Hakuno, Y., & Minagawa, Y. (2013). Nyujionsei chosyu ni okeru shichokaku jyoho riyo no hattatsu henka: Hahaoya to tasha no baai [developmental changes in infants use of audio-visual information for speech produced by mother and strangers]. *Journal of the Phonetic Sociology of Japan*, 17, 77e85.
- Morris, J. S., Friston, K. J., Büchel, C., Frith, C. D., Young, A. W., Calder, A. J., & Dolan, R. J. (1998). A neuromodulatory role for the human amygdala in processing emotional facial expressions. *Brain: a journal of neurology*, 121(1), 47-57. <https://doi.org/10.1093/brain/121.1.47>
- Morton, J., & Johnson, M. H. (1991). CONSPEC and CONLERN: a two-process theory of infant face recognition. *Psychological review*, 98(2), 164-181. <https://doi.org/10.1037/0033-295X.98.2.164>
- Moscovitch, M., Winocur, G., & Behrmann, M. (1997). What is special about face recognition? Nineteen experiments on a person with visual object agnosia and dyslexia but normal face recognition. *Journal of cognitive neuroscience*, 9(5), 555-604. <https://doi.org/10.1162/jocn.1997.9.5.555>
- Mower, G. D., Berry, D., Burchfiel, J. L., & Duffy, F. H. (1981). Comparison of the effects of dark rearing and binocular suture on development and plasticity of cat visual cortex. *Brain research*, 220(2), 255-267. [https://doi.org/10.1016/0006-8993\(81\)91216-6](https://doi.org/10.1016/0006-8993(81)91216-6)
- Muchnik, C., Efrati, M., Nemeth, E., Malin, M., & Hildesheimer, M. (1991). Central auditory skills in blind and sighted subjects. *Scandinavian audiology*, 20(1), 19-23. <https://doi.org/10.3109/01050399109070785>
- Mumford, D. (1994). Neuronal architectures for pattern-theoretic problems. *Large-scale neuronal theories of the brain*, 125-152.
- Munhall, K. G., Jones, J. A., Callan, D. E., Kuratate, T., & Vatikiotis-Bateson, E. (2004). Visual prosody and speech intelligibility: Head movement improves auditory speech perception. *Psychological science*, 15(2), 133-137. <https://doi.org/10.1111/j.0963-7214.2004.01502010.x>
- Munoz, D. P., Broughton, J. R., Goldring, J. E., & Armstrong, I. T. (1998). Age-related performance of human subjects on saccadic eye movement tasks. *Experimental brain research*, 121(4), 391-400. <https://doi.org/10.1007/s002210050473>
- Mugitani, R., & Kobayashi, T. (2003). The role of speech production experience in 8-month-old infants' lip-voice matching. In *Poster session presented at the XIth European Conference on Developmental Psychology*
- Mugitani, R., Kobayashi, T., & Hiraki, K. (2004). Nihonjin nyuji ni okeru boin no shichokakuonsei koukei matching no kentoh [Audiovisual Lip-Voice Matching on Vowels by Japanese Infant]. *Journal of the Phonetic Sociology of Japan*, 8, 85e95.

- Mugitani, R., Kobayashi, T., Ishizuka, K., Amano, S., & Hiraki, K. (2006). Nihongogakusyunyujji no onseikoukei matching no hattatsu ni kansuru boin/i/wo mochiita kentoh [Audiovisual Matching in Lips and Voice on Vowel/i/by Japanese-Learning infants]. *Journal of the Phonetic Society of Japan*, 10, 96e108.
- Mugitani, R., Kobayashi, T., & Hiraki, K. (2008). Audiovisual matching of lips and non-canonical sounds in 8-month-old infants. *Infant Behavior and Development*, 31(2), 307-310. <https://doi.org/10.1016/j.infbeh.2007.12.002>
- 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. <https://doi.org/10.1111/j.1467-7687.2009.00898.x>
- Navarra, J., & Soto-Faraco, S. (2007). Hearing lips in a second language: visual articulatory information enables the perception of second language sounds. *Psychological research*, 71(1), 4-12. <https://doi.org/10.1007/s00426-005-0031-5>
- Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human perception and performance*, 24(3), 756. <https://doi.org/10.1037/0096-1523.24.3.756>
- 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. <https://doi.org/10.1006/jmla.2000.2698>
- Nelson, D. A. (2000). A preference for own-subspecies' song guides vocal learning in a song bird. *Proceedings of the National Academy of Sciences*, 97(24), 13348-13353. <https://doi.org/10.1073/pnas.240457797>
- Newport, E. L., & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive psychology*, 48(2), 127-162. [https://doi.org/10.1016/S0010-0285\(03\)00128-2](https://doi.org/10.1016/S0010-0285(03)00128-2)
- Newport, E. L., Hauser, M. D., Spaepen, G., & Aslin, R. N. (2004). Learning at a distance II. Statistical learning of non-adjacent dependencies in a non-human primate. *Cognitive psychology*, 49(2), 85-117. <https://doi.org/10.1016/j.cogpsych.2003.12.002>
- Ng, W. J., & Lindsay, R. C. (1994). Cross-race facial recognition: Failure of the contact hypothesis. *Journal of Cross-Cultural Psychology*, 25(2), 217-232. <https://doi.org/10.1177/0022022194252004>
- Nomura, M. (1984). *Body language wo yomu* [Understanding body language]. Tokyo: Heibonsha
- Nowak, L. G., & Bullier, J. (1997). The timing of information transfer in the visual system. *Extrastriate Cortex in Primates*, 12, 205-241. Springer, Boston, MA. https://doi.org/10.1007/978-1-4757-9625-4_5
- O'Connor, C., & Joffe, H. (2020). Intercoder reliability in qualitative research: debates and practical guidelines. *International Journal of Qualitative Methods*, 19. <https://doi.org/10.1177/1609406919899220>

