Contrôle de la production de la parole chez l'enfant de 4 ans : l'anticipation comme indice de maturité motrice

Guillaume Barbier

▶ **To cite this version:**

Guillaume Barbier. Contrôle de la production de la parole chez l'enfant de 4 ans : l'anticipation comme indice de maturité motrice. Sciences cognitives. Université Grenoble Alpes, 2016. Français. NNT : . tel-01417677v2

HAL Id: tel-01417677

https://theses.hal.science/tel-01417677v2

Submitted on 23 Dec 2016 (v2), last revised 25 Apr 2017 (v3)

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.

Communauté UNIVERSITÉ Grenoble Alpes

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

Spécialité : **Sciences Cognitives**

Arrêté ministériel : 7 août 2006

Présentée par

Guillaume BARBIER

Thèse dirigée par **Pascal Perrier** et
codirigée par **Yohan Payan** et **Lucie Ménard**

préparée au sein du **Laboratoire GIPSA-lab, Département Parole et Cognition** et du **Laboratoire de Phonétique de l'Université du Québec à Montréal** dans **l'École Doctorale Ingénierie pour la Santé, la Cognition et l'Environnement**

Contrôle de la production de la parole chez l'enfant de 4 ans : l'anticipation comme indice de maturité motrice

Thèse soutenue publiquement le **8 septembre 2016**,
devant le jury composé de :

Anne GUÉRIN-DUGUÉ
Prof., Université Grenoble Alpes, Grenoble, France, Présidente du Jury
Melissa REDFORD
Prof., University of Oregon, Eugene, Oregon, USA, Rapporteur
Cécile FOUGERON
DR CNRS, Laboratoire de Phonétique et Phonologie, Université Sorbonne Nouvelle - Paris III, Paris, France, Rapporteur
David OSTRY
Prof., McGill University, Montréal, Québec, Canada, Examinateur
Yohan PAYAN
DR CNRS, Labo. TIMC-IMAG, Grenoble, France, Co-directeur de thèse
Lucie MÉNARD
Prof., Université du Québec à Montréal, Canada, Co-directrice de thèse
Pascal PERRIER
Prof., Grenoble-INP, GIPSA-lab, Grenoble, France, Directeur de thèse

Communauté UNIVERSITÉ Grenoble Alpes

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

Spécialité : **Sciences Cognitives**

Arrêté ministériel : 7 août 2006

Présentée par

Guillaume BARBIER

Thèse dirigée par **Pascal Perrier** et
codirigée par **Yohan Payan** et **Lucie Ménard**

préparée au sein du **Laboratoire GIPSA-lab, Département Parole et Cognition** et du **Laboratoire de Phonétique de l'Université du Québec à Montréal** dans **l'École Doctorale Ingénierie pour la Santé, la Cognition et l'Environnement**

Contrôle de la production de la parole chez l'enfant de 4 ans : l'anticipation comme indice de maturité motrice

Thèse soutenue publiquement le **8 septembre 2016**,
devant le jury composé de :

Anne GUÉRIN-DUGUÉ
Prof., Université Grenoble Alpes, Grenoble, France, Présidente du Jury
Melissa REDFORD
Prof., University of Oregon, Eugene, Oregon, USA, Rapporteur
Cécile FOUGERON
DR CNRS, Laboratoire de Phonétique et Phonologie, Université Sorbonne Nouvelle - Paris III, Paris, France, Rapporteur
David OSTRY
Prof., McGill University, Montréal, Québec, Canada, Examinateur
Yohan PAYAN
DR CNRS, Labo. TIMC-IMAG, Grenoble, France, Co-directeur de thèse
Lucie MÉNARD
Prof., Université du Québec à Montréal, Canada, Co-directrice de thèse
Pascal PERRIER
Prof., Grenoble-INP, GIPSA-lab, Grenoble, France, Directeur de thèse

Résumé – Abstract

Résumé

Ce travail de thèse étudie la production de la parole à l'âge de 4 ans, en comparaison avec l'adulte, sous l'angle du contrôle moteur. Un intérêt particulier a été porté à deux indices : la variabilité liée à la répétition de voyelles isolées et la coarticulation anticipatoire intra et extra-syllabique dans le cadre de la production de séquences de type V_1-C-V_2. Des données acoustiques et articulatoires ont été enregistrées via l'échographie linguale et le dispositif HOCUS. Les données acoustiques ont été analysées pour 20 enfants et 10 adultes et les données articulatoires ont été analysées pour 6 de ces enfants et 2 de ces adultes. En accord avec de nombreuses études de la littérature, la variabilité liée à la répétition est plus importante chez les enfants que chez les adultes. Une forte anticipation de V_2 dans la réalisation de V_1 a été observée pour tous les adultes. Chez les enfants, cette anticipation n'est pas systématique, et lorsqu'elle est observée, elle est d'amplitude plus faible que chez les adultes. Ainsi, seulement 5 enfants parmi les 20 étudiés montrent des signes d'anticipation, majoritairement dans la dimension antéropostérieure, manifestée dans l'espace acoustique par la dimension F2. La coarticulation anticipatoire intra-syllabique semble également être d'amplitude plus faible chez les enfants. Enfin, les gestes de parole des enfants sont plus lents que ceux des adultes. Dans leur ensemble, ces résultats sont interprétés comme la preuve de l'immaturité du système moteur pour la production de la parole chez l'enfant, selon deux perspectives : des patrons moteurs insuffisamment stables pour la production de gestes vocaliques et un manque d'efficacité dans l'anticipation les gestes futurs lors de la production d'une séquence de sons de parole. Selon une perspective empruntée aux théories du contrôle moteur optimal, nous proposons que la coarticulation anticipatoire repose sur l'usage de modèles internes, représentations sensori-motrices du système moteur de la parole dans le système nerveux central, et que son amplitude reflète la maturation progressive de ces représentations sensori-motrices à mesure que la production de la parole se développe.

Mots-clés : Développement, production de parole, contrôle moteur, coarticulation anticipatoire, planification, échographie linguale.

Abstract

This thesis work investigates speech production in 4-year-old children, in comparison with adults, from a speech motor control perspective. It focuses on two indices: measures of token-to-token variability in the production of isolated vowels and on anticipatory intra and extra-syllabic coarticulation within V_1-C-V_2 sequences. Acoustic and articulatory data were recorded thanks to ultrasound tongue imaging within the HOCUS system. Acoustic data from 20 children and 10 adults have been analyzed. Ultrasound data have been analyzed from a subset of these participants: 6 children and 2 adults. In agreement with former studies, token-to-token variability was greater in children than in adults. Strong anticipation of V2 in V1 realization was found in all adults. In children, anticipation was not systematic, and when observed, it was of smaller amplitude than in adults. In more details, only 5 children among the 20 studied showed a small amount of anticipation, mainly along the antero-posterior dimension, manifested in the acoustic F2 dimension. Anticipatory intra-syllabic coarticulation also seems to be of smaller amplitude in children than in adults. Last, children's speech gestures are slower than those of adults. These results are interpreted as evidence for the immaturity of children's speech motor control from two perspectives: insufficiently stable motor control patterns for vowel production, and a lack of effectiveness in anticipating forthcoming gestures. In line with theories of optimal motor control, we assume that anticipatory coarticulation is based on the use of internal models, i.e. sensori-motor representations of the speech production apparatus in the central nervous system, and that the amplitude of anticipatory coarticulation reflects the increasing maturation of these sensori-motor representations as speech develops.

Keywords: Development, speech production, speech motor control, anticipatory coarticulation, speech planning, ultrasound tongue imaging.

Remerciements

Je tiens en tout premier lieu à remercier Pascal Perrier. Merci tout d'abord de m'avoir accordé ta confiance et de t'être lancé dans cette aventure. Merci pour ton encadrement rigoureux, stimulant et toujours agréable, pour ton enthousiasme, ta patience, ta persévérance, ton soutien, tes grandes qualités scientifiques et humaines, qui font que même dans les moments de difficulté et de doute, cela a toujours été et reste aujourd'hui un plaisir de travailler avec toi.

Je tiens également à remercier Lucie Ménard pour la qualité de son encadrement, son sens de l'organisation et son esprit de synthèse, ses compétences en psychologie expérimentale, ses compétences scientifiques et ses grandes qualités humaines. Lucie, je ne saurais comment te remercier de m'avoir offert ce cadre de recherche, dans un projet impliquant Pascal, Yohan, Mark et Joe autour d'un sujet passionnant, de m'avoir offert les moyens expérimentaux de recueillir ces belles données, et de mener ce projet à bien. A bien des étapes du projet, ton encadrement et ton soutien ont été plus que nécessaires.

Merci à Yohan Payan, pour son encadrement et son implication dans le projet. Même si nous n'avons pas eu l'occasion de beaucoup travailler ensemble, cela a été un plaisir de travailler avec toi et j'espère que nous aurons l'occasion à l'avenir de développer l'aspect modélisation de ce travail.

Merci ensuite à deux personnes qui ont été centrales dans ce projet, bien que ne faisant pas officiellement partie de l'encadrement de cette thèse : Mark Tiede et Joe Perkell. Merci pour toutes les discussions et tous vos apports, de la revue de la littérature à l'analyse des résultats, en passant par la formulation des hypothèses et le prétraitement des données. Merci particulièrement à Mark pour son aide technique, mais pas seulement, et à tous les deux pour ces riches réunions régulières qui nous ont permis de construire peu à peu ce projet.

Merci également à toute l'équipe du Laboratoire de Phonétique de l'Université du Québec à Montréal. Merci particulièrement à Amélie Prémont, Paméla Trudeau-Fisette, Christine Turgeon et Dominique Côté pour leur aide précieuse concernant le recueil des données, leurs connaissances techniques et leur contact avec les enfants.

Je tiens également à remercier sincèrement les membres du jury, Anne Guérin, Melissa Redford, Cécile Fougeron et David Ostry, d'avoir accepté d'évaluer ce travail. C'est un honneur de vous avoir eu comme jury, et je vous remercie pour l'intérêt que vous avez porté à ce travail.

Mes remerciements vont ensuite à nos participants. Merci aux participants adultes d'avoir joué le jeu et d'avoir complété une tâche destinée aux enfants de 4 ans, et merci aux enfants et aux parents d'avoir accepté de participer à cette étude, souvent avec enthousiasme et curiosité. Merci aussi à Carole Langlois du CPE Le Sablier pour l'intérêt qu'elle a montré au projet dès le départ et pour son implication.

Je tiens ensuite à remercier Louis-Jean Boë, qui m'a beaucoup apporté. Merci pour ta passion contagieuse pour la phonétique. Si je ne t'avais pas rencontré en 2009, mon intérêt pour la phonétique ne serait probablement jamais né, et toutes ces rencontres et ces beaux projets nés depuis n'auraient peut-être jamais vu le jour. Merci énormément.

Merci également à deux professeures qui ont su éveiller en moi la curiosité et m'ont donné envie de continuer mes études en sciences du langage et en phonétique : Elisabetta Carpitelli et Anne Vilain.

Merci du fond du cœur à Anne Guérin.

Merci ensuite à tous les jeunes du GIPSA, passés et présents, les stagiaires, doctorants et post-docs, notre ingénieur X., pour l'ambiance, la cohésion, les pauses café, toutes les soirées, sorties, etc... Merci en particulier à ceux qui ont partagé notre bureau, Olha, Maël Raphaël, et Diandra. Merci également à tous les personnels du GIPSA, chercheurs, ingénieurs et administratifs. Merci d'avoir fait en sorte que ce séjour au laboratoire soit le plus agréable possible.

Merci ensuite à tous mes amis, particulièrement à ceux qui m'ont aidé à traverser les moments les plus durs comme les plus festifs, qui ont été présents quand j'avais le plus besoin d'eux, et à tous ceux qui ont pu être présents à la soutenance.

Merci à ma famille. Merci en particulier à mes parents et à mon frère.

Merci à Sabine, pour ses illustrations, pour son soutien sans faille, et surtout pour m'avoir supporté ces deux dernières années malgré les deadlines à minuit et les moments de stress, conférences, soutenance, et j'en passe.

Merci enfin aux différentes sources de financement qui ont permis de mener à bien ce projet de thèse : le ministère pour la bourse de thèse, le FQRNT (via Lucie) pour le projet scientifique et le cadre, et enfin le GIPSA-lab, l'Université Grenoble Alpes et Jean-Luc Schwartz pour m'avoir permis de finir cette thèse dans de bonnes conditions.

Table des Matières

Résumé 5

Abstract 6

Remerciements 7

Table des Matières 10

Introduction 12

Chapitre 1 - Le développement de la production de la parole 18

1.1. La parole et son développement précoce 18

1.2. Le développement tardif de la production de la parole 34

1.3. Cadre théorique général 46

1.4. La coarticulation : de l'enfant à l'adulte 64

1.5. La construction des représentations sensori-motrices 82

1.6. Bilan et hypothèses 106

Chapitre 2 - Méthodologie 111

2.1. Participants 111

2.2. Acquisition des données 111

2.3. Prétraitements et analyse des données 121

2.4. Traitement des données et analyses statistiques 134

Chapitre 3 - Résultats 144

3.1. Stabilité de la production des voyelles en contexte isolé 145

3.2. Coarticulation intra-syllabique 157

3.3. Coarticulation extra-syllabique 177

3.4. Durées des segments et des séquences 208

3.5. Relations entre nos principaux indices 211

Chapitre 4 - Discussion **214**

4.1. Stabilité de la production des voyelles en contexte isolé 214

4.2. Coarticulation intra-syllabique 217

4.3. Coarticulation extra-syllabique 222

4.4. Durées des segments et des séquences 227

4.5. Relations entre nos principaux indices 229

4.6. Limites et perspectives 231

Conclusion **235**

Références bibliographiques **242**

Introduction

La parole, ou communication parlée, est un sujet d'étude passionnant. D'abord parce que c'est notre mode de communication privilégié, que nous utilisons sans effort apparent et quotidiennement, et parce que derrière cette apparente simplicité se cache l'une des facultés les plus complexes que l'être humain possède. La production et la perception de la parole mettent en jeu de nombreuses facultés cognitives sous-jacentes, que ce soit la catégorisation, la mémorisation, l'abstraction, ou le contrôle moteur. De ce point de vue, la parole est donc une faculté extrêmement complexe, et la compréhension des mécanismes mis en œuvre dans son utilisation n'est pas facilitée par cette extrême complexité. Le processus d'apprentissage de cette faculté est encore davantage passionnant, et revêt un intérêt tout particulier, car son étude permet de mettre en lumière les composantes fondamentales et les processus mis en œuvre dans cette faculté. L'étude de sa mise en place et de son développement, pas à pas, du nouveau-né à l'adulte, permet de rendre apparente sa structure, aussi bien physique que cognitive, et met en évidence les composantes fondamentales de cette faculté.

Cette capacité à percevoir et à produire un ensemble de sons distincts et à les utiliser dans le but de communiquer constitue un sujet d'étude passionnant. Ce serait l'une des capacités humaines les plus complexes et remarquables. Pourtant, les enfants qui entrent dans l'ère de la parole semblent le faire avec un naturel déconcertant. La tâche, pour le volet production, est pourtant terriblement complexe : apprendre à contrôler, de façon coordonnée, une centaine de muscles oro-faciaux, afin de produire une sensation, chez soi et chez l'interlocuteur, qui renvoie aux mêmes représentations mentales abstraites (les unités de la phonologie) afin de communiquer un message. Il y a en effet de nombreuses raisons pour lesquelles cette tâche pourrait faillir. Et pourtant, pour l'immense majorité d'entre nous, cette faculté se met en place naturellement au cours des premières années de vie. Peut-être sous la pression exercée par la nécessité de communiquer avec ses pairs, puisque l'Homme est un animal social. Il existe en effet de nombreuses pressions évolutives qui rendent l'apprentissage de la communication indispensable. Communiquer au sein du groupe est en effet crucial pour les individus, dans leur propre intérêt comme dans celui du groupe. La communication est donc cruciale, et c'est probablement pour cette raison que la communication parlée est

acquise si rapidement et si naturellement au cours de l'ontogenèse, et c'est également probablement pour cette raison que la communication, entre les êtres vivants, est universelle.

En effet, tous les êtres vivants communiquent, que les messages soient transmis par voie visuelle, auditive, chimique, ou électrique. Les végétaux communiquent entre eux, les animaux également, et nous n'échappons pas à cette règle. Le mode de communication que nous avons adopté est majoritairement auditif et visuel. C'est un mode de communication relativement efficace dans notre environnement, qui a été adopté par de nombreuses espèces : reptiles, oiseaux et mammifères, entre autres. En effet, tous les mammifères terrestres possèdent, entre autres, des cordes vocales, des poumons et une langue, et tous utilisent cet instrument vocal afin de communiquer. Il est donc probable que ce mode de communication vocal remonte à notre dernier ancêtre commun, c'est-à-dire bien avant l'apparition de l'Homme. La particularité de la communication parlée, propre à l'espèce humaine, en comparaison avec la communication vocale animale, réside en une différence de degré plutôt que de nature. Nous sommes en effet capables de produire une grande variété de contrastes sonores et nous sommes également capables de combiner ces différents sons pour produire du sens (c'est la fameuse double articulation du langage), en référer à des objets du monde où à des actions. Cette grande capacité à produire et à percevoir des objets sonores distincts, à les catégoriser, à les combiner et à les associer à des représentations abstraites, en fait peut-être la différence la plus notable avec les autres formes de communications non humaines.

Chez l'Homme, l'apprentissage de cette faculté démarre aussi tôt que possible (et ne s'arrête jamais) : dès 3 mois avant la naissance, le système auditif de l'enfant en devenir est fonctionnel et ce dernier est déjà exposé à une ou plusieurs langues et commence déjà à traiter des signaux de parole humaine. Côté production, l'acte de communication est également particulièrement précoce, puisque dès les premières secondes après la naissance, le nouveau-né déclare sa présence par un cri. Dès qu'il en est capable, et même s'il ne dispose que d'un contrôle moteur très rudimentaire, l'enfant utilise ses muscles oro-faciaux pour communiquer oralement. L'appétence pour la parole, l'envie de communiquer, est forte. Il est tentant d'établir de ce point de vue un parallèle entre la parole et la marche, puisque l'enfant est dans un premier temps dans l'incapacité de

marcher de son propre chef à cause du manque de tonus musculaire et probablement également à cause d'un contrôle moteur encore rudimentaire. En observant l'envie de communiquer des jeunes enfants, il serait tentant de penser que s'ils disposaient du tonus musculaire et surtout du contrôle moteur adéquat, ils ne se priveraient pas de cette faculté, et ce dès les premiers instants. Mais la production de la parole est une tâche bien plus complexe que ne l'est la marche, et, comme nous allons le voir, pas uniquement du point de vue du nombre de muscles à contrôler.

En effet, si les deux tâches requièrent le contrôle coordonné d'un grand nombre de muscles, il existe des spécificités de la tâche de production de parole qui augmentent considérablement sa complexité du point de vue du contrôle et de son apprentissage. Une des spécificités de la tâche de production de parole est que les objectifs moteurs sont associés à une tâche cognitive : l'observation du résultat de l'action (particulièrement la satisfaction des objectifs) n'est pas directement observable dans l'espace distal. Contrairement à la marche, dont l'apprentissage n'est pas supervisé (c'est-à-dire que les objectifs moteurs ne sont pas spécifiés à l'apprenant), l'enfant, dans le cadre de la production de la parole, n'a pas directement accès à l'information concernant la réussite de son action dans ses afférences motrices. Dans le cadre de la production de la parole, l'apprentissage est supervisé, les catégories phonologiques d'une langue donnée sont spécifiées à l'apprenant, et ce dernier n'a connaissance de la réussite de son action qu'une fois qu'il a reçu le retour de son interlocuteur. L'apprentissage de la production de la parole ne se met donc pas en place uniquement grâce aux retours sensoriels de nos afférences motrices, mais repose sur l'interaction avec un interlocuteur. Une autre spécificité des mouvements rapides tels que ceux de la production de parole est le problème des délais des retours sensoriels en comparaison avec la rapidité d'exécution des mouvements. Le contrôle, initialement fondé sur les retours sensoriels (contrôle de type *feedback*) s'avère rapidement trop lent et trop instable au vu de la rapidité d'exécution des mouvements, d'où la nécessité de disposer, aussi vite que possible, d'un contrôle prédictif efficace (contrôle de type *feedforward*). Au-delà de ces principales spécificités, il existe un certain nombre de questions propres au cadre du développement de la production de la parole.

Si la production de la parole est sans conteste une activité relevant du contrôle moteur, et, à ce titre, se soumettant aux principes généraux qui régissent les activités motrices, il n'en n'est pas moins vrai que la production de la parole soulève des questions spécifiques. La question de l'apprentissage du contrôle des synergies musculaires associées aux objectifs cognitifs de la tâche de production de parole dépasse en effet le cadre du seul contrôle moteur, et nécessite d'intégrer des connaissances notamment linguistiques et cognitives. En effet, la tâche de production de la parole est hiérarchisée (contrôle moteur, phonétique, phonologie, sémantique, syntaxe), et la prise en compte de chacun de ses différents niveaux et de leurs interactions est nécessaire pour une meilleure compréhension des mécanismes mis en jeu à chacun de ses niveaux. Par exemple, la pression exercée par la nécessité d'acquérir un lexique varié peut influencer la vitesse d'acquisition de l'inventaire phonologique. A l'inverse, un contrôle moteur fin des muscles oro-faciaux est nécessaire pour produire des contrastes phonétiques, qui sont eux-mêmes nécessaires dans le but d'acquérir un lexique. La question de l'émergence des catégories phonologiques et la construction des buts moteurs associés (soulevant à leur tour des questions concernant l'émergence des unités phonologiques, leur nature au cours du développement et sur l'espace de contrôle : auditif, somatosensoriel, articulatoire ou moteur), est une question centrale du développement phonétique. La question des rôles des contrôles de type *feedback* et *feedforward* au cours du développement est également une question centrale, du point de vue du contrôle moteur. Enfin, la question de la construction des représentations sensori-motrices associées aux buts de la production de parole est probablement la question à la fois la plus passionnante et la plus difficile d'accès concernant le développement phonétique.

Devant la multitude des questions soulevées et des angles d'attaque pour étudier la mise en place de cette faculté, nous avons choisi de nous centrer sur la question de l'apprentissage de l'articulation des sons de parole, du point de vue du contrôle moteur. C'est dans le cadre du développement du contrôle moteur pour la production de la parole que ce situe ce travail de thèse. Comment l'enfant apprend à contrôler ses différents muscles, pour atteindre des configurations articulatoires amenant à la modulation du signal acoustique, est en effet une des questions centrales dans le domaine du développement de la production de parole.

Pour comprendre comment l'enfant apprend, petit à petit, à contrôler son instrument vocal, l'observation des productions enfantines, mais aussi le recueil du signal de parole produit par les enfants, constituent une première étape. Une seconde étape vise à observer et à mesurer directement les gestes de parole des enfants, à travers des mesures dites articulatoires. Ce type de mesure peut être effectué pour les articulateurs visibles, tels que les lèvres et la mâchoire, via des dispositifs de capture de mouvements. Ces dispositifs sont adaptables, et déjà adaptés, aux études développementales et ont permis de mieux comprendre le développement du contrôle moteur de ces articulateurs visibles. Mais un articulateur joue un rôle primordial dans la modulation du signal de parole : la langue. Or cet articulateur n'est pas directement observable depuis l'extérieur, c'est un articulateur non visible, et le recueil de données cinématiques concernant ce type d'articulateur constitue un défi technique, particulièrement dans le cadre d'études développementales. Plusieurs dispositifs techniques ont été proposés afin de permettre la mesure des mouvements linguaux : la cinéradiographie, l'électromagnétométrie, ou encore l'échographie. L'échographie linguale fait partie des rares dispositifs techniques qui peuvent être adaptés à la mesure des productions enfantines. C'est grâce à cet outil que nous allons effectuer des mesures directes des gestes linguaux produits par les enfants.

La problématique générale de ce travail de thèse est celle de la production de la parole chez l'enfant, plus spécifiquement celle de l'apprentissage des relations sensori-motrices pour la production de parole et de la maturité du contrôle moteur de la langue. L'objectif de ce travail de thèse est de contribuer à qualifier et à quantifier la maturité du contrôle moteur de la langue pour la production de parole à l'âge de 4 ans, via la mesure articulatoire des mouvements linguaux et la mesure du signal de parole associé, en comparant ces mêmes mesures au sein d'une population adulte. Pour ce faire, nous avons choisi d'étudier deux indices principaux, la stabilité du contrôle moteur, à travers la mesure de la variabilité lors de la répétition d'une même tâche, et la maturité du contrôle moteur, à travers la mesure de l'anticipation de sons de parole dans le cadre de la production d'une séquence de sons de parole. Afin de saisir plus précisément les enjeux de ce travail de thèse et de comprendre les étapes qui ont mené à la satisfaction de nos objectifs de travail, nous proposons le plan de lecture suivant :

Le chapitre 1 constitue un état de l'art et permet d'exposer l'ensemble des connaissances qui ont contribué à l'élaboration de ce travail de thèse et qui permet la compréhension des enjeux de nos travaux. La section 1.1. propose une vue d'ensemble des connaissances portant sur le développement précoce des capacités de perception et de production de la parole. La section 1.2. propose une vue d'ensemble des connaissances portant sur le développement tardif des capacités de perception et de production de la parole. La section 1.3. expose notre cadre théorique, et permet en particulier de comprendre le choix de nos indices et l'interprétation de nos résultats. La section 1.4. traite du problème spécifique de la coarticulation, et en particulier du développement de la coarticulation linguale au cours de l'ontogenèse. La section 1.5. traite de la construction des relations sensori-motrices au cours du développement, en faisant le lien entre études du contrôle moteur du bras et de la main et études portant sur la production (perturbée ou non) de parole. La section 1.6. fait le bilan de cet état de l'art, expose notre problématique et récapitule nos principales hypothèses de travail.

Le chapitre 2 décrit la méthodologie de ce travail expérimental, et expose les outils et méthodes qui ont conduit à l'obtention des résultats. Nous y présentons la population étudiée (section 2.1.), le dispositif d'acquisition des données (section 2.2.), les prétraitements et les analyses des données acoustiques et articulatoires (section 2.3.), ainsi que les analyses statistiques que nous avons effectuées (section 2.4.).

Le chapitre 3 expose les résultats de notre étude expérimentale. Nous y présentons les résultats portant sur la stabilité de la production de voyelles isolées (section 3.1.), ceux portant sur la coarticulation intra-syllabique (section 3.2.), ceux portant sur la coarticulation extra-syllabique (section 3.3.), ceux portant sur les durées des segments et des séquences (section 3.4.) et les relations entre nos principaux indices (section 3.5.).

Le chapitre 4 constitue une discussion, et propose, pour chacun de nos résultats et dans le même ordre que celui proposé dans le chapitre 3, une lecture de ces résultats, notre interprétation de ces résultats et une confrontation avec les études de la littérature.

Dans la dernière partie de cette thèse, nous revenons sur nos principaux résultats et leur interprétation à travers notre cadre théorique, et rappelons les principaux apports de ce travail de thèse.

Chapitre 1

Le développement de la production de la parole

1.1. La parole et son développement précoce

1.1.1. La parole : un objet d'étude complexe

La parole, ou communication parlée, est la faculté de communiquer grâce à un ensemble de sons, produits par le contrôle coordonné de multiples organes et articulateurs (poumons, plis vocaux, langue, lèvres, mâchoire, voile du palais, etc.) mettant en jeu plus d'une centaine de muscles oro-faciaux (Levelt, 1993). Cette activité est probablement l'une des tâches les plus complexes que l'humain accomplisse, du point de vue du contrôle moteur, c'est-à-dire du point de vue du contrôle des muscles par le système nerveux central. Cette faculté, si complexe et passionnante à étudier, qui repose sur de multiples capacités, comme celle de coordonner ses activités musculaires pour produire des sons de parole, mais aussi celle de décoder un flux de parole continu pour comprendre un message, se met pourtant en place rapidement et naturellement au cours du développement ontogénétique[1].

En effet, les nouveau-nés et les jeunes enfants semblent avoir un attrait particulier pour la communication parlée (Masapollo, Polka & Ménard, 2015) et semblent développer cette capacité sans effort apparent. Si la faculté de communiquer avec ses semblables par des vocalisations se développe relativement rapidement chez l'Homme, c'est probablement en partie parce que l'Homme est un animal social, et qu'il existe une forte pression sur la capacité à vivre en groupe. Or la capacité à vivre en groupe dépend fortement de la capacité de ses membres à se reconnaître, à communiquer et à se comprendre afin d'entreprendre des actions communes, ou à se mettre d'accord sur la stratégie à suivre. D'où le fait, probablement, que les capacités à reconnaître des visages (Pascalis et al.,

[1] Le développement ontogénétique, ou ontogenèse, fait référence au développement de l'individu. La phylogenèse traite de l'évolution de l'espèce.

1995 ; Pascalis et al., 2002 ; Damon et al., 2015), donc à identifier les membres de son groupe social, et à communiquer avec ceux-ci, se mettent en place rapidement au cours du développement.

Nous verrons, dans les sections qui suivent, comment, au cours de l'ontogenèse, se développent les capacités de production et de perception de parole chez les nouveau-nés et les jeunes enfants, afin de mieux comprendre les étapes et la chronologie de la mise en place de cette faculté.

1.1.2. Phénomène spécial ou fruit de contraintes évolutives

Puisque la parole est une faculté si complexe, et qu'elle vient pourtant si naturellement aux enfants, certains scientifiques ont émis l'hypothèse que la parole serait spéciale, voire innée : il existerait un module spécialisé dédié à cette faculté dans le cerveau humain (Liberman et al., 1967 ; Whalen & Liberman, 1987), et un ou des gènes spécialisés responsables du développement de cette faculté chez notre espèce (Enard et al., 2009). Le développement fulgurant de cette capacité lors des premières années de vie serait expliqué, selon ce point de vue, par le fait que les individus sont préprogrammés pour cette faculté. C'est la vision innéiste de l'acquisition de la parole, portée entre autres par Chomsky (Chomsky, 2000) et Pinker (Pinker, 1995). A l'opposé de cette vision spécialiste et innéiste de la parole et de son acquisition, il existe un point de vue théorique selon lequel la parole serait un phénomène naturel et non spécial, qui aurait évolué, qui serait le fruit de contraintes physiques, auditives, cognitives, et qui serait le fruit de l'évolution conjointe de plusieurs facultés cognitives (attention, mémoire, capacité de représentation abstraite et de symbolisme, contrôle moteur, etc.) ayant chacune leurs contraintes et limites, et qui contraindraient, chacune à leur manière, façonneraient, mettraient en forme la parole (Schwartz, 2009). Selon ce second point de vue, la parole n'est donc pas spéciale, mais est un objet naturel qui est le fruit de centaines de milliers (millions?) d'années d'évolution, contraint par de multiples éléments, qui la façonnent, et qui s'est « greffé » sur les capacités cognitives existantes. Par exemple, les contraintes biomécaniques et perceptives ont une forte influence sur les choix possibles et efficaces des sons de parole utilisés dans un système phonologique donné et sur la mise en forme de ces systèmes phonologiques (Liljencrants & Lindblom, 1972 ; Schwartz et al., 1997 ;

Schwartz et al., 2012), les limites de mémoire et d'attention exercent quant à elles des contraintes quant à la longueur des mots et des combinaisons de sons de parole. Enfin, la perception de la parole repose sur des mécanismes généraux d'analyse temporelle et fréquentielle du signal acoustique de parole (Stevens & Klatt, 1974), mécanismes qui ne semblent pas spécifiques à la parole humaine, puisqu'ils sont partagés par de nombreuses espèces animales (Kuhl & Miller, 1978). Selon ce second point de vue, c'est la parole qui s'est adaptée à la nature, qui s'est greffée sur les possibilités que cette dernière offrait et qui s'est adaptée aux contraintes qu'elle imposait. Selon ce point de vue, la communication parlée chez l'Homme ne serait que le fruit, contraint par de multiples aspects, de l'évolution de la communication par vocalisation commune à tous les animaux à ventilation pulmonaire (comprenant notamment les mammifères terrestres, mais aussi les oiseaux), et ne comporte rien, de ce point de vue, de spécial. L'évolution de la parole lors de la phylogenèse, de la communication par vocalisations (mais pas seulement, la forme orale de la communication n'étant qu'une facette de la communication, cette dernière faisant également intervenir également les mimiques oro-faciales et la communication non verbale dans son ensemble) à la forme actuelle que l'on lui connaît, a certainement été rendue possible par le développement des capacités cognitives sous-jacentes qui constituent également ses limites (évolution de la capacité d'abstraction, du symbolisme, de la mémoire, de l'attention, du contrôle moteur, etc.). Sans entrer davantage dans ce débat passionnant, nous prenons le parti théorique du second point de vue et considérons la parole comme une sorte de méta-faculté ou d'arché-faculté cognitive, rendue possible par l'évolution des capacités cognitives sous-jacentes sur lesquelles elle repose. Le fait que le réseau neuronal impliqué dans cette faculté soit largement distribué dans le système nerveux central et que cette fonction implique, pour la production comme pour la perception, l'activité de nombreuses régions cérébrales, par ailleurs également impliquées dans d'autres fonctions cognitives, constitue un argument en faveur du second point de vue théorique.

1.1.3. Pourquoi étudier son développement ?

Puisque la parole est un objet d'étude complexe, faisant intervenir de multiples composantes, la compréhension de l'organisation et du fonctionnement de cette capacité est rendue difficile par la complexité même de ce phénomène. L'étude de sa

mise en place et de son développement, pas à pas, étape par étape, du nouveau-né à l'adulte, permet de rendre apparente sa structure, aussi bien physique que cognitive. L'étude de l'acquisition et du développement de la parole peut ainsi nous renseigner sur l'organisation de la parole adulte. En effet, un certain nombre d'avancées théoriques importantes concernant l'organisation de la parole sont issues d'études développementales. Par exemple, concernant la perception, l'étude d'Eimas et al. (1971) a permis de mettre en évidence le phénomène de surdité phonologique, montrant que la parole humaine est construite en fonction de contraintes perceptives universelles, faisant écho à ce que nous évoquions dans la section précédente. L'étude de Werker & Tees (1984), toujours en perception, a permis de mettre en évidence que la perception de la parole est catégorielle et construite par notre expérience ; celle de Meltzoff & Borton (1979) (et Kuhl & Meltzoff, 1982) a permis de mettre en évidence le rôle de la modalité visuelle dans le développement de la perception de la parole. Concernant la production, la théorie de MacNeilage & Davis (1990) formulée à partir de l'observation du babillage, sur lequel nous reviendrons, a permis de mettre en avant le rôle des contraintes biomécaniques universelles, en particulier celles qui s'exercent sur les mouvements mandibulaires, façonnant et contraignant la production de sons de parole, en particulier en ce qui concerne le cadre syllabique et le rythme naturel de la parole. Nous reviendrons en détail sur ces études et sur d'autres, dans les sections suivantes, puisqu'elles ont permis de mieux comprendre la mise en place et le développement de cette faculté au cours de l'ontogenèse, et ont également apporté un éclairage important concernant l'organisation de la parole adulte.