- Odom J.V., Green M. (1984). Visually evoked potential (VEP) acuity: test-ability in a clinical pediatric population. *Acta Ophthalmologica*, 62(6), 993-998. <https://doi.org/10.1111/j.1755-3768.1984.tb08451.x>
- O'Neill, J. J. (1954). Contributions of the visual components of oral symbols to speech comprehension. *Journal of Speech and Hearing Disorders*, 19(4), 429-439. <https://doi.org/10.1044/jshd.1904.429>
- Ong, J. H., Burnham, D., & Escudero, P. (2015). Distributional learning of lexical tones: A comparison of attended vs. unattended listening. *PloS one*, 10(7). <https://doi.org/10.1371/journal.pone.0133446>
- Otsuka, Y., Konishi, Y., Kanazawa, S., Yamaguchi, M. K., Abdi, H., & O'Toole, A. J. (2009). Recognition of moving and static faces by young infants. *Child development*, 80(4), 1259-1271. <https://doi.org/10.1111/j.1467-8624.2009.01330.x>
- Palmer, S. B., Fais, L., Golinkoff, R. M., & Werker, J. F. (2012). Perceptual narrowing of linguistic sign occurs in the 1st year of life. *Child development*, 83(2), 543-553. <https://doi.org/10.1111/j.1467-8624.2011.01715.x>
- Parraguez, V. H., Sales, F., Valenzuela, G. J., Vergara, M., Catalán, L., & Serón-Ferré, M. (1998). Diurnal changes in light intensity inside the pregnant uterus in sheep. *Animal reproduction science*, 52(2), 123-130. [https://doi.org/10.1016/s0378-4320\(98\)00094-3](https://doi.org/10.1016/s0378-4320(98)00094-3)
- Pascalis, O., de Schonen, S., Morton, J., Deruelle, C., & Fabre-Grenet, M. (1995). Mother's face recognition by neonates: A replication and an extension. *Infant Behavior and Development*, 18(1), 79-85. [https://doi.org/10.1016/0163-6383\(95\)90009-8](https://doi.org/10.1016/0163-6383(95)90009-8)
- Pascalis, O., & Bachevalier, J. (1998). Face recognition in primates: a cross-species study. *Behavioural processes*, 43(1), 87-96. [https://doi.org/10.1016/S0376-6357\(97\)00090-9](https://doi.org/10.1016/S0376-6357(97)00090-9)
- Pascalis, O., De Haan, M., Nelson, C. A., & De Schonen, S. (1998). Long-term recognition memory for faces assessed by visual paired comparison in 3- and 6-month-old infants. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24(1), 249. <https://doi.org/10.1037/0278-7393.24.1.249>
- Pascalis, O., de Haan, M., & Nelson, C. A. (2002). Is face processing species-specific during the first year of life? *Science*, 296, 1321-1323. <https://doi.org/10.1126/science.1070223>
- Pascalis, O., Scott, L. S., Kelly, D. J., Shannon, R. W., Nicholson, E., Coleman, M., & Nelson, C. A. (2005). Plasticity of face processing in infancy. *Proceedings of the National Academy of Sciences of the United States of America*, 102, 5297-5300. <https://doi.org/10.1073/pnas.0406627102>
- Pascalis, O., & Kelly, D. J. (2009). The origins of face processing in humans: Phylogeny and ontogeny. *Perspectives on Psychological Science*, 4(2), 200-209. <https://doi.org/10.1111/j.1745-6924.2009.01119.x>
- Pascalis, O., Loevenbruck, H., Quinn, P. C., Kandel, S., Tanaka, J. W., & Lee, K. (2014). On the links among face processing, language processing, and narrowing during development. *Child development perspectives*, 8(2), 65-70. <https://doi.org/10.1111/cdep.12064>

- Pascalis, O., Dole, M., & Løevenbruck, H. (2017). More evidence of the linkage between face processing and language processing. *British Journal of Psychology*, *108*(1), 31-33. <https://doi.org/10.1111/bjop.12228>
- Patterson, M. L., & Werker, J. F. (1999). Matching phonetic information in lips and voice is robust in 4.5-month-old infants. *Infant Behavior and Development*, *22*(2), 237-247. [https://doi.org/10.1016/S0163-6383\(99\)00003-X](https://doi.org/10.1016/S0163-6383(99)00003-X)
- Patterson, M. L., & Werker, J. F. (2002). Infants' ability to match dynamic phonetic and gender information in the face and voice. *Journal of experimental child psychology*, *81*(1), 93-115. <https://doi.org/10.1006/jecp.2001.2644>
- Patterson, M. L., & Werker, J. F. (2003). Two-month-old infants match phonetic information in lips and voice. *Developmental Science*, *6*(2), 191-196. <https://doi.org/10.1111/1467-7687.00271>
- Peelle, J. E., & Sommers, M. S. (2015). Prediction and constraint in audiovisual speech perception. *Cortex*, *68*, 169-181. <https://doi.org/10.1016/j.cortex.2015.03.006>
- Peery, D., & Bodenhausen, G. V. (2008). Black+ White= Black: Hypodescent in reflexive categorization of racially ambiguous faces. *Psychological Science*, *19*(10), 973-977. <https://doi.org/10.1111%2Fj.1467-9280.2008.02185.x>
- Peleg, D., & Goldman, J. A. (1980). Fetal heart rate acceleration in response to light Stimulation as a clinical measure of fetal well-being. A preliminary report. *J. Perinat. Med*, *8*, 38. <https://doi.org/10.1515/jpme.1980.8.1.38>
- Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Statistical learning in a natural language by 8-month-old infants. *Child development*, *80*(3), 674-685. <https://doi.org/10.1111/j.1467-8624.2009.01290.x>
- Perrachione, T. K., Del Tufo, S. N., & Gabrieli, J. D. E. (2011). Human Voice Recognition Depends on Language Ability. *Science*, *333*(6042), 595-595. <https://doi.org/10.1126/science.1207327>
- Perrachione, Tyler K., & Wong, P. C. M. (2007). Learning to recognize speakers of a non-native language: Implications for the functional organization of human auditory cortex. *Neuropsychologia*, *45*(8), 1899-1910. <https://doi.org/10.1016/j.neuropsychologia.2006.11.015>
- Perrett, D. I., Rolls, E. T., & Caan, W. (1982). Visual neurones responsive to faces in the monkey temporal cortex. *Experimental brain research*, *47*(3), 329-342. <https://doi.org/10.1007/BF00239352>
- Perrett, D. I., Oram, M. W., Harries, M. H., Bevan, R., Hietanen, J. K., Benson, P. J., & Thomas, S. (1991). Viewer-centred and object-centred coding of heads in the macaque temporal cortex. *Experimental brain research*, *86*(1), 159-173. <https://doi.org/10.1007/BF00231050>
- Petitto, L. A., Berens, M. S., Kovelman, I., Dubins, M. H., Jasinska, K., & Shalinsky, M. (2012). The “Perceptual Wedge Hypothesis” as the basis for bilingual babies’ phonetic processing advantage: New insights from fNIRS brain imaging. *Brain and language*, *121*(2), 130-143. <https://doi.org/10.1016/j.bandl.2011.05.003>