1.1.4. L'émergence et le développement précoce de la parole

En sciences développementales, le développement précoce désigne en général le développement du nouveau-né jusqu'à la fin de la première année de vie. En ce qui concerne le développement de la production et de la perception de la parole, on parle généralement de développement précoce de la naissance à l'émergence des premiers mots, voire à la fin de la seconde année de vie. Le développement, de 2 à 6 ans, est défini comme la période de la petite enfance ; de 6 à 10 ans, c'est la période de l'enfance ; et au-delà, la préadolescence (10-14 ans) puis l'adolescence (14-20 ans) et enfin l'âge adulte. De nombreux chercheurs se sont intéressés au développement précoce, jusqu'aux premiers

mots et aux premières phrases. Mais en ce qui concerne la production de parole, le développement est loin d'être achevé à un an et demi, et le chemin vers une parole fluente comme celle de l'adulte est encore long. Nous aborderons, dans les sections qui suivent, le développement précoce des capacités de perception et de production de parole des nouveau-nés à la fin de la première année de vie.

1.1.4.1. Le développement précoce de la perception

Le premier élément que nous tenons à mettre en avant concernant les capacités précoces de perception de parole des nouveau-nés est l'existence, probablement même avant la naissance, du réseau neuronal fonctionnel impliqué dans la perception de la parole. Dehaene-Lambertz et al., (2002) ont mis en évidence l'existence de ce réseau fonctionnel chez des nouveau-nés âgés de 3 mois seulement, en faisant écouter à ces nouveau-nés de la parole naturelle et de la parole temporellement inversée, afin de mettre en évidence les activations neuronales spécifiques à la perception de la parole, en Imagerie par Résonance Magnétique fonctionnelle (IRMf). Cette étude a montré de nombreuses similarités entre les activations neuronales provoquées par la perception de la parole chez l'adulte et chez le nouveau-né, soulignant que les éléments indispensables à la perception de la parole, d'un point de vue cognitif, sont déjà en place, et ce de façon très précoce.

De nombreuses études comportementales ont également mis en avant les capacités surprenantes des nouveau-nés concernant la perception de la parole, notamment grâce aux paradigmes expérimentaux de la préférence du regard ou de la succion non nutritive. Puisque les nouveau-nés ne sont pas en mesure de fournir de réponses verbales, le temps qu'ils passent à fixer une image plutôt qu'une autre, ou le nombre de fois qu'ils tètent en réponse à un stimuli ou à un autre, sont des indicateurs permettant notamment d'effectuer des tâches de discrimination.

Ainsi, dès la naissance, les nouveau-nés montrent une préférence pour la voix de la mère (DeCasper & Fifer, 1980) et une préférence pour la langue maternelle (Mehler et al., 1988), soulignant leurs capacités à percevoir la prosodie, et particulièrement le suivi de la fréquence fondamentale (on verra que cette capacité influence directement leurs premières productions, dans la section suivante concernant la production). Ils montrent

également, dès la naissance, une préférence pour des sons de parole plutôt que pour des sons de non parole (Jusczyk et al., 1977). La perception du rythme est également très précoce, puisque les nouveau-nés sont capables de discriminer différentes classes rythmiques de langues dès la naissance (Pike, 1946 ; Abercrombie 1967 ; Nazzi et al., 1998).

Dès 1 ou 2 mois, les nouveau-nés sont capables de discriminer des contrastes phonétiques portant sur le VOT ou sur la nasalité (Eimas et al., 1971 ; Trehub & Rabinovitch, 1972), mais aussi des contrastes phonétiques portant sur le lieu d'articulation (Eimas, 1974 ; Miller & Morse, 1976 ; Trehub, 1976 ; Jusczyk, 1977), le mode d'articulation (Eimas & Miller, 1980 ; Hillenbrand et al, 1979), ou encore les contrastes entre voyelles (Trehub, 1973 ; 1976 ; Kuhl 1979 ; 1983 ; Grieser & Kuhl, 1989 ; Kuhl et al., 1992), y compris pour des contrastes absents de leur langue maternelle. Ces expériences ont permis de mettre en évidence que les nouveau-nés ont une perception catégorielle des sons de parole, comme les adultes, soulignant qu'une réalité physique constituant un continuum peut amener, en termes de représentations phonétiques, à une perception catégorielle, par des effets de seuils et de saturation (cf. Liberman, Delattre & Cooper (1958) pour cette notion de perception catégorielle et Stevens (1989) pour les notions de seuils quantiques et de saturation amenant à l'origine d'une partie de l'organisation de nos systèmes phonologiques), et que ces frontières catégorielles sont spécifiques à une langue donnée, donc construites par l'apprentissage (Abramson & Lisker, 1970).

Entre 2 et 5 mois, les nouveau-nés sont capables de catégorisation en dépit de variations dans l'intonation, sont capables de reconnaitre une syllabe dans des énoncés différents, et sont capables de reconnaître leur prénom (Mandel et al., 1995).

Dès 6-8 mois, les enfants sont spécialisés pour reconnaitre des contrastes appartenant à leur langue maternelle, et deviennent insensibles aux contrastes absents de leur langue maternelle : c'est le début de la surdité phonologique. Cela souligne que cet âge est une étape clé vers la spécialisation pour la langue maternelle (LM). Cela indique également que les frontières catégorielles sont apprises, et sont spécifiques à chaque langue. On voit ainsi, en perception, émerger des catégories propres à une langue maternelle donnée, dès 11 mois pour les consonnes (e.g. Eimas et al., 1971 ; Trehub & Rabinovitch,

1972) et dès 6 mois pour les voyelles (Kuhl et al., 1992). Ce phénomène appelé surdité phonologique a été mis en évidence par Werker & Tees (1984) qui ont testé, chez des enfants âgés de 6, 10 et 12 mois, les capacités de discrimination portant sur des contrastes appartenant à leur LM ou n'appartenant pas à leur LM. Leurs résultats ont permis de mettre en évidence que les enfants sont capables, à 6 mois, de discriminer des contrastes portant sur des sons de parole appartenant et n'appartenant pas à leur LM, alors que les enfants de 12 mois ne sont plus capables de discriminer les contrastes phonétiques qui n'appartiennent pas à leur LM. Cette étude montre donc qu'une spécialisation vers la langue maternelle s'opère entre 6 et 12 mois.

De façon intéressante, l'étude de Bruderer et al. (2015) a mis en évidence que l'apprentissage des contrastes phonétiques, en perception, chez des enfants de 6 mois, peut être perturbé si la mobilité buccale est entravée. Ce résultat mérite d'être souligné, car il fait partie du maigre ensemble de travaux montrant que la production influence aussi la perception.

Dès 8-10 mois, la spécialisation vers la langue maternelle se fait plus forte, avec notamment la préférence pour les formes de mots respectant l'accentuation de la LM et la préférence pour les formes respectant les contraintes phonotactiques de la LM (préférence qui émerge entre 6 et 9 mois selon Jusczyk et al., 1993). A cet âge, les enfants sont capables d'utiliser les propriétés distributionnelles du signal de parole et effectuent des calculs de probabilité concernant les co-occurences de syllabes dans des mots (Saffran et al., 1996). A cet âge, ils sont également capables de reconnaitre un mot, même inconnu, dans un flux continu de parole (Kooijman, Hagoort & Cutler, 2005 ; Junge, Cutler, & Hagoort, 2014).

Entre 10 et 13 mois, les enfants sont capables de détecter les frontières entre les mots (dès 3 jours, ils semblent capables de détecter les frontières prosodiques, d'après Christophe et al., 1994), la surdité phonologique est complète, ils sont capables d'associer quelques mots à leur sens en contexte, et apprennent de nouveaux mots par association à des référents. Leurs capacités de perception leur offrent alors tous les atouts permettant l'apprentissage d'un vocabulaire grandissant, leur permettant réellement d'entrer dans la communication parlée.

1.1.4.2. Le développement précoce de la production

S'il est possible de décomposer le développement des capacités de production de parole des jeunes enfants en étapes successives, il faut cependant prendre cette décomposition avec prudence. Par exemple, Koopmans-Van Beinum & van der Stelt (1986) ont étudié les premières vocalisations de l'enfant et ses premiers mouvements articulatoires entre 1 et 18 mois, en distinguant ce qui relève de la phonation (contrôle des poumons et de la vibration des cordes vocales, ou gestes glottiques) de ce qui relève des articulations supra-glottiques (mouvements produits par les lèvres, la langue, le velum ou la mandibule) et ont mis en évidence plusieurs étapes clés lors du développement.

Ainsi, la première étape de ce développement consiste en un contrôle volontaire des poumons et de la mise en tension des cordes vocales, amenant à la vibration des cordes vocales (que l'on sait être fortement dépendante du niveau de pression sous-glottique, et donc du contrôle du débit d'air expiré), autour de 2-3 mois, puisqu'elle devient rythmée à cette étape, alors qu'elle est continue avant 2 mois ([aaa?] puis [a?a?a?]). L'entrée dans l'ère de la communication vocale se fait effectivement dès les premières secondes après la naissance. Si le premier cri d'un nouveau-né à la naissance n'est pas une prouesse de contrôle moteur, il peut être en revanche révélateur de l'influence précoce de la langue maternelle sur les productions des nouveau-nés. Mampe et al. (2009) ont montré que les cris de nouveau-nés français et allemands sont influencés par leur langue maternelle, en particulier par le patron prosodique. En effet, le français et l'allemand sont respectivement une langue iambique et une langue trochaïque, mettant l'accent sur la dernière ou la première syllabe, et les cris des nouveau-nés respectent ces patrons prosodiques. Cet exemple semble montrer la très forte et très précoce influence de la perception sur la production.

Puis apparaissent les premiers gestes supra-glottiques entre 3 et 5 mois, mais ces gestes restent imprécis et présentent des constrictions larges ([aRe]). Le contrôle de la coordination oro-pharyngée, c'est-à-dire la coordination entre gestes glottiques et supra-glottiques, apparaît généralement entre 5 et 7 mois (Lalevée & Vilain, 2008). L'étude de Lalevée & Vilain (2008), portant sur le VOT (*Voice Onset Time* ou délai d'apparition du voisement) montre que les enfants semblent maîtriser dès 6 mois la coordination entre gestes glottiques et supra-glottiques.

Aux alentours de 7-8 mois se met en place une étape clé du développement phonétique : c'est le babillage canonique ([bababa, dadada]). Ce dernier est défini comme l'émergence de proto-syllabes, rendue possible par le contrôle d'une série de mouvements d'ouverture et de fermeture du conduit vocal. A ce stade, « le bébé est alors capable de coordonner les gestes glottiques et supra-glottiques à l'intérieur d'une unité rythmique, ce qui constitue la structure de base de la communication humaine » d'après Lalevée et Vilain (2004). Dans le cadre de la théorie « Frames, then Content » (MacNeilage, 1998), il est proposé que le cadre proto-syllabique, cette structure de base du langage, résulte de l'oscillation mandibulaire. A ce stade et selon cette théorie, la mandibule serait le seul articulateur contrôlé volontairement, contrairement aux autres articulateurs supra-glottiques pour lesquels le contrôle serait davantage complexe et ne serait maîtrisé que plus tardivement, et le geste d'ouverture-fermeture de la mandibule serait responsable de cette structure proto-syllabique. Les successions d'ouverture et de fermeture du cycle mandibulaire (l'ouverture correspondant aux productions vocaliques et la fermeture aux productions consonantiques), mèneraient à la structure de type CVCVCV du babillage, soit à l'alternance continue de consonnes et de voyelles, qui constitue l'une des structures de base de la parole adulte et une étape universelle du babillage. Selon cette théorie, ce sont donc les contraintes biomécaniques universelles, en particulier celles qui s'exercent sur l'oscillation mandibulaire, qui donnent à la parole humaine sa structure de base et qui contraignent les premières productions. Plusieurs données expérimentales viennent appuyer ce postulat. A ce stade, les lèvres ne semblent pas être contrôlées volontairement, puisqu'elles viennent s'écraser passivement l'une contre l'autre pour produire une occlusion bilabiale (Munhall & Jones, 1998). Cependant, quelques critiques ont été formulées à l'encontre de cette théorie (notamment Whalen et al., 2011 ; Whalen et al., 2012). En effet, il semblerait que le nouveau-né dispose, dès les premiers jours de vie, d'un contrôle actif des lèvres, de la mandibule mais aussi de la langue, afin de téter et de déglutir. Même si ce contrôle est très imprécis voire grossier, nous ne pouvons pas nier qu'il existe. Cependant, si un contrôle volontaire des articulateurs tels que les lèvres et la langue existe, même rudimentaire, il pourrait ne pas être, à ce stade, suffisant pour produire des contrastes phonétiques. Enfin, l'existence de ce contrôle rudimentaire des lèvres et de la langue ne remet pas en cause l'apport principal de la théorie F/C de

MacNeilage : le fait que ce soit bien l'oscillation mandibulaire qui soit à l'origine de ce cadre proto-syllabique et de la contrainte universelle du rythme de la parole.

Par ailleurs, tous les enfants n'ont pas la chance de disposer de la modalité auditive, et il semblerait que les enfants sourds apprenant la langue des signes passent également par une étape partageant certaines caractéristiques du babillage de la communication parlée, et suivant le même calendrier développemental (Petitto & Marentette, 1991 ; Petitto et al., 2001).

Puis, aux alentours de 8-10 mois, les lieux d'articulation deviennent plus variés et un début de contrôle volontaire de la langue semble émerger ([badi]). Sur cette question précise de l'indépendance du contrôle de la langue, qui serait, au stade du babillage canonique, portée passivement par la mandibule, l'étude de Canault (2007) apporte un éclairage supplémentaire. Canault (2007) montre en effet qu'entre 8 et 12 mois, les mouvements de la langue se dissocient de l'influence mandibulaire. En d'autres termes, vers 8 mois, la langue est plutôt passive et portée par la mandibule, puis avec l'âge, on observe davantage de mouvements antéropostérieurs, indépendants de l'influence de la mandibule.

Enfin, aux alentours de 12 mois, les premières lexies (ou pseudo-mots) sont produites. Le contenu segmental est alors varié, les mouvements des articulateurs supra-glottiques sont donc différenciés et permettent la production de contrastes phonétiques. C'est donc véritablement au cours de cette première année de vie que se mettent en place les bases fondamentales de la communication parlée. Mais si, à un an, l'enfant est capable de produire des sons contrastés et de produire des séquences de sons rythmés, qui partagent les caractéristiques de base de la parole adulte, il ne dispose ni de l'inventaire phonétique complet de sa (ses) langue(s) maternelle(s) ni d'un contrôle moteur aussi fin que celui de l'adulte, et le chemin vers une maîtrise de ses articulateurs similaire à celle de l'adulte est encore long, comme nous allons le voir dans la section suivante. En réalité, on peut plutôt considérer que cette première année de vie pose les bases et permet d'entrer dans la première étape de production de ce que l'on peut effectivement appeler de la parole.

Ce développement chronologique par étapes successives démontre bien l'idée défendue dans la section 1.1.3, puisqu'il permet de mettre en évidence différentes composantes de la production de parole, à savoir ici la phonation (vibration des cordes vocales générant un signal périodique) et l'articulation (modulation de ce signal périodique en générant différentes configurations du conduit vocal, grâce aux articulateurs supra-glottiques). Mais il faut prendre cette décomposition avec prudence, car au cours du développement, même s'il est possible d'identifier des étapes clé, plusieurs articulateurs entrent en jeu et leurs contrôles ne sont jamais strictement indépendants et successifs. Ainsi, si l'on peut donner l'impression, à la lecture de cette section, que le contrôle des articulateurs s'effectue de manière indépendante et successive, et que les cordes vocales sont maîtrisées d'abord, puis la mandibule (et le velum), puis les lèvres, puis la langue, rien n'est si simple en réalité. S'il est vrai que le contrôle des cordes vocales est plus précoce que celui de la mandibule ou celui de langue, établir une chronologie du contrôle, articulateur par articulateur, reste délicat.

1.1.4.2.1. Le développement du contrôle, articulateur par articulateur

Concernant le contrôle du velum, Rossato et al. (2003) ont proposé une théorie expliquant qu'un contrôle partiel du velum pourrait permettre de produire la distinction entre consonnes orales et nasales, alors qu'un contrôle complet du velum serait nécessaire à la production du contraste entre voyelles orales et nasales. Cela permet d'expliquer le nombre important de consonnes nasales dans les inventaires phonétiques des jeunes enfants. Un contrôle partiel du velum, permettant la production de consonnes nasales, serait présent dès 6 mois, alors que le contrôle complet de ce dernier n'apparaîtrait qu'autour de 11 mois, d'après Lalevée & Vilain (2004).

Concernant le contrôle des lèvres, ce dernier ne semble pas être mis en place au début du babillage (Munhall & Jones, 1998). Le contrôle volontaire des lèvres semble progressif, puisque les patrons des mouvements des lèvres à 1 et 2 ans montrent un contrôle volontaire, mais imprécis (manquant de consistance) et que ces patrons ne se rapprochent de ceux de l'adulte que vers 6 ans, d'après Green et al. (2002). Mais si l'on en croit d'autres études (e.g. Smith & Goffman, 1998) le contrôle des lèvres et de la

mandibule ne serait similaire à celui de l'adulte que tard durant l'adolescence, soulignant que le développement du contrôle moteur est particulièrement progressif et lent.

Concernant le contrôle de la langue, ce dernier semble émerger entre 8 et 14 mois (Canault, 2007) mais l'établissement d'un profil ontogénétique clair de ce contrôle est relativement délicat, puisque la langue est un articulateur caché, dont la mesure des mouvements n'est pas évidente. Nous verrons dans la section 1.2 que le développement de ce contrôle est très lent et ne semble être semblable à l'adulte qu'au début de l'adolescence pour certains (Kent, 1984), à la fin de l'adolescence pour d'autres (Smith, 2006 ; 2010).

1.1.4.2.2. Le développement des contrastes phonétiques

Comme nous venons de le voir, du point de vue du contrôle moteur, le contrôle indépendant de chacun des différents articulateurs supra-glottiques et la coordination entre articulateurs se mettent en place progressivement au cours de la première année de vie. L'acquisition de l'une des structures de base de la parole humaine, à savoir l'alternance entre consonnes et voyelles au sein d'une unité de base rythmée, la syllabe CV, est acquise autour de 7-8 mois, mais la diversification du contenu segmental n'est observée que plus tardivement, vers 10-12 mois, et correspondrait à l'émergence d'une parole organisée autour du segment (MacNeilage & Davis, 2000). Les premières étapes, comprenant le babillage, sont des étapes universelles du développement. Ce n'est qu'après ces étapes que l'enfant, en production, se spécialise vers une langue maternelle donnée.

Si un certain nombre d'éléments, tels que l'attention conjointe, l'intention de communication, l'appétence pour le langage, c'est-à-dire cette volonté de communiquer, la prosodie, le rythme, sont maîtrisés précocement, et sont des éléments indispensables à l'entrée dans la communication parlée, la capacité à contraster les différents sons de parole se met en place plus tardivement, et peut être vue comme la finalité de ce processus. Le contenu segmental commence à être diversifié vers 10 mois et l'inventaire phonétique (c'est-à-dire l'ensemble des sons de parole produits, correspondant aux catégories phonologiques d'une langue donnée) n'est complet qu'entre 2 et 5 ans selon les langues (e.g. Gangji et al. (2015) pour le swahili ; Naidoo et al. (2005) pour le zoulou ;

Ferrante et al. (2008) pour le portugais du brésil). Pendant cette période, l'enfant apprend à produire des sons de parole contrastés, appartenant à différentes catégories phonétiques, au début avec beaucoup de superposition entre les catégories phonétiques et beaucoup de variabilité, puis, graduellement, produisant des contrastes de plus en plus importants et de plus en plus stables, jusque la production d'une parole parfaitement intelligible.

Concernant les capacités précoces des enfants à produire des sons de parole contrastés, l'étude de Kuhl & Meltzoff (1996), étudiant les capacité d'enfants de 12, 16 et 20 semaines à produire les voyelles /i/, /a/ et /u/ lors d'une tâche d'imitation, a mis en évidence que cette capacité augmente avec l'âge, et que les enfants de 20 semaines produisent des contrastes plus importants que les enfants plus jeunes. Cette étude met en avant le fait que le contrôle des différents articulateurs tels que la langue ou les lèvres, n'est pas inexistant à l'âge de 5 ou 6 mois, et permet déjà la production de sons de parole contrastés (la production de /i/ et /a/ requièrent un contrôle actif de la langue et de la mandibule, et la production de /u/ nécessite un contrôle de la langue et des lèvres). Mais ces résultats sont à prendre avec précaution, car les voyelles /i/, /a/ et /u/ sont des voyelles cardinales et représentent donc les contrastes phonétiques les plus importants, et si les productions vocales des enfants se recoupent moins à 20 semaines qu'à 12, cela ne veut pas dire pour autant que leur intelligibilité est bonne. Cependant, cette étude a le mérite de souligner l'importance de l'imitation dans le processus d'apprentissage, puisque c'est l'adulte qui guide le nouveau-né vers une spécialisation pour sa langue maternelle, donc vers un système phonologique donné. En effet, l'apprentissage est supervisé et les cibles phonétiques sont spécifiées par les locuteurs, guidant l'apprentissage de l'enfant vers une langue cible en faisant émerger des prototypes perceptuels correspondant aux différents sons de parole.

Mais la spécialisation vers la langue maternelle et l'apprentissage de la production de tous les contrastes phonologiques d'une langue donnée est une tâche complexe, nécessitant un contrôle fin des articulateurs et de leur coordination spatiale et temporelle, et n'est maîtrisée que tardivement. Par exemple, l'étude d'Eguchi & Hirsh (1969) a montré, en étudiant la perception par des adultes, de voyelles produites par des enfants âgés entre 3 et 13 ans, que l'intelligibilité des productions augmente avec l'âge.

Ainsi, pour toutes les voyelles de l'anglais américain confondues (8 voyelles sont étudiées), le taux de reconnaissance des voyelles produites par des enfants de 3 ans n'est que de 54%. Selon les auteurs, l'intelligibilité des voyelles augmente avec l'âge jusqu'à 7 ans, soulignant ainsi que le contrôle de la production, pour les voyelles tout au moins, n'est pas maîtrisé totalement avec cet âge.

Il faut néanmoins souligner que de nombreux facteurs influencent le développement de l'inventaire phonétique. Hallé & Cristia (2012) proposent par exemple que la nécessité d'acquérir un lexique (un vocabulaire) varié exerce une pression sur l'acquisition de l'inventaire phonétique.

Zmarich & Bonifacio (2005) ont étudié l'évolution des inventaires consonantiques de 13 enfants italiens entre 18 et 27 mois, et observent que les consonnes occlusives apparaissent avant les fricatives (Locke, 1983 ; ce qui est cohérent avec la théorie F/C de MacNeilage, mais aussi avec la vision selon laquelle une occlusion complète est plus facile à réaliser qu'une constriction partielle dont on doit maitriser le degré de constriction, ce qui explique la forte présence des consonnes occlusives dans l'inventaire du babillage, mais aussi des langues du monde), et que les lieux d'articulation antérieurs sont favorisés dans les inventaires phonétiques des jeunes enfants. Les inventaires consonantiques des jeunes enfants s'étoffent entre 18 et 27 mois, ce qui montre que le contrôle de leurs articulateurs supra-glottiques et de leur coordination s'affine durant cette période. Selon les langues maternelles, mais aussi selon le développement de l'individu, on estime que l'inventaire phonétique des enfants est complet entre 2 et 5 ans (Gangji et al., 2015 ; Naidoo et al., 2005 ; Ferrante et al., 2008). Cependant, le contrôle de certains contrastes sonores, comme les fricatives ou le /R/ de l'anglais, nécessitent un contrôle moteur fin et ne sont acquis que plus tardivement. A ce stade, le contrôle indépendant de tous les articulateurs est acquis, et la coordination entre ces articulateurs l'est également, mais la finesse de ce contrôle est sans commune mesure avec celle de l'adulte, comme nous le verrons dans la section suivante.

1.1.5. En somme...

Pour résumer, les grandes étapes de la capacité de perception au cours de la première année de vie sont : à la naissance, les capacités de sensibilité à la langue maternelle et à la voix maternelle, autour de 2 mois, les capacités de perception catégorielle portant sur divers contrastes, la segmentation, autour de 6-8 mois, des unités de parole, la sensibilité aux régularités phonotactiques de la LM, puis la spécialisation vers la LM et la surdité phonologique autour de 10 mois, et enfin la reconnaissance de mots familiers et la possibilité d'apprendre de nouveaux mots, en relation avec un référent et en fonction du contexte, entre 10 et 13 mois. Concernant le volet production, les grandes étapes sont le contrôle de la phonation, vers 2 mois, puis le contrôle de la coordination oro-pharyngée, entre 3 et 5 mois, le babillage canonique vers 7 mois, le babillage varié vers 8-10 mois, l'apparition d'un contenu segmental varié, puis l'apparition des premières lexies autour de 12-13 mois. Les grandes étapes du développement phonétique des capacités de production et de perception de parole au cours de la première année de vie ont été résumées, entre autres par Kuhl (2004). La figure 1.1.1. ci-dessous récapitule les principales étapes du développement phonétique précoce.

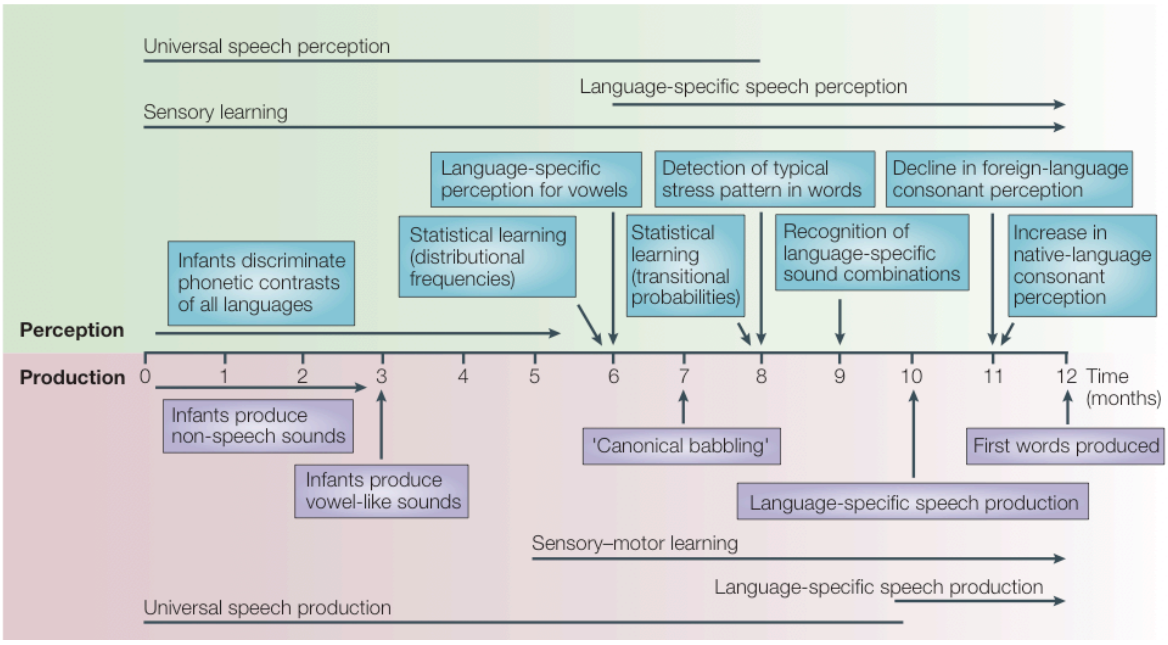


Figure 1.1.1. Récapitulatif des étapes universelles de perception et de production de parole au cours de la première année de vie, d'après Kuhl (2004).

En résumé, comme nous avons pu l'observer, les capacités précoces de perception et de production de parole des nouveau-nés ont fait l'objet d'un nombre important d'investigations. En revanche, comme nous allons le voir dans la section suivante, les capacités de perception et de production qui concernent le développement après la première année de vie n'ont pas fait l'objet d'autant d'études expérimentales. Le développement tardif des capacités de perception et de production de parole est un sujet pour lequel nous disposons de moins de données empiriques, et pour lequel il est donc plus difficile d'établir un développement chronologique, par étapes successives, comme nous l'avons fait pour le développement précoce. Et pourtant, rien n'est fini à la fin de la première année de vie, quand apparaissent les premiers mots. C'est plutôt le début d'un formidable apprentissage, qui va s'étendre jusqu'au début où à la fin de l'adolescence, et dont la première année de vie n'a fait finalement que jeter les bases.

1.2. Le développement tardif de la production de la parole

De nombreuses études se sont attachées à décrire et à comprendre le développement précoce de la production de la parole, car c'est au cours des premiers mois de vie que ce développement est le plus spectaculaire. Il n'en reste pas moins que ce développement, si impressionnant soit-il des premières vocalisations aux premiers mots (aux alentours de 18 mois), en passant par le babillage canonique (autour de 7 mois), n'est absolument pas achevé à la fin de la première année de vie. En réalité, de nombreux chercheurs considèrent que les productions des touts petits font partie d'une forme préalable de parole, mais que l'on ne peut pas considérer comme de la parole à proprement parler, tant son organisation et ses manifestations sont différentes de la parole adulte. En effet, on peut voir dans le babillage et les premiers mots une étape préalable à la mise en place de la production de la parole, et considérer que l'entrée dans l'ère de la parole ne se fait qu'à partir du moment où l'enfant devient capable de produire et de percevoir tous les sons de son (ses) système(s) phonologique(s) et de planifier et d'exécuter des séquences de sons ayant une signification ou une visée communicative. L'étude du développement tardif de la production de parole constitue donc, de ce point de vue, l'observation de la mise en place réelle et du développement de cette faculté, par opposition à l'étude de la mise en place des prémices de cette faculté, que constituent les études portant sur le développement précoce.

Nous décrirons, dans la section suivante, le développement des capacités de production et de perception de la parole au cours de la petite enfance (entre 2 et 6 ans) et de l'enfance. Comme nous le verrons, la production de la parole est une tâche complexe et délicate, pour laquelle le contrôle s'affine jusqu'à la fin de l'adolescence.

1.2.1. Les études acoustiques du développement tardif de la production

De nombreuses études acoustiques se sont attachées à mesurer les propriétés du signal acoustique des productions vocaliques enfantines. Les voyelles étant des objets acoustiques plus faciles à caractériser dans le domaine physique que les consonnes, on comprendra que la plupart des études développementales se soient attachées à mesurer les caractéristiques physiques des objets vocaliques au cours du développement. En

effet, la mesure des formants F1, F2 et F3 offre à elle seule une bonne caractérisation des objets vocaliques, et la mesure de ces formants peut être effectuée à partir d'une étude spectrographique du signal de parole recueilli, à l'aide d'une transformée de Fourier. D'autres mesures sont également faciles d'accès, par exemple la fréquence fondamentale, ou F0, correspondant à la vibration des cordes vocales, ou encore la durée des vocalisations.

Historiquement, la première étude acoustique offrant des mesures de formants portant sur des voyelles produites par des enfants, en comparaison avec les productions d'hommes et de femmes adultes est celle de Peterson & Barney (1952). Cette étude compare les valeurs de F1, F2 et F3 des voyelles /i a æ u/ de 15 enfants de 9 ans, à celles de 33 hommes adultes et de 28 femmes adultes. Depuis, de nombreuses études acoustiques se sont attachées à décrire l'évolution des valeurs formantiques au cours du développement, et ce de façon plus systématique. Par exemple, l'étude d'Eguchi & Hirsh (1969) porte sur les productions de 84 enfants âgés de 3 à 13 ans, et celle de Lee, Potamianos & Narayanan (1999) porte sur les productions de 436 enfants âgés de 5 à 18 ans. D'autres études acoustiques apportent des informations quant aux productions d'enfants moins âgés, comme celle de McGowan et al. (2014) qui s'intéresse au développement entre 30 et 48 mois, ou encore celle de Lyasko & Grigor'ev (2015) qui porte sur les productions d'enfants âgés de 3 mois à 7 ans.

Outre les travaux mentionnés au paragraphe précédent, on dénombre une trentaine d'études acoustiques portant sur les productions vocaliques au cours du développement, parmi lesquelles Kent & Forner (1979) ; Buhr (1980) ; Kent & Murray (1982) ; Pentz & Gilbert (1983) ; Hodge (1989) ; de Boysson-Bardies et al. (1989) ; Childers & Wu (1991) ; Zahorian & Jagharghi (1993) ; Busby & Plant (1995) ; Higawara (1997) ; Hillenbrand & al. (1995) ; Sussman et al. (1996) ; Sussman et al. (1999) ; Kuhl & Meltzoff (1996) ; Gilbert, Robb & Chen (1997) ; Robb, Chen & Gilbert (1997) ; Matyear et al. (1998) ; Whiteside & Hodgson (2000) ; Assman & Katz (2000) ; Perry, Ohde & Ashmead (2001) ; Nijland et al. (2002) ; Casal et al. (2002) ; Pettinato et al. (2016).

Ces études montrent une diminution des valeurs de formants (F1 et F2) au cours du développement, ainsi qu'une diminution de la fréquence fondamentale avec l'âge et un

raccourcissement de la durée des segments vocaliques, mais aussi consonantiques (voir Lee, Potamianos & Narayanan (1999) pour l'évolution de la durée des voyelles entre 5 et 18 ans et l'évolution de la durée de la fricative /s/ entre 8 et 18 ans). Elles montrent également une diminution de la variabilité avec l'âge.