- Pickens, J., Field, T., Nawrocki, T., Martinez, A., Soutullo, D., & Gonzalez, J. (1994). Full-term and preterm infants' perception of face-voice synchrony. *Infant Behavior and Development*, *17*(4), 447-455. [https://doi.org/10.1016/0163-6383\(94\)90036-1](https://doi.org/10.1016/0163-6383(94)90036-1)
- Pinel, P., Lalanne, C., Bourgeron, T., Fauchereau, F., Poupon, C., Artiges, E., ... & Dehaene, S. (2015). Genetic and environmental influences on the visual word form and fusiform face areas. *Cerebral Cortex*, *25*(9), 2478-2493. <https://doi.org/10.1093/cercor/bhu048>
- Pitcher, D., Walsh, V., & Duchaine, B. (2011). The role of the occipital face area in the cortical face perception network. *Experimental brain research*, *209*(4), 481-493. <https://doi.org/10.1007/s00221-011-2579-1>
- Poncelet, J., Rinck, F., Bourgeat, F., Schaal, B., Rouby, C., Bensafi, M., & Hummel, T. (2010). The effect of early experience on odor perception in humans: psychological and physiological correlates. *Behavioural brain research*, *208*(2), 458-465. <https://doi.org/10.1016/j.bbr.2009.12.011>
- Pons, F., Lewkowicz, D. J., Soto-Faraco, S., & Sebastián-Gallés, N. (2009). Narrowing of intersensory speech perception in infancy. *Proceedings of the National Academy of Sciences*, *106*(26), 10598-10602. <https://doi.org/10.1073/pnas.0904134106>
- Puce, A., Allison, T., Asgari, M., Gore, J. C., & McCarthy, G. (1996). Differential sensitivity of human visual cortex to faces, letterstrings, and textures: a functional magnetic resonance imaging study. *Journal of neuroscience*, *16*(16), 5205-5215. <https://doi.org/10.1523/JNEUROSCI.16-16-05205.1996>
- Press, C., Taylor-Clarke, M., Kennett, S., & Haggard, P. (2004). Visual enhancement of touch in spatial body representation. *Experimental brain research*, *154*(2), 238-245. <https://doi.org/10.1007/s00221-003-1651-x>
- Proverbio, A. M., Massetti, G., Rizzi, E., & Zani, A. (2016). Skilled musicians are not subject to the McGurk effect. *Scientific reports*, *6*(1), 1-11. <https://doi.org/10.1038/srep30423>
- Puce, A., Syngeniotis, A., Thompson, J. C., Abbott, D. F., Wheaton, K. J., & Castiello, U. (2003). The human temporal lobe integrates facial form and motion: evidence from fMRI and ERP studies. *Neuroimage*, *19*(3), 861-869. [https://doi.org/10.1016/S1053-8119\(03\)00189-7](https://doi.org/10.1016/S1053-8119(03)00189-7)
- Purves, D., White, L. E., & Riddle, D. R. (1996). Is neural development Darwinian?. *Trends in neurosciences*, *19*(11), 460-464. [https://doi.org/10.1016/S0166-2236\(96\)20038-4](https://doi.org/10.1016/S0166-2236(96)20038-4)
- Qian, M. K., Quinn, P. C., Heyman, G. D., Pascalis, O., Fu, G., & Lee, K. (2017). Perceptual individuation training (but not mere exposure) reduces implicit racial bias in preschool children. *Developmental Psychology*, *53*(5), 845. <https://doi.org/10.1037/dev0000290>
- Quartz, S. R., & Sejnowski, T. J. (1997). The neural basis of cognitive development: A constructivist manifesto. *Behavioral and brain sciences*, *20*(4), 537-556. <https://doi.org/10.1017/s0140525x97001581>
- Quinn, P. C., Uttley, L., Lee, K., Gibson, A., Smith, M., Slater, A. M., & Pascalis, O. (2008). Infant preference for female faces occurs for same- but not other-race faces. *Journal of Neuropsychology*, *2*, 15-26. <https://doi.org/10.1348/174866407X231029>

- Quinn, P. C., Lee, K., Pascalis, O., & Tanaka, J. W. (2015). Narrowing in categorical responding to other-race face classes by infants. *Developmental Science*, 1–10. <https://doi.org/10.1111/desc.12301>
- Quinn, P. C., Lee, K., Pascalis, O., & Xiao, N. G. (2020). Emotional expressions reinstate recognition of other-race faces in infants following perceptual narrowing. *Developmental psychology*, 56(1), 15. <https://doi.org/10.1037/dev0000858>
- Rakic, P. (1995). A small step for the cell, a giant leap for mankind: a hypothesis of neocortical expansion during evolution. *Trends in neurosciences*, 18(9), 383-388. [https://doi.org/10.1016/0166-2236\(95\)93934-P](https://doi.org/10.1016/0166-2236(95)93934-P)
- Ramon-Casas, M., Swingle, D., Sebastián-Gallés, N., & Bosch, L. (2009). Vowel categorization during word recognition in bilingual toddlers. *Cognitive psychology*, 59(1), 96-121. <https://doi.org/10.1016/j.cogpsych.2009.02.002>
- Ramus, F., Nespore, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 73(3), 265-92. [https://doi.org/10.1016/S0010-0277\(99\)00058-X](https://doi.org/10.1016/S0010-0277(99)00058-X)
- Rennels, J. L., & Davis, R. E. (2008). Facial experience during the first year. *Infant Behavior and Development*, 31(4), 665-678. <https://doi.org/10.1016/j.infbeh.2008.04.009>
- Rhodes, G., Brennan, S., & Carey, S. (1987). Identification and ratings of caricatures: Implications for mental representations of faces. *Cognitive psychology*, 19(4), 473-497. [https://doi.org/10.1016/0010-0285\(87\)90016-8](https://doi.org/10.1016/0010-0285(87)90016-8)
- Rhodes, G., Brake, S., Taylor, K., & Tan, S. (1989). Expertise and configural coding in face recognition. *British journal of psychology*, 80(3), 313-331. <https://doi.org/10.1111/j.2044-8295.1989.tb02323.x>
- Rhodes, G., & Moody, J. (1990). Memory representations of unfamiliar faces: Coding of distinctive information. *New Zealand Journal of Psychology*, 19(2), 70-78.
- Rhodes, G., Ronke, K., & Tang, S. (1990). Asymmetries in face perception: Component processes, face specificity and expertise effects. *Cortex*, 26(1), 13-32. [https://doi.org/10.1016/S0010-9452\(13\)80072-X](https://doi.org/10.1016/S0010-9452(13)80072-X)
- Rhodes, G., Byatt, G., Tremewan, T., & Kennedy, A. (1997). Facial distinctiveness and the power of caricatures. *Perception*, 26(2), 207-223. <https://doi.org/10.1068/p260207>
- 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. <https://doi.org/10.1111/j.1467-7687.2005.00403.x>
- Rizzolatti, G., & Buchtel, H. A. (1977). Hemispheric superiority in reaction time to faces: a sex difference. *Cortex*, 13(3), 300-305. [https://doi.org/10.1016/S0010-9452\(77\)80039-7](https://doi.org/10.1016/S0010-9452(77)80039-7)
- Robbins, R., & McKone, E. (2007). No face-like processing for objects-of-expertise in three behavioural tasks. *Cognition*, 103(1), 34-79. <https://doi.org/10.1016/j.cognition.2006.02.008>
- Rosenblum, L. D., & Saldaña, H. M. (1992). Discrimination tests of visually influenced syllables. *Perception & psychophysics*, 52(4), 461-473. <https://doi.org/10.3758/BF03206706>

- Rosenblum, L. D., Schmuckler, M. A., & Johnson, J. A. (1997). The McGurk effect in infants. *Perception & psychophysics*, 59(3), 347-357. <https://doi.org/10.3758/BF03211902>
- Rosenblum, L. D., Yakel, D. A., & Green, K. P. (2000). Face and mouth inversion effects on visual and audiovisual speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, 26(2), 806–819. <https://doi.org/10.1037/0096-1523.26.2.806>
- Rosenblum, L. D. (2008). Speech Perception as a Multimodal Phenomenon. *Current Directions in Psychological Science*, 17(6), 405-409. <https://doi.org/10.1111/j.1467-8721.2008.00615.x>
- Rossion, B. (2008). Picture-plane inversion leads to qualitative changes of face perception. *Acta psychologica*, 128(2), 274-289. <https://doi.org/10.1016/j.actpsy.2008.02.003>
- Rossion, B. & Retter, T.L. (2020). Face Perception. In: D. Poeppel, M.S. Gazzaniga & G. Mangun (Eds). *The Cognitive Neurosciences*, Sixth Edition. The MIT Press, ch.11, 129-139.
- Rotshtein, P., Vuilleumier, P., Winston, J., Driver, J., & Dolan, R. (2007). Distinct and convergent visual processing of high and low spatial frequency information in faces. *Cerebral Cortex*, 17(11), 2713-2724. <https://doi.org/10.1093/cercor/bhl180>
- Rouger, J., Lagleyre, S., Fraysse, B., Deneve, S., Deguine, O., & Barone, P. (2007). Evidence that cochlear-implanted deaf patients are better multisensory integrators. *Proceedings of the National Academy of Sciences*, 104(17), 7295-7300. <https://doi.org/10.1073/pnas.0609419104>
- Ryan, E.B., Carranza, M. A., & Moffie, R. W. (1977). Reactions Toward Varying Degrees of Accentedness in the Speech of Spanish-English Bilinguals. *Language and Speech*, 20(3), 267-273. <https://doi.org/10.1177/002383097702000308>
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274(5294), 1926-1928. <https://doi.org/10.1126/science.274.5294.1926>
- Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological science*, 8(2), 101-105. <https://doi.org/10.1111%2Fj.1467-9280.1997.tb00690.x>
- Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, 70(1), 27-52. [https://doi.org/10.1016/S0010-0277\(98\)00075-4](https://doi.org/10.1016/S0010-0277(98)00075-4)
- Sai, F. Z. (2005). The role of the mother's voice in developing mother's face preference: Evidence for intermodal perception at birth. *Infant and Child Development*, 14(1), 29-50. <https://doi.org/10.1002/icd.376>
- Salapatek, P., Aslin, R. N., Simonson, J., & Pulos, E. (1980). Infant saccadic eye movements to visible and previously visible targets. *Child Development*, 1090-1094. <https://doi.org/10.2307/1129548>
- Salman, M. S., Sharpe, J. A., Eizenman, M., Lillakas, L., To, T., Westall, C., To, T., Dennis, M., & Steinbach, M. J. (2006). Saccadic in children. *Journal of Child Neurology*, 21(12), 1025-1031. <https://doi.org/10.1016/j.visres.2005.06.011>

- Sangrigoli, S., & De Schonen, S. (2004). Recognition of own-race and other-race faces by three-month-old infants. *Journal of Child Psychology and Psychiatry*, 45(7), 1219-1227. <https://doi.org/10.1111/j.1469-7610.2004.00319.x>
- Sangrigoli, S., Pallier, C., Argenti, A. M., Ventureyra, V. A. G., & de Schonen, S. (2005). Reversibility of the other-race effect in face recognition during childhood. *Psychological Science*, 16(6), 440-444.
- Sato, Y., Sogabe, Y., & Mazuka, R. (2010). Discrimination of phonemic vowel length by Japanese infants. *Developmental Psychology*, 46(1), 106. <https://doi.org/10.1037/a0016718>
- Schaal, B., Marlier, L., & Soussignan, R. (2000). Human foetuses learn odours from their pregnant mother's diet. *Chemical senses*, 25(6), 729-737. <https://doi.org/10.1093/chemse/25.6.729>
- Schatz, T., Feldman, N. H., Goldwater, S., Cao, X. N., & Dupoux, E. (2021). Early phonetic learning without phonetic categories: Insights from large-scale simulations on realistic input. *Proceedings of the National Academy of Sciences*, 118(7). <https://doi.org/10.1073/pnas.2001844118>
- Schmolsky, M. T., Wang, Y., Hanes, D. P., Thompson, K. G., Leutgeb, S., Schall, J. D., & Leventhal, A. G. (1998). Signal timing across the macaque visual system. *Journal of neurophysiology*, 79(6), 3272-3278. <https://doi.org/10.1152/jn.1998.79.6.3272>
- Schwartz, J. L., Berthommier, F., & Savariaux, C. (2004). Seeing to hear better: evidence for early audio-visual interactions in speech identification. *Cognition*, 93(2), 69-78. <https://doi.org/10.1016/j.cognition.2004.01.006>
- Scott, L. S., Pascalis, O., & Nelson, C. A. (2007). A domain-general theory of the development of perceptual discrimination. *Current directions in psychological science*, 16(4), 197-201. <https://doi.org/10.1111/j.1467-8721.2007.00503.x>
- Scott, L. S., & Monesson, A. (2009). The origin of biases in face perception. *Psychological Science*, 20(6), 676-680. <https://doi.org/10.1111/j.1467-9280.2009.02348.x>
- Searcy, J. H., & Bartlett, J. C. (1996). Inversion and processing of component and spatial-relational information in faces. *Journal of Experimental Psychology: Human Perception and Performance*, 22(4), 904. <https://doi.org/10.1037/0096-1523.22.4.904>
- Sebastián-Gallés, N., Albareda-Castellot, B., Weikum, W. M., & Werker, J. F. (2012). A bilingual advantage in visual language discrimination in infancy. *Psychological Science*, 23(9), 994-999. <https://doi.org/10.1177/0956797612436817>
- Seigneuric, A., Durand, K., Jiang, T., Baudouin, J. Y., & Schaal, B. (2010). The nose tells it to the eyes: crossmodal associations between olfaction and vision. *Perception*, 39(11), 1541-1554. <https://doi.org/10.1068/p6740>
- Sekiyama, K., & Tohkura, Y. I. (1991). McGurk effect in non-English listeners: Few visual effects for Japanese subjects hearing Japanese syllables of high auditory intelligibility. *The Journal of the Acoustical Society of America*, 90(4), 1797-1805. <https://doi.org/10.1121/1.401660>