La plupart de ces données acoustiques a été synthétisée dans la méta analyse de Vorperian & Kent (2007), qui a compilé les résultats de 14 de ces études de la littérature. Ces données acoustiques ont également été confrontées à des travaux de modélisation (i.e. Ménard et al., 2007 ; Boë et al., 2013), qui permettent, puisque les produits du modèle sont en cohérence avec les données empiriques, d'interpoler les valeurs de formants à tout âge, y compris aux âges pour lesquels on ne dispose pas de données expérimentales. Enfin, les données acoustiques concernant les valeurs de formants sont en accord avec la théorie (Fant, 1975 ; Mol, 1970), puisque la fréquence des résonances du conduit vocal sont globalement inversement proportionnelles à la taille totale de ce conduit vocal : un conduit vocal de petite taille produit des résonances de fréquences élevées, alors qu'un grand conduit vocal produit des résonances aux fréquences plus basses.

L'ensemble de ces études montre l'évolution de la fréquence fondamentale, des formants et de la durée des segments vocaliques et consonantiques avec l'âge. La diminution de la durée des segments avec l'âge peut être interprétée comme un indice de maturité motrice, puisque les enfants dépendraient de moins en moins des retours sensoriels avec l'âge, et se reposeraient davantage sur la composante prédictive de la production de la parole, ayant appris les conséquences acoustiques de leurs manœuvres articulatoires. Mais les résultats concernant l'évolution de la F0 et des valeurs des formants nous renseignent uniquement sur l'évolution de l'Espace Vocalique Maximal (Boë, 1989) et ne nous renseignent pas directement sur les capacités des enfants à produire des sons de parole et à contraster ces sons de parole. Pour cela, une mesure de l'écart type des valeurs de formants, et ce pour chaque voyelle, est nécessaire. Parmi ces études, très peu ont reporté cette variabilité, ou se sont directement posé la question de l'apprentissage des contrastes phonétiques et de la stabilité du contrôle moteur sous-jacent.

L'étude de McGowan et al. (2014), qui porte sur le développement des productions vocaliques d'enfants âgés de 30 à 48 mois, s'intéresse précisément à cette question : les catégories vocaliques sont-elles stables en production et se recoupent-elles ? A partir de quel âge se met en place un système vocalique stable en production ? Leurs résultats montrent qu'à 30 mois, soit 2 ans et demi, le système vocalique est déjà en place et les catégories vocaliques ne se recoupent pas. L'évolution entre 30 et 48 mois est marginale, indiquant que le système vocalique et les catégories vocaliques se mettent donc en place avant 2 ans et demi.

Concernant les consonnes, l'étude de l'évolution de l'inventaire consonantique (Zmarich & Bonifacio, 2005) montre qu'à 18 mois, soit un an et demi, seule la production des consonnes plosives est maitrisée, et l'apparition des consonnes fricatives commence entre 21 et 24 mois. Le contrôle des fricatives étant particulièrement délicat, il semble être maîtrisé relativement tardivement. L'étude de Li (2012), qui porte sur l'apprentissage, en Anglais et en Japonais, du contraste /s/ ~ /ʃ/, montre qu'à 2 ans, les enfants produisent des fricatives, mais que ces productions ne sont pas contrastées, différenciées. Entre 2 et 5 ans, leurs productions deviennent de plus en plus contrastées, différenciées. Mais cet apprentissage n'est pas achevé à 5 ans, puisque, comme le montrent Romeo, Hazan & Pettinato (2013), la différenciation acoustique du contraste /s/ ~ /ʃ/ n'est pas réalisée de manière similaire à l'adulte, même à l'âge de 13 ans.

En résumé, beaucoup d'études acoustiques se sont attachées à mesurer l'évolution des productions vocaliques au cours du développement, et ont permis de caractériser l'évolution de l'Espace Vocalique Maximal. Cependant, l'étude des seuls indices de la fréquence fondamentale, des valeurs de formants, et de la durée des segments ne nous renseignent pas sur tous les aspects de l'apprentissage, notamment sur l'apprentissage des contrastes phonétiques en production, et sur la maturation du contrôle moteur pour la production de parole. La production d'objets vocaliques qui correspondent aux catégories d'une langue maternelle cible vont émerger d'une interaction complexe entre l'évolution des capacités de perception, de discrimination et de catégorisation des objets vocaliques et l'évolution des capacités motrices, le tout sous l'influence de facteurs de plus haut niveau, notamment linguistiques et cognitifs.

Concernant la maturation du contrôle moteur pour la production de parole, certaines études acoustiques apportent tout de même des éléments de réponse. Concernant le développement des contrastes vocaliques, l'étude de Kuhl & Meltzoff (1996) montre que les enfants sont capables, à 20 semaines, de produire des contrastes vocaliques en imitant un adulte. L'étude de McGowan et al. (2014) montre que le développement des contrastes vocaliques et la mise en place du système vocalique complet s'effectue avant 2 ans et demi (30 mois). Concernant le développement des contrastes consonantiques, ces derniers apparaissent dès 8-10 mois concernant les plosives, dès 21 mois concernant les fricatives (d'après des transcriptions phonétiques et non des mesures acoustiques), mais le contrôle moteur sous-jacent à ces productions, suffisant pour produire un contraste, est encore très approximatif et s'affinera encore pendant de nombreuses années (Romeo, Hazan & Pettinato, 2013). Notons qu'il existe peu d'études expérimentales portant sur le développement tardif de la production et de la perception de la parole, et que la quantification et la qualification de la maturité du contrôle moteur, au cours du développement typique, reste délicat, particulièrement si l'on se base seulement sur les études acoustiques. Même si l'on peut observer des tendances universelles, telles que le babillage canonique ou l'inventaire des sons du babillage (dans une certaine mesure), il faut garder en tête que la (les) langue(s) maternelle(s) cible(s) ont un effet sur l'apprentissage, par exemple, les sons de parole les plus fréquents dans la langue cible auront tendance à être appris plus précocement que les sons les moins fréquents (Edwards, Beckman & Munson, 2015 ; Monnin, 2010 ; de Boysson-Bardies & Vihman, 1991).

Il existe cependant un certain nombre d'études acoustiques qui caractérisent les productions des jeunes enfants et que nous n'avons pas évoqué dans cette partie. Les études portant sur la coarticulation apportent en effet des éléments de réponse importants quant à la maturation de la coordination spatio-temporelle des gestes de parole, et seront abordées dans la section 1.4 ; les études portant sur les capacités des enfants à compenser lors de tâches de perturbation, qui nous informent sur leurs représentations sensori-motrices et leur capacité à utiliser les degrés de liberté de leur système moteur, dans une optique d'équivalence motrice, seront abordées dans la section 1.5.

1.2.2. Les études du développement tardif de la perception

Nous ne pouvons pas traiter de l'évolution des capacités de production au cours de la petite enfance et du développement plus tardif sans parler des capacités de perception. En effet, la perception et la production s'influencent mutuellement, et ce de manière très forte. Cependant, très peu d'études se sont attachées à étudier les capacités de perception de parole chez les enfants après la première année de vie, alors que cette littérature est foisonnante en ce qui concerne le nouveau-né.

L'étude de Hazan & Barrett (2000) porte sur la perception des contrastes /k/ ~ /g/ (contraste reposant sur le voisement, ou VOT), /d/ ~ /g/ (contraste portant sur le lieu d'articulation), /s/ ~ /z/ (voisement) et /s/ ~ /ʃ/ (lieu d'articulation) chez 84 enfants âgés de 6 à 12 ans, via des tests de catégorisation effectués sur des continuums artificiels créés pour chacun des contrastes étudiés (contenus dans des mots constituant des paires minimales). Leurs résultats montrent que les courbes d'identification sont de plus en plus abruptes avec l'âge, indiquant que les frontières catégorielles des enfants sont de mieux en mieux définies au cours du développement. Les catégories phonétiques ne sont pas aussi bien définies à 6 ans qu'à l'âge adulte, car les frontières catégorielles sont plus floues (moins abruptes, donc moins bien définies). De plus, leurs résultats montrent que les courbes d'identification, même à 12 ans, ne sont pas similaires à celles des adultes. Cette étude souligne le fait que la perception catégorielle s'affine au cours du développement tardif, et que même après 12 ans, ces capacités s'affinent encore.

Walley & Flege (1999) rapportent des résultats similaires concernant les capacités de discrimination d'enfants de 5 et 9 ans pour des contrastes vocaliques. En étudiant la perception du contraste /i/ ~ /ɪ/ et /ɪ/ ~ /y/ de l'anglais américain, ils montrent que les courbes d'identification des enfants sont moins abruptes que celles des adultes.

Rappelons que l'étude des frontières catégorielles ne concerne pas les capacités auditives des enfants mais bien leurs capacités cognitives. Ces études montrent que, du point de vue cognitif, les capacités de perception de parole des enfants évoluent au cours du développement. Or les capacités perceptuelles de discrimination des sons de parole, via les sensations, dans le but de créer des représentations sensori-motrices, sont un élément important qui va jouer un rôle clé dans le processus de production.

1.2.3. Méthodes et outils pour étudier le développement de la production

Il existe diverses méthodes et divers outils qui permettent l'étude du développement de la production de la parole. L'observation et les transcriptions phonétiques, qui consistent en l'écoute et la transcription en caractères phonétiques des sons de parole produits par les enfants, sont chronologiquement les premiers outils qui ont été utilisés, et permettent une analyse descriptive des productions enfantines (Taine, 1870 ; Darwin, 1877 ; Lewis, 1936 ; Grégoire, 1937). Les enregistrements des signaux acoustiques et les analyses spectrales (rendues possibles notamment par la transformée de Fourier), ont permis l'analyse quantitative, basée sur des indices acoustiques, des productions enfantines (cf. section 1.2.1).

Mais il existe également des outils qui permettent la mesure directe des gestes de parole, des mesures que nous appelons articulatoires. Ainsi, les dispositifs de capture de mouvement ont permis l'étude des articulateurs visibles tels que les lèvres et la mâchoire. Enfin, des dispositifs tels que la cinéradiographie, l'électromagnétométrie (EMA), l'échographie, ou encore l'électropalatographie (EPG) ou l'optopalatographie (OPG) (cf. section 2.2.2.1 pour une revue de ces différents dispositifs et leur adaptabilité aux études développementales) ont permis d'effectuer des mesures sur des articulateurs non visibles, tel que la langue.

Nous décrirons, dans la section qui suit, les études développementales qui utilisent des mesures articulatoires des productions enfantines. Nous verrons que ces études, qui permettent d'effectuer des mesures quantitatives des gestes de parole, apportent un nouvel éclairage quant au développement du contrôle moteur pour la production de parole.

1.2.4. Les études portant sur le contrôle des articulateurs visibles : mouvements des lèvres et de la mâchoire

L'étude de Green et al. (2002) concerne les mouvements des articulateurs visibles, les lèvres et la mâchoire, chez 5 enfants âgés de 11 à 13 mois, 9 enfants âgés de 23 à 29 mois, 10 enfants âgés de 6 à 7 ans et 10 jeunes adultes (âgés de 27 à 35 ans). La tâche consiste en la répétition de séquences [baba], [papa] et [mama] pendant laquelle les mouvements

de la mâchoire, de la lèvre inférieure et de la lèvre supérieure sont enregistrés via un système de capture de mouvements de type Optotrak.

Leurs résultats montrent que le contrôle de la mandibule est déjà relativement précis (semblable à celui de l'adulte) à un an, alors que le contrôle des lèvres est très imprécis par rapport à l'adulte. Ce résultat va dans le sens de MacNeilage (1998). Leurs résultats montrent également que le contrôle des lèvres est très imprécis à 1 et 2 ans, et que les patrons se rapprochent des patrons adultes à l'âge de 6 ans, sans toutefois atteindre leur degré de consistance. La figure 1.2.1. ci-dessous récapitule l'ensemble de ces résultats.

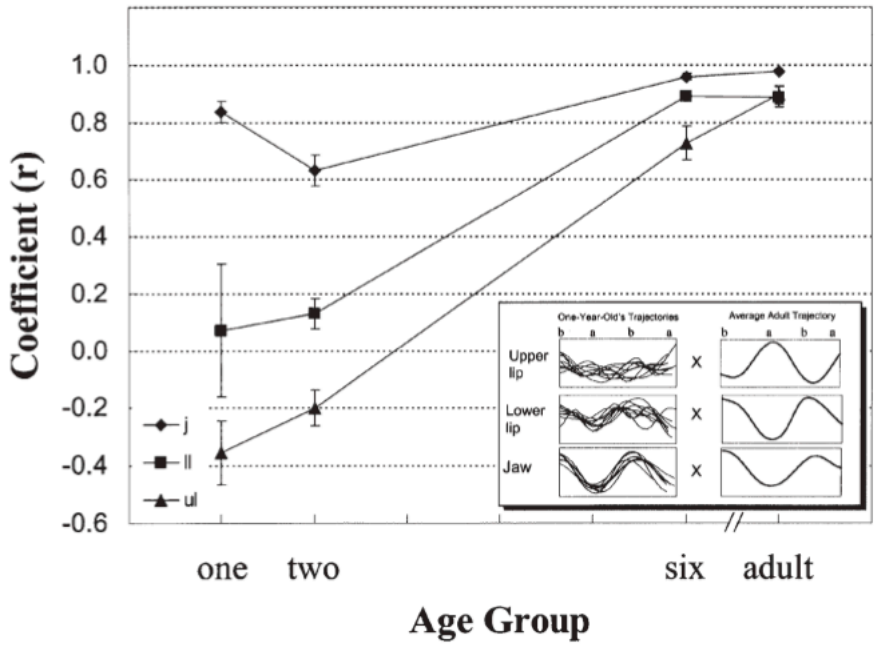


Figure 1.2.1. Résultats principaux de l'étude de Green et al. (2002), illustrant la maturité du contrôle de la mandibule (losanges), de la lèvre inférieure (carrés) et de la lèvre supérieure (triangles) aux âges de 1 ans, 2 ans, 6 ans et à l'âge adulte.

Cette étude montre que le contrôle de la mandibule est relativement précoce, alors que le contrôle des lèvres suit un développement plus lent, et ne semble être maîtrisé que plus tardivement, après l'âge de 6 ans d'après cette étude.

D'autres études articulatoires ont étudié la maturité du contrôle de la mandibule et des lèvres, avec une méthodologie similaire (on citera notamment Sharkey & Folkins 1985 ; Smith & Goffman, 1998 ; Goffman & Smith, 1999 ; Walsh & Smith, 2002 ; Walsh et al., 2006).

L'étude de Smith & Zelaznik (2004) s'intéresse au contrôle de la mandibule, de la lèvre inférieure et de la lèvre supérieure chez 180 participants, âgés de 4 à 22 ans (15 filles et 15 garçons, soit 30 individus, pour chacun des groupes d'âge suivants : 4 ans, 5 ans, 7 ans, 12 ans, 14 ans et jeunes adultes (20-22 ans) lors de la production de phrases comme « *Buy Bobby a puppy* ». La figure 1.2.2. ci-dessous illustre les résultats de cette étude concernant la diminution de la variabilité avec l'âge (variabilité spatio-temporelle liée à la répétition).

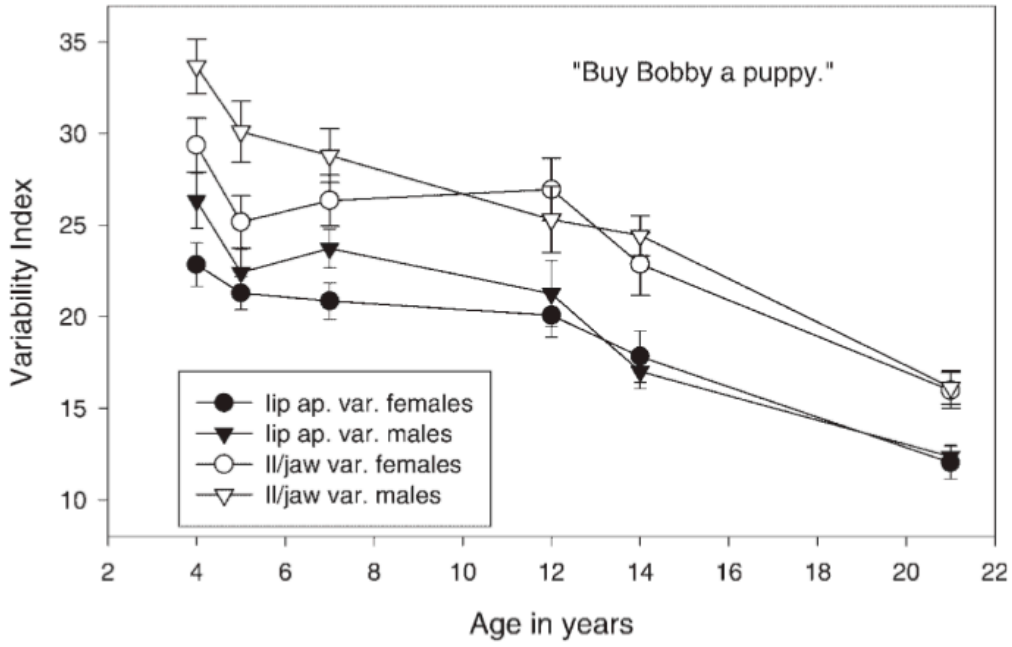


Figure 1.2.2. Résultats de l'étude de Smith & Zelaznik (2004) concernant la diminution de la variabilité des patrons de mouvement articulatoires des lèvres et de la mandibule avec l'âge.

Leurs résultats montrent que la durée totale de la phrase, ainsi que la variabilité, diminuent avec l'âge. Leurs résultats montrent également, de façon importante, que chez les enfants, le contrôle de la mandibule et des lèvres n'est pas similaire à l'adulte, même à l'âge de 14 ans. Si la diminution de la variabilité est très importante entre 4 et 7 ans, et est encore relativement importante entre 7 et 12 ans, elle n'est pas nulle entre 14 et 20 ans, soulignant que le processus de maturation motrice est encore en construction.

Ces études articulatoires, qui permettent la mesure directe des mouvements liés aux gestes de parole, mettent en avant le fait que le développement du contrôle est plus lent et plus progressif que l'on ne l'aurait pensé au premier abord, et que ce contrôle n'est pas achevé avant la fin de l'adolescence.

1.2.5. Les études portant sur le contrôle de la langue

Quelques études développementales se sont attachées à mesurer les mouvements des articulateurs non visibles, tel que la langue, via l'EMA (Cheng et al., 2007) ou l'échographie linguale (Zharkova et al., 2011 ; 2012 ; 2014 ; Noiray et al., 2013).

L'étude de Cheng et al. (2007) porte sur la coordination entre langue et mandibule lors de la production des consonnes /t/ et /k/ portées dans des phrases, et mesurées via l'articulographie électromagnétique (EMA), pour 48 participants âgés de 6 à 38 ans. Leurs résultats montrent que le contrôle indépendant de la langue et de la mandibule progresse fortement entre 6 et 8-11 ans, mais que les productions s'affinent également après 11 ans. Leurs résultats vont dans le sens des résultats de Smith & Zelaznik (2004), entre autres, et montrent que le développement du contrôle moteur est un processus lent et progressif.

Les études de Zharkova et al. (2011) ; (2012) ; (2014) et celle de Noiray et al. (2013) reposent sur des mesures articulatoires obtenues par échographie linguale, et portent sur la coarticulation intra-syllabique. L'étude de Noiray et al. (2013) concerne la production de parole d'enfants de 4 et 5 ans, et n'offre pas d'autre mesure que celle de la coarticulation intra-syllabique, et sera donc uniquement discutée dans la section 1.3. Les études de Zharkova et al. (2011), (2012) et (2014) concernent également la coarticulation intra-syllabique, chez des enfants âgés de 6 à 9 ans pour Zharkova et al. (2011) & (2012) et chez des préadolescents âgés de 10 à 12 ans pour Zharkova et al. (2014), mais les études de 2011 et 2012 offrent également une mesure de la variabilité en répétition. L'étude de Zharkova et al. (2011) montre que la variabilité en répétition pour la production de la consonne /ʃ/, et des voyelles /i/, /a/ et /u/ est plus élevée chez les enfants que chez les adultes (presque 2 fois pour les voyelles, un peu moins pour le /ʃ/). L'étude de Zharkova et al. (2012) montre le même résultat, mais pour la consonne /s/. Ces résultats concernant la variabilité en répétition viennent appuyer le constat déjà effectué par certaines études acoustiques (Kent & Forner 1980) et cinématiques (Sharkey & Folkins 1985), indiquant qu'au cours du développement, la variabilité des mouvements liée à la répétition diminue, amenant la production vers une certaine stabilité et une certaine consistance.

Mais, comme nous venons de le voir, la littérature portant sur le développement tardif du contrôle de la langue est très peu fournie, et une description, plus systématique, de l'apprentissage de la production des contrastes phonétiques et de la variabilité des productions reste toujours à fournir. Si les outils permettant cette étude sont maintenant à notre disposition (et l'étude du développement du contrôle des lèvres et de la mandibule au cours du développement est déjà bien fournie), il reste néanmoins à mettre en place une véritable volonté de décrire le développement tardif, du point de vue du contrôle lingual. Aucune étude ne s'est focalisée, par exemple, sur la capacité des enfants à produire des gestes différenciés et correctement catégorisés par des locuteurs adultes, ni sur leur capacité à anticiper des cibles au-delà des frontières syllabiques, ni sur leur capacité à compenser lors de tâches de perturbation (les quelques études développementales produites à ce jour sur ce sujet, notamment Ménard et al., 2008 ; Shiller et al., 2010 ; MacDonald et al., 2012 ne comportent que des mesures acoustiques, à l'exception de la récente étude de Ménard et al., 2016). Or tout n'est pas observable et mesurable dans le seul domaine acoustique, et étant donné le peu de connaissances que nous avons sur la morphologie des instruments vocaux des jeunes enfants, il serait intéressant de comprendre comment les enfants apprennent à contrôler leurs gestes de parole, d'un point de vue articulatoire, dans leurs propres instruments vocaux.

Cependant, un certain nombre d'études vise à caractériser la maturité du contrôle moteur pour la production de la parole à travers un cadre spécifique, celui de la coarticulation. En effet, si l'étude de la stabilité des productions rend compte de la maturité des productions enfantines, l'étude de l'organisation spatio-temporelle de la parole renseigne encore davantage sur les processus de maturation motrice. Un autre cadre théorique permet également de contribuer à l'évaluation de la maturité du contrôle moteur, celui des tâches de perturbation. Nous avons choisi de ne pas aborder ces études dans cette section, mais de leur consacrer respectivement les sections 1.4. et 1.5. suivantes.

1.2.6. En somme...

Pour conclure, l'étude du développement tardif des capacités de production de parole montre que les processus sous-jacents de contrôle moteur suivent un développement lent et progressif, dont la maîtrise n'est semblable à celle de l'adulte qu'à la fin de l'adolescence probablement (Smith & Zelaznik, 2004). La variabilité diminue avec l'âge, les contrastes sont de plus en plus importants et stables, et le contrôle des articulateurs supra-glottiques comme la langue, les lèvres ou la mâchoire, s'affine progressivement, mais n'est toujours pas semblable à celui de l'adulte à l'âge de 14 ans (Smith & Zelaznik, 2004 ; Cheng et al., 2007). Certes, tous les articulateurs ne suivent pas le même profil développemental, d'un point de vue du contrôle, et la mâchoire semble être contrôlée plus précocement que les autres articulateurs portés. Les capacités de perception catégorielle s'affinent également au cours de l'enfance et de l'adolescence, et ne sont toujours pas semblables à celles de l'adulte à l'âge de 12 ans.

En résumé, l'étude du développement tardif révèle une facette de l'apprentissage de la parole bien différente de celle que l'on a pu apercevoir au cours du développement précoce. En effet, à la lecture des études portant sur le développement précoce, on pourrait être amené à penser que les bases de la parole sont présentes à la fin de la première année de vie, et que cette capacité ne s'affine que légèrement, jusqu'à la fin de la petite enfance, âge auquel les enfants « savent parler ». Or les études portant sur le développement tardif, et particulièrement celles offrant des mesures fines, via la mesure des mouvements articulatoires, montrent que le développement est bien plus progressif et lent qu'il n'y paraît. La production de parole est en effet une tâche complexe, nécessitant le contrôle coordonné d'une centaine de muscles oro-faciaux et de multiples articulateurs, pour laquelle le contrôle moteur ne semble être comparable à celui de l'adulte qu'à la fin de l'adolescence.

1.3. Cadre théorique général

Après avoir brièvement décrit les étapes du développement précoce et tardif de la production de la parole, nous proposons, avant d'aller plus loin, d'exposer dans cette partie le cadre théorique dans lequel cette recherche s'inscrit. Nous commencerons par introduire quelques notions d'anatomie concernant l'instrument vocal et son développement au cours de l'ontogenèse, avant d'introduire le problème spécifique de la tâche de production de parole, dans le cadre d'un système périphérique en plein développement. Nous exposerons ensuite quelques notions clés du cadre théorique général dans lequel ce travail s'inscrit, à savoir celui du contrôle moteur, puisque la tâche de production de parole relève sans conteste du contrôle moteur, puis nous finirons par exposer brièvement notre cadre théorique spécifique à la production de la parole.

1.3.1. Le contrôle d'un système périphérique en plein développement

1.3.1.1. L'anatomie de l'instrument vocal adulte

L'instrument qui permet la production de la parole est complexe, puisque composé de multiples organes et structures (langue, mandibule, etc.), dont l'activité doit être coordonnée afin de produire des sons de parole. L'instrument vocal est principalement composé de tissus mous, comme la langue, les lèvres, le velum (ou voile du palais), mais comporte également - et est délimité - par des structures osseuses, notamment le palais dur, la dentition, la mandibule, l'os hyoïde, ou encore les cartilages du larynx. Toutes ces structures forment un seul et même instrument, impliqué notamment dans les tâches de mastication, déglutition, et de production de parole.

Le contrôle coordonné de ces structures, impliquant les contributions d'une centaine de muscles oro-faciaux, entraine la déformation du conduit vocal (le conduit aérien formé par l'instrument vocal), et mène, dans le cas de la production de sons voisés (comme les voyelles par exemple), à la modulation de l'onde sonore périodique générée par la vibration des plis vocaux en agissant comme un résonateur, ce qui permet de générer une multitude de contrastes phonétiques de ce type.

Légende :

A. palais dur
B. palais mou

1. cavité nasale et septum na
2. fente orale
3. cavité orale (rose)
4. os hyoïde
5. épiglotte
6. pli vestibulaire
7. pli vocal
8. cavité laryngée
9. glande thyroïde
10. choanes
11. ostium pharyngé de la trom
12. naso-pharynx (jaune)
13. isthme du gosier (vert) et
14. oropharynx (bleu)
15. œsophage

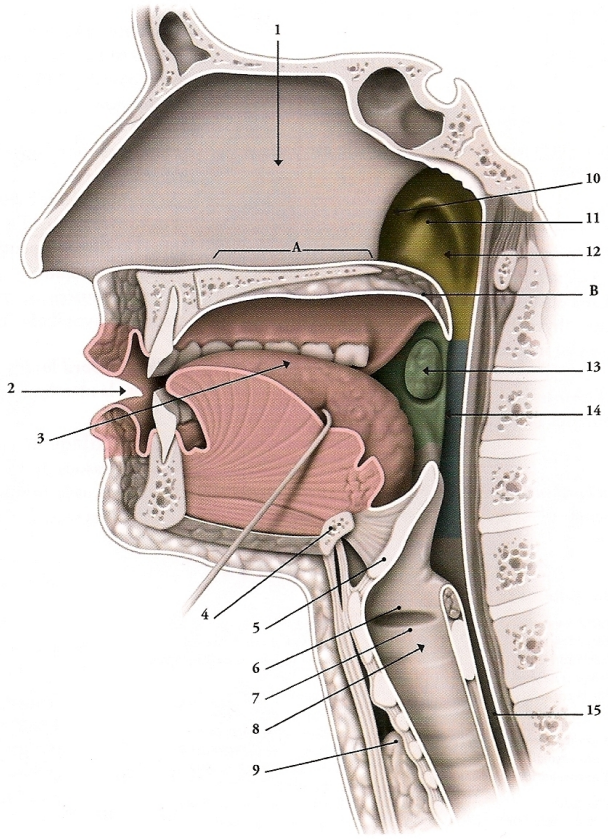
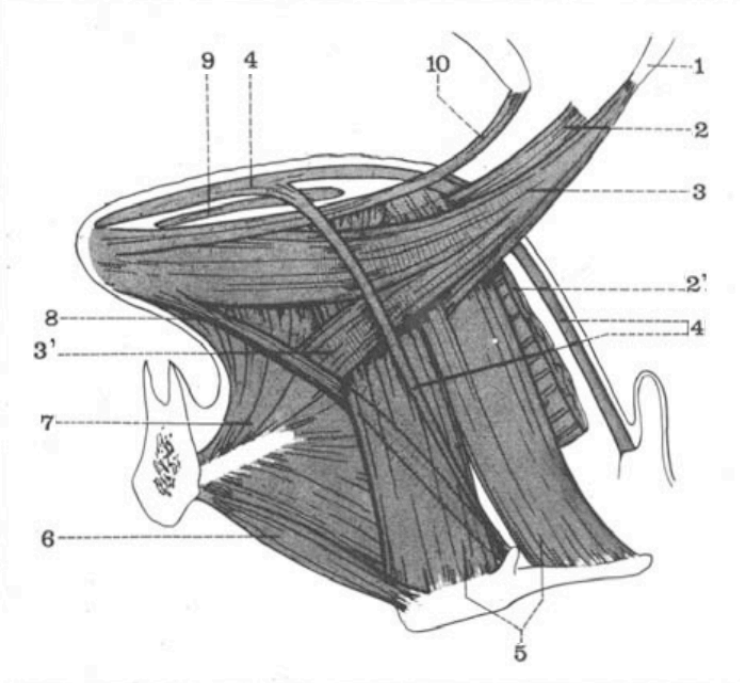


Figure 1.3.1. Anatomie des voies aéro-digestives supérieures. Tiré de Kamina, 2009.

Parmi ces articulateurs, il en est un qui joue un rôle central : la langue. Sa structure musculaire relativement complexe et ses propriétés hydrostatiques lui confèrent une grande mobilité, ce qui permet la génération d'un grand nombre de sons de parole. Plusieurs études anatomiques ont permis de mettre en évidence la structure interne de la langue humaine (Abd-el-Malek, 1939 ; Miyawaki, 1974 ; Takemoto, 2001). La langue est un organe musculo-membraneux viscoélastique quasi-incompressible, composé de 9 muscles principaux. Cinq muscles extrinsèques, Génioglosse, Styloglosse, Hyoglosse, Géniohyoïde et Mylohyoïde, et 4 muscles intrinsèques, Superior Longitudinalis, Inferior Longitudinalis, Verticalis et Transversalis. Le rôle de ces muscles dans la production de parole a été exploré par des études électromyographiques (EMG), mesurant directement l'activité électrique des muscles dans des tâches de production de parole (Miyawaki et al., 1975 ; Baer et al., 1988 ; Honda et al., 1993). Enfin, les travaux portant sur la modélisation de la langue permettent également de mieux comprendre les différentes contributions de l'activité individuelle de chacun de ses muscles à la déformation linguale (Perkell, 1974 ; Kiritani *et al.*, 1975 ; Kakita, Fujimura & Honda, 1985 ; Payan & Perrier, 1997 ; Dang &

Honda, 2002 ; 2004 ; Wilhelms-Tricarico, 1995 ; 1996 ; 2000 ; Buchaillard, 2007 ; Buchaillard et al., 2009).



Légende :

1 - Apophyse styloïde
2 - Pharyngoglosse
2' - Constricteur supérieur du pharynx
3 et 3' - Styloglosse
4 - Longitudinalis supérieur
5 - Hyoglosse
6 - Géniohyoïde
7 - Génioglosse
8 - Longitudinalis inférieur
9 - Transversalis
10 - Palatoglosse

Figure 1.3.2. Illustration de l'organisation musculaire de la langue dans le plan sagittal médian, d'après Bouchet & Cuilleret (1991).

La langue repose sur le plancher lingual, composé des muscles Géniohyoïdien, qui s'insère sur la face antérieure interne de la mandibule (apophyses génie) et sur l'os hyoïde, et Mylohyoïdien, qui s'insère sur chacune des faces internes de la mandibule et sur l'os hyoïde. Ces muscles ne jouent pas de rôle majeur dans les déformations linguales, mais constituent son socle et contribuent à contrôler la position de la racine de la langue, via la modification de la position de l'os hyoïde.

1.3.1.2. La croissance de l'instrument vocal au cours de l'ontogenèse

La croissance de l'instrument vocal, de la naissance à l'âge adulte, n'est pas linéaire, mais représente une véritable anamorphose. Ainsi, l'instrument vocal des jeunes enfants n'entretient pas une relation homothétique avec celui de l'adulte. Plusieurs études anatomiques (Goldstein, 1980 ; Fitch & Giedd, 1999 ; Lieberman et al., 2001 ; Vorperian et al., 2005 ; Vorperian et al., 2009 ; Barbier, 2010 ; Captier et al., 2010 ; Barbier et al., 2012 ; 2015) ont montré que l'instrument vocal de l'enfant n'est pas une copie conforme, à échelle réduite, de l'instrument vocal de l'adulte, mais que cet instrument, au cours de

l'ontogenèse, subit un remodelage complexe au cours duquel chacune des structures constitutives de l'instrument vocal suit un rythme de croissance qui lui est propre. Le crâne et la mandibule, par exemple, suivent deux rythmes de croissance particulièrement différents. Or c'est la croissance de l'architecture osseuse qui rythme, en grande partie, celle des tissus mous. Le crâne subit une croissance importante au cours des premières années de vie, pour atteindre ses dimensions adultes très tôt dans le développement (à 4 ans il représente 86 % du volume adulte), alors que la mandibule, ainsi que l'avancement de la face, suit un rythme très lent et progressif, pour atteindre la configuration adulte aux alentours de 17 ans. Il en résulte un remodelage qui représente une véritable anamorphose, les structures rattachées au crâne, celles qui sont attachées au rachis, et celles qui sont attachées à la face ayant chacune un rythme de croissance qui leur est propre. Plusieurs autres aspects méritent d'être évoqués : la dimension verticale du conduit vocal, correspondant à la cavité pharyngale, est réduite chez le jeune enfant en comparaison à l'adulte ; les enfants disposent de lèvres, d'un voile du palais, et dans une moindre mesure, d'une langue, de dimensions relativement importantes au vu des dimensions globales de leur instrument vocal ; les fibres musculaires de la langue ne sont pas orientées de la même façon chez l'enfant et chez l'adulte, ce qui implique que les mêmes commandes musculaires amèneraient à des déformations linguales différentes, ou pour le dire autrement, que les activités musculaires des enfants et des adultes doivent être nécessairement différentes pour aboutir aux mêmes types de déformations linguales (comme l'a montré l'étude anatomique des insertions extrinsèques des muscles linguaux chez l'adulte et chez l'enfant Barbier, 2011 ; Barbier et al., 2012).