- Sekiyama, K., & Tohkura, Y. I. (1993). Inter-language differences in the influence of visual cues in speech perception. *Journal of Phonetics*, 21(4), 427-444. [https://doi.org/10.1016/S0095-4470\(19\)30229-3](https://doi.org/10.1016/S0095-4470(19)30229-3)
- Sekiyama, K. (1994). Differences in auditory-visual speech perception between Japanese and Americans: McGurk effect as a function of incompatibility. *Journal of the Acoustical Society of Japan (E)*, 15(3), 143-158. <https://doi.org/10.1250/ast.15.143>
- Sekiyama, K. (1997). Cultural and linguistic factors in audiovisual speech processing: The McGurk effect in Chinese subjects. *Perception & Psychophysics*, 59(1), 73-80. <https://doi.org/10.3758/BF03206849>
- Sekiyama, K., & Burnham, D. (2008). Impact of language on development of auditory-visual speech perception. *Developmental science*, 11(2), 306-320. <https://doi.org/10.1111/j.1467-7687.2008.00677.x>
- Sekiyama, K., Hisanaga, S., & Mugitani, R. (2021). Selective attention to the mouth of a talker in Japanese-learning infants and toddlers: Its relationship with vocabulary and compensation for noise. *Cortex*, 140, 145-156. <https://doi.org/10.1016/j.cortex.2021.03.023>
- Sergent, J. (1984). An investigation into component and configural processes underlying face perception. *British journal of psychology*, 75(2), 221-242. <https://doi.org/10.1111/j.2044-8295.1984.tb01895.x>
- Sergent, J., Ohta, S., & Macdonald, B. (1992). Functional neuroanatomy of face and object processing: a positron emission tomography study. *Brain*, 115(1), 15-36. <https://doi.org/10.1093/brain/115.1.15>
- Setoh, P., Lee, K. J., Zhang, L., Qian, M. K., Quinn, P. C., Heyman, G. D., & Lee, K. (2019). Racial categorization predicts implicit racial bias in preschool children. *Child Development*, 90(1), 162-179. <https://doi.org/10.1111/cdev.12851>
- Shafto, C. L., Conway, C. M., Field, S. L., & Houston, D. M. (2012). Visual sequence learning in infancy: Domain-general and domain-specific associations with language. *Infancy*, 17(3), 247-271. <https://doi.org/10.1111/j.1532-7078.2011.00085.x>
- Shams, L., & Seitz, A. R. (2008). Benefits of multisensory learning. *Trends in cognitive sciences*, 12(11), 411-417. <https://doi.org/10.1016/j.tics.2008.07.006>
- Shapiro, P. N., & Penrod, S. (1986). Meta-analysis of facial identification studies. *Psychological bulletin*, 100(2), 139. <https://doi.org/10.1037/0033-2909.100.2.139>
- Shepherd, J. W., & Deregowski, J. B. (1981). Races and faces—a comparison of the responses of Africans and Europeans to faces of the same and different races. *British Journal of Social Psychology*, 20(2), 125-133. <https://doi.org/10.1111/j.2044-8309.1981.tb00485.x>
- Shinozaki, J., Hiroe, N., Sato, M. A., Nagamine, T., & Sekiyama, K. (2016). Impact of language on functional connectivity for audiovisual speech integration. *Scientific reports*, 6(1), 1-13. <https://doi.org/10.1038/srep31388>

- Simion, F., Valenza, E., Macchi Cassia, V., Turati, C., & Umiltà, C. (2002). Newborns' preference for up-down asymmetrical configurations. *Developmental Science*, 5(4), 427-434. <https://doi.org/10.1111/1467-7687.00237>
- Simion, F., Regolin, L., & Bulf, H. (2008). A predisposition for biological motion in the newborn baby. *Proceedings of the National Academy of Sciences*, 105(2), 809-813. <https://doi.org/10.1073/pnas.0707021105>
- Simpson, E. A., Varga, K., Frick, J. E., & Fragaszy, D. (2011). Infants experience perceptual narrowing for nonprimate faces. *Infancy*, 16(3), 318-328. <https://doi.org/10.1111/j.1532-7078.2010.00052.x>
- Singh, S., & Black, J. W. (1966). Study of twenty-six intervocalic consonants as spoken and recognized by four language groups. *The Journal of the Acoustical Society of America*, 39(2), 372-387. <https://doi.org/10.1121/1.1909899>
- Singh, L., Loh, D., & Xiao, N. G. (2017). Bilingual infants demonstrate perceptual flexibility in phoneme discrimination but perceptual constraint in face discrimination. *Frontiers in psychology*, 8, 1563. <https://doi.org/10.3389/fpsyg.2017.01563>
- Slater, A., & Morison, V. (1985). Shape constancy and slant perception at birth. *Perception*, 14(3), 337-344. <https://doi.org/10.1068/p140337>
- Slater, A., Mattock, A., & Brown, E. (1990). Size constancy at birth: Newborn infants' responses to retinal and real size. *Journal of experimental child psychology*, 49(2), 314-322. [https://doi.org/10.1016/0022-0965\(90\)90061-C](https://doi.org/10.1016/0022-0965(90)90061-C)
- Slater, A., Quinn, P. C., Brown, E., & Hayes, R. (1999). Intermodal perception at birth: Intersensory redundancy guides newborn infants' learning of arbitrary auditory-visual pairings. *Developmental Science*, 2(3), 333-338. <https://doi.org/10.1111/1467-7687.00079>
- Slater, A., & Quinn, P. C. (2001). Face recognition in the newborn infant. *Infant and Child Development: An International Journal of Research and Practice*, 10(1-2), 21-24. <https://doi.org/10.1002/icd.241>
- Spangler, S. M., Schwarzer, G., Freitag, C., Vierhaus, M., Teubert, M., Fassbender, I., ... & Keller, H. (2013). The other-race effect in a longitudinal sample of 3-, 6- and 9-month-old infants: Evidence of a training effect. *Infancy*, 18(4), 516-533. <https://doi.org/10.1111/j.1532-7078.2012.00137.x>
- Spencer, J., O'Brien, J., Johnston, A., & Hill, H. (2006). Infants' discrimination of faces by using biological motion cues. *Perception*, 35(1), 79-89. <https://doi.org/10.1068/p5379>
- Stevenage, S. V. (1995). Can caricatures really produce distinctiveness effects?. *British Journal of Psychology*, 86(1), 127-146. <https://doi.org/10.1111/j.2044-8295.1995.tb02550.x>
- Strand, J., Cooperman, A., Rowe, J., & Simenstad, A. (2014). Individual differences in susceptibility to the McGurk effect: links with lipreading and detecting audiovisual incongruity. *Journal of Speech, Language, and Hearing Research*, 57(6), 2322-2331. https://doi.org/10.1044/2014_JSLHR-H-14-0059