L'instrument vocal des jeunes enfants est donc résolument différent de celui des adultes, et comporte, de plus, des orientations musculaires différentes. Se pose alors la question des stratégies articulatoires mises en œuvre par les enfants pour parvenir à produire des sons de parole, dans le cadre de ce système périphérique en plein développement.

1.3.1.3. De la production de parole dans un système périphérique en plein développement

L'apprentissage et le développement de la production de la parole, dans ce contexte d'un système périphérique en plein développement, pose de nombreuses questions parmi

lesquelles celles de la relation entre anatomie et acoustique. Nous évoquerons très rapidement ces questions, qui ne constituent pas le socle de réflexion de ce travail de thèse, puisque les questions qu'aborde ce travail de thèse relèvent du domaine du contrôle moteur, en visant à étudier les questions de stabilité du contrôle et d'émergence de la planification de tâches séquentielles.

Néanmoins, la question du remodelage de l'instrument vocal au cours de l'ontogenèse soulève trois questions majeures, du point de vue de la relation entre anatomie et acoustique : (1) existe t-il des contraintes anatomiques et biomécaniques spécifiques à l'instrument vocal des jeunes enfants, qui les empêcheraient de produire des contrastes phonétiques ? (2) quelles sont les stratégies articulatoires employées par jeunes enfants ? (3) Le remodelage de l'instrument vocal, au cours de l'ontogenèse, est-il un frein au développement de la production de la parole, en obligeant constamment l'enfant à réorganiser ses représentations sensori-motrices ?

Concernant la première question, soit on considère qu'il existe de réelles contraintes anatomiques et biomécaniques spécifiques aux instruments vocaux des enfants (un larynx haut limiterait la mobilité de la base de la langue, mobilité cruciale lors de la production du son [u] par exemple, pour dégager de l'espace dans la cavité pharyngale et créer un résonateur suffisant pour la production de ce son (de Boer & Fitch, 2010)) comme Lieberman & Crelin (1971) l'affirment concernant les potentialités acoustiques des conduits vocaux de nos ancêtres hominidés et des primates non-humains ; soit on considère qu'il n'existe pas de contrainte anatomique insurmontable (Boë et al., 2002 ; 2007 ; 2013), considérant notamment que la très grande déformation de l'instrument vocal (rendue possible par le contrôle des muscles oro-faciaux) permet de compenser les contraintes anatomiques statiques. Plusieurs arguments semblent étayer ce second point de vue. Dès un âge très précoce, les enfants sont capables de produire des contrastes importants et ont un Espace Vocalique Maximal aussi large que celui de l'adulte, malgré la petite dimension de leur cavité pharyngale et la position haute de leur larynx. Ce ne serait donc pas, *a priori*, les contraintes morphologiques qui entraveraient la production de différents sons de parole suffisamment contrastés. Notons que les conclusions tirées par les partisans des deux théories opposées sont issues de travaux de modélisation, donc sujettes à débat (cf. Lieberman & Crelin, 1971 ; Lieberman, 2007 ; Boë et al., 2002 ; 2007 ;

Ménard, et al., 2007). Gardons en tête que les potentialités des conduits vocaux et leur utilisation effective sont deux choses bien distinctes.

S'il n'existe pas de contrainte théorique absolue à la production de sons de parole suffisamment contrastés, pour les nouveau-nés et les jeunes enfants, il reste en revanche une difficulté inhérente à la tâche d'un point de vue du contrôle. Cette tâche semble en effet être plus complexe pour les jeunes enfants que pour les adultes, et ce pour plusieurs raisons, dont : (1) la taille importante de leur langue, relativement au volume total de leur instrument vocal, qui complique cette tâche : les cibles dans l'espace articulatoire sont restreintes et il est nécessaire, malgré le peu d'espace disponible pour l'articulation des sons de parole, de maintenir des contrastes suffisants (2) les caractéristiques acoustiques de leurs propres productions sont très différentes de celles des adultes, dont ils disposent pourtant majoritairement comme modèle (imitation) ou comme retour (renforcement), et malgré cette grande variabilité, les enfants (et les adultes) doivent pourtant trouver ce qui fait l'unité de ces sons de parole (Ménard et al., 2002).

Concernant la seconde question, la modélisation articulatoire de la croissance de l'instrument vocal (Boë & Maeda, 1998 ; Ménard et al., 2002 ; Boë et al., 2013) a fourni des arguments en faveur de l'absence de contrainte anatomique insurmontable, mais a également fourni des éléments de réponse concernant les stratégies articulatoires enfantines, notamment en prédisant des lieux d'articulation plus antérieurs chez les enfants que chez les adultes, pour des raisons anatomiques et acoustiques simples, notamment celle de leur petite cavité pharyngale. Mais l'investigation expérimentale concernant ce sujet n'a pu être fournie que très récemment, car l'acquisition de données articulatoires chez l'enfant est longtemps restée un sujet délicat. L'utilisation de l'échographie linguale en parole a permis à ce champ d'investigation un certain développement ces dernières années (i.e. Ménard & Noiray, 2011 ; Noiray et al., 2013 ; Zharkova et al., 2014) et nous avons espoir que cette question sera probablement bien documentée dans les années à venir.

Concernant la troisième question, l'adaptation de la définition des buts auditifs, au cours de l'ontogenèse, ne constitue pas, de notre point de vue, un frein majeur au

développement phonétique. Le travail de modélisation de Callan, Kent, Guenther & Vorperian (2000), proposant un modèle de production robuste aux changements morphologiques, va d'ailleurs dans ce sens, en proposant que le retour auditif de leurs propres productions suffirait aux enfants pour adapter leurs buts acoustiques, malgré le remodelage de leur instrument vocal et donc le changement des propriétés acoustiques de leurs productions. Nous ajouterons que les représentations sensori-motrices sont particulièrement dynamiques et plastiques. Il suffit en effet de quelques dizaines d'exemplaires en production pour modifier la perception catégorielle (Shiller et al., 2009). Pris ensemble, ces résultats suggèrent que la réorganisation des représentations sensori-motrices s'opère donc plus vite que le remodelage de l'instrument vocal et ne constitue pas un frein majeur au développement phonétique.

En résumé, il n'existe pas, de notre point de vue, de contrainte anatomique spécifique à l'instrument vocal des jeunes enfants engendrant une impossibilité de contraster suffisamment ou de produire l'ensemble des sons de parole, en accord avec Boë et al. (2002) ; (2007) ; (2013). Les stratégies articulatoires des enfants doivent nécessairement être différentes de celles des adultes, compte tenu des différences anatomiques et physiologiques, et enfin le remodelage de l'instrument vocal au cours de l'ontogenèse, engendrant l'adaptation des représentations sensori-motrices, ne constitue pas, de notre point de vue, un frein majeur au développement phonétique. Mais comme nous allons le voir, si la production de parole est affaire de conduit vocal, elle est avant tout affaire de cerveau, de capacités cognitives, et particulièrement de contrôle moteur.

1.3.2. Le cadre théorique général : celui du contrôle moteur

La tâche de production de parole relève sans conteste du contrôle moteur. De ce point de vue, elle est soumise aux contraintes générales du contrôle moteur des gestes rapides et précis, en particulier celles s'exerçant sur le contrôle sériel vers des buts séquentiels.

Le contrôle moteur, ou contrôle neuromusculaire, traite de la façon dont le Système Nerveux Central (SNC) envoie les commandes motrices au système nerveux périphérique dans le cadre de la production volontaire de mouvements. On peut décomposer ce problème en plusieurs niveaux, celui de la spécification des buts moteurs, celui de l'envoi des commandes motrices au système périphérique, celui de l'activité musculaire, celui des

conséquences de cette activité musculaire en interaction avec les propriétés physiques et mécaniques du système périphérique, et enfin celui de la résultante des quatre premiers niveaux, c'est à dire le mouvement effectif.

Nous introduisons dans cette section quelques notions de contrôle moteur, puisque la tâche de production de parole, tâche cognitive par essence, est soumise aux contraintes des principes généraux du contrôle moteur. Les idées clés dans ce domaine ont pu naturellement offrir, au cours des années, des éléments de réflexion à intégrer au problème spécifique de la production de parole.

1.3.2.1. Les modèles internes

Pour commencer, une notion clé issue du contrôle moteur, relativement bien admise aujourd'hui en parole et offrant une solution puissante et relativement satisfaisante à de nombreux problèmes dans ce domaine est celle des **modèles internes**.

Il s'est posé dans le domaine du contrôle moteur une question centrale, celle de savoir comment contrôler un système pour lequel on ne dispose que de l'information concernant un état passé de ce système. Plus précisément, comment contrôler un système dans le but de produire des mouvements rapides, alors que le retard important des informations afférentes de ce système ne permet pas des corrections rapides. Hickok (Hickok, 2012) utilise une métaphore illustrant ce problème : comment conduire une voiture uniquement à partir de l'information disponible dans les rétroviseurs ? C'est en quelque sorte le problème auquel fait face le SNC dans le cadre de la génération de mouvements rapides. En effet, le SNC ne dispose que d'informations concernant l'état actuel et l'état passé du système périphérique (via la proprioception et les retours auditifs, dans le cadre de la production de parole). Le temps de trajet et de traitement de l'information sensorielle (auditive et somatosensorielle) au niveau cortical est en effet relativement élevé. Le principal problème que pose le contrôle d'un tel système réside dans le rapport entre la vitesse d'exécution du mouvement et la vitesse des retours sensoriels. C'est probablement pour cette raison que, dans un premier temps, l'enfant réduit la vitesse d'exécution de ses gestes, pour ramener proportionnellement le retard des retours sensoriels à une quantité raisonnable et réduire l'instabilité générée par le délai entre exécution des gestes et informations afférentes du système périphérique.

Dans ce cadre, comment le SNC procède t-il pour planifier les mouvements rapides ? La solution des **modèles internes** a été proposée pour répondre à ce problème. Cette solution propose que le SNC effectue des prédictions concernant l'état futur du système périphérique, en d'autres termes que le SNC prédit les conséquences sensorielles de l'activité motrice. Ces prédictions ne peuvent être effectuées qu'après une phase d'apprentissage concernant les relations qu'entretiennent commandes motrices et retours sensoriels, et reposent sur deux niveaux. Le niveau du contrôle *feedforward*, exploitant le concept de **modèle inverse**, qui permet de définir les commandes motrices nécessaires à l'atteinte d'un but, et le niveau du **modèle** prédictif **direct**, qui permet de prédire les conséquences de l'activité motrice et permet d'exploiter un *feedback* interne plus rapide que le *feedback* biologique externe.

Cet outil a été introduit dans le domaine du contrôle moteur dans le cadre de la modélisation computationnelle (Kawato et al., 1987 ; Kawato et al., 1990 ; Jordan & Rumelhart, 1992). L'attrait de cette proposition et sa puissance explicative concernant le contrôle des mouvements biologiques est tel qu'il a rapidement été intégré à la plupart des modèles théoriques de contrôle moteur. Initialement proposés sous la forme de modèles internes généralisés, implantés sous la forme de réseaux de neurones artificiels, ces modèles permettent au SNC l'apprentissage des relations entre commandes motrices et retours sensoriels.

1.3.2.1.1. Le modèle direct

Selon Kawato et al. (1987), les premiers gestes, au cours de l'apprentissage, doivent être exécutés lentement, puisqu'il ne peuvent être contrôlés que par un contrôle de type ***feedback***. C'est l'exécution de ces premiers mouvements lents contrôlés par *feedback* qui va permettre au SNC d'apprendre les relations entre commandes motrices et conséquences sensorielles, stockées sous la forme d'un **modèle direct** (*forward model*). Ce modèle fonctionne comme un « système moteur virtuel » puisqu'il fait le lien entre commandes motrices (en entrée) et conséquences sensorielles (en sortie), imitant d'une certaine manière ce que fait le système moteur physique. Ce modèle direct permet la génération d'un « résultat virtuel » de l'action réalisée (appelé décharge corollaire ou *corollary discharge*), c'est-à-dire une estimation des conséquences sensorielles de l'acte

moteur. La comparaison de ce « résultat virtuel » avec les conséquences sensorielles effectives (appelé divergence sensorielle ou *sensory discrepancy*) permet, petit à petit, un apprentissage efficace de ce modèle direct. Grâce à ce modèle, si l'apprentissage est efficace et que le modèle direct imite suffisamment précisément le système moteur physique, le SNC peut exécuter des gestes rapides en se reposant sur les prédictions effectuées par le modèle direct, générant une **boucle rapide de *feedback* interne**, sans avoir à attendre le retour sensoriel biologique externe.

1.3.2.1.2. Le modèle inverse

Selon Kawato et al. (1987) et Jordan & Rumelhart (1992), le contrôle par *feedback* interne (généré par le modèle direct), n'est pas, à lui seul, suffisant pour rendre compte de la génération de mouvements rapides. En effet, une fois qu'un apprentissage suffisant est effectué, il semblerait que le SNC soit capable de générer des mouvements uniquement à partir du **but désiré** (dans le cadre du contrôle moteur du bras, une trajectoire désirée, ou une posture, dans le cadre de la production de parole, un but acoustique et somatosensoriel). Cette idée est en cohérence avec de nombreux résultats expérimentaux, entre autres ceux observant la compensation immédiate lors de tâches de perturbation (cf. section 1.5). Cela nous amène au second type de modèle interne : les **modèles inverses** (*inverse models* exploités dans les contrôleurs de type *feedforward*). Le modèle inverse entretient également une relation entre commandes motrices et conséquences sensorielles, mais, comme son nom l'indique, dans la direction opposée de celle du modèle direct : les buts sensoriels constituent l'entrée de ce modèle et les commandes motrices la sortie. L'apprentissage de ce modèle inverse ne reposerait pas uniquement sur l'ensemble de l'apprentissage réalisé dans le cadre de différentes tâches motrices, mais serait nourri par les multiples simulations réalisées par le modèle direct. Dans ce cadre, l'efficacité et la précision du modèle inverse dépendent drastiquement de la maturité du modèle direct. Une fois l'apprentissage de ce modèle inverse réalisé, il est possible de générer, à partir du but sensoriel désiré, les commandes motrices qui vont être envoyées au système périphérique. La génération de mouvement s'appuie alors fortement sur les prédictions, et ce type de contrôle est appelé **contrôle *feedforward***, par opposition au contrôle par *feedback*.

Quand le modèle inverse envoie au système périphérique les commandes motrices, il envoie également au modèle direct une copie de ces commandes motrices, appelée **copie d'efférence** (dans une partie de la littérature, la décharge corollaire et la copie d'efférence sont appelées indistinctement copie d'efférence). Le couplage du contrôle par *feedback* interne et du contrôle *feedforward* au sein d'un même système permet à la fois de rendre compte de la génération de mouvements à partir de buts sensoriels désirés (contrôle *feedforward*, modèle inverse) et de la prédiction des conséquences sensorielles de l'action motrice (contrôle par *feedback* interne, modèle direct). Après un apprentissage suffisant, la production serait majoritairement guidée par la composante prédictive, le contrôle de type *feedforward*. Le contrôle de type *feedback* interne rapide serait avant tout utile à l'apprentissage. Les retours sensoriels réels sont eux aussi indispensables à l'apprentissage, mais restent utiles, même après un apprentissage intensif, particulièrement dans le cadre de perturbations qui viendraient donner un caractère moins prédictible aux conséquences auditives et/ou somatosensorielles des actions réalisées (cf. section 1.5).

Nous proposons d'illustrer le fonctionnement d'un tel modèle (cf. figure 1.3.3. ci-dessous), reposant sur l'action couplée d'un modèle direct et d'un modèle inverse, d'après Jordan et Rumelhart (1992), qui partage avec Kawato et al. (1987) les grands principes fondamentaux de cette approche, et qui représente de façon simple et épurée ce fonctionnement. La différence majeure entre les deux travaux est que Jordan et Rumelhart (1992) ne suppose pas la nécessité d'un apprentissage des propriétés dynamiques du système moteur.

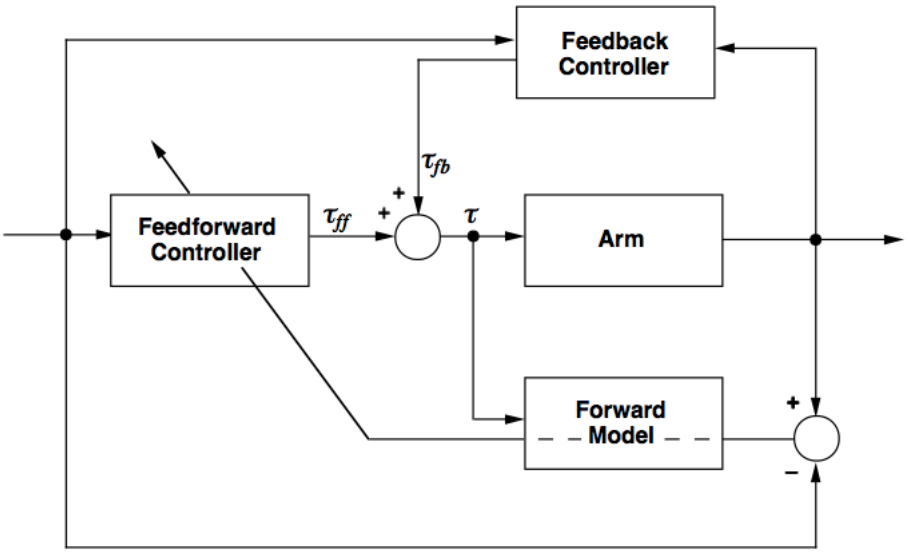


Figure 1.3.3. Illustration du fonctionnement du modèle proposé par Jordan & Rumelhart (1992). Le contrôleur *feedforward* repose sur l'idée du modèle inverse et le *forward model* correspond au modèle direct. Le but spécifié constitue l'entrée du modèle, le modèle inverse permet d'estimer les commandes motrices, ces commandes motrices sont envoyées au système périphérique (le bras, *arm*) et une copie d'efférence de ces commandes est envoyée au modèle direct, ce qui permet le calcul de la divergence sensorielle et une boucle de *feedback* interne rapide, en opposition à la boucle de *feedback* biologique externe (*feedback controller*).

1.3.2.2. L'équivalence motrice et la planification motrice optimale

Se pose maintenant, plus en détail, la question de la façon dont s'opère l'apprentissage des modèles internes. Le problème principal de cet apprentissage est celui de la **relation non univoque entre commandes motrices et conséquences sensorielles**, c'est-à-dire le fait que plusieurs ensembles de commandes motrices mènent aux mêmes conséquences sensorielles, et que, pour prendre le problème dans l'autre sens, un même but acoustique peut être atteint en utilisant différents ensembles de commandes motrices. Le problème de l'inversion, c'est-à-dire celui de retrouver les commandes motrices à partir des conséquences acoustiques, est un problème dit mal posé (du fait de la relation non univoque entre commandes motrices et conséquences sensorielles, ne menant pas à une solution unique mais à plusieurs solutions équivalentes). Ce problème est propre à tous les systèmes possédant des **degrés de liberté en excès**, et est dû aux relations non univoques entre commandes motrices et conséquences sensorielles, également

désignées sous le terme d'**équivalence motrice**. Ce type de problème ne peut être résolu qu'en apportant des contraintes supplémentaires. Kawato et al. (1987) et Jordan & Rumelhart (1992) ont par exemple proposé comme contrainte supplémentaire le maximum de fluidité (*smoothness*) dans la trajectoire. Un certain nombre de propositions ont été formulées quant au coût à minimiser : minimisation du *jerk* (dérivée de l'accélération) (Hogan, 1984), minimisation du changement de force (Uno, Kawato & Suzuki, 1989). Toutes ces propositions reposent sur l'idée de la minimisation d'un coût tout au long de la trajectoire, afin de minimiser l'effort, et reposent sur la notion centrale de **planification motrice optimale** (*optimal motor planning*). Pour une revue détaillée de la littérature portant précisément sur ces questions, ainsi que sur leur mise en œuvre dans le cadre de la production de parole, nous invitons à la lecture de Perrier (2012).

1.3.3. Le cadre théorique spécifique à la production de la parole

1.3.3.1. Les notions clés

La production de la parole est sans conteste une tâche qui relève du contrôle moteur, elle peut donc être étudiée dans le cadre général du contrôle volontaire du mouvement. Il existe en effet de nombreuses similarités entre le contrôle volontaire des muscles du bras et celui des muscles oro-faciaux pour la production de parole.

Les deux tâches semblent être **orientées vers un but** (goal-directed), la notion de **modèle inverse** semble donc être pertinente pour la production de parole. Nous reviendrons plus en détail dans la section suivante sur l'espace dans lequel sont définis les buts dans la tâche de production de parole.

Les deux tâches reposent également sur un **apprentissage sensori-moteur** (sensorimotor learning), et de nombreuses études expérimentales viennent confirmer le rôle du contrôle de type ***feedforward*** (les tâches de perturbation par exemple, Kelso & Tuller, 1983), le rôle des **retours sensoriels** (les expériences effectuées sur les personnes sourdes et implantées cochléaires notamment, Matthies et al., 1996 ; Perkell et al., 2000), et le rôle des **retours internes rapides** (par exemple la manipulation en temps réel des formants, Houde & Jordan, 1998 ; 2002) dans des tâches de production de parole.

Les propositions théoriques formulées par Kawato et al. (1987) et Jordan & Rumelhart (1992) semblent donc être adaptables au cadre spécifique de la production de parole. Nous proposons d'adapter le modèle proposé par Jordan et Rumelhart (1992) au cadre spécifique du contrôle moteur de la production de parole (cf. figure 1.3.4. ci-dessous).

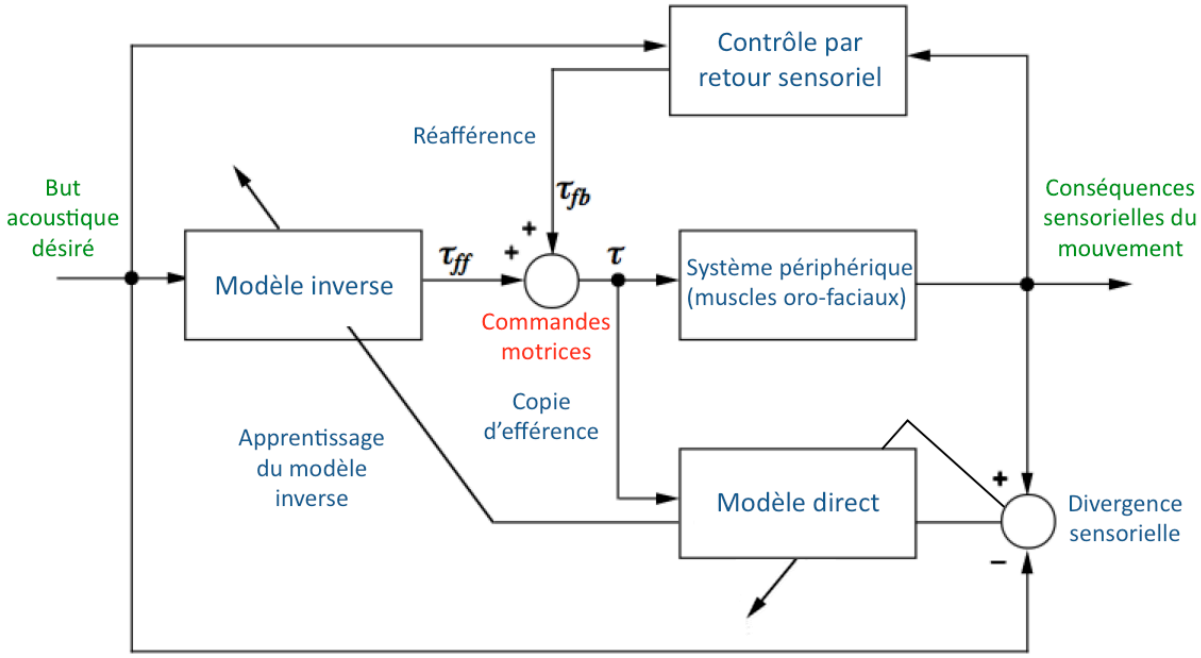


Figure 1.3.4. Le modèle théorique de Jordan & Rumelhart (1992) appliqué au cadre spécifique de la production de la parole. Le but, spécifié en termes acoustiques et somatosensoriels constitue l'entré du modèle, le modèle inverse permet d'estimer les commandes motrices, ces commandes motrices sont envoyées au système périphérique oro-facial et une copie d'efférence de ces commandes est envoyée au modèle direct, ce qui permet le calcul de la divergence sensorielle et une boucle de *feedback* interne rapide, en opposition à la boucle de *feedback* biologique externe (contrôle par retour sensoriel). Les flèches traversant les modèles indiquent un apprentissage.

La plupart des notions et outils clés apportés par le domaine du contrôle moteur ont d'ailleurs été assez rapidement intégrés au problème spécifique de la production de la parole (Perkell et al., 1997 ; Perkell, 2012 pour une vue d'ensemble), ainsi qu'aux modèles théoriques de production de parole (e.g. modèle DIVA de Guenther et Perkell, Guenther et al., 1998 ; 2004, Guenther, 2006, modèle GEPPETO, Perrier, 2014), mais aussi à des modèles portant spécifiquement sur la question de l'apprentissage de la production

(Guenther, 1995 ; Bailly, 1997). Nous n'entrerons pas davantage dans le détail de ces théories, et invitons à la lecture de Perkell (2012) et Hickok (2012) pour une vue d'ensemble de ces mécanismes dans la production de parole ainsi que la présentation de nombreuses données expérimentales qui viennent conforter ce cadre théorique. Notons tout de même que de nombreuses critiques ont été formulées à l'encontre de cette idée de modèles internes généralisés. Cet outil permet néanmoins de fournir un cadre permettant de comprendre l'idée de représentations internes du système moteur qui permettent de prédire les conséquences sensorielles des commandes motrices et permettent de planifier (au sens de la planification motrice optimale, sans entrer dans les détails de cette optimisation) la tâche et de résoudre le problème inverse en exploitant les degrés de liberté en excès.

1.3.3.2. La nature des buts de la production de parole

La question de la variabilité de la parole naturelle, en apparente opposition avec la nature discrète des unités de la phonologie, a amené au questionnement concernant les corrélats physiques des unités de parole, c'est-à-dire la question de l'invariant, aussi bien en perception qu'en production. La question de la recherche des invariants en production de parole est indissociable des débats entre les différentes théories de la perception de la parole, notamment le débat entre les théories motrices de la perception de la parole (Liberman et al, 1967 ; Liberman & Mattingly, 1985), qui supposent que l'unité minimale de la parole est de nature motrice, et les théories auditives (Stevens & Klatt, 1974 ; Diehl & Kluender, 1989), qui supposent que l'unité minimale de la parole est de nature acoustique.

Concernant la définition des buts de la tâche de production de la parole, un débat similaire a eu lieu entre tenants d'une théorie motrice de la production de la parole (Browman & Goldstein, 1986 ; 1989) et tenants d'une théorie acoustique de la production de la parole (Perkell et al., 1997 ; Guenter et al., 1998). La Phonologie Articulatoire (Browman & Goldstein, 1986 ; 1989) fait l'hypothèse que les buts de la production de la parole sont des constrictions à réaliser au sein du conduit vocal, l'invariant serait donc de nature motrice. Les théories auditives de la production émettent l'hypothèse que les

objectifs de la production sont définis dans l'espace acoustique (Stevens, 1989 ; Perkell et al., 1997 ; 2000 ; Perkell, 2012).

Un certain nombre d'arguments vont en faveur des théories auditives de la production : dans le cadre de la production de voyelles, les retours somatosensoriels seraient trop faibles pour connaître précisément le degré et le lieu de constriction (Guenther et al., 1998) ; les expériences de perturbation de l'articulation (cf. section 1.5.4.2.1.) montrent que les participants tendent à maintenir des objectifs acoustiques, et non des lieux de constriction (e.g. Savariaux et al., 1995) ; et enfin le principe d'équivalence motrice va à l'encontre des théories telles que la Phonologie Articulatoire, puisqu'il existe de multiples configurations articulatoires menant au même produit acoustique, ce qui laisse penser que les locuteurs tentent bien d'atteindre une cible définie dans le domaine acoustique. Notons que ce dernier argument ne permet pas véritablement de trancher en faveur d'une théorie ou d'une autre, puisque dans un certain nombre de configurations, les contraintes en termes de lieux de constriction et les contraintes acoustiques sont difficilement dissociables.

Mais la nature des unités de la production ne semble pas être purement acoustique. Tremblay, Shiller & Ostry (2003) ont montré de façon convaincante la contribution de la composante somatosensorielle à l'établissement des buts de la production. De la même manière, Perkell et al. (2004) et Ghosh et al. (2010) ont montré, pour la production des consonnes sibilantes (fricatives non voisées /s/ et /ʃ/), que les objectifs étaient de nature acoustique et somatosensorielle.

Au vu du nombre de résultats expérimentaux et d'arguments théoriques allant en faveur de la nature acoustique et somatosensorielle des objectifs de la production de parole, nous adopterons ce point de vue théorique. Pour mieux comprendre les contributions des modalités auditives et somatosensorielles à la définition de la tâche de production de parole, nous invitons à la lecture de Lametti, Nasir & Ostry (2012) et Hickok (2012).

1.3.3.3. L'ordre sériel, la coarticulation et la planification motrice optimale

Maintenant que nous avons parlé de la nature des unités minimales de la production, la question de la planification et de l'exécution d'une séquence de sons de parole se pose.

Concernant la question de la **planification motrice optimale**, cette suggestion théorique a été adoptée par un certain nombre de propositions théoriques spécifiques à la production de parole (Perkell et al., 2002 ; Winkler, Ma & Perrier, 2010 ; Ma, Perrier & Dang, 2015). La question de l'excès de degrés de liberté, ou équivalence motrice, a amené en parole aux modèles de coarticulation basés sur l'exploitation de la variabilité sub-phonémique (e.g. Keating, 1990), qui repose implicitement ou explicitement sur la question du contrôle optimal, et la nécessité de la minimisation d'un coût. Des données expérimentales (e.g. Whalen, 1990) suggèrent effectivement que la coarticulation est largement planifiée, venant conforter l'hypothèse d'un contrôle séquentiel basé sur la planification motrice optimale. La minimisation de l'effort articulatoire le long d'une séquence de sons de parole vient par ailleurs renforcer cette idée (Perkell et al., 2002). Concernant la coarticulation, c'est d'ailleurs dans ce cadre théorique que nous nous positionnons (cf. section 1.3.2.) et c'est à travers ce cadre que nous tenterons d'expliquer nos résultats expérimentaux.

1.3.3.4. Spécificités de la tâche de production de parole

S'il existe, comme nous venons de le voir, un certain nombre de similarités entre contrôle des muscles du bras et contrôle des muscles oro-faciaux pour la production de sons de parole, il existe en revanche quelques spécificités du contrôle pour la production de parole. Il ne semblerait pas nécessaire, dans le cas de la production de gestes rapides (dont la production de parole fait partie), que le SNC intègre de façon précise les propriétés dynamiques du système moteur (Perrier, Lœvenbruck & Payan, 1996 ; Perrier, Ostry & Laboissière, 1996), en accord avec la proposition de Jordan et Rumelhart (1992). D'ailleurs, l'apprentissage des propriétés dynamiques du système musculaire oro-facial, extrêmement complexes, semblerait extrêmement difficile. Par ailleurs, les buts de la production de parole subissent l'influence de facteurs de plus haut niveau, notamment linguistiques (phonologiques, lexicaux) et cognitifs (intégration sensorimotrice, représentations mentales de haut niveau), et ne sont pas strictement physiques (Perrier, 2012). De façon générale, gardons en tête que la production de parole, au niveau phonétique, ne peut que très difficilement se détacher de l'influence des facteurs de plus haut niveau (linguistiques et cognitifs) et ne peut, de ce point de vue, être vue comme

une simple tâche de contrôle moteur de bas niveau, comme peut l'être le contrôle du bras et de la main.

1.3.4. En somme...

Nous avons abordé la question de l'anatomie de l'instrument vocal et de son développement au cours de l'ontogenèse, et discuté des possibles contraintes anatomiques des instruments vocaux des jeunes enfants. Nous avons ensuite introduit quelques notions clés de contrôle moteur, qui s'avèrent particulièrement adaptées pour expliquer le cadre spécifique de la production de la parole, parmi lesquelles les modèles internes, les copies d'efférence, l'équivalence motrice et la planification motrice optimale. Nous avons ensuite exposé notre cadre théorique spécifique à la production de parole, reprenant les différents éléments apportés par le domaine du contrôle moteur. Parmi ces notions, celles de modèles internes, d'équivalence motrice et de planification optimale revêtent un intérêt particulier pour nous, parce qu'elles permettent, entre autres, d'expliquer les causes sous-jacentes de deux phénomènes que nous allons aborder plus en détail dans les sections qui suivent, à savoir le problème de la coarticulation et celui de l'équivalence motrice dans le cadre perturbations de la production de la parole. Nous verrons, au cours des deux sections suivantes, en quoi ces phénomènes de coarticulation et d'équivalence motrice nous renseignent sur les représentations sensori-motrices associées à la production de la parole, et en quoi, au cours du développement, des expériences portant sur ces deux paradigmes nous renseignent sur la maturité des représentations sensori-motrices au cours du développement.

1.4. La coarticulation : de l'enfant à l'adulte

1.4.1. Définition

La coarticulation est un terme spécifique au domaine de la parole, qui désigne un phénomène non spécifique à ce domaine, mais très étudié en parole, notamment parce que ce phénomène soulève la question de la nature des unités de parole, sous-jacentes aux mouvements continus des articulateurs et au flux continu de parole, qu'il est difficile de segmenter. Ce phénomène désigne l'action par laquelle une cible est modifiée par la présence de cibles adjacentes. Dans le domaine de la parole, la coarticulation est définie comme l'articulation de deux sons de parole ou plus, produits conjointement, de façon que les sons de parole s'influencent mutuellement. Le phénomène de coarticulation ne s'exerce pas uniquement au niveau des segments adjacents et peut être observé sur de grandes distances (e.g. Recasens, 1989), à savoir jusqu'à trois segments vocaliques (Grosvald, 2009), deux syllabes, et ne s'arrête pas aux frontières syllabiques ni lexicales.

Par exemple, le phonème /g/ est réalisé de manière très différente selon qu'il est produit dans une séquence /aga/ ou /igi/. Cela semble être davantage vrai pour les consonnes que pour les voyelles, qui sont plus contraintes par leur lieu de constriction, donc moins enclines à accepter différentes réalisations en fonction du contexte (e.g. Gay, Lindblom & Lubker, 1981). L'étude de ce phénomène a soulevé de nombreuses questions, notamment celle du passage d'une organisation segmentale et discrète des sons de parole en un flux cinématique et acoustique continu, amenant à discuter la nature même des unités de la parole. En effet, si un /g/ est réalisé de manière si différente en fonction du contexte vocalique, quelle est sa réalité phonétique ? Peut-on le caractériser et si oui, dans quel espace ? Puisqu'il est si variable acoustiquement, mais également articulatoirement (en ce qui concerne son lieu de constriction), qu'est-ce qui fait son unité ? En bref, le phénomène de coarticulation a soulevé de nombreuses questions, et a posé, entre autres, le problème de l'invariance phonémique (phonologique) versus la réalisation, en contexte, de ces unités phonologiques très variables si on tente de les mesurer dans de la parole naturelle (interface phonétique - phonologie).

Il existe différents modèles de coarticulation en parole, expliquant chacun de façon différente les processus sous-jacents mis en œuvre lors de la production séquentielle de sons de parole et abordant le problème de la variabilité des réalisations. Nous en présenterons quelques uns dans la partie suivante, qui reflètent les différentes façons d'aborder ce problème.

1.4.2. Les différents modèles théoriques de la coarticulation

Il existe de nombreux modèles théoriques tentant d'expliquer les processus sous-jacents amenant au comportement observé : la coarticulation des sons de parole. Une façon de penser la coarticulation est d'imaginer que les phases temporelles d'exécution des gestes correspondant aux sons de parole sont superposées, se recoupent. Ainsi, le modèle de Browman & Goldstein (1986) *Articulatory Phonology*, est un modèle de **coproduction**, selon lequel la coarticulation résulterait de l'interaction dynamique entre les gestes. Sur une partition gestuelle où chacune des lignes représente un articulateur, la coarticulation est simplement la résultante de la partition gestuelle, qui est modifiée en fonction du contexte. La coarticulation est dans ce cas le produit de l'interaction dynamique des articulateurs. Dans le modèle d'Öhman (Öhman, 1966 ; 1967), qui est également un modèle de coproduction, les voyelles et les consonnes n'ont pas le même statut et sont deux objets de nature différente. Les consonnes sont considérées comme des perturbations insérées entre les voyelles, qui, elles, sont rapprochées, donc coarticulées, selon un principe proche de celui de la superposition temporelle. Parmi les nombreux modèles de coproduction, qui supposent que la composante temporelle est la plus importante dans le processus de coarticulation, on peut citer le Modèle d'Expansion du Mouvement (MEM, Abry & Lallouache, 1995). La spécificité de ce modèle est de prendre en compte la distance entre deux cibles et de moduler l'amplitude de la coarticulation selon ce principe.

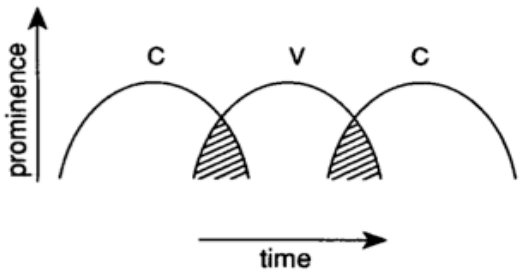


Figure 1.4.1. Illustration des modèles de coproduction, tiré de Nittrouer (1993).

Une autre façon radicalement différente de penser la coarticulation a été proposée par Henke (1966). Le modèle dit « *look-ahead* » de Henke tient compte de la **planification**. Dans ce modèle, la coarticulation n'est pas la résultante « passive » de la coproduction, dictée par la composante temporelle de la partition gestuelle, mais est un procédé plus « intentionnel ». Il faudra attendre Whalen (1990) pour montrer expérimentalement que la coarticulation résulte bien du processus de planification. Parmi les modèles qui tiennent compte de la planification, on citera le modèle de Keating (1990). Selon ce modèle, la coarticulation serait la résultante d'un processus d'optimisation, visant à minimiser l'effort articulatoire lors de la production d'une séquence de sons de parole. Le « *Window model of coarticulation* » (Keating, 1990) est un modèle look-ahead de coarticulation selon lequel la coarticulation résulterait du processus d'économie articulatoire pour atteindre des buts acoustiques ou articulatoires définis. Dans ce modèle, la coarticulation repose sur l'exploitation de la variabilité sub-phonémique (au sein d'une même catégorie phonétique) lors de la production de séquences de sons de parole. Ce modèle prend donc en compte la dimension « goal-directed » de la parole (c'est-à-dire la notion de cibles phonémiques) ainsi qu'un principe de minimisation de coût pour le passage d'une cible à l'autre afin d'expliquer la variabilité de la parole naturelle. En revanche, ce modèle ne spécifie pas explicitement l'espace dans lequel les cibles sont définies, ni quel est le coût à minimiser.

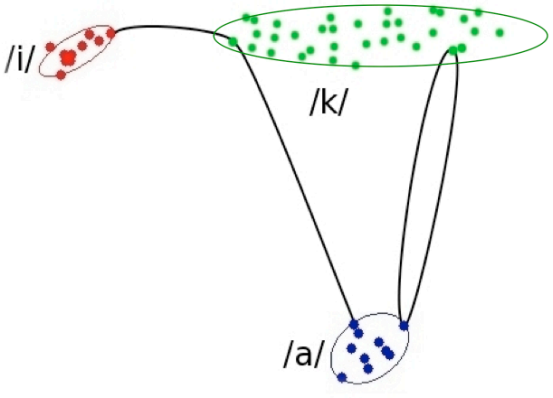


Figure 1.4.2. Illustration des modèles de planification motrice optimale. Noter que dans le cadre de ces modèles, l'optimisation s'effectue dans l'espace moteur, et non pas dans l'espace acoustique. La trajectoire résultante dans l'espace acoustique peut ne pas être la trajectoire la plus courte ou la plus lisse dans cet espace.

Dans la famille des modèles de contrôle optimal, il existe tout un ensemble de modèles visant à voir la coarticulation comme le résultat de la minimisation d'un coût de passage d'une cible à une autre. Cette idée de contrôle optimal est à l'origine née dans le domaine du contrôle moteur (Kawato et al., 1987 ; Jordan & Rumelhart, 1992). Ces modèles diffèrent sur le coût à minimiser, mais s'accordent sur la notion d'optimalité. Le trajet le plus court, le plus fluide ou le moins coûteux, est toujours choisi. Ces modèles reposent sur la notion de planification, puisque la coarticulation est la résultante de la planification volontaire (*look-ahead*) optimale de la parole. Dans cette famille, on citera par exemple les modèles DIVA (Guenther et al., 2004) et GEPETTO (Perrier, 2014), dont nous avons exposé quelques éléments fondamentaux à la section 1.3.

1.4.3. Coarticulation intra-syllabique versus extra-syllabique

La coarticulation ne s'exerce pas de manière homogène sur toute la séquence de sons de parole. Certains segments sont plus affectés que d'autres, certaines unités sont produites davantage conjointement, en particulier parce que la parole est organisée autour d'unités de différents niveaux : le phonème, la syllabe, le mot, la phrase. Nous ne nous intéresserons qu'aux deux premiers niveaux. La syllabe étant une unité d'organisation relativement forte de la parole, il est naturel que la cohésion des éléments au sein de cette unité soit forte.

Afin de mesurer la coarticulation au sein de la syllabe, ou coarticulation intra-syllabique, un indice acoustique, Les Equations de Locus (cf. section 2.3.2.3) a été développé. Selon Sussman et al. (1999), ces Equations de Locus, utilisées dans le cadre d'études développementales, permettent la mesure de l'acquisition progressive du contrôle segmental et de la maîtrise des patrons de coarticulation adulte. L'étude de Sussman et al. (1999) porte sur le développement des Equations de Locus chez un enfant suivi de 7 à 40 mois, lors de la production de séquences CV contenant les consonnes /b d g/. Les patrons de coarticulation CV sont en construction jusqu'à l'apparition des premiers mots, autour de 14-18 mois, et restent relativement stables entre 18 et 40 mois. Les patrons de coarticulation ressemblent donc à ceux de l'adulte dès 18 mois, en ce qui concerne ce type particulier de coarticulation. Etant donné que la coarticulation intra-syllabique de type plosive-voyelle semble être maîtrisée relativement précocement, autour de

l'apparition des premiers mots, ce type de coarticulation, dont les Equations de Locus sont une mesure, ne semblent donc pas âtre un bon indice de maturité du contrôle moteur en ce qui concerne le développement tardif.

Or, sans entrer dans un vif débat, nous pensons que les mesures déduites grâce à cet indice reflètent simplement des contraintes très locales qui peuvent s'expliquer selon les modèles comme celui de la cohésion dans les commandes (*C-center model*, Browman & Goldstein, 2000 ; Byrd, 1995) ou par des contraintes dynamiques (Öhman, 1966 ; 1967) ou articulatoires (Rochet-Capellan & Schwartz, 2007). Les Equations de Locus dans les séquences CV sont censées refléter uniquement la coarticulation anticipatoire entre deux phonèmes voisins. Or, la syllabe est une unité forte, au sein de laquelle, quel que soit le modèle explicatif choisi, l'organisation des segments n'est pas faite de manière indépendante. On mesure donc, avec les Equations de Locus, bien malgré nous, une forte composante due à la coarticulation persévérante.

De plus, les plosives ont, selon leur lieu d'articulation, des contraintes biomécaniques et perceptives plus ou moins fortes, définissant ce que l'on peut appeler un indice de soumission ou de résistance à la coarticulation (Recasens, Pallarès & Fontdevila, 1997). Les plosives bilabiales /p b/ n'exercent presque aucune contrainte sur la position de la langue, puisqu'elles sont produites avec une occlusion des lèvres, la contrainte sur la position de la langue est donc minimale ; les plosives vélaires /k g/ exercent une contrainte plus importante sur les gestes linguaux, puisqu'ils imposent un contact entre le dos de langue et le palais, mais la région autorisée pour effectuer ce contact est large, la restriction sur la position de la langue n'est donc pas très exigeante ; enfin, les plosives alvéolaires ou dentales /t d/ exigent, quant à elles, une position de la langue très contrainte, puisqu'un contact entre la pointe de la langue et les dents ou les alvéoles doit être effectué, et la région de ce contact est relativement restreinte, la contrainte sur la position de la langue est dans ce cas maximale. Ce sont ces contraintes biomécaniques et perceptives universelles que l'on retrouve via les Equations de Locus : la pente est maximale pour les bilabiales, un peu moins forte pour les vélaires et encore moins forte pour les alvéolaires et dentales. Certes, on mesure bien via ces équations la relation qu'entretient la consonne avec la voyelle, et le rapprochement entre ces deux unités.

Mais cette mesure reflète des contraintes inhérentes à toute parole composée d'alternances entre voyelles et consonnes, contraintes biomécaniques et perceptives universelles et de bas niveau, et reflète donc un processus, qui certes traite de la variabilité des objets consonantiques et vocaliques et des relations qu'ils entretiennent entre eux, mais ne reflète pas toute la complexité de la construction de relations sensori-motrices des unités de la parole. En effet, ces contraintes, qui sont donc de bas niveau, doivent être intégrées à un stade très précoce dans le développement, et leur étude ne sera donc que très peu utiles en ce qui concerne le développement tardif de la production de la parole.

De notre point de vue, la coarticulation intra-syllabique, particulièrement de type plosive-voyelle, reflète donc davantage des contraintes universelles de bas niveau, telles que les contraintes biomécaniques et la coarticulation persévérante. Dans le but d'étudier la capacité des enfants à anticiper, lors de la production d'une séquence de sons de parole, les sons de parole suivants, nous pensons que la coarticulation anticipatoire extra-syllabique repose davantage sur la planification des sons de parole et est moins le produit de contraintes universelles de bas niveau. De notre point de vue, la coarticulation extra-syllabique constitue donc un meilleur indice dans le but de mesurer la capacité des enfants à associer à chaque son de parole, une certaine variabilité autorisée, et à utiliser cette variabilité lors de la production d'une séquence de sons de parole dans le but de minimiser un coût. Cependant, comme nous allons le voir dans la section suivante, un certain nombre d'études s'est focalisé sur la coarticulation intra-syllabique, et très peu d'études ont à ce jour étudié le développement de la coarticulation extra-syllabique.

1.4.4. La coarticulation chez l’enfant

Afin de caractériser la maturité des productions enfantines, l'étude de la stabilité des productions semble être un indice solide. Un second indice très important est celui permettant de quantifier l’organisation spatio-temporelle de la parole. Atteindre une cible, et être capable de bien définir cette cible afin de la produire de façon identique à chaque répétition demande une certaine maturité motrice. Mais apprendre quelle est la variabilité autorisée autour d'une cible donnée, et être capable d'utiliser cette variabilité

lors de la planification puis de l'exécution d'une séquence de cibles, dans le but de minimiser l'effort articulatoire, est encore bien plus complexe.

De notre point de vue, anticiper la (les) cible(s) phonémique(s) suivante(s) demande en effet d'avoir une certaine idée de la variabilité autorisée autour de chacune des cibles phonémiques de la séquence, afin que l'interlocuteur catégorise correctement le phonème en question, et d'utiliser cette variabilité sub-phonémique afin de minimiser le coût du passage d'une cible phonémique à une autre. La variabilité sub-phonémique ou intra-catégorielle en répétition peut être utilisée pour mesurer la stabilité du contrôle musculaire permettant la production de sons de parole. Une grande variabilité indique ainsi une immaturité motrice, puisque une des caractéristiques de la parole adulte est sa grande consistance (ou grande stabilité en répétition). Une grande variabilité en production est typique des productions enfantines. Au contraire, dans le cadre de la coarticulation anticipatoire, l'utilisation de cette variabilité sub-phonémique à des fins de planification optimale est un gage de maturité. On observe en effet chez l'adulte une forte amplitude de coarticulation anticipatoire, et donc une forte variabilité sub-phonémique dans ce cadre. Chez l'enfant, on s'attend au contraire à observer une moindre amplitude de coarticulation que chez l'adulte, car on suppose que l'enfant n'a pas la maturité sensori-motrice nécessaire pour tirer parti de cette variabilité, qui lui aurait pourtant permis de produire des séquences de sons de parole pour un moindre effort articulatoire.

1.4.4.1. Revue de la littérature sur la coarticulation linguale chez l'enfant

On dénombre, à ce jour et à notre connaissance, une vingtaine d'études portant sur la coarticulation linguale anticipatoire chez l'enfant. La majorité de ces études a été produite durant les années 1990 et repose sur des mesures acoustiques des signaux de parole enfantins, même si quelques études articulatoires commencent à voir le jour. Une première lecture rapide de cette littérature amène à penser que ces résultats se contredisent les uns les autres et qu'il est difficile d'en tirer un schéma développemental clair, ou de conforter un cadre théorique, quel qu'il soit.

En effet, à première vue, ces résultats sont contradictoires, car les résultats des études de Goodell & Studdert-Kennedy (1993) ; Nittrouer et al. (1989) ; Nittrouer et al. (1996) ; Siren

& Wilcox (1995) et Zharkova et al. (2011) suggèrent que l'amplitude des patrons de coarticulation décroît avec l'âge ; les résultats des études de Sussman et al. (1999) ; Nittrouer (1993) ; Repp (1986) ; Turnbaugh et al. (1985) ; Sereno et al. (1987) ; Sereno & Lieberman (1987) ; Katz, Kripke & Tallal (1991) et Noiray, Ménard & Iskarous (2013) suggèrent que l'amplitude des patrons de coarticulation est stable avec l'âge ; et enfin, les résultats des études de Repp (1986) ; Hodge (1989) ; Kent (1984) ; Sereno et al. (1987) et Zharkova et al. (2012) ; (2014) suggèrent que l'amplitude des patrons de coarticulation croît avec l'âge. Il est tentant de conclure que ces résultats contradictoires sont le fruit d'effets divers, de corpus différents, de types de mesures et d'indices différents, ou que ces résultats contradictoires sont simplement le fruit d'un développement non uniforme, constitué d'étapes successives au cours desquelles l'enfant se focalise sur certaines composantes de la parole. Une vue d'ensemble de ces études est proposée dans le tableau 1.4. ci-dessous.

Tableau 1.4. Vue d'ensemble des études portant sur le développement de la coarticulation.

Etude	Partici-pants	Age	Séquences étudiées	Indices mesurés	Détails additionnels	Résultats
Sussman et al., 1999	n=1	7 à 40 mois	/b,d,g/ + V	Equations de Locus	Différents patrons de coarticulation pour chaque plosive, qui tendent vers des patrons adultes autour de 16 mois	Patrons et amplitudes similaires à l'adulte à partir de 16 mois et jusqu'à 40 mois
Goodell & Studdert-Kennedy, 1993	n=12	6 filles de 20 à 27 mois, suivies longitudinalement 10 mois et 6 femmes adultes	/bə/ + /b,d,g/ + /i,a,u/	Suivi du F2	Première étude longitudinale, similarités avec Sussman et al., 1999.	L'amplitude de la coarticulation décroit avec l'âge, et semble être stable autour de 3 ans
Repp, 1986	n=3	4 ans, 9 ans, adulte	/s,t/ + /i,a,u/	Equations de Locus et suivi de F2 dans les voyelles	A également mesuré la coarticulation extra-syllabique dans des séquences de type /ə.CV/	Autant de coarticulation intra-syllabique chez l'enfant de 4 ans que chez l'adulte. Pas de marque de coarticulation extra-syllabique à 4 ans, mais une nette marque à 9 ans
Hodge, 1989	n=40	3, 5, 9 ans et adultes (n=10 par groupe)	/ə.stV/ avec V = /i,u/	F2 en Bark	La mesure en Barks permet une certaine normalisation des données et permet de s'affranchir des problèmes de relativité des mesures entre conduits vocaux de différentes tailles	Seuls les enfants de 9 ans et les adultes anticipent la voyelle accentuée dans le schwa
Nittrouer, 1993	n=40	3, 5, 7 ans et adultes (n=10 par groupe)	/t,k/ + /i,a,u/	Suivi du F2 Equations de Locus	La plupart des effets recherchés ne sont pas significatifs ou très peu puissants statistiquement	Amplitude de la coarticulation similaire à l'adulte dès 3 ans

Kent, 1984	n=6	4 ans et adultes (n=3) par groupe	/boks/	Valeurs des formants du /o/ juste avant le /k/	F2 monte avant la réalisation du /k/ pour les adultes, pas pour les enfants	Moins de coarticulation chez l'enfant
Turnbaugh et al., 1985	n=9	3, 5 ans et adultes (n=3 par groupe)	/b,d,g/ + /i,u/	Suivi du F2 Equations de Locus	Coarticulation labiale (anticipation de l'arrondissement des lèvres)	Patrons de coarticulation similaires à ceux de l'adulte dès 3 ans.
Sereno et al., 1987	n=12	enfants (entre 3 et 7 ans, n = 8), adultes (n=4)	/s,t,d/ + /i,u/	Suivi du F2 Equations de Locus	Coarticulation labiale	Patrons de coarticulation similaires à ceux de l'adulte dès 3 ans.
Sereno & Lieberman,1987	n=19	enfants (entre 3 et 7 ans, n = 14), adultes (n=5)	/k/ + /i,a/	Suivi du F2 Equations de Locus	-	Patrons de coarticulation similaires à ceux de l'adulte dès 3 ans
Noiray et al., 2013	n=6	4-5 ans	/p,t,k/ + /i,a,u/	Equations de Locus Point le plus haut de la langue	Données échographiques sans restriction ou correction de mouvement et données acoustiques	Patrons de coarticulation similaires à ceux de l'adulte dès 4-5 ans
Nittrouer et al., 1989	n=30	3, 5 et 7 ans (n=10 par groupe)	/ʃ,s/ + /i,u/	CoG et F2	Mesures temporelles absolues (30 ms et 100 ms avant le début de la voyelle)	L'amplitude de la coarticulation décroit avec l'âge
Nittrouer et al., 1996	n=40	3, 5, 7 ans et adultes (n=10 par groupe)	/ʃ,s/ + /i,a/	Suivi du F2 dans la fricative et dans la voyelle	Pas de différence significative entre les groupes d'âge d'enfants	L'amplitude de la coarticulation est plus forte chez les enfants que chez les adultes
Katz et al., 1991	n=40	3, 5, 8 ans et adultes (n=10 par groupe)	/s/ + /i,u/	CoG et F2	-	Patrons de coarticulation similaires à ceux de l'adulte dès 3 ans
Siren & Wilcox, 1995	n=40	3, 5, 7 ans et adultes (n=10 par groupe)	/ʃ,s/ + /i,u/	CoG et F2	Mesures temporelles absolues (30 ms et 100 ms avant le début de la voyelle)	L'amplitude de la coarticulation décroit avec l'âge
Zharkova et al., 2011	n=20	6-9 ans (n=10) et adultes (n=10)	/ʃ/ + /i,a,u/	Distance Euclidienne entre les contours de langue	Aucune mesure acoustique. Différences de l'ordre du dixième de millimètre alors que la précision de la mesure échographique est plutôt de l'ordre du millimètre.	Plus de coarticulation chez les enfants que chez les adultes.
Zharkova et al., 2012	n=20	6-9 ans (n=10) et adultes (n=10)	/s/ + /i,a,u/	idem	Aucune mesure acoustique. Le contraste mesuré est plus important que celui de Zharkova et al., 2011. Mesures effectuées au milieu de la fricative.	Pas d'effet de coarticulation mis en évidence chez les enfants, alors que c'est le cas pour les adultes.
Zharkova et al., 2014	n=30	10-12 ans et adultes (n=15 par groupe)	/s, ʃ/ + /i,a/	idem	Moins de coarticulation chez les adolescents que chez les adultes.	L'anticipation de la voyelle dans /ʃ/ est observée à la fin de /ʃ/ pour les adolescents alors qu'elle l'est plus tôt pour les adultes.

Nous avons regroupé dans le tableau 1.4. les études portant sur le développement précoce de la coarticulation (Sussman et al., 1999 et Goodell & Studdert-Kennedy, 1993), ceux portant sur la coarticulation extra-syllabique (Repp, 1986 et Hodge, 1989), ceux portant sur la coarticulation intra-syllabique de type plosive-voyelle (de Nittrouer, 1993 à Noiray et al., 2013) et ceux portant sur la coarticulation intra-syllabique de type fricative-voyelle (de Nittrouer et al., 1989 à Zharkova et al., 2014).

Ces résultats, contradictoires au premier abord, sont par ailleurs venus alimenter un débat autour de l'organisation de la parole au cours du développement. Pour certains auteurs, une amplitude de coarticulation moindre serait le signe de l'immaturité du contrôle, puisqu'une indépendance segmentale accompagnée de connaissances incomplètes sur les degrés de liberté du système moteur de production de parole amènerait à de tels patrons : très peu de coarticulation. A l'inverse, certains auteurs, partisans de la théorie selon laquelle la parole enfantine serait organisée autours d'unités larges, puis de plus en plus restreintes au fil du développement, pour atteindre une organisation segmentale semblable à l'adulte, voient au contraire l'immaturité du contrôle dans des patrons très coarticulés. Puisqu'il n'y a, selon ce point de vue, pas d'indépendance du segment, une forte amplitude de coarticulation signifierait une parole immature, puisque les gestes sont entremêlés et que la différenciation des gestes articulatoires est un marqueur de parole adulte. On imagine combien chacune des études susmentionnées a pu apporter des éléments en faveur de l'une ou de l'autre des théories. Nous reviendrons plus tard sur cet aspect, dans la section 1.4.5.

Mais une lecture plus approfondie de ces études et de leurs résultats suggère en réalité que ces résultats ne sont pas si contradictoires qu'il n'y paraît au premier abord. En effet, si l'on ne considère que les études portant sur la coarticulation intra-syllabique CV de type plosive-voyelle, à savoir Sussman et al. (1999) ; Nittrouer (1993) ; Repp (1986) ; Turnbaugh et al. (1985) ; Sereno et al. (1987) ; Sereno & Lieberman (1987) ; Noiray et al. (2013), toutes ces études s'accordent sur le fait que l'amplitude de la coarticulation CV est similaire à celle de l'adulte dès 3 ans, et reste similaire à l'amplitude de coarticulation adulte après 3 ans (même après 18 mois si l'on en croit les résultats de Sussman et al. (1999) ne portant que sur une enfant). D'après ces résultats, et en particulier ceux de Sussman et al. (1999) et ceux de Goodell & Studdert-Kennedy (1993), qui ont effectué des études longitudinales

du développement de la coarticulation linguale, l'amplitude de la coarticulation intra-syllabique serait similaire chez l'enfant et chez l'adulte, et ce dès l'apparition des premiers mots, correspondant à l'émergence du contrôle segmental de la parole, selon Sussman et al. (1999). En effet, la syllabe CV semble être une des unités de base de la parole (d'après les données typologiques et les patrons développementaux du babillage et des premières lexies, MacNeilage & Davis, 1990 ; 2000), et apprendre à produire conjointement consonne et voyelle au sein d'une syllabe est une aptitude élémentaire qu'il est nécessaire d'acquérir tôt.

Ensuite, si l'on ne considère que les études portant sur la coarticulation intra-syllabique CV de type fricative-voyelle, à savoir Nittrouer et al. (1989) ; Nittrouer et al. (1996) ; Katz et al. (1991) ; Siren & Wilcox (1995) ; Zharkova et al. (2011) ; (2012) ; (2014), les études de Nittrouer et celle de Siren & Wilcox ainsi que Zharkova et al. (2011) tendent à montrer que l'amplitude de la coarticulation décroît avec l'âge, alors que l'étude de Katz et al. (1991) tend à montrer que l'amplitude de la coarticulation est stable avec l'âge, et les études de Zharkova et al. (2012) et (2014) tendent à montrer que l'amplitude de la coarticulation croît avec l'âge. Or selon nous, parmi les études portant sur la coarticulation intra-syllabique des fricatives et des voyelles, quatre comportent selon nous des biais évidents : chaque étude qui tend à montrer qu'il existe plus de coarticulation chez l'enfant que chez l'adulte, dans ce type de séquence fricative-voyelle, comporte au moins un biais évident selon nous. Tout d'abord, les études de Nittrouer et al. (1989) et (1996) et l'étude de Siren & Wilcox (1995) utilisent des mesures temporelles absolues, ce qui est un problème selon nous. En effet, les fricatives produites par les enfants ayant tendance à avoir des durées supérieures aux fricatives produites par les adultes (Lee, Potamianos & Narayanan, 1999), effectuer des mesures 30 ms ou 100 ms avant le début de la voyelle, donc de manière absolue, ne revient pas au même dans de la parole adulte et dans de la parole enfantine. En effet, dans une parole au débit rapide, 30 ms avant le début de la voyelle peut être un instant proche du milieu de la fricative, alors que pour une parole au débit lent, 30 ms avant le début de la voyelle reflète la fin de la fricative. De plus, Nittrouer et al. (1996) utilise le suivi du F2 dans des fricatives non voisées, mesure qui peut s'avérer délicate. Second biais de ces études selon nous, les mesures acoustiques sont également effectuées de manière absolue, à savoir directement en Hertz (Hz), or on

sait que les conduits vocaux des enfants, des femmes et des hommes adultes sont très différents. Les différences observées en Hz pourraient être uniquement le fruit de ces différences morphologiques. Enfin, l'étude de Zharkova et al. (2011), qui ne comporte que des mesures articulatoires et vient conforter les études précédemment citées, fait l'état de différences de l'ordre du dixième de millimètre, car le contraste étudié est très fin, et les différences mesurées sont bien en dessous de la précision de l'outil utilisé, à savoir l'échographie linguale. Il est donc très difficile d'apprécier les conclusions de cette étude. Pour conclure, les seules études qui ne comportent pas, selon nous, de biais majeurs et évidents sont celles qui montrent qu'il y a moins (Zharkova et al., 2012 ; 2014) ou autant (Katz et al., 1991) de coarticulation chez l'enfant que chez l'adulte pour ce type particulier de séquences fricative-voyelle.

Enfin, si l'on ne considère que les études portant sur la coarticulation extra-syllabique, à savoir Repp (1986) ; Hodge (1989) ; Goodell & Studdert-Kennedy (1993), les résultats sont encore une fois bien moins contradictoires qu'il n'y paraît. L'étude de Goodell & Studdert-Kennedy suggère que les très jeunes enfants coarticulent davantage que les adultes, et ce jusqu'autour de 2 ans et demi, où les patrons sont, selon les auteurs, similaires à ceux observés chez l'adulte. Ces résultats peuvent être expliqués par un manque d'indépendance des gestes articulatoires à un stade précoce du développement. Les études de Repp (1986) et Hodge (1989), qui s'intéressent à un développement plus tardif, montrent que les enfants de 3 ans ne montrent aucun signe d'anticipation au-delà de la syllabe, alors que les enfants de 9 ans et les adultes montrent des signes d'anticipation. Ces deux études sont les seules à s'être intéressées, à notre connaissance, à la coarticulation extra-syllabique au cours du développement tardif et montrent que ce type de coarticulation, inexistante chez les très jeunes enfants, se développe avec l'âge.

Pour conclure, les résultats de la littérature portant sur le développement de la coarticulation linguale anticipatoire ne sont pas si contradictoires qu'il n'y paraît au premier abord et peuvent être intégrés dans notre cadre théorique. La coarticulation intra-syllabique de type plosive-voyelle semble être maîtrisée relativement précocement, au sortir des premiers mots (aux alentours de 16-20 mois), et l'amplitude des patrons de coarticulation semblent rester stables après cette étape ; la coarticulation intra-syllabique de type fricative-voyelle semble être maîtrisée plus tardivement et les données tendent à

montrer que pour ce type particulier de séquences, l'amplitude de la coarticulation augmente avec l'âge ; et enfin, concernant la coarticulation au-delà de la frontière syllabique, faisant appel à des représentations sensori-motrices plus fines et à des capacités de production de parole plus élaborées, les enfants ne semblent atteindre les patrons adultes que plus tard dans le développement, autour de 8 ou 9 ans, et montrent avant cet âge une absence d'effet de coarticulation ou une amplitude de coarticulation beaucoup plus réduite.

Pour résumer, l'ensemble de ces études apporte des informations cohérentes si l'on exclue 4 de ces études, et il est possible de résumer les choses en ces mots : la coarticulation intra-syllabique est stable avec l'âge très tôt, dès les premiers mots, au moins en ce qui concerne les syllabes CV simples de type plosive-voyelle. En ce qui concerne les syllabes de type fricative-voyelle, il est possible qu'elles soient maitrisées plus tard et les données tendent plutôt à montrer que les enfants coarticulent moins que les adultes voyelles et fricatives, mais il n'y a pas de consensus sur ce sujet. En ce qui concerne la coarticulation extra-syllabique, les données tendent à montrer qu'il y a de plus en plus de coarticulation avec l'âge, ce qui va dans notre sens.

Seules deux études, portant sur la coarticulation intra-syllabique de type plosive-voyelle, celle de Goodell & Studdert-Kennedy (1993) et celle de Sussman et al. (1999), tendent à montrer qu'il existe davantage de coarticulation chez les très jeunes enfants (autour de 2 ans et avant) que chez les adultes. Mais ces résultats ne sont pas incohérents de notre point de vue, car ils concernent le développement plus précoce, reflétant le passage du babillage à l'entrée dans la parole, qui partage les caractéristiques de base de la parole adulte. Or à ce stade, au moment d'entrer véritablement dans l'ère de la parole, il se pourrait que le très jeune enfant organise la production de sa parole autour d'unités plus larges, telles que la syllabe, la lexie ou encore des patrons articulatoires ou *Vocal Motor Schemes* (Vihman, 1996 ; 1998 ; Velleman & Vihman, 2007). Plusieurs études tendent en effet à corroborer cette hypothèse (Vihman, 1981 ; 1993 ; Vihman et al., 1986) qui permettrait d'expliquer pourquoi on retrouve davantage de coarticulation dans la parole enfantine avant 20 mois.

1.4.5. L'apprentissage des patrons de coarticulation et la nature des unités de la parole : approche phonologique des productions précoces et tardives

En ce qui concerne le domaine d'organisation de la parole enfantine, l'étude des patrons de coarticulation permet d'apporter des informations, en particulier en comparant coarticulation intra-syllabique et extra-syllabique. En effet, si l'on observe une coarticulation intra-syllabique forte et peu ou pas de coarticulation extra-syllabique, cela suggèrerait que le domaine d'organisation de la parole des enfants serait centré autour de la syllabe. Si l'on observe un degré similaire de coarticulation intra et extra-syllabique, cela indiquerait que les domaines d'organisation de la parole des adultes et des enfants seraient organisés autour des mêmes unités, à savoir les segments. Si l'on observe un degré plus important de coarticulation extra-syllabique et un degré similaire de coarticulation intra-syllabique, cela indiquerait que le domaine d'organisation de la parole enfantine serait centré autour d'unités plus larges que la syllabe. Enfin, si l'on observe moins de coarticulation intra et extra-syllabique chez l'enfant, cela indiquerait que son apprentissage est centré autour du phonème, du segment, unités qu'il apprendrait à combiner afin de produire des séquences de sons de parole.

Les résultats de Goodell & Studdert-Kennedy (1993) (forte amplitude de coarticulation intra et extra-syllabique au cours du développement précoce) suggèrent que le domaine d'organisation de la parole serait, autour de 20 mois, centré autour d'unités plus larges que le phonème, voire plus larges que la syllabe, du type de ce que propose Vihman (Vihman, 1996 ; Velleman & Vihman, 2007). Ces résultats viennent donc plutôt conforter l'hypothèse selon laquelle l'organisation de la parole enfantine serait centrée autour d'unités larges, de type syllabe (Nittrouer et al., 1996) ou même très larges, du type du mot entier ou d'un patron sous-jacent de même ampleur qu'un mot (comme les *Vocal Motor Schemes* de Vihman). Les enfants n'apprendraient donc à différencier les gestes segmentaux que tardivement et l'organisation de leur parole ne deviendrait segmentale que plus tard au cours du développement. Nous pensons effectivement que lors du babillage et des premiers mots, la production de la parole enfantine n'est pas organisée comme la parole adulte, à un niveau segmental. L'idée de Vihman nous paraît séduisante et peut expliquer en partie les résultats des études de Goodell & Studdert-Kennedy (1993)

et ceux de Sussman et al. (1999), indiquant qu'avant la production des premiers mots, les patrons de coarticulation de /dV/ et /gV/ sont d'amplitude supérieure à ceux de l'adulte. Ces résultats pourraient s'expliquer par une organisation de la parole d'ordre supérieur à l'organisation segmentale, au cours de laquelle l'indépendance des segments et donc des gestes articulatoires n'est pas encore acquise.