- Strelnikov, K., Rouger, J., Lagleyre, S., Fraysse, B., Deguine, O., & Barone, P. (2009). Improvement in speech-reading ability by auditory training: Evidence from gender differences in normally hearing, deaf and cochlear implanted subjects. *Neuropsychologia*, *47*(4), 972-979. <https://doi.org/10.1016/j.neuropsychologia.2008.10.017>
- Sugden, N. A., Mohamed-Ali, M. I., & Moulson, M. C. (2014). I spy with my little eye: Typical, daily exposure to faces documented from a first-person infant perspective. *Developmental psychobiology*, *56*(2), 249-261. <https://doi.org/10.1002/dev.21183>
- Sugita, Y. (2008). Face perception in monkeys reared with no exposure to faces. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 394-398. <https://doi.org/10.1073/pnas.0706079105>
- Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *The journal of the acoustical society of america*, *26*(2), 212-215. <https://doi.org/10.1121/1.1907309>
- Summerfield, Q., & McGrath, M. (1984). Detection and resolution of audio-visual incompatibility in the perception of vowels. *The Quarterly Journal of Experimental Psychology Section A*, *36*(1), 51-74. <https://doi.org/10.1080/14640748408401503>
- Summerfield, Q. (1992). Lipreading and audio-visual speech perception. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *335*(1273), 71-78. <https://doi.org/10.1098/rstb.1992.0009>
- Sundara, M., Polka, L., & Molnar, M. (2008). Development of coronal stop perception: Bilingual infants keep pace with their monolingual peers. *Cognition*, *108*(1), 232-242. <https://doi.org/10.1016/j.cognition.2007.12.013>
- Swingle, D. (2009). Contributions of infant word learning to language development. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1536), 3617-3632. <https://doi.org/10.1098/rstb.2009.0107>
- Reid, V. M., Dunn, K., Young, R. J., Amu, J., Donovan, T., & Reissland, N. (2017). The human fetus preferentially engages with face-like visual stimuli. *Current Biology*, *27*(12), 1825-1828. <https://doi.org/10.1016/j.cub.2017.05.044>
- Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *The Quarterly journal of experimental psychology*, *46*(2), 225-245. <https://doi.org/10.1080/14640749308401045>
- Tanaka, J. W., Curran, T., Porterfield, A. L., & Collins, D. (2006). Activation of preexisting and acquired face representations: the N250 event-related potential as an index of face familiarity. *Journal of Cognitive Neuroscience*, *18*(9), 1488-1497. [10.1162/jocn.2006.18.9.1488](https://doi.org/10.1162/jocn.2006.18.9.1488)
- Tereno, S., Soares, I., Martins, E., Sampaio, D., & Carlson, E. (2007). La théorie de l'attachement: son importance dans un contexte pédiatrique. *Devenir*, *19*(2), 151-188.
- Tham, D. S. Y., Bremner, J. G., & Hay, D. (2017). The other-race effect in children from a multiracial population: A cross-cultural comparison. *Journal of experimental child psychology*, *155*, 128-137. <https://doi.org/10.1016/j.jecp.2016.11.006>

- Tham, D. S. Y., Woo, P. J., & Bremner, J. G. (2019). Development of the other-race effect in Malaysian-Chinese infants. *Developmental psychology*, 61(1), 107-115. <https://doi.org/10.1002/dev.21783>
- Thomas, S. M., & Jordan, T. R. (2004). Contributions of oral and extraoral facial movement to visual and audiovisual speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, 30(5), 873. <https://doi.org/10.1037/0096-1523.30.5.873>
- Thompson, P. (1980). Margaret Thatcher: A new illusion. *Perception*, 9(4), 483-484. <https://doi.org/10.1068/p090483>
- Thompson, R. F. (2009). Habituation: a history. *Neurobiology of learning and memory*, 92(2), 127. <https://doi.org/10.1016/j.nlm.2008.07.011>
- Tiippana, K. (2014). What is the McGurk effect? *Frontiers in Psychology*, 5, 725. <https://doi.org/10.3389/fpsyg.2014.00725>
- Timney, B., Mitchell, D. E., & Giffin, F. (1978). The development of vision in cats after extended periods of dark-rearing. *Experimental Brain Research*, 31(4), 547-560. <https://doi.org/10.1007/BF00239811>
- Tootell, R. B., Silverman, M. S., Switkes, E., & De Valois, R. L. (1982). Deoxyglucose analysis of retinotopic organization in primate striate cortex. *Science*, 218(4575), 902-904. <https://doi.org/10.1126/science.7134981>
- Toro, J. M., & Trobalón, J. B. (2005). Statistical computations over a speech stream in a rodent. *Perception & psychophysics*, 67(5), 867-875. <https://doi.org/10.3758/BF03193539>
- Trainor, L. J., Austin, C. M., & Desjardins, R. N. (2000). Is infant-directed speech prosody a result of the vocal expression of emotion?. *Psychological science*, 11(3), 188-195. <https://doi.org/10.1111/1467-9280.00240>
- Trehub, S. E. (1973). Infants' sensitivity to vowel and tonal contrasts. *Developmental Psychology*, 9(1), 91. <https://doi.org/10.1037/h0034999>
- Trehub, S. E. (1976). The discrimination of foreign speech contrasts by infants and adults. *Child development*, 466-472. <https://doi.org/10.2307/1128803>
- Tremblay, C., Champoux, F., Voss, P., Bacon, B. A., Lepore, F., & Théoret, H. (2007). Speech and non-speech audio-visual illusions: A developmental study. *PloS one*, 2(8), e742. <https://doi.org/10.1371/journal.pone.0000742>
- Tsao, F. M., Liu, H. M., & Kuhl, P. K. (2006). Perception of native and non-native affricate- fricative contrasts: Cross-language tests on adults and infants. *The Journal of the Acoustical Society of America*, 120(4), 2285-2294. <https://doi.org/10.1121/1.2338290>
- Tsao, D. Y., Moeller, S., & Freiwald, W. A. (2008). Comparing face patch systems in macaques and humans. *Proceedings of the National Academy of Sciences*, 105(49), 19514-19519. <https://doi.org/10.1073/pnas.0809662105>
- 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. *Third international conference on spoken language processing*, 1695-1698.