Mais comme Cutler & Mehler (1993) le soulignent, l'enfant est naturellement attiré, en perception, vers des unités de type segmentales, avant même la fin de la première année de vie. Concernant la production, Hallé & Cristia (2012) supportent l'idée (en accord avec MacNeilage) qu'au cours du babillage, les enfants disposent de représentations sous-spécifiées en termes de voyelles et de consonnes, plutôt spécifiées en termes de syllabe, et qu'à partir des premiers mots et du développement lexical, ces représentations deviennent peu à peu spécifiées phonétiquement, c'est-à-dire d'ordre segmental. Les auteurs concluent : « *representations for production definitely develop from whole-word to analytic representations into phoneme-like elements at around 2 years of age* ». Il semblerait donc que l'organisation de la parole enfantine soit, en production comme en perception, organisée autour des segments, et ce de manière relativement précoce dans le calendrier développemental (autour de 2 ans).

Si l'idée que la parole enfantine, lors du babillage et du développement précoce, est organisée autours d'unités larges et sous-spécifiées en termes phonétiques nous semble vraisemblable, il semble invraisemblable d'appliquer ce raisonnement au développement tardif, tel que le fait Nittrouer. Les conclusions que tire Nittrouer de ses travaux sont que les enfants de 7 ans n'ont pas encore acquis l'indépendance des fricatives et des voyelles. Il est, selon nous, tout à fait possible d'expliquer la forte coarticulation du babillage par le manque d'indépendance des segments et donc des gestes articulatoires, mais il est difficilement concevable d'attribuer les résultats de Nittrouer et al. (1989) et (1996) à un manque d'indépendance du segment, à un âge aussi avancé que celui de 7 ans.

Si l'on adopte une vision strictement segmentale (phonémique) de l'organisation de la parole et que l'on applique cette vision au développement, alors on devrait observer des enfants qui apprennent d'abord à produire et à contraster des sons uniques, puis qui apprennent à produire conjointement plusieurs sons pour former leurs premiers mots. On

devrait donc observer de plus en plus de coarticulation, à mesure que l'enfant apprend à produire les sons conjointement, et donc à les rapprocher les uns des autres pour les coproduire. Or il n'en n'est rien. Aucun enfant n'apprend d'abord à contraster tous les sons de parole de sa (ses) langue(s) maternelle(s), tous les enfants les combinent, babillent, intègrent la structure rythmique de la parole avant de maîtriser tous les contrastes phonétiques, et c'est au travers de cette pratique que vont émerger les contrastes phonétiques (cf. section 1.1., théorie Frames, then Content) opposant les segments qui deviennent peu à peu les unités de base du contrôle et de la perception (avant la fin de la première année de vie pour la perception et autour de 2 ans pour la production). Une vision purement segmentale de l'organisation de la parole, dès le départ, ne peut donc tout simplement pas expliquer les données développementales.

Une autre vision, à l'opposé, suggère que l'organisation segmentale de la parole est le point d'arrivée et non le point de départ du développement (e.g. Vihman, 1998). Selon ce point de vue, les jeunes enfants, au cours du babillage et de la production des premiers mots, organiseraient davantage leur parole autour d'unités larges, telles que la syllabe ou le mot, ou encore des patrons articulatoires ou *Vocal Motor Schemes* comme suggéré par Vihman. A travers une approche similaire, mais appliquée à un autre stade du calendrier développemental, celui du développement phonétique tardif, Nittrouer propose que le domaine d'organisation de la parole enfantine soit la syllabe. Nous avons évoqué au cours des paragraphes précédents en quoi cette hypothèse, au vu des connaissances dont nous disposons concernant le développement phonétique et phonologique et au vu du calendrier développemental, nous semble peu vraisemblable.

Pour conclure, l'observation, à un stade très précoce, de la forte coarticulation intra-syllabique pour /gV/ et /dV/ dans les études de Goodell & Studdert-Kennedy (1993) et de Sussman et al. (1999), devrait s'accompagner de l'absence de coarticulation extra-syllabique pour aller dans le sens d'une organisation syllabique de la parole enfantine. En effet, si on observe une forte coarticulation au sein de la syllabe et l'absence de coarticulation entre syllabes, cela irait dans le sens d'une organisation syllabique de la parole enfantine. Or les seules données empiriques dont on dispose montrent le contraire, puisque Goodell & Studdert-Kennedy (1993) observent de fortes amplitudes de coarticulation intra-syllabique et extra-syllabique. Ces résultats suggèrent donc que

l'organisation de la parole enfantine est centrée, à un stade précoce, autour d'unités plus larges que la syllabe, allant dans le sens de ce que propose Vihman.

Nous ne débattrons pas davantage de phonologie, du domaine d'organisation de la parole enfantine, ni de l'existence des syllabes, phonèmes et mots dans les tâches de planification et d'exécution de parole. Nous adopterons un point de vue strictement phonétique. De notre point de vue, apprendre à coarticuler, c'est apprendre à produire conjointement, dans une séquence, différents sons de parole, et cela requiert un certain degré de maîtrise des aspects spatiaux et temporels de la production de parole. De notre point de vue, apprendre à produire de façon séquentielle des sons de parole demande donc une certaine connaissance de la variabilité autorisée pour un son de parole donné, et d'utiliser cette variabilité permise afin de modifier la réalisation de ses unités de parole en fonction du contexte. En cohérence avec ce cadre théorique, une fois que l'organisation de la parole des jeunes enfants est d'ordre segmental, c'est-à-dire après l'apparition des premiers mots, nous devrions observer une amplitude croissante de la coarticulation avec l'âge. C'est d'ailleurs sur le développement tardif que portent nos travaux, et nous n'entrerons pas davantage dans le détail du développement précoce, ni dans le débat portant sur la nature des unités de parole avant un stade où il est considéré que le domaine d'organisation de la parole des enfants est le segment.

1.4.6. En somme...

La coarticulation est un phénomène complexe qu'il est possible d'aborder selon différents angles d'attaque : coproduction basée sur la superposition temporelle de la définition des cibles phonémiques ou minimisation d'un coût, effectué grâce à la planification optimale basée sur l'exploitation de la variabilité autorisée autour des cibles phonémiques. Nous avons pris le parti théorique de la seconde proposition. Dans ce cadre, à partir du moment où le phonème comme unité de base de la parole a émergé, nous devrions observer une amplitude croissante de la coarticulation avec l'âge, reflétant la maturation croissante des représentations sensori-motrices des sons de parole et la capacité croissante des enfants à utiliser la variabilité allouée à un son de parole donné dans le but de minimiser l'effort articulatoire tout au long de la production d'une séquence de sons de parole.

Concernant la coarticulation extra-syllabique, qui reflète davantage le processus de planification motrice et est moins tributaire des contraintes biomécaniques, l'amplitude de la coarticulation semble effectivement croitre avec l'âge. Concernant la coarticulation intra-syllabique, la très large majorité des données expérimentales est en cohérence avec notre cadre théorique, et quelques études développementales restent contradictoires. Mais comme nous l'avons vu, certaines études comportent des biais évidents, la majorité de ces études repose uniquement sur des données acoustiques, et ces études comportent peu de participants. Si la définition d'un profil ontogénétique clair reste difficile à effectuer à partir de ces études développementales parcellaires et quelque peu contradictoires, la très large majorité des données développementales portant sur la coarticulation peut, cependant, être intégrée dans un cadre théorique cohérent de la production de parole, que nous avons appliqué au développement.

1.5. La construction des représentations sensori-motrices

Au cours de son développement, l'enfant apprend le contrôle de ses mouvements en contractant ou relâchant volontairement ses muscles dans diverses tâches et en prêtant attention aux retours sensoriels que ses actions produisent. La première phase de cet apprentissage consiste en une exploration libre, au cours de laquelle l'enfant prête attention aux retours sensoriels produits par ses actions. Les toutes premières explorations vocales des nouveau-nés et le babillage canonique en sont de merveilleux exemples. Pendant cette période, l'enfant explore librement son conduit vocal et apprend progressivement quelles sont les conséquences acoustiques et somatosensorielles de ses mouvements articulatoires. C'est très probablement dès cette période que se construisent des « cartes » des relations entre commandes motrices et retours sensoriels, ce que nous appelons les représentations sensori-motrices de la parole.

Se pose ensuite la question de l'action orientée vers un but. A quel moment l'enfant oriente réellement son action vers l'atteinte d'un but sensoriel ? A quel moment commence t-il à utiliser les connaissances de son appareil vocal et ses représentations sensori-motrices pour prédire les conséquences de ses gestes articulatoires, et atteindre le but sensoriel désiré, ou corriger sa production pour aller vers le but désiré ? Comme pour la perception, pour laquelle les catégories phonétiques propres à une langue maternelle semblent se construire sur la base distributionnelle des propriétés acoustiques de la parole adulte (Saffran et al., 1996), nous pensons qu'en ce qui concerne la production, le caractère répétitif et les régularités du babillage canonique servent en partie à l'enfant à extraire des régularités des signaux acoustiques et des conséquences somatosensorielles de ses propres productions, lui permettant d'établir des objectifs moteurs sur la base des régularités sensorielles de ses propres productions.

Il semblerait que, même à un stade très précoce, les enfants soient capables de gestes relativement intentionnels. En effet, les nouveau-nés sont capables d'imiter des gestes oro-faciaux tels que la protrusion de la langue (Meltzoff & Moore, 1977) ou imiter des sons de parole (Kuhl & Meltzoff, 1996). Ces résultats montrent qu'il existe très probablement des mécanismes très précoces permettant de lier perception et action.

Selon un schéma développemental largement accepté, ce sont les parents qui spécifient à l'apprenant le but de la tâche, et l'apprenant qui reproduit. L'imitation serait donc centrale dans le processus d'acquisition. Mais un autre schéma développemental est possible, celui du renforcement. Howard & Messum (2011) ; (2014) montrent (dans le cadre d'un modèle computationnel de production de parole) que l'apprentissage de la production de parole peut être réalisé en l'absence totale de processus d'imitation. L'enfant explore librement son conduit vocal, et l'adulte apporte un feedback aux productions de l'enfant. L'apprentissage de la production de sons de parole est donc réalisable sans imitation. En revanche, l'apprentissage est toujours supervisé et nécessite un professeur, qui spécifie le but de la tâche, que ce soit en amont ou en aval de la production.

L'apprentissage des représentations sensori-motrices permet ensuite d'effectuer des prédictions quant aux conséquences des gestes articulatoires, ce qui est central dans le processus de production (cf. section 1.3 sur les modèles directs et inverses). Comme nous l'avons vu, le processus de production de parole (mais aussi de production de gestes rapides en général) repose en effet en grande partie sur les prédictions, car il est nécessaire de programmer avant même de produire, soit en se basant sur les modèles internes, soit en se basant sur des patrons moteurs mémorisés, car les boucles de retour sensoriel sont trop lentes pour permettre à la production d'être uniquement guidée par les retours. Dans le cadre théorique que nous avons présenté dans la partie précédente, ces mécanismes sont effectués soit par le modèle direct (composante prédictive) soit par le modèle inverse (patrons moteurs mémorisés), tous deux nourris par l'apprentissage, et permettant respectivement le contrôle direct (*feedforward*) et le contrôle par retour (*feedback interne*).

Dans les parties qui suivent, nous définirons d'abord ce que nous appelons les représentations sensori-motrices de la production de parole, nous présenterons ensuite les différents paradigmes expérimentaux qui permettent leur étude, puis nous ferons un rapide état de l'art des études développementales concernant le contrôle moteur du bras et de la main, avant d'aborder la question de la maturation des relations sensori-motrices dans le cadre spécifique de la production de la parole à travers une revue de la littérature portant sur des études développementales de production perturbées ou non.

1.5.1. Définition

Il est possible de définir ce que l'on appelle les représentations sensori-motrices comme étant *les représentations neuronales associées à la relation liant commandes motrices et sensations acoustiques et somatosensorielles*. En d'autres termes, elles désignent le stockage (la mémorisation) des informations liées à la relation entrée-sortie de la boucle perceptuo-motrice de la parole, liant les commandes motrices (entrée) aux retours acoustiques et somatosensoriels (sortie).

Dans le cadre théorique qui est le nôtre (qui a été présenté précédemment, cf. section 1.3), ces représentations neuronales émergent de l'apprentissage des modèles inverse et direct. Ce sont ces deux modules qui opèrent la rétention des informations liant commandes motrices et signaux acoustiques et somatosensoriels. L'apprentissage repose d'abord sur l'exploration, qui permet dans un premier temps de construire le modèle direct, puis fait intervenir ce modèle direct dans le but de multiplier les expériences (virtuelles), afin d'élaborer le modèle inverse. C'est le modèle inverse qui permet ensuite le contrôle prédictif (*feedforward*).

Ces représentations neuronales associées aux tâches de parole, si elles se construisent probablement dès les premières vocalisations, mettent un certain nombre d'années à se consolider. C'est l'aboutissement de la construction, lente et progressive, de ces représentations neuronales associées aux tâches de parole, qui permet aux adultes ayant une bonne connaissance des conséquences perceptives de leurs gestes articulatoires, de produire de la parole audible, même quand les retours sensoriels sont appauvris (Kelso & Tuller, 1983). La production à l'âge adulte est très probablement principalement guidée par cette composante prédictive en l'absence de perturbation. Néanmoins, les retours sensoriels sont nécessaires au maintien des modèles internes (Matthies et al., 1996). Ces représentations ne sont en effet pas immuables, elles sont plastiques et dynamiques (Shiller et al., 2009), et nécessitent une réactualisation régulière afin de permettre d'effectuer des prédictions efficaces.

1.5.1.1. Conception et implémentation

Il existe diverses façons de concevoir le stockage de ces représentations sensori-motrices dans le Système Nerveux Central : des modèles internes généralisés, implémentés tels des réseaux de neurones artificiels par Kawato et al. (1987) et Jordan & Rumelhart (1992), qui permettraient théoriquement la généralisation des connaissances tirées de l'expérience à des situations nouvelles via des mécanismes d'interpolation et d'extrapolation (un réseau de neurones artificiels donne en effet toujours une sortie, quelle que soit l'entrée donnée) ; aux modèles fondés sur des modèles de mémoire épisodique (Hintzman, 1984) et la théorie de l'exemplarité (Pierrehumbert, 2001) ; en passant par la possibilité d'une multitude de modèles internes locaux (Wolpert & Kawato, 1998).

Concrètement, aujourd'hui, malgré le développement important des neurosciences et des sciences computationnelles, quasiment comme au premier jour des sciences cognitives, le seul constat qu'il est possible de faire sans entrer dans le détail des débats est le suivant : il existe, quelque part dans le cerveau, sous une forme qu'il reste à déterminer, des mécanismes effectuant le stockage des informations relatives aux relations entre commandes motrices et retours acoustiques et somatosensoriels. Il existe un certain nombre d'hypothèses concernant la localisation d'un possible module dédié à cette fonction, ou d'un réseau neuronal impliqué dans cette fonction. Des propositions ont été formulées quant au fonctionnement de ce module ou de ce réseau, impliquant éventuellement des systèmes multiples (stochastiques ou non) permettant le stockage de ces informations. Gardons en tête que le mystère des représentations neuronales reste à ce jour entier et que nombre des suggestions qui vont suivre nécessitent confirmation.

La présence de cellules de Purkinje, localisées dans le cervelet, qui communiquent avec le cortex moteur, notamment concernant la vitesse et l'accélération des mouvements, a permis de proposer que le cervelet serait un bon candidat à l'hébergement des modèles internes (Gomi et al., 1998 ; Wolpert, Miall & Kawato, 1998). Polit & Bizzi (1978) ; Bizzi et al. (1971) ; (1976) ; (1991) suggèrent également que le cervelet joue un rôle important dans la planification des mouvements en général et en particulier en ce qui concerne la

composante prédictive et les trajectoires apprises (contrôle *feedforward*, modèle inverse). De plus, l'activité du cervelet lors de tâches motrices (Imamizu et al., 2000 ; Kawato et al., 2003), et en particulier lors de tâches de production de parole (Murphy et al., 1997 ; Ackermann et al., 1998 ; Riecker et al., 2000 ; Grabski, 2012 ; Acher, 2014), confirme son implication dans les tâches relevant du contrôle moteur. Gardons donc en tête qu'au vu de l'état actuel des connaissances, le cervelet semble être un bon candidat à l'hébergement des modèles internes.

1.5.2. Les paradigmes expérimentaux qui permettent l'étude des relations sensori-motrices de la parole et de leur développement

L'étude du comportement observable permet de formuler des hypothèses sur le mode de fonctionnement de ce qui a produit ce comportement. Dans notre cas, les mesures des indices dans le signal acoustique et les mesures des mouvements des articulateurs renseignent sur le mode de fonctionnement des représentations neuronales associées aux tâches de parole. De façon plus générale, de nombreuses études se sont attachées à qualifier mais aussi à quantifier la maturité du contrôle moteur (pour des tâches de parole mais pas seulement, comme on va le voir par la suite), via la mesure du comportement, qu'il s'agisse de tâches de pointage manuel, de tâches de parole perturbée ou non, afin de mesurer la stabilité et l'adaptabilité des gestes, et donc, d'en inférer des hypothèses sur le fonctionnement ou la maturation du contrôle moteur sous-jacent à ces tâches.

En ce qui concerne le développement du contrôle moteur pour la production de la parole, il existe, à notre connaissance, au moins trois grands paradigmes expérimentaux qui permettent de mesurer, plus ou moins indirectement, la maturité des représentations sensori-motrices associées aux tâches de parole. Le premier de ces paradigmes est l'étude expérimentale de la production de parole en conditions normales. En choisissant judicieusement les indices mesurés dans de la parole naturelle, on peut en tirer de nombreuses informations sur l'organisation de la parole. Pour le dire simplement, c'est le cadre le plus simple par lequel l'étude du comportement peut nous renseigner sur la « boîte noire » qui a amené à ce comportement. Par exemple, la mesure de la stabilité du comportement, lors de la production d'une même tâche répétée plusieurs fois, permet de

renseigner sur la stabilité du contrôle moteur pour la réalisation de cette tâche. Le second paradigme expérimental est celui regroupant les tâches de perturbation de la production de parole. Nous faisons entrer dans ce cadre aussi bien les tâches de perturbation articulatoire telles que celles qui impliquent l'usage du *bite-block*[2] et du *lip-tube*[3] et les tâches de perturbation ou de suppression du retour auditif ou somatosensoriel. Bien que fondamentalement différentes et apportant des éléments de réflexion sur des sujets bien distincts (respectivement la capacité d'adaptation d'un locuteur dans un contexte donné, en faisant appel à ses capacités à exploiter les degrés de liberté de son système de production afin d'atteindre un but malgré les contraintes appliquées à ses mouvements ; et la capacité d'un locuteur, à adapter son comportement en fonction de retours auditifs perturbés, étudiant donc entre autres le rôle du retour auditif dans les tâches de production), elles permettent l'étude des comportements acquis, ainsi que la capacité à trouver des solutions lorsque le système ne fonctionne pas en conditions normales, et sur la façon dont on procède afin d'arriver à ces solutions, qu'il s'agisse d'équivalence motrice ou d'adaptation à un retour perturbé. Le troisième paradigme expérimental qui permet d'étudier les représentations neuronales associées à la production de parole est celui de la coarticulation anticipatoire, car elle repose sur la prédiction, le *planning*. Les tâches de coarticulation peuvent, dans une certaine mesure, être rapprochées des tâches de perturbation. En effet, elles font appel, comme les tâches de perturbation, à l'expérience passée du locuteur et à l'exploitation de ses représentations sensori-motrices de la parole, afin, dans ce cas précis, de minimiser l'effort articulatoire et non pas d'atteindre le but acoustique ou somatosensoriel désiré. La différence fondamentale entre ces deux paradigmes étant que dans le cas de la perturbation, on oblige le locuteur à utiliser des stratégies d'équivalence motrice alors que dans le cas de la coarticulation, c'est le locuteur qui utilise ses degrés de liberté en excès afin de minimiser l'effort articulatoire, sans contrainte exercée sur son système périphérique.

[2] Le bite-block est un dispositif visant à entraver les mouvements mandibulaires. Il permet par exemple de perturber la production de voyelles fermées (comme /i/) en obligeant la mandibule du locuteur à être en position ouverte.

[3] Le lip-tube est un dispositif visant à entraver le protrusion des lèvres. Il permet par exemple de perturber la production de voyelles arrondies (comme /u/) en imposant une aire aux lèvres importante.

1.5.3. Le développement du contrôle moteur : le cadre bracchio-manuel

En dehors du domaine de la parole, il existe un intérêt général à comprendre le développement du contrôle moteur. Les travaux effectués dans ce sens apportent des informations importantes concernant le développement du contrôle des mouvements volontaires, et il serait inenvisageable d'ignorer ces connaissances, car la production de la parole est avant tout une question de contrôle moteur. Comme pour toute faculté complexe, c'est-à-dire mettant en jeu de multiples composantes, l'étude de sa mise en place, étape par étape, permet d'en disséquer le fonctionnement afin de mieux comprendre son organisation chez l'adulte.

1.5.3.1. Tâches de pointage et de préhension

Dans le but de comprendre comment se mettent en place et se développent les représentations sensori-motrices pour la production de la parole, il est important de comprendre, dans un premier temps, comment se mettent en place et se construisent ces représentations dans un cadre général. Le cadre le plus étudié en contrôle moteur, car comportant le système périphérique le moins complexe et le plus facile d'observation, est celui du contrôle du bras. L'étude du développement du contrôle des mouvements du bras et de la main apportent des éléments de réponse quant aux questions soulevées dans notre introduction, à savoir comment mûrit la capacité à planifier des gestes, comment se développe l'exécution de ces gestes, et enfin comment évolue la capacité à compenser pour des perturbations, donc à corriger ses prédictions en tenant compte des retours sensoriels. De plus, l'angle d'attaque de ce travail de thèse étant la maturation de la production de la parole du point de vue du contrôle moteur, il est important d'apporter les informations nécessaires à la compréhension de notre cadre et de nos hypothèses de travail. Nous présentons, dans la section suivante, quelques études développementales visant à comprendre la maturation du contrôle moteur du bras et de la main, via les paradigmes du pointage et de la préhension (*grasping*).

Les études portant sur les capacités des enfants à atteindre une cible virtuelle via le contrôle d'une souris d'ordinateur (i.e. Lambert & Bard, 2005 ; Hourcade et al., 2004) montrent que l'imprécision, la lenteur et la variabilité diminuent avec l'âge. L'étude de Hourcade et al. (2004), qui étudie le développement des capacités à atteindre des cibles

virtuelles, de différentes tailles et à différentes distances, montre que les enfants de 4 et 5 ans (n = 13 par groupe) sont plus lents que les adultes (n = 13), ont des stratégies plus variables, et montrent davantage de difficultés à atteindre la cible virtuelle lorsqu'ils sont à une position finale proche de celle-ci et qu'ils doivent effectuer des mouvements de correction afin d'atteindre la cible. L'étude de Lambert & Bard (2005) utilise un protocole expérimental similaire et montre que les enfants de 6, 8 et 10 ans (n = 10 par groupe) sont plus lents que les adultes (n = 10), leur trajectoire est en moyenne plus longue, ils effectuent davantage de changements de direction, et leur variabilité est plus importante. Les auteurs précisent que les patrons des mouvements des enfants ne sont pas semblables à ceux des adultes, même à 10 ans : ils présentent davantage de variabilité en répétition et sont bien moins lisses (*smooth*) que ceux des adultes. Le processus de maturation motrice est donc lent et progressif.

Les études portant sur les capacités des enfants à contrôler le niveau de force délivré lors d'une tâche de préhension (i.e. Forssberg et al., 1991) montrent le même type de patron développemental. Dans ce type de tâche, l'objectif est de maintenir une pression suffisante entre son pouce et son index afin de tenir un objet. L'étude de Forssberg et al. (1991), comportant des participants allant de 8 mois à l'âge adulte, montre que la variabilité observée dans les patrons de force diminue avec l'âge, et apporte une information supplémentaire : il semble y avoir une étape développementale importante entre 4 et 6 ans, étape au cours de laquelle les patrons très variables des enfants de 4 ans viennent à ressembler davantage à ceux de l'adulte autour de 6 ans.

Dans leur ensemble, ces études concernant le contrôle moteur du bras et de la main montrent une **maturation lente et progressive**, dont l'arrêt de l'apprentissage semble être au début de l'adolescence (12 ans) pour certains, ou à la fin de l'adolescence (18-20 ans) pour d'autres. D'autre part, un **changement important apparaît entre 4 et 6 ans** : la variabilité liée à la répétition d'une même tâche diminue de façon drastique, le contrôle moteur s'affine donc de façon importante durant cette période. Enfin, les difficultés des jeunes enfants à atteindre une cible lorsqu'ils sont à une position finale proche de celle-ci et qu'ils doivent effectuer des mouvements de correction pour l'atteindre, montrent que les enfants planifient un geste globalement dans la bonne direction, mais imprécis (contrôle *feedforward* imprécis) et qu'**en ce qui concerne les prédictions fines et locales,**

les enfants exploitent essentiellement un **contrôle par feedback,** impliquant une **stratégie d'essai-erreur**, ce qui laisse penser que les modèles internes des jeunes enfants ne sont pas aussi efficaces que ceux des adultes et des enfants plus âgés.

1.5.3.2. Tâches de pointage et de préhension perturbées

L'étude de ce même type de tâche (pointage et préhension), en développement, avec l'ajout d'une condition perturbée, peut apporter de précieuses informations quant à la maturité des représentations sensori-motrices. En effet, comme nous l'avons vu au cours de la section 1.3., l'apprentissage des modèles internes permet d'effectuer des prédictions sur les conséquences des commandes motrices sur le système moteur périphérique. Or dans le cas de perturbations, c'est exactement ce que l'on demande au Système Nerveux Central : utiliser la composante prédictive afin de trouver une stratégie qui permet de compenser l'effet de la perturbation. Ces tâches perturbées permettent donc d'étudier, moins indirectement que dans le cas de tâches non perturbées, les capacités prédictives des représentations sensori-motrices apprises.

Plusieurs études développementales se sont attachées à étudier les capacités des enfants à compenser lors d'une tâche de pointage en condition de perturbation, en appliquant une force sur le bras du participant dans une direction donnée, venant perturber le résultat du mouvement et venant contredire les prédictions effectuées par le participant sur les conséquences de son geste. L'étude de l'adaptation à cette perturbation renseigne donc sur les capacités des enfants à prendre en compte l'effet de cette perturbation dans leur planification motrice et à intégrer cette perturbation dans leurs nouvelles prédictions, afin d'obtenir le comportement de sortie désiré.

Les études concernant le développement de la capacité à compenser lors d'une tâche de pointage perturbée par un champ de force appliqué au bras (Jansen-Osmann et al., 2002 ; Takahashi et al., 2003 ; Konczak et al., 2003) montrent toutes que les enfants, y compris aux stades les plus précoces qui ont été étudiés (4 ans), sont capables de compenser pour ces perturbations. L'adaptation à ce champ de force réalisée par les enfants de 4 ans est cependant moins efficace que celle qui est réalisée par les enfants plus âgés. Ces résultats montrent que les enfants sont capables, dès un jeune âge, d'effectuer des prédictions efficaces, indiquant que les modèles internes appris sont capables, au moins

pour ces tâches simples, de prédire l'action réalisée afin de redéfinir la cible et d'atteindre cette cible dans les conditions imposées (cf. figure 1.5.1. ci-dessous).

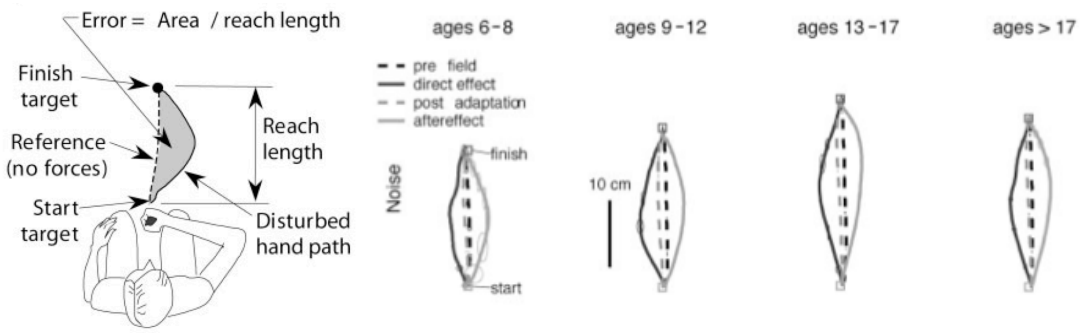


Figure 1.5.1. Résultats de l'étude de Takahashi et al. (2003) concernant les trajectoires avant, pendant et après perturbation à l'aide d'un champ de force, pour les 4 groupes d'âge étudiés.

L'étude de Vasudevan et al. (2011), qui concerne la locomotion, montre que les enfants, à partir de 11 ans, sont capables de compenser spatialement et temporellement la perturbation du rythme de marche, mais montre également que les enfants de moins de 6 ans ne sont pas capables de compenser spatialement cette perturbation. Ces résultats mettent en évidence la maturation du cervelet entre 6 et 11 ans.

L'étude de Bard, Hay & Fleury (1990) concerne des tâches de pointage de direction et d'amplitude avec et sans retour visuel, chez des enfants de 6, 8 et 10 ans (n=8 par groupe). Leurs résultats montrent que la précision est meilleure en présence du retour visuel qu'en l'absence de ce retour, pour tous les groupes d'âge. Notons également que la précision est équivalente dans les différents groupes d'âge. Ces résultats ne permettent pas de mettre en évidence une amélioration significative du contrôle moteur entre ces stades développementaux. Leurs résultats montrent également que, sans retour visuel, pour les tâches de direction, il n'y a pas de différence significative entre les groupes d'âge, mais que pour les tâches d'amplitude, les enfants de 6 ans sont moins efficaces que les enfants plus âgés. Ces résultats montrent que les enfants, dès l'âge de 6 ans, sont capables d'effectuer une tâche de pointage, même sans retour visuel, et ce avec une précision acceptable. En revanche, lorsqu'il s'agit de tâches plus fines ou plus locales, comme la tâche d'amplitude, nécessitant un contrôle par correction visuelle, les jeunes enfants sont moins performants que les enfants plus âgés.

Ces derniers résultats nous paraissent importants car ils permettent de mettre en évidence respectivement, que les modèles internes des jeunes enfants sont mis en place, car ils permettent d'effectuer la tâche sans retour visuel, mais que ces mêmes modèles, pour effectuer des prédictions locales, même grâce à l'aide du retour visuel, ne sont pas aussi mûrs que ceux de l'adulte puisqu'ils ne permettent pas de corriger localement de façon précise.

Dans leur ensemble, ces résultats montrent que les enfants sont capables de compenser pour des tâches de perturbation simples (comme les champs de force), et sont capables d'effectuer des tâches de pointage sans retour visuel. En revanche, pour des tâches plus fines et locales (comme la tâche de pointage d'amplitude ou la redéfinition des cibles lors d'une marche au rythme perturbé), les jeunes enfants (6 ans et moins) sont moins performants que les enfants plus âgés, et la précision augmente avec l'âge. Ces résultats semblent indiquer que les **représentations sensori-motrices** (modèles internes) des **jeunes enfants** sont suffisamment **efficaces,** dans une certaine mesure, pour effectuer des **prédictions globales,** mais sont relativement **peu efficaces** lorsqu'il s'agit d'effectuer des **prédictions fines et locales**.

1.5.4. Le développement des représentations sensori-motrices associées à la production de parole

1.5.4.1. Tâches de production de parole

Nous revenons rapidement dans cette section sur les études développementales de production de parole présentées aux sections 1.2 (développement tardif) et 1.4 (coarticulation) afin d'évaluer leur apport concernant la maturation des représentations sensori-motrices. Les résultats de ces expériences développementales mettent en avant un développement non uniforme, très progressif et relativement lent (du point de vue du contrôle moteur), et différent pour chaque articulateur. Le contrôle de la mandibule et des lèvres semble être acquis plus précocement que le contrôle de la langue (Green et al., 2002 ; Canault, 2007), et il existe un désaccord sur l'âge auquel ce contrôle est semblable à l'adulte. Le contrôle de la langue, plus complexe à maîtriser comme à étudier, semble être relativement lent (Cheng et al., 2007) et le manque de données expérimentales (et

longitudinales) rend difficile l'établissement d'un profil ontogénétique complet. Quoi qu'il en soit, même avec un contrôle imprécis, l'enfant ressent les conséquences sensorielles de ses actes moteurs, et ce dès les premières productions, et exploite ces sensations dans le but de construire des représentations phonétiques. Même si seules les lèvres et la mandibule sont contrôlées au début du babillage (contrôle auquel il faut ajouter celui de la phonation et du contrôle de la coordination oro-pharyngée), les retours sensoriels produits sont d'une grande importance. Ils vont probablement, dès les premières vocalisations, construire le jeu de relations sensori-motrices et guider l'enfant vers l'exploration de différents gestes, produisant des conséquences perceptives et somatosensorielles différentes. Lorsque le jeune enfant devient capable, petit à petit, de coordonner le contrôle de ses différents articulateurs volontairement dans le but de produire des sons de parole, on peut imaginer que les retours sensoriels produits sont à ce moment de très grande importance, puisque l'enfant prête alors une attention toute particulière aux conséquences de ses actes moteurs, souvent renforcées par un adulte imitant (et guidant) ou répétant les productions de l'enfant (Howard & Messum, 2011).

L'étude de la variabilité en répétition, par exemple, donne une idée de l'état de ces relations sensori-motrices. Les productions des très jeunes enfants sont tout d'abord extrêmement variables, puis s'organisent selon les catégories phonétiques de la langue maternelle ou des langues maternelles parlées, tout d'abord avec beaucoup de superposition entre catégories, puis de moins en moins, allant vers une production catégorielle bien établie. Dès les premières productions vocaliques contrastées (Kuhl & Meltzoff, 1996) en imitation, les conséquences perceptives des actes moteurs sont forcément intégrées, même dans leur forme la plus rudimentaire. Puis l'évolution de l'inventaire phonétique, devenant de plus en plus complet pour être établi, selon les langues maternelles, quelque part entre 2 et 3 ans, montre que les enfants, même à un jeune âge, ont bien intégré les différents produits sensoriels et les ont bien associés à des actes moteurs, déjà relativement précis puisque suffisamment contrastés. Enfin, l'étude des patrons de coarticulation, et le fait que ces patrons de coarticulation, pour le cas des syllabes CV plosive-voyelle, soient établis dès 18 mois, soit l'apparition des « premiers mots », montrent que les jeunes enfants disposent, déjà à cet âge, d'une certaine idée de la variabilité autorisée autour d'un son de parole donné, ce qui suppose qu'ils disposent

déjà d'unités d'ordre segmental, c'est-à-dire que l'organisation de leur production n'est pas uniquement centrée autour de la syllabe. L'étude de la coarticulation et de son développement renseigne donc en partie sur la maturité des représentations sensori-motrices associées à la production de parole. Mais un champ expérimental spécifique vient éclairer davantage nos connaissances concernant la maturité de ces représentations : celui des tâches de production de parole perturbées.