- Tummeltshammer, K. S., Wu, R., Sobel, D. M., & Kirkham, N. Z. (2014). Infants track the reliability of potential informants. *Psychological Science*, 25(9), 1730-1738. <https://doi.org/10.1177/0956797614540178>
- Tüttenberg, S. C., & Wiese, H. (2021). Recognising other-race faces is more effortful: The effect of individuation instructions on encoding-related ERP Dm effects. *Biological Psychology*, 158, 107992. <https://doi.org/10.1016/j.biopsycho.2020.107992>
- Uttley, L., de Boisferon, A. H., Dupierrix, E., Lee, K., Quinn, P. C., Slater, A. M., & Pascalis, O. (2013). Six-month-old infants match other-race faces with a non-native language. *International Journal of Behavioral Development*, 37(2), 84-89. <https://doi.org/10.1177/0165025412467583>
- Valentine, T., & Bruce, V. (1986). The effects of distinctiveness in recognising and classifying faces. *Perception*, 15(5), 525-535. <https://doi.org/10.1068/p150525>
- Valentine, T. (1988). Upside-down faces: A review of the effect of inversion upon face recognition. *British journal of psychology*, 79(4), 471-491. <https://doi.org/10.1111/j.2044-8295.1988.tb02747.x>
- Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology Section A*, 43(2), 161-204. <https://doi.org/10.1080/14640749108400966>
- Valentine, T., & Endo, M. (1992). Towards an exemplar model of face processing: The effects of race and distinctiveness. *The Quarterly Journal of Experimental Psychology*, 44(4), 671-703. <https://doi.org/10.1080/14640749208401305>
- Valenza, E., Simion, F., Cassia, V. M., & Umiltà, C. (1996). Face preference at birth. *Journal of experimental psychology: Human Perception and Performance*, 22(4), 892-903. <https://doi.org/10.1037/0096-1523.22.4.892>
- Vallortigara, G., & Rogers, L. (2005). Survival with an asymmetrical brain: advantages and disadvantages of cerebral lateralization. *Behavioral and brain sciences*, 28(4), 589-633. <https://doi.org/10.1017/S0140525X05000105>
- Vallortigara, G., & Regolin, L. (2006). Gravity bias in the interpretation of biological motion by inexperienced chicks. *Current Biology*, 16(8), 279-280. <https://doi.org/10.1016/j.cub.2006.03.052>
- Van Engen, K. J., Xie, Z., & Chandrasekaran, B. (2017). Audiovisual sentence recognition not predicted by susceptibility to the McGurk effect. *Attention, Perception, & Psychophysics*, 79(2), 396-403. <https://doi.org/10.3758/s13414-016-1238-9>
- Vogel, M., Monesson, A., & Scott, L. S. (2012). Building biases in infancy: the influence of race on face and voice emotion matching. *Developmental science*, 15(3), 359-372. <https://doi.org/10.1111/j.1467-7687.2012.01138.x>

- Ujii, Y., Asai, T., & Wakabayashi, A. (2015). The relationship between level of autistic traits and local bias in the context of the McGurk effect. *Frontiers in Psychology, 6*, 891. <https://doi.org/10.3389/fpsyg.2015.00891>
- Ujii, Y., Kanazawa, S., & Yamaguchi, M. K. (2020). The Other-Race-Effect on Audiovisual Speech Integration in Infants: A NIRS Study. *Frontiers in Psychology, 11*, 971. <https://doi.org/10.3389/fpsyg.2020.00971>
- Vilain, A., Dole, M., Loevenbruck, H., Pascalis, O., & Schwartz, J. L. (2019). The role of production abilities in the perception of consonant category in infants. *Developmental science, 22*(6), e12830. <https://doi.org/10.1111/desc.12830>
- Vouloumanos, A., Druhen, M. J., Hauser, M. D., & Huizink, A. T. (2009). Five-month-old infants' identification of the sources of vocalizations. *Proceedings of the National Academy of Sciences, 106*(44), 18867-18872. <https://doi.org/10.1073/pnas.0906049106>
- Wada, Y., & Yamamoto, T. (2001). Selective impairment of facial recognition due to a haematoma restricted to the right fusiform and lateral occipital region. *Journal of Neurology, Neurosurgery & Psychiatry, 71*(2), 254-257. <http://dx.doi.org/10.1136/jnnp.71.2.254>
- Walker-Andrews, A. S. (1986). Intermodal perception of expressive behaviors: Relation of eye and voice? *Developmental psychology, 22*(3), 373. <https://doi.org/10.1037/0012-1649.22.3.373>
- Walker-Andrews, A. S., & Gibson, E. J. (1986). What develops in bimodal perception. *Advances in infancy research, 4*, 171-181.
- Walker-Andrews, A. S., Bahrick, L. E., Raglioni, S. S., & Diaz, I. (1991). Infants' bimodal perception of gender. *Ecological Psychology, 3*(2), 55-75. https://doi.org/10.1207/s15326969eco0302_1
- Walton, G. E., Bower, N. J., & Bower, T. G. (1992). Recognition of familiar faces by newborns. *Infant Behavior & Development, 5*(2), 265-269. [https://doi.org/10.1016/0163-6383\(92\)80027-R](https://doi.org/10.1016/0163-6383(92)80027-R)
- Walton, G. E., & Bower, T. G. R. (1993). Amodal representation of speech in infants. *Infant Behavior and Development, 16*(2), 233-243. [https://doi.org/10.1016/0163-6383\(93\)80019-5](https://doi.org/10.1016/0163-6383(93)80019-5)
- Wanrooij, K., Boersma, P., & Van Zuijen, T. (2014). Fast phonetic learning occurs already in 2-to-3-month old infants: an ERP study. *Frontiers in psychology, 5*, 77. <https://doi.org/10.3389/fpsyg.2014.00077>
- Weikum, W. M., Vouloumanos, A., Navarra, J., Soto-Faraco, S., Sebastián-Gallés, N., & Werker, J. F. (2007). Visual language discrimination in infancy. *Science, 316*(5828), 1159-1159. <https://doi.org/10.1126/science.1137686>
- Weiss, P. H., & Fink, G. R. (2009). Grapheme-colour synaesthetes show increased grey matter volumes of parietal and fusiform cortex. *Brain, 132*(1), 65-70. <https://doi.org/10.1093/brain/awn304>
- Werker, J. F., Gilbert, J. H., Humphrey, K., & Tees, R. C. (1981). Developmental aspects of cross-language speech perception. *Child development, 349-355*. <https://doi.org/10.2307/1129249>

- Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant behavior and development*, 7(1), 49-63. [https://doi.org/10.1016/S0163-6383\(84\)80022-3](https://doi.org/10.1016/S0163-6383(84)80022-3)
- Wheeler, A., Anzures, G., Quinn, P. C., Pascalis, O., Omrin, D. S., & Lee, K. (2011). Caucasian infants scan own-and other-race faces differently. *PloS one*, 6(4), e18621. <https://doi.org/10.1371/journal.pone.0018621>
- White, T. P., Wigton, R. L., Joyce, D. W., Bobin, T., Ferragamo, C., Wasim, N., ... & Shergill, S. S. (2014). Eluding the illusion? Schizophrenia, dopamine and the McGurk effect. *Frontiers in Human Neuroscience*, 8, 565. <https://doi.org/10.3389/fnhum.2014.00565>
- Wilcox, B. M., & Clayton, F. L. (1968). Infant visual fixation on motion pictures of the human face. *Journal of Experimental Child Psychology*, 6(1), 22-32. [https://doi.org/10.1016/0022-0965\(68\)90068-4](https://doi.org/10.1016/0022-0965(68)90068-4)
- Wilson, S. M., Saygin, A. P., Sereno, M. I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature neuroscience*, 7(7), 701-702. <https://doi.org/10.1038/nn1263>
- Winograd, E. (1981). Elaboration and distinctiveness in memory for faces. *Journal of Experimental Psychology: Human Learning and Memory*, 7(3), 181. <https://doi.org/10.1037/0278-7393.7.3.181>
- Wobbrock, J. O., Findlater, L., Gergle, D., & Higgins, J. J. (2011). The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 143-146. <https://doi.org/10.1145/1978942.1978963>
- Wu, R., & Kirkham, N. Z. (2010). No two cues are alike: Depth of learning during infancy is dependent on what orients attention. *Journal of Experimental Child Psychology*, 107(2), 118-136. <https://doi.org/10.1016/j.jecp.2010.04.014>
- Xiao, W. S., Xiao, N. G., Quinn, P. C., Anzures, G., & Lee, K. (2013). Development of face scanning for own-and other-race faces in infancy. *International journal of behavioral development*, 37(2), 100-105. <https://doi.org/10.1177/0165025412467584>
- Xiao, W. S., Fu, G., Quinn, P. C., Qin, J., Tanaka, J. W., Pascalis, O., & Lee, K. (2015). Individuation training with other-race faces reduces preschoolers' implicit racial bias: A link between perceptual and social representation of faces in children. *Developmental Science*, 18(4), 655-663. <https://doi.org/10.1111/desc.12241>
- Xiao, N. G., Quinn, P. C., Liu, S., Ge, L., Pascalis, O., & Lee, K. (2017a). Older but not younger infants associate own-race faces with happy music and other-race faces with sad music. *Developmental Science*, 21(2), e12537. <https://doi.org/10.1111/desc.12537>
- Xiao, N. G., Wu, R., Quinn, P. C., Liu, S., Tummeltshammer, K. S., Kirkham, N. Z., Ge, L., & Lee, K. (2017b). Infants rely more on gaze cues from own-race than other-race adults for learning under uncertainty. *Child development*, 89(3), e229-e244. <https://doi.org/10.1111/cdev.12798>

- Xiao, N. G., Mukaida, M., Quinn, P. C., Pascalis, O., Lee, K., & Itakura, S. (2018). Narrowing in face and speech perception in infancy: Developmental change in the relations between domains. *Journal of Experimental Child Psychology*, *176*, 113-127. <https://doi.org/10.1016/j.jecp.2018.06.007>
- Yeung, H. H., & Werker, J. F. (2009). Learning words' sounds before learning how words sound: 9-month-olds use distinct objects as cues to categorize speech information. *Cognition*, *113*(2), 234-243. <https://doi.org/10.1016/j.cognition.2009.08.010>
- Yeung, H. H., & Werker, J. F. (2013). Lip movements affect infants' audiovisual speech perception. *Psychological Science*, *24*(5), 603-612. <https://doi.org/10.1177/0956797612458802>
- Yin, R. K. (1969). Looking at upside-down faces. *Journal of experimental psychology*, *81*(1), 141-145. <https://doi.org/10.1037/h0027474>
- Yoshida, K. A., Pons, F., Maye, J., & Werker, J. F. (2010). Distributional phonetic learning at 10 months of age. *Infancy*, *15*(4), 420-433. <https://doi.org/10.1111/j.1532-7078.2009.00024.x>
- Young, A. W., Hellowell, D., & Hay, D. C. (1987). Configurational information in face perception. *Perception*, *42*(11), 1166-1178. <https://doi.org/10.1068/p160747>
- Young, A. W., Newcombe, F., Haan, E. H. D., Small, M., & Hay, D. C. (1993). Face perception after brain injury: Selective impairments affecting identity and expression. *Brain*, *116*(4), 941-959. <https://doi.org/10.1093/brain/116.4.941>
- Yovel, G., & Kanwisher, N. (2004). Face perception: domain specific, not process specific. *Neuron*, *44*(5), 889-898. <https://doi.org/10.1016/j.neuron.2004.11.018>
- Yue, X., Tjan, B. S., & Biederman, I. (2006). What makes faces special? *Vision research*, *46*(22), 3802-3811. <https://doi.org/10.1016/j.visres.2006.06.017>
- Yuen, I., Davis, M. H., Brysbaert, M., & Rastle, K. (2010). Activation of articulatory information in speech perception. *Proceedings of the National Academy of Sciences*, *107*(2), 592-597. <https://doi.org/10.1073/pnas.0904774107>
- Zamm, A., Schlaug, G., Eagleman, D. M., & Loui, P. (2013). Pathways to seeing music: enhanced structural connectivity in colored-music synesthesia. *Neuroimage*, *74*, 359-366. <https://doi.org/10.1016/j.neuroimage.2013.02.024>
- Zhou, X., Elshiekh, A., & Moulson, M. C. (2019). Lifetime perceptual experience shapes face memory for own-and other-race faces. *Visual Cognition*, *27*(9-10), 687-700. <https://doi.org/10.1080/13506285.2019.1638478>