1.5.4.2. Tâches de production de parole perturbées

Dans le domaine de la production de parole aussi, les tâches de production introduisant une perturbation ont apporté leur contribution à la compréhension des rôles des processus *feedback* et *feedforward* dans le processus de production de parole. Certaines études s'étant focalisées sur le rôle de ces processus au cours du développement, nous nous attacherons davantage à décrire ces études développementales, qui revêtent un intérêt tout particulier pour nous.

1.5.4.2.1. Tâches de production de parole perturbées chez l'adulte

1.5.4.2.1.1. Les expériences de perturbation de l'articulation

Dès le début des années 1980, des chercheurs se sont penchés sur les questions de l'équivalence motrice : comment les locuteurs parviennent-ils, avec des stratégies articulatoires différentes, à obtenir un produit acoustique similaire ? Quelles sont ces différentes stratégies articulatoires qui mènent au même produit acoustique ? Ces questions soulèvent également celle de la nature du but de la production : acoustique ou articulatoire ? Ce paradigme permet également d'étudier, même s'il n'est pas toujours le premier centre d'intérêt des auteurs, le rôle du retour proprioceptif dans la tâche de production, car dans le cas de ces perturbations articulatoires, le retour proprioceptif est mis à mal.

Chez l'adulte, Gay, Lindblom & Lubker (1981) ont montré (de façon pionnière) que l'adulte est capable de compenser pour une perturbation de type *bite-block* (ou blocage de la mandibule). En étudiant les différentes stratégies articulatoires (via la radiographie) adoptées par 5 locuteurs américains anglophones pour produire les voyelles /i u o/ avec un *bite-block* de 22,5 mm (mâchoire ouverte) et la voyelle /a/ avec un *bite-block* de 2,5 mm

(mâchoire quasi fermée), ils ont montré que tous les locuteurs compensaient immédiatement et complètement pour les voyelles /a/ et /o/ et moins nettement pour les voyelles /i/ et /u/. Pour ces dernières, la compensation n'est pas totale, mais les données radiographiques montrent que le lieu de constriction a été préservé. Cette étude montre également que, même avec un retour proprioceptif différent, les locuteurs sont capables de produire les bonnes voyelles, allant dans le sens de l'existence d'un modèle inverse.

Fowler & Turvey (1980) ont reproduit l'expérience de Gay, Lindblom & Lubker, publiée en 1981, en mesurant les temps de réponse. L'idée étant que la compensation immédiate mise en évidence par Gay et collègues pourrait être due à une reprogrammation motrice opérée avant la production. En comparant deux groupes de sujets, le premier auquel on n'a donné aucune indication spécifique sur le temps, et le second auquel il était explicitement demandé de produire le plus rapidement possible, il est montré que la compensation est immédiate et n'est pas liée au temps de préparation motrice.

Enfin, l'expérience de Savariaux et al. (1995), introduit un nouvel outil de perturbation de l'articulation, le *lip-tube*. L'insertion d'un tube labial, empêchant la constriction bilabiale, vient perturber la production des voyelles arrondies, produites, en conditions normales, grâce à une constriction bilabiale. L'étude de Savariaux et al. (1995) s'intéresse aux productions, par 11 locuteurs adultes, de la voyelle /u/ avec et sans *lip-tube* de 2,5 cm de diamètre. Des données acoustiques et radiographiques ont été recueillies. Sept des 11 participants montrent une compensation partielle dès le premier essai, mais aucun ne montre une compensation totale. Durant les 19 essais suivants, les 7 participants montrant une compensation partielle tendent à améliorer leurs produits acoustiques, mais un seul participant parvient à une compensation totale (sur le plan acoustique), et 4 participants ne montrent aucune compensation. Notons tout de même que la tâche est bien plus compliquée que celle du *bite-block*, ce qui explique que tous les sujets ne parviennent pas à compenser pour cette perturbation. En effet, lors des tâches de *bite-block*, une compensation quasi totale (acoustiquement) est presque toujours possible grâce à la compensation effectuée par la langue (qui s'élève plus ou moins selon la position de la mandibule) ; dans les tâches de *lip-tube*, aucun articulateur ne peut compenser totalement l'absence de constriction labiale, une reconfiguration complète du conduit vocal est donc nécessaire. La tâche est donc d'une nature très différente.

Ces résultats vont en faveur de l'existence d'un modèle inverse. En effet, les locuteurs adultes semblent capables, dans le cas des expériences de type *bite-block*, de compenser ces perturbations de l'articulation complètement et surtout immédiatement. Pour l'expérience de type *lip-tube*, une compensation partielle est effectuée dès le premier essai par sept des onze locuteurs. Cette adaptation immédiate ne pouvant résulter d'une stratégie d'essai-erreur, cela suggère que les locuteurs se reposent sur leurs représentations sensori-motrices, afin d'adopter une configuration différente de leur conduit vocal, dans ces conditions perturbées, qui mènerait à un produit acoustique proche du but désiré.

1.5.4.2.1.2. Les expériences de perturbation des retours sensoriels

L'expérience de Kelso & Tuller (1983) teste directement la question du rôle des retours sensoriels dans une expérience de perturbation, avec plusieurs conditions comme (1) condition normale (2) ajout d'un bruit blanc à 90 dB (3) anesthésie locale des muqueuses de la surface de la langue et de la cavité orale et (4) anesthésie bilatérale de l'articulation temporo-mandibulaire et (5) ajout d'un *bite-block*. Leurs résultats montrent que, sous toutes les conditions, des voyelles correctes d'un point de vue acoustique sont produites, et de plus que la compensation est immédiate.

Cette expérience souligne le rôle du processus *feedforward* (ou processus prédictif) dans la tâche de production de parole. Elle ne permet cependant pas d'affirmer de façon catégorique que les retours sensoriels ne jouent plus aucun rôle dans le processus de production de parole à l'âge adulte, une fois l'apprentissage achevé. Il est en effet très difficile de vérifier que les participants ne bénéficiaient d'aucun retour sensoriel, car même avec un bruit blanc à 90 dB, le retour auditif existe encore et les vibrations transmises par l'articulation temporo-mandibulaire au système auditif ne sont pas négligeables. De plus, il est impossible de vérifier que l'anesthésie locale a bien inhibé tous les retours sensoriels, et de plus la capacité de proprioception demeure, quoi qu'il advienne des retours sensoriels. Par ailleurs, les études portant sur des personnes devenues sourdes à l'âge adulte montrent de façon évidente que le retour auditif joue un rôle important dans le maintien du processus de production de parole, même à l'âge

adulte (Perkell et al., 2000). Cela n'enlève rien aux résultats de cette étude, qui a permis de mettre en évidence le rôle du processus *feedforward* dans la production de parole.

Un autre type d'étude permet de renseigner sur le rôle des processus *feedback* et *feedforward* dans les tâches de production de parole, cette fois sans tenter d'entraver les retours sensoriels, mais en les modifiant. Houde & Jordan (1998) ont manipulé les retours auditifs (les valeurs des deux premiers formants) lors de la production de voyelles et ont montré que les adultes compensent, via l'élévation ou l'abaissement de la langue, dans la direction opposée de la perturbation, afin de compenser cette dernière et d'atteindre le but (acoustique) désiré. Ces résultats soulignent le rôle du retour auditif dans le processus de production de parole (et ne sont pas pour autant contradictoires avec les résultats de Kelso & Tuller, 1983) même à l'âge adulte, et apporte un argument en faveur des théories auditives de la production de parole (Stevens & Klatt, 1974 ; Perkell, et al. 1997 ; Perkell, 2012, cf. section 1.3), indiquant que les cibles, au moins en ce qui concerne les voyelles, sont définies dans l'espace acoustique.

L'étude pionnière de Houde & Jordan (1998) a depuis inspiré de nombreuses expérimentations, créant un paradigme de recherche à part entière, celui de l'apprentissage ou de l'adaptation sensori-motrice (*sensorimotor adaptation* ou *sensorimotor learning*), dont on citera par exemple les études de Max, Wallace & Vincent (2003) ; Jones & Munhall (2005) ; Purcell & Munhall (2006) ; Villacorta, Perkell & Guenther (2007) ; Mitsuya et al. (2013) ou encore Bourguignon, Baum & Shiller (2008). L'étude de Shiller et al. (2009) a apporté un élément supplémentaire, en montrant que suite à une adaptation sensori-motrice, la perception catégorielle s'en trouve modifiée. Leur expérience montre donc que la production influence la perception, en jouant un rôle dans la définition et le maintien des frontières catégorielles des sons de parole.

De façon surprenante, ce paradigme qui paraît être en mesure d'apporter de précieuses informations concernant le rôle du retour auditif au cours du développement n'a encore donné que peu de fruits dans ce domaine, puisqu'on ne recense à ce jour que deux études développementales publiées utilisant ce paradigme, celle de Shiller, Gracco & Rvachew (2010) et celle de MacDonald et al. (2012).

Nous citerons également, à titre informatif puisque nous n'en discuterons pas outre mesure dans ce travail de thèse, un autre type de perturbation des retours sensoriels qui a apporté un éclairage important sur la question des représentations des sons de parole. Les études d'Ostry et collègues (eg. Tremblay, Shiller & Ostry 2003 ; Nasir & Ostry, 2006 ; 2008 ; Ito, Tiede & Ostry, 2009 ; Ostry et al., 2010), portant sur la manipulation du retour somatosensoriel, en perturbant légèrement la position de la mandibule, ont montré que les adultes compensent dans la direction opposée de la perturbation, indiquant que le retour somatosensoriel joue un rôle important dans leurs représentations des sons de parole (la perturbation, de quelques millimètres, appliquée à la mandibule, n'engendre pas de conséquences acoustiques perceptibles, de plus, les locuteurs ne compensent que pour la tâche de production de parole). Pour le dire simplement, les locuteurs ont une idée bien arrêtée sur la façon dont doit sonner un son (expériences de perturbation du retour auditif) mais ont aussi une idée bien arrêtée de ce qu'ils doivent ressentir lors de la production de ce son. Ce type d'étude n'ayant jamais été adapté aux études développementales, nous ne disposons d'aucune information concernant le rôle du retour somatosensoriel au cours du développement.

Ces résultats, dans leur ensemble, montrent que la composante prédictive (*feedforward*) joue un rôle central dans le processus de production de parole, mais que les retours sensoriels (auditifs et somatosensoriels) jouent également un rôle important dans la définition, mais aussi dans la redéfinition et le maintien des cibles phonémiques.

1.5.4.2.2. Tâches de production de parole perturbées chez l'enfant

Comme nous venons de le voir, les expériences de perturbation des retours sensoriels et de l'articulation, chez l'adulte, viennent conforter l'existence d'une composante prédictive (modèle inverse) dans le processus de production de parole. Ce paradigme de perturbation, appliqué aux études développementales, a permis de comprendre davantage la maturation des représentations sensori-motrices de la parole, sur lesquelles reposent cette composante prédictive. Comment les enfants construisent-ils ces représentations sensori-motrices leur permettant de prédire les conséquences acoustiques de leurs manœuvres articulatoires et d'utiliser ces connaissances afin

d'exploiter les degrés de liberté de leur système moteur et de compenser pour une perturbation donnée ?

1.5.4.2.2.1. Tâches de perturbation de l'articulation

La première expérience développementale de perturbation de l'articulation, via le *bite-block*, est proposée par Gibson & McPhearson (1980). Cette étude s'intéresse aux productions de 5 enfants suédois de 6 et 7 ans, des voyelles /i a u o/ dans deux conditions, normale et avec l'ajout du *bite-block*. Les productions acoustiques des enfants sont évaluées par un test perceptif de catégorisation (effectué par des adultes). Ces résultats montrent que seule la voyelle /a/ est immédiatement et complètement compensée. Les autres voyelles sont perçues comme plus ouvertes, ce qui indique que les enfants ont compensé la perturbation dans la bonne direction, mais pas de manière suffisante. Les auteurs en concluent que les enfants de 6 et 7 ans n'ont pas encore appris leur système vocalique suffisamment précisément pour mettre en place des stratégies de réorganisation neuromusculaire.

Une seconde expérience développementale utilisant le paradigme du *bite-block* est proposée par Oller & MacNeilage (1983). Les auteurs étudient les productions de deux enfants, de 4 et 8 ans, lors de la production des voyelles /i/ et /æ/ de l'anglais américain, avec bite-block pour le /i/ et avec les dents serrées pour le /æ/. Les productions sont évaluées par un test perceptif de catégorisation (effectué par des adultes). Leurs résultats montrent que les deux enfants ont compensé la perturbation, mais que celle-ci est compensée de façon plus complète par l'enfant de 8 ans. Ces résultats vont dans le sens d'un modèle interne qui s'affine entre 4 et 8 ans.

L'expérience de Smith & McLean-Muse (1987) s'intéresse à la comparaison de la compensation d'une perturbation de type *bite-block* (7 mm) entre un groupe d'enfants de 4-5 ans et des adultes. Leurs résultats montrent que le contrôle des articulateurs visibles (lèvres et mâchoire) est aussi précis chez les enfants de 4-5 ans que chez les adultes.

L'expérience de Baum & Katz (1988) propose de s'intéresser aux productions de 5 enfants de 4 à 5 ans et de 5 enfants de 7 à 8 ans, lors de la production des voyelles /i a u/ avec et sans *bite-block* de 12 mm. Leurs résultats montrent que la compensation est totale

(aucune différence significative entre les valeurs de formants lors des deux conditions) et immédiate (aucune différence significative dès les premières impulsions glottiques) pour tous les participants.

L'expérience d'Edwards (1992), qui concerne les productions de 10 enfants âgés de 3 à 5 ans (dont 6 enfants au développement typique et 4 enfants ayant des troubles de l'élocution) lors de la production des voyelles /i ɪ æ/ en contexte /hVt/ avec et sans *bite-block* (10 mm), montre que la compensation n'est pas effectuée de façon complète par les 6 enfants au développement typique.

Enfin, les expériences de Ménard et al. (2008) et Ménard, Perrier & Aubin (2016), reprennent l'outil de perturbation de l'articulation (*lip-tube*) introduit par Savariaux et al. (1995), en s'intéressant aux capacités des jeunes enfants (4 ans) à compenser pour une telle perturbation. L'étude de Ménard et al. (2008) s'intéresse aux productions de 12 enfants de 4 ans lors de la production de la voyelle /u/ avec et sans *lip-tube* de 15 mm. Leurs résultats montrent que les enfants de 4 ans ne sont, dans l'ensemble, pas capables de compenser totalement la perturbation, mais qu'il existe une forte variabilité interindividuelle en ce qui concerne cette aptitude. L'expérience de Ménard, Perrier & Aubin (2016), étend cette expérience en comparant les productions d'enfants de 4 ans et d'adultes et ajoute le recueil de données articulatoires (échographie linguale). Leurs résultats montrent que les adultes améliorent de façon constante leurs productions, essai après essai, alors que les enfants ne semblent pas montrer de stratégie cohérente et explorent plutôt aléatoirement différentes stratégies (cf. figure 1.5.2 ci-dessous).

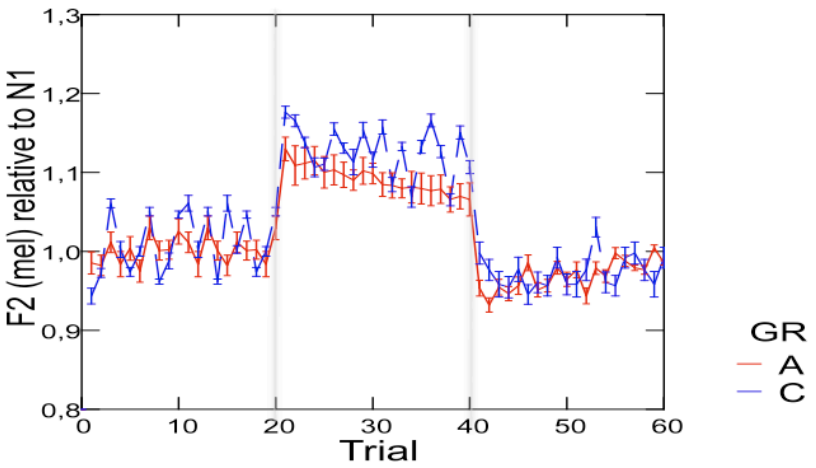


Figure 1.5.2. Résultats de l'étude de Ménard, Perrier & Aubin (2016) concernant le second formant (F2, ici en Mel). Les 20 premiers essais représentent une ligne de base ; les essais 21 à 40 constituent la condition perturbée. Les productions des adultes (en rouge), dans cette seconde phase, s'améliorent essai après essai alors que les productions des enfants (en bleu) ne montrent pas d'amélioration systématique.

Les résultats de Ménard et al. (2008) & (2016) sont, de notre point de vue, à mettre en perspective avec ceux de Bard et al. (1990), concernant une tâche de pointage perturbée, car ils mettent en avant le fait que les modèles internes des jeunes enfants sont peu efficaces pour effectuer des prédictions, en particulier lorsqu'elles sont fines et locales (et proches du but désiré).

Ces études, dans leur ensemble, montrent que les enfants sont capables, au moins partiellement, de compenser pour une perturbation, mais que les jeunes enfants (4 ans) compensent de façon moins complète que les enfants plus âgés. Ces résultats, quelque peu contradictoires (mais pas tant que cela si on considère la difficulté de la tâche, via les différents phonèmes cibles et les différentes tailles de *bite-block*) montrent, dans l'ensemble, que les enfants de 4 ans disposent de mécanismes prédictifs et sont capables de tirer partie des degrés de liberté de leur système moteur, mais ne sont pas capables d'exploiter au maximum ces connaissances pour compenser totalement les perturbations ou effectuer des corrections fines et locales.

1.5.4.2.2.2. Tâches de perturbation du retour auditif

Deux études ont, à ce jour et à notre connaissance, utilisé le paradigme de perturbation en temps réel du retour auditif, ou *sensorimotor adaptation*, apporté par Houde & Jordan (1998), pour étudier le rôle du retour auditif au cours du développement.

L'étude de Shiller, Gracco & Rvachew (2010) traite de la production des fricatives /s/ ~ /ʃ/ chez des enfants âgés de 9 à 11 ans (n = 11) et des adultes (n = 13). Leurs résultats montrent qu'il existe davantage de variabilité dans les productions de /s/ et de /ʃ/ des enfants par rapport aux adultes et que la distinction /s/ ~ /ʃ/ est moins nette chez les enfants : on observe de moins bons contrastes en production (davantage de recoupement entre les deux catégories phonétiques dans le domaine acoustique), ainsi qu'une perception catégorielle moins abrupte, puisque la frontière catégorielle est moins nettement définie (cf. figure 1.5.3. ci-dessous).

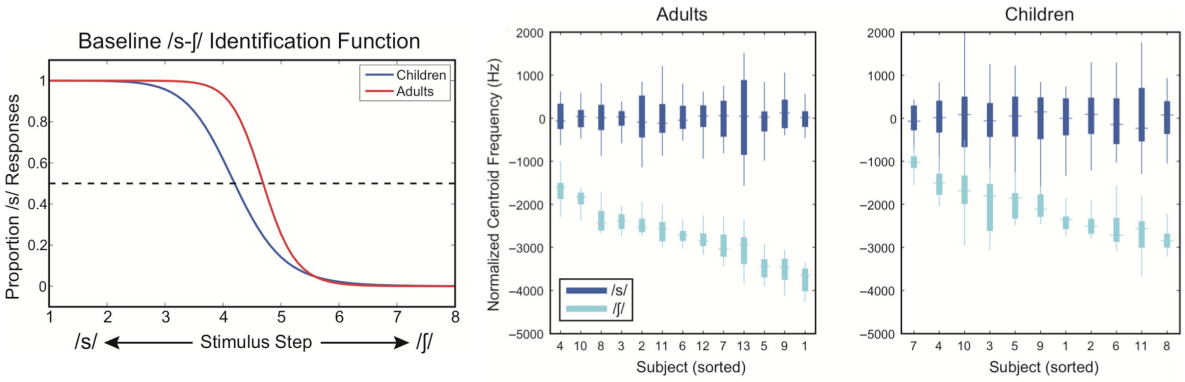


Figure 1.5.3. Illustration des résultats de perception et de production (sans perturbation) de l'étude de Shiller, Gracco & Rvachew (2010). Le panneau de gauche montre la courbe d'identification de /s/ ~ /ʃ/ sur un continuum artificiel, pour les enfants (bleu) et les adultes (rouge). On observe que la frontière catégorielle des adultes est plus abrupte que celle des enfants, plus clairement définie. Les panneaux de droite montrent les productions de /s/ et /ʃ/ par les adultes et les enfants. On observe davantage de variabilité dans les productions enfantines et davantage de recoupement entre les catégories par rapport aux productions adultes.

En modifiant, en temps réel, les centres de gravité (CoG) des bruits de friction, ils montrent que les enfants de 9 à 11 ans, comme les adultes, modifient leur production dans le sens opposé à la perturbation, afin de compenser cette dernière. Ils observent également une modification des frontières catégorielles suite à l'apprentissage sensori-

moteur, importante chez les participants adultes et moins importante chez les enfants (cf. figure 1.5.4. ci-dessous). Si les résultats concernant la variabilité des productions des fricatives par les enfants et le contraste /s/ ~ /ʃ/ n'en sont pas moins intéressants, nous retiendrons le résultat principal de cette étude : les enfants de 9 à 11 ans sont capables de compenser leur articulation pour une perturbation en temps réel du retour auditif.

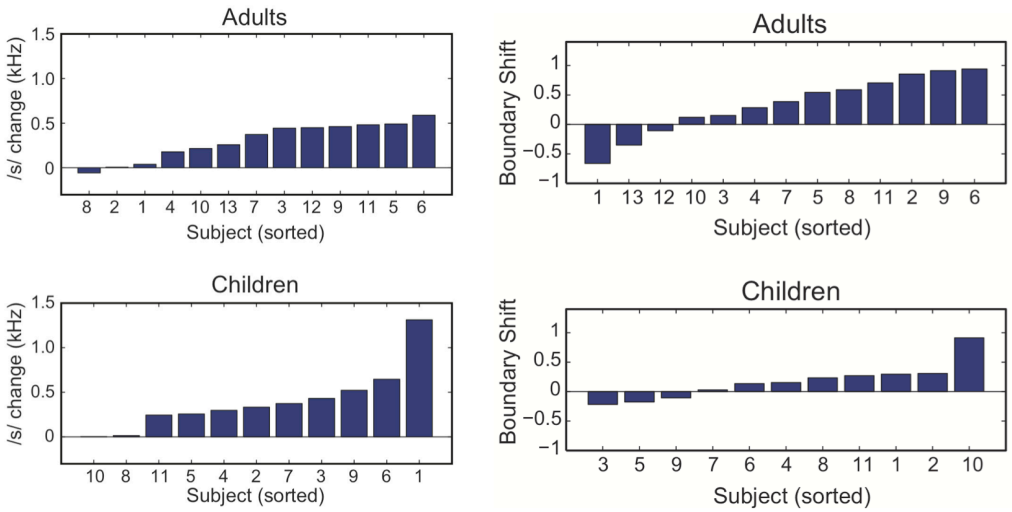


Figure 1.5.4. Illustration des résultats de Shiller, Gracco & Rvachew (2010) portant sur la compensation de la perturbation en production (panneau de gauche) et sur l'effet de cet apprentissage sur le déplacement de la frontière catégorielle du contraste /s/ ~ /ʃ/ (panneau de droite). Les enfants compensent effectivement la perturbation, mais l'impact sur le déplacement de la frontière catégorielle est moindre que chez les adultes.

L'étude de MacDonald et al. (2012) concerne la perturbation, en temps réel, des deux premiers formants lors de la production de voyelles chez 26 adultes, 26 enfants de 4 ans et 20 enfants de 2 ans. Cette étude montre que les adultes compensent dans la direction opposée de la perturbation, afin de compenser cette dernière, les enfants de 4 ans font de même, avec toute fois beaucoup plus de variabilité dans les données, et les enfants de 2 ans ne semblent pas compenser la perturbation (cf. figure 1.5.5. ci-dessous).

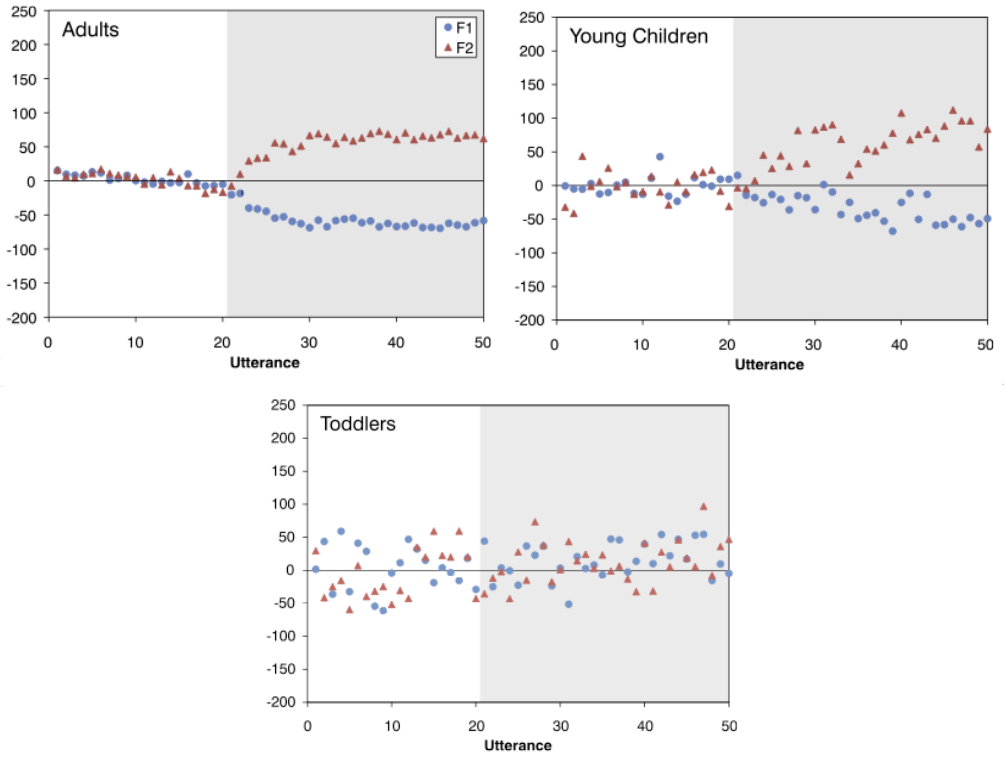


Figure 1.5.5. Illustration des résultats de l'étude de MacDonald et al. (2012) pour les groupes Adultes (*adults*), enfants de 4 ans (*young children*) et les enfants de 2 ans (*toddlers*). Les adultes compensent dans la direction opposée à la perturbation, les enfants de 4 ans font de même, avec davantage de variabilité, et les enfants de 2 ans ne semblent pas compenser la perturbation.

Il est possible de formuler de nombreuses critiques à l'égard de cette expérience : l'amplitude de la perturbation est la même pour tous les groupes d'âge (F1 augmenté de 200 Hz et F2 diminué de 250 Hz), or on sait combien les Espaces Vocaliques Maximaux des jeunes enfants, en Hz, sont plus larges que ceux des adultes ; la variabilité dans les données des enfants de 2 ans est bien trop importante (environ 100 Hz) pour garantir que la perturbation a un réel effet de changement de catégorie vocalique ; la variabilité anatomique interindividuelle, très importante chez les jeunes enfants, joue également un rôle à cet égard. Tous ces biais peuvent, en partie, expliquer pourquoi les enfants de 2 ans ne compensent pas cette perturbation, et il serait hâtif de conclure, comme semblent le faire les auteurs, que les enfants de 2 ans ne tiennent pas compte des retours auditifs ou qu'ils ne disposent pas de modèles internes leur permettant de compenser pour cette perturbation. Nous retiendrons, ce qui, de notre point de vue constitue le principal apport de cette étude et le moins discutable : les enfants de 4 ans sont capables de compenser pour ce type de perturbation.

1.5.5. En somme...

Les études concernant le développement du contrôle bracchio-manuel et celles concernant le développement du contrôle oro-facial apportent ensemble d'importantes informations concernant la maturation des représentations sensori-motrices au cours du développement ontogénétique. Les expériences de perturbation apportent des informations particulièrement intéressantes concernant ces représentations, car l'adaptation à ces expériences requiert la redéfinition des prédictions, et permet donc d'étudier la maturité de ces prédictions et la capacité des enfants à intégrer une contrainte physique extérieure (dans le cas des perturbations du retour auditif ou des champs de force par exemple) ou une modification de leur système périphérique (dans le cas des perturbations de l'articulation).

Dans leur ensemble, ces études montrent que les jeunes enfants disposent de capacités à prédire les conséquences sensorielles de leurs gestes articulatoires, de façon relativement précoce (ils en sont capables dès l'âge de 4 ans) mais que ces représentations sensori-motrices, permettant d'effectuer ces prédictions, ne sont pas aussi fines et complètes que celles des adultes, qui ont été construites par des millions d'exemplaires et d'essais. Ces résultats suggèrent que les modèles directs et inverses des jeunes enfants sont encore en phase de maturation.

En conclusion, nous savons que les représentations sensori-motrices associées à la production de parole sont mises en place dès les premières productions, mais que ces représentations semblent immatures chez les jeunes enfants, et s'affinent lentement, au cours des années et des millions d'exemplaires de sons de parole produits. En effet, très tôt, les enfants vont se reposer sur les connaissances qu'ils ont acquises, et établir des hypothèses, faire des prédictions sur les conséquences sensorielles de leurs actes moteurs. Même si leur expérience est limitée, donc la précision de leurs prédictions faible, les enfants se reposent sur leurs modèles internes. Les résultats des expériences de perturbation vont dans ce sens : les enfants savent globalement dans quelle direction aller pour obtenir le résultat désiré. Mais l'étude de ces représentations sensori-motrices reste délicate, car nous n'avons accès qu'au comportement, et ce paradigme expérimental reste peu exploré, particulièrement en développement.

1.6. Bilan et hypothèses

1.6.1. Résumé de l'état de l'art

Nous avons décrit, au cours des sections introductives 1.1 et 1.2, le développement de la production et de la perception de la parole lors de la première année de vie, puis au cours du développement plus tardif. Nous avons vu que si les bases, qui permettent l'entrée dans l'ère de la communication parlée, sont acquises très précocement, la production de parole est une tâche complexe, pour laquelle le contrôle moteur s'affine progressivement, jusqu'à la fin de l'adolescence. Dans la section 1.3, nous avons exposé le cadre théorique dans lequel ce travail s'inscrit, en exposant les spécificités de la tâche de production de parole au cours du développement ontogénétique, mais aussi en introduisant les principes théoriques clés du contrôle moteur, cadre théorique général dans lequel ce travail s'inscrit, puis en exposant notre cadre théorique spécifique à la production de la parole. Enfin, au cours des sections 1.4 et 1.5, nous avons exposé une revue de la littérature apportant des informations concernant la maturation du contrôle moteur, selon deux points de vue complémentaires. Les études portant sur la coarticulation (section 1.4) nous renseignent sur les connaissances que les enfants ont de la variabilité associée à chaque son de parole, et de l'utilisation qu'ils font de cette variabilité lors de la production d'une séquence de sons de parole. Les études portant sur la perturbation de la production (section 1.5) nous renseignent quant à elles sur les capacités qu'ont les enfants à utiliser les degrés de liberté de leur système moteur dans le cadre de l'équivalence motrice. Chacune à leur façon, ces études nous renseignent sur la maturation du contrôle moteur pour la production de parole, via l'étude des différentes stratégies motrices qui permettent la production d'un son de parole donné, que ce soit pour minimiser l'effort articulatoire (coarticulation) ou pour atteindre le but désiré malgré une perturbation (équivalence motrice).

1.6.2. Définition de la problématique

Notre problématique générale est celle de la production de la parole chez l'enfant, plus spécifiquement celle de l'apprentissage des relations sensori-motrices pour la production

de la parole, via l'étude du comportement, renseignant sur la maturation progressive du contrôle moteur de la langue.

1.6.2.1. Définition des objectifs de travail

L'objectif de ce travail de thèse est de contribuer à qualifier et à quantifier la maturité du contrôle moteur de la langue pour la production de parole à l'âge de 4 ans dans le cadre du développement typique (et non pathologique), via la mesure articulatoire des mouvements linguaux et la mesure du signal de parole associé, en comparaison aux mêmes mesures chez une population d'adultes.

Nous avons choisi d'étudier deux indices principaux. Le premier indice concerne la stabilité de la production de voyelles en contexte isolé, renseignant sur la stabilité et la consistance du contrôle moteur sous-jacent. Le second indice concerne la capacité à anticiper les cibles futures, dans le cadre de la production d'une séquence de sons de parole, renseignant sur la maturité du contrôle moteur sous-jacent dans la capacité à exploiter la variabilité allouée à un son de parole donné dans l'objectif de minimiser l'effort articulatoire le long de la séquence de sons de parole. Plus précisément, nous avons choisi d'étudier l'anticipation de la seconde voyelle (V_2) dans la réalisation de la première voyelle (V_1) dans le cadre de la production d'une séquence de sons de parole de type V_1-C-V_2.

La comparaison des comportements adultes et enfantins, pour les mêmes tâches et dans les mêmes conditions, via des mesures acoustiques et articulatoires concernant nos deux indices principaux, renseigne sur la maturité du contrôle moteur pour la production de la parole chez les enfants âgés de 4 ans.

1.6.2.2. Choix de la problématique

Le choix de cette problématique vient du constat suivant : pour les chercheurs en parole, mais aussi pour les médecins et les orthophonistes, il y a un manque considérable de données expérimentales concernant le développement tardif de la production de la parole. Ce constat est encore plus flagrant vis-à-vis des études articulatoires. Quelques études commencent à voir le jour, mais les efforts à produire dans le but de dresser un schéma développemental typique de la production et de la perception de la parole, de la

naissance à l'âge adulte, sont encore colossaux. Ce travail a pour objectif d'apporter une petite contribution à cet édifice qu'il reste à construire, en étudiant une étape importante dans le cours du développement phonétique, celle de l'âge de 4 ans. Cet âge semble marquer le début d'une étape clé du développement, le passage d'un contrôle moteur encore approximatif à celui d'un contrôle imprécis, mais davantage semblable à celui de l'adulte dans de nombreuses caractéristiques (cf. sections 1.2, 1.4 et 1.5).

La question du développement de la production de la parole est une question passionnante, qu'il est possible d'aborder selon plusieurs angles d'attaque (psycholinguistique, contrôle moteur) et à différents niveaux (syntaxique, lexical, phonologique, phonétique). Concernant notre angle d'attaque, qui est celui du contrôle moteur, et à notre choix de niveau d'analyse, celui du développement phonétique, force est de constater qu'il n'existe encore que très peu de travaux.

La question de l'évolution de la stabilité des productions vocaliques au cours du développement a été relativement bien documentée, à travers les nombreuses études acoustiques (cf. section 1.2.1.). La question de la mise en place des patrons de coarticulation au cours du développement a également été relativement bien documentée, concernant la coarticulation intra-syllabique, à travers de nombreuses études acoustiques (cf. section 1.4.). En revanche, la question de la coarticulation au-delà des frontières syllabiques, c'est-à-dire la coarticulation extra-syllabique, a très peu été documentée. Or cette question revêt un intérêt tout particulier, car la coarticulation extra-syllabique repose davantage sur les stratégies de planification motrice et est moins tributaire des conséquences biomécaniques de bas niveau. L'étude de ce type de coarticulation renseigne donc davantage sur la maturation du contrôle moteur que ne peut le faire la question de la coarticulation intra-syllabique. Enfin, l'étude des productions enfantines à un niveau articulatoire renseigne plus directement que l'étude du signal acoustique de parole, résultante des mouvements des articulateurs, et constitue donc un niveau d'analyse moins indirect, rendant compte des stratégies articulatoires et donc motrices des productions enfantines.

Susan Nittrouer écrivait à ce propos, en 1993 « *Clearly, the most appropriate method for studying the gestural organization of children's speech would be procedures recording*

articulatory movements directly. ». Heureusement, depuis 1993, les méthodes et outils permettant l'étude directe des mouvements des articulateurs non visibles, comme la langue, ont évolué, et c'est désormais chose possible. L'échographie linguale permet en effet l'étude articulatoire des mouvements linguaux et constitue un outil adapté aux études développementales. C'est cet outil que nous avons choisi pour étudier les productions des enfants de 4 ans (cf. section 2.2.2.1. pour notre choix des outils expérimentaux).

1.6.2. Résumé de nos hypothèses

Concernant notre étude expérimentale, dont les méthodes et outils et les résultats seront exposés et discutés dans les chapitres suivants, nous formulons les quatre hypothèses *a priori* suivantes :

Hypothèse 1 - Nous devrions observer, dans les productions des enfants par rapport à celles des adultes, **davantage de variabilité liée à la répétition d'une même tâche**. Cette hypothèse est formulée en accord avec les données de la littérature existantes (cf. sections 1.2. et 1.5.3.) et en accord avec notre cadre théorique (cf. section 1.3.), qui suggèrent ensemble que la stabilité du contrôle moteur chez l'enfant de 4 ans est moindre que chez l'adulte. Cette hypothèse sera testée à travers les mesures acoustiques et articulatoires concernant la répétition de voyelles en contexte isolé, et concerne la **stabilité du contrôle moteur** sous-jacent.

Hypothèse 2 - Nous devrions observer des **patrons de coarticulation intra-syllabique d'amplitude similaire entre enfants et adultes**. Cette hypothèse est formulée en accord avec les données existantes dans la littérature (cf. section 1.4.4) et avec notre cadre théorique, qui suggèrent ensemble que les syllabes CV de type plosive + voyelle constituent des structures de base de la production de parole, pour lesquelles il existe des contraintes biomécaniques fortes, et que la production de ces structures de base est maitrisée relativement précocement du fait même de la structure du babillage et des premiers mots. Cette hypothèse sera testée à travers la mesure d'un indice acoustique de l'amplitude de la coarticulation dans ce type de séquences, à savoir les équations de locus, et sera également étudiée d'un point de vue articulatoire.

Hypothèse 3 - Nous devrions observer, dans les productions des enfants par rapport à celles des adultes, **une faible amplitude de coarticulation anticipatoire extra-syllabique, voire une absence de coarticulation anticipatoire au-delà de la frontière syllabique**. Cette hypothèse est formulée en accord avec les données de la littérature (cf. section 1.4.4) et notre cadre théorique, qui suggèrent ensemble que pour ce type particulier de coarticulation, une amplitude similaire à celle de l'adulte ne semble être observable que relativement tard (aux alentours de 9 ans, d'après Repp (1986) et Hodge (1989)). En effet, ce type de coarticulation semble refléter essentiellement les stratégies de planification motrice, qui mettent en jeu les représentations sensori-motrices associées aux buts de la production de parole, dans le cadre de la minimisation de l'effort articulatoire. Cette hypothèse concerne la **maturité du contrôle moteur** sous-jacent, en particulier la maturité des modèles internes (ou représentations du système de production de parole dans le Système Nerveux Central), composante prédictive du contrôle de la production de la parole. Elle sera testée à travers les mesures acoustiques et articulatoires de l'anticipation de V_2 dans la réalisation de V_1 dans le cadre de la production de séquences de type V_1-C-V_2.

Hypothèse 4 - Nous devrions observer des **gestes légèrement plus lents** chez les enfants que chez les adultes. Cette hypothèse est formulée en accord avec les données de la littérature (cf. section 1.2. et section 1.5.3.) et notre cadre théorique (cf. section 1.4.), qui suggèrent ensemble que les gestes des jeunes enfants sont plus lents que ceux des adultes, car le contrôle repose dans un premier temps essentiellement sur les retours sensoriels. La relative lenteur des gestes des jeunes enfants s'expliquerait par le fait que l'apprentissage de leur modèle direct est encore en construction, et ne permet pas encore de substituer le feedback rapide interne au feedback biologique externe. A l'âge de 4 ans, les gestes de parole devraient donc être encore légèrement plus lents que ceux des adultes. Cette question est importante, car le débit d'élocution a un effet sur l'amplitude des patrons de coarticulation et les différences entre enfants et adultes concernant ces patrons de coarticulation pourraient être attribuées à ce facteur. Il est donc important de contrôler cet indice afin d'écarter cette explication alternative des patrons de coarticulation. Cette hypothèse sera testée à travers la mesure de la durée des séquences V_1-C-V_2 et de chacun des segments au sein de cette séquence.

Chapitre 2

Méthodologie

2.1. Participants

Nous avons collecté des données acoustiques et articulatoires concernant 20 enfants (de 4 ans 4 mois à 4 ans et 11 mois, moyenne d'âge 4.65 ans et écart type de 0.20 ans) et 10 jeunes adultes (de 18 à 30 ans, moyenne d'âge 25.8 ans et écart type de 3.94 ans). Afin de recruter nos participants, une annonce a été passée pour que des personnes bénévoles viennent participer à l'expérience au Laboratoire de Phonétique de l'UQÀM. Le projet a également été présenté à plusieurs Centres de Petite Enfance (CPE, l'équivalent des écoles maternelles en France) dont le CPE de l'UQÀM et le CPE Le Sablier qui ont activement participé au projet.

Chaque participant a été informé au préalable du déroulement de l'expérience et a signé un formulaire de consentement (dans le cas des enfants, leurs parents ont signé ce formulaire). Tous les participants (ou leurs parents) ont rempli un questionnaire renseignant sur l'âge, le sexe, la langue maternelle, la langue utilisée quotidiennement (tous nos participants sont de langue maternelle français québécois et sont élevés dans des familles monolingues), le lieu de résidence, le niveau d'études des parents, les professions des parents. De plus, le questionnaire comportait quatre questions visant à exclure les participants présentant un trouble du langage ou de l'audition, un retard développemental, une naissance prématurée ou une prise en charge par un orthophoniste.

2.2. Acquisition des données

2.2.1. Test d'audition

Tous nos participants, adultes comme enfants, ont passé, préalablement à l'expérience de production de parole, un test d'audition (audiogramme). Ce test consiste en l'écoute binaurale, à 20 dB, de tons purs à 250, 500, 1000, 2000, et 4000 Hz. Les participants

devaient répondre (oralement ou manuellement) lorsqu'ils percevaient un son. Tous nos participants présentent une audition normale.

2.2.2. Dispositif technique

2.2.2.1. Dispositifs de mesure des mouvements articulatoires

De nombreux dispositifs techniques ont été développés afin de rendre possible le recueil de données cinématiques sur les articulateurs non visibles, tels que la langue. Parmi ces dispositifs, très peu sont adaptés au recueil des productions enfantines. La cinéradiographie, un film constitué d'images radiographiques de la tête et du cou, permet de visualiser, généralement de profil, la langue, le velum et autres articulateurs cachés (e.g. Perkell, 1969, Bothorel et al., 1986). Ce dispositif, non-invasif, est potentiellement nocif puisqu'il délivre au participant des doses élevées de radiations (rayons X). L'électromagnétométrie, ou EMA, a été développée à la fin des années 1980 afin de poursuivre les études articulatoires sans porter atteinte à la santé des participants (Perkell et al., 1992). Elle repose sur la localisation de petites bobines par induction dans un champ magnétique variable. Ces petites bobines doivent être collées aux dents et sur la langue des participants, ce qui est très peu adapté aux études développementales. Enfin, l'électropalatographie, ou EPG (i.e. Gibbon & Hardcastle, 1989), qui consiste à insérer un faux palais équipé d'électrodes et à détecter les contacts entre la langue et ces électrodes, représente également un dispositif technique assez lourd, peu adapté aux études développementales concernant de très jeunes enfants, et qui ne permet que la détection des contacts et ne permet pas de réelles mesures cinématiques. En définitive, parmi tous les dispositifs existants, l'échographie linguale semble être le dispositif le plus adapté au recueil et à l'étude des gestes linguaux des enfants. Il est facile d'utilisation, d'un coût modéré, peu contraignant pour les participants, non-invasif et sans danger pour l'enfant.

L'échographie linguale, qui est utilisée classiquement dans le plan sagittal médian[4], est un outil non-invasif, au dispositif technique relativement simple, qui permet de visualiser, en

[4] Le plan sagittal médian est le plan passant par le milieu de la face, via l'arrête nasale, le phylum et le milieu du menton.

temps réel, la surface ou le contour supérieur de la langue. Cet outil a commencé à être utilisé dans le domaine de la parole dans les années 1980 (Sonies et al., 1981), notamment avec des échographes doppler pulsés (e.g. Keller & Ostry, 1983), comportant un seul cristal piézo-électrique, capable de transformer une énergie électrique en énergie mécanique (en vibrant à très haute fréquence), faisant entrer en vibration les particules adjacentes et propageant alors des ultrasons (vibrations de longueur d'onde comprise entre 20 kHz et plusieurs MHz, non audibles pour l'oreille humaine). Ce même cristal piézo-électrique réceptionne les ultrasons réfléchis par les obstacles auxquels ils se sont heurtés. La mesure des temps de trajet aller-retour des ultrasons pendant une fenêtre temporelle fixée, permet de calculer la distance parcourue par les ultrasons. Il est ainsi possible de reconstruire, relativement à la position du cristal, une image des objets sur lesquels se sont réfléchis les ultrasons. On peut en réalité collecter de nombreuses mesures grâce à l'ultrasonographie : densité des tissus, élasticité, température, ou encore activités électriques et changements de pression sanguine, tels que ceux associés à l'activité cérébrale (Macé et al., 2011). Mais nous nous intéresserons ici uniquement à l'échographie linguale, et dans ce domaine, seule l'imagerie Doppler traditionnelle, qui permet de mesurer le contour supérieur de la langue, est utilisée. De nos jours, les sondes échographiques utilisent de multiples cristaux piézo-électriques alignés, ce qui permet de reconstruire et d'observer le contour d'un objet en deux dimensions. On peut même utiliser plusieurs lignes de cristaux dans différentes directions, ce qui permet de reconstruire une surface tridimensionnelle (Bressman et al., 2005), mais le coût encore important des dispositifs 3D est un frein à leur développement. Pour une présentation plus complète de l'utilisation de cet outil pour imager la langue, voir Stone (2005).

Malgré sa facilité d'utilisation, son coût raisonnable et son apparition sur le marché dès les années 1980, l'imagerie échographique linguale reste à ce jour encore trop peu utilisée en parole, et ce pour une simple et bonne raison : il faut lui ajouter des dispositifs souvent complexes pour en tirer des mesures fiables. En effet, un problème de taille rencontré avec l'imagerie échographique linguale est que le sujet peut bouger par rapport à la sonde. Si un déplacement du sujet ou de la sonde se produit, alors les images enregistrées dans une séquence ne le sont pas dans un même et unique référentiel absolu, et il devient alors délicat d'en extraire des mesures pertinentes et d'en tirer des

conclusions sur des hypothèses de recherche. Plusieurs dispositifs expérimentaux ont été proposés pour pallier ce problème, entre autres le système HATS, pour *Head and Transducer Support System* (Stone & Davis, 1995), qui consiste en un dispositif assez lourd composé d'un casque et d'un pied de micro rigide afin d'empêcher les mouvements relatifs du participant et de la sonde. Plusieurs dispositifs plus ou moins similaires existent, ayant tous pour but de contraindre au maximum les mouvements du participant et la sonde, pour les garder autant que faire se peut dans un référentiel unique.

Un autre type de dispositif existe, le dispositif HOCUS, pour *Haskins Optically Corrected Ultrasound System* (Whalen et al., 2005), qui offre une solution élégante au problème en ne contraignant ni les mouvements du sujet, ni les mouvements de la sonde. Ce dispositif propose de coupler des mesures de capture de mouvement et l'imagerie échographique, en disposant des diodes émettrices d'ondes infrarouges à la fois sur la sonde et sur le visage du participant (ou sur des lunettes) et en suivant la position de ces diodes (à une fréquence élevée) grâce à trois caméras infrarouge. Ce couplage de dispositifs permet, après acquisition des données, de réaligner les données échographiques dans un seul et même référentiel, relatif aux repères anatomiques du participant. Ce système ne contraint donc pas le participant et permet de corriger les mouvements relatifs de la sonde et du participant (mouvements de tête de bas en haut, rotation, déplacement par rapport à la sonde). Pour que cette correction soit efficace, il faut que les données soient prises dans le plan sagittal médian, ou, au minimum, que l'angle par rapport à ce plan ne soit pas trop important. En effet, compte tenu de la distance moyenne entre la sonde et la langue, chez un adulte, une inclinaison de 5° par rapport au plan sagittal médian correspond à un décalage de 0.7 mm une fois le contour projeté sur le plan correspondant (Whalen et al., 2005). Si les données sont prises trop loin du plan sagittal médian ou si l'angle par rapport à ce dernier est trop important, alors le système HOCUS ne peut que signaler ce problème et les données doivent être éliminées manuellement. Ce système nous semble particulièrement adapté au recueil des productions enfantines et a été retenu pour notre expérience. Il sera décrit avec plus de détails dans la section suivante. Les étapes de la passation de l'expérience jusqu'à l'analyse des données seront décrites dans la suite de ce chapitre.

2.2.2.2. Le dispositif technique adopté

En ce qui concerne notre travail, le dispositif technique nécessaire à l'acquisition de données est constitué d'une sonde échographique 2D (angle de vue de 90°, fréquence d'échantillonnage de 29.97 Hz, modèle Sonosite 180+, 256 teintes de gris, en mode écho Doppler), d'un microphone directionnel attaché à l'oreille des participants (fréquence d'échantillonnage 44.1 kHz), le tout relié à l'entrée analogique d'une caméra numérique qui envoie ensuite à un ordinateur le flux d'images échographiques et le signal audio synchrones. L'enregistrement se commande depuis un ordinateur, le flux étant enregistré via Abobe Première Pro. Avant la passation de l'expérience, on demande au participant de parler afin de régler le gain du signal audio ainsi que la profondeur de champ de l'image échographique et la position de la sonde pour obtenir le contour le plus complet de la langue.

En plus du matériel nécessaire pour l'acquisition des flux audio et vidéo, il faut recueillir les données relatives au suivi de mouvement, dans le cadre du dispositif HOCUS. Pour ce faire, le Laboratoire de Phonétique de l'UQÀM dispose d'un Optotrak Certus 3020, un système de marque Northern Digital Inc. (NDI) de capture de mouvement basé sur la localisation, via trois caméras infrarouge, de diodes émettrices d'infrarouge que l'on peut disposer sur les participants. Ce système peut localiser jusqu'à 512 diodes avec, selon le constructeur, une précision de l'ordre du centième de millimètre, ce à une fréquence d'échantillonnage de 166 Hz pour le modèle que nous avons utilisé.

Dans notre configuration, six diodes de l'Optotrak étaient disposées sur le visage du participant. Trois d'entre elles placées sur le front serviront de référentiel pour réaligner les données du participant. Deux autres étaient placées sur les lèvres, une sur la lèvre supérieure et une sur la lèvre inférieure, et une dernière sur la mandibule, ces trois dernières diodes étant placées dans le plan sagittal médian. Cette disposition nous permet de vérifier, lors de la passation de l'expérience, que la sonde reste bien dans le plan sagittal médian. Elle nous permet également de vérifier visuellement, une fois les données acquises, qu'elles ont bien été acquises, tout au long de la passation, dans le plan sagittal médian et permet également de constater l'ampleur des mouvements du participant lors de l'acquisition. La dernière diode disposée sur la mandibule permet

d'extraire, dans les mouvements globaux de la langue, ceux qui sont et ceux qui ne sont pas corrélés avec les mouvements de la mandibule, la langue étant portée par cette dernière. Quatre diodes sont disposées sur le front de la sonde échographique, afin de localiser cette dernière et de savoir pour chaque image, sur quel plan, dans un espace en trois dimensions, a été acquise cette image. Enfin, trois diodes sont disposées sur un triangle, que le participant doit tenir fermement entre ses dents durant 5 secondes tout en restant immobile. Cette séquence est effectuée dans le but de mesurer le plan occlusal[5], plan de référence définissant le plan horizontal. Le point d'origine du référentiel est l'incisive supérieure. L'enregistrement de ces données se fait via l'interface NDI First Principles, sur l'ordinateur de contrôle qui permet également de lancer l'enregistrement des données audio et des images échographiques.

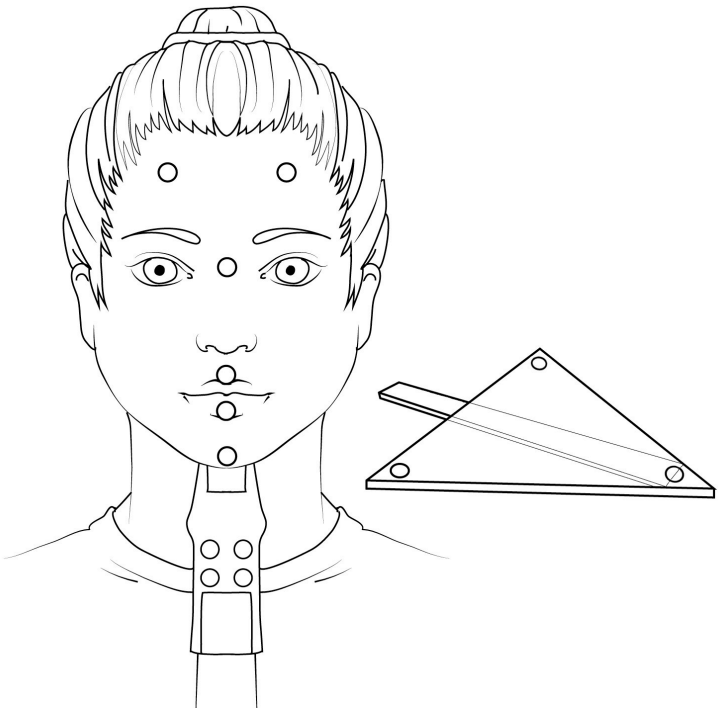


Figure 2.1. Disposition des diodes infrarouges sur le visage d'un participant, sur la sonde et sur le triangle qui sert à obtenir le plan occlusal. La disposition des diodes est toujours la même pour faciliter le traitement des données. Illustration de Sabine Burfin.

Ce système - installé au Laboratoire de Phonétique de l'UQÀM par les soins de Mark K. Tiede des laboratoires Haskins, que nous remercions pour avoir pris le temps de nous

[5] Le plan occlusal correspond à la surface virtuelle sur laquelle les surfaces occlusales maxillaires et mandibulaires d'un participant à la dentition normale se rencontrent.

expliquer son fonctionnement et les programmes d'analyse des données issus de ce dispositif - nous a permis de recueillir des données échographiques de la langue, alignées dans le plan sagittal médian, en lien avec des repères anatomiques du sujet, avec le signal audio synchronisé.

2.2.3. Passation de l'expérience

Les 10 participants adultes ainsi que 5 enfants ont passé l'expérience dans la chambre sourde du Laboratoire de Phonétique de l'UQÀM. Onze enfants ont passé l'expérience au Centre de Petite Enfance « Le Sablier » à Montréal, et 6 enfants au Centre de Petite Enfance de l'UQÀM.

L'expérience se déroule comme un jeu de marionnettes, lors duquel le bloc caméra de l'Optotrak sert de théâtre de marionnettes. Pour éviter toute surcharge de la mémoire de travail, les marionnettes sont présentées par paires. Il faut alors simplement retenir le nom de deux marionnettes qui peuvent surgir. Le but de la tâche est de nommer la marionnette à chaque fois qu'elle apparait. Le corpus est le même pour chaque participant, mais l'ordre d'apparition des paires de marionnettes est aléatoire et différent pour chaque participant. Cette tâche a été conçue dans le but d'être plaisante pour des enfants de 4 ans, de conserver leur attention au maximum afin d'obtenir un maximum de données (la passation d'un participant dure environ 20 minutes et il est difficile d'aller au-delà avec des enfants de 4 ans). Nous nous sommes aussi efforcés de nous placer, autant que possible, dans un cadre où les participants doivent mémoriser, planifier et exécuter des séquences de parole, et non pas imiter un expérimentateur adulte.

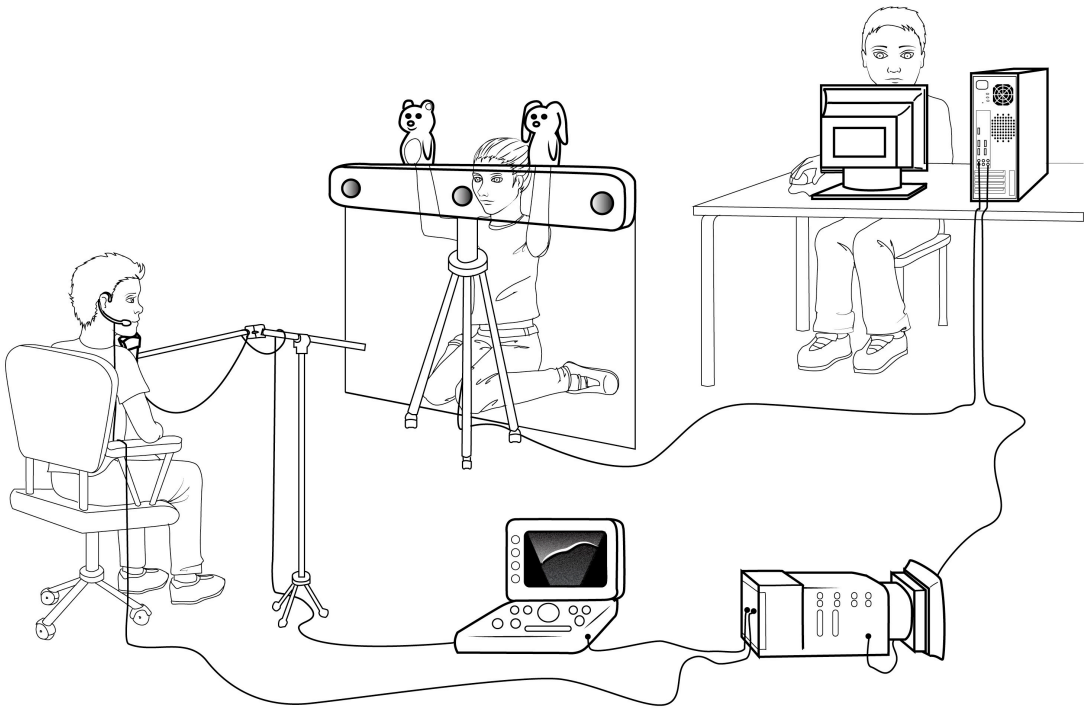


Figure 2.2. Schéma du dispositif expérimental. Le participant, assis, porte un micro directionnel et son menton est posé sur la sonde échographique. Il interagit avec le marionnettiste caché derrière le dispositif de capture de mouvements. Les signaux échographiques et audio synchronisés (via la caméra) ainsi que les données de capture de mouvement sont enregistrés par un expérimentateur situé au poste de contrôle. Illustration de Sabine Burfin.

Pour le bon déroulement de cette expérience, il est nécessaire qu'au moins quatre expérimentateurs soient présents. Un expérimentateur joue le rôle du marionnettiste, il interagit avec le participant, le stimule, lui explique les règles du jeu, vérifie avec quelques exemples qu'il a bien compris les règles et incarne les différentes marionnettes au cours de la passation. Un autre expérimentateur est au poste de contrôle, et lance l'acquisition des données échographiques et audio via Adobe Première et l'acquisition des données de mouvement via NDI First Principles. Dans cette tâche, l'expérimentateur doit systématiquement lancer l'acquisition des données échogaphiques avant celle de l'Optotrak, pour que chaque image de la séquence puisse être localisée dans le référentiel 3D, et de la même manière couper l'enregistrement de l'ultrason après celle de l'Optotrak. Les marionnettes sont présentées par paires, et chaque séquence est constituée d'au moins huit répétitions de chacun des noms des deux marionnettes. Un troisième expérimentateur se place derrière l'enfant, vérifie que l'image échographique est bien prise dans le plan sagittal médian et que la plus grande partie du contour

supérieur de la langue est bien visible. Il replace la tête du participant si besoin ou la maintient légèrement si le participant a tendance à beaucoup bouger. Enfin, un quatrième expérimentateur vérifie que, de face, l'image est bien dans le plan sagittal médian. Il a aussi pour tâche de compter le nombre de répétions de chaque nom de marionnettes, d'informer le marionnettiste quand la séquence est au complet, et d'informer le poste de contrôle que l'on peut clore cette séquence et ouvrir la suivante.

Figure 2.3. (a) Disposition des diodes infrarouges sur le visage d'un participant et sur la sonde. (b) Le bloc caméra de l'Optotrak utilisé comme théâtre de marionnettes.

Chaque passation commence par une courte explication du déroulement de l'expérience, le remplissage du questionnaire et du formulaire de consentement, et l'installation du dispositif technique (diodes à disposer, chaise et pied de micro tenant la sonde à régler, gel à déposer sur la sonde, trouver le bon alignement pour voir le maximum du contour de langue). Puis on demande aux participants une ou plusieurs séquences de 5 secondes, pendant lesquelles ils ne doivent pas bouger, avec la sonde positionnée sous le menton, qui nous serviront de référence pour réaligner les données par la suite. En plus de ces séquences fixes, on demande à chaque participant de serrer fermement entre ses dents le triangle sur lequel sont disposées trois diodes, afin de mesurer le plan occlusal du participant. L'expérience se déroule ensuite et dure environ 20 minutes. A la fin de l'expérience, on demande aux participants de boire du jus de pomme en maintenant la sonde sous leur menton, ceci afin d'obtenir une trace du palais durant les séquences de déglutition. Chaque participant adulte a été rémunéré 20$ et chaque enfant s'est vu offrir une peluche d'une valeur marchande comprise entre 20 et 25$.

2.2.4. Corpus

Le corpus que nous avons demandé à nos participants de produire est constitué d'une part de **voyelles isolées /i e ɛ a u/**, afin d'observer la variabilité liée à la répétition d'une même voyelle et nous donner une idée de l'Espace Acoustique Maximal (Boë et al., 1989) de chaque participant et de la répartition des catégories vocaliques au sein de cet espace. Mesurer la variabilité liée à la répétition d'un même phonème (entre 8 et 10 fois) renseigne sur la stabilité du contrôle.

Le corpus est également constitué de séquences Voyelle-Consonne-Voyelle symétriques et dissymétriques, que nous noterons **V_1-C-V_2, pour lesquelles V_1 était /a/ ou /ɛ/, C était /b/, /d/, ou /g/, et V_2 était /a/ ou /i/.** Toutes les combinaisons possibles de ces trois sons ont été incluses. Ce corpus a été défini afin de permettre la mesure de l'anticipation de V_2 dans V_1, en choisissant pour V_1 des voyelles pour lesquelles une certaine variabilité est permise, et pour V_2 des voyelles extrêmes, /i/ étant la voyelle la plus antérieure et la plus haute (ou la plus fermée) et /a/ la voyelle la plus basse (ou la plus ouverte). Ce corpus a été également défini dans le but de mesurer la coarticulation anticipatoire au sein de la syllabe CV_2 plosive-voyelle, que l'on sait varier considérablement selon les contraintes consonantiques, mais que l'on sait également être d'amplitude similaire chez l'adulte et chez les enfants de 4 ans (Noiray et al., 2013).

Pour résumer, ce corpus permet : (1) via la mesure de la variabilité entre répétitions des voyelles isolées, d'évaluer la stabilité du contrôle et la précision des représentations sensorimotrices et perceptives des catégories vocaliques, (2) via la mesure de la coarticulation intra-syllabique, d'évaluer le degré d'indépendance du contrôle des constituants phonémiques au sein de la syllabe CV_2, (3) via la mesure de l'anticipation extra-syllabique, de V_2 à V_1, associée aux mesures temporelles renseignant sur le débit d'élocution, d'évaluer l'empan de la planification motrice de chaque participant. La coarticulation extra-syllabique permet ainsi d'évaluer, pour chaque participant, ce que nous appellerons la maturité des représentations sensori-motrices associés aux sons de parole. Par le terme maturité, nous comprenons les capacités, en comparaison avec celles de l'adulte, qui concernent la variabilité des sons de parole compatible avec une communication efficace, ainsi que l'utilisation des degrés de liberté du système moteur

pour exploiter cette variabilité lors de la production de séquences de sons de parole, tout en respectant les limites de cette variabilité.

2.3. Prétraitements et analyse des données

2.3.1. Etiquetage du signal audio

Préalablement au prétraitement des données acoustiques, le signal audio a été sous-échantillonné à 16 kHz afin de permettre une détection de formants plus précise. La première étape de prétraitement des données consiste à étiqueter le signal audio. Cette étape permet de repérer, sur le signal audio, les instants qui sont pertinents pour caractériser la production des sons étudiés. L'étiquetage du signal audio a été effectué manuellement via le logiciel *Praat* (Boersma & Weenink, 1996), logiciel qui présente plusieurs avantages, notamment celui d'être gratuit, Open source et facile d'utilisation. Chaque voyelle isolée et chaque séquence V_1-C-V_2 a été étiquetée selon les règles suivantes. Le début d'une voyelle est défini sur le signal temporel par le premier passage par zéro descendant suivant l'émergence claire d'une structure formantique sur le spectrogramme (la trace la plus claire d'une structure formantique étant l'apparition d'un F2) ; la fin d'une voyelle est définie par le premier passage par zéro descendant du signal temporel après la disparition du F2 sur le spectrogramme. Compte tenu du fait que toutes les consonnes que nous avons étudiées sont précédées et suivies d'une voyelle, nous avons déterminé différentes phases dans la production de la consonne : (1) la phase allant de la fin de la voyelle V_1 jusqu'au début du bruit de plosion, que nous appellerons la tenue consonantique (notée $ dans l'étiquetage sous *Praat*, cf. Figure 2.4), (2) le relâchement consonantique correspondant au burst acoustique (identifiable sur le spectrogramme par une composante non périodique de forte énergie dans les hautes fréquences), (3) la phase de transition vers la voyelle suivante, caractérisée par le VOT de Klatt allant du début du relâchement consonantique jusqu'au début de V_2 (voir figure 2.4).

Cet étiquetage permet également d'identifier les images échographiques d'intérêt qui seront analysées par la suite afin de caractériser les propriétés articulatoires de la production des différents sons de parole. Nous avons choisi quatre instants clés dans les séquences V_1-C-V_2. Ces quatre instants sont le milieu de V_1 (situé au milieu de l'instant de

début et de l'instant de fin de la voyelle), la fin de V_1, le relâchement consonantique et enfin le milieu de V_2.

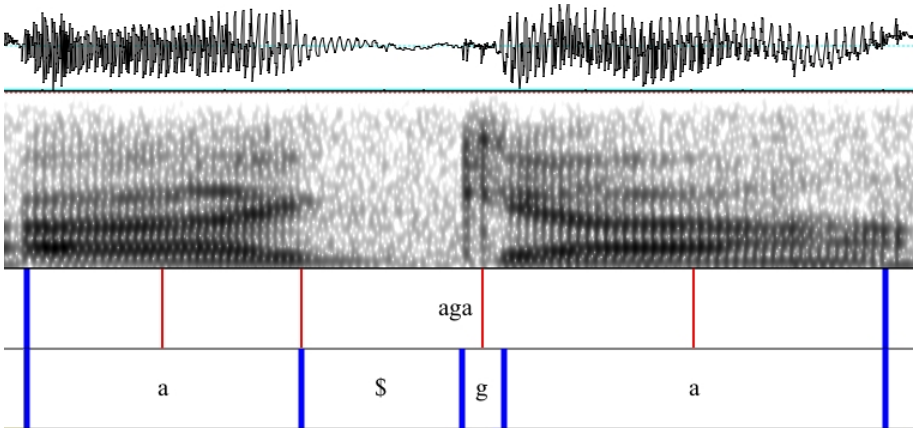


Figure 2.4. Illustration de l'étiquetage du signal audio via *Praat*. La première ligne sert à définir la longueur totale de la séquence V_1-C-V_2 et à l'identifier, la seconde ligne sert à définir les limites de chaque segment individuel (de gauche à droite, voyelle V_1, tenue consonantique représentée par le symbole $, relâchement consonantique correspondant à la définition du VOT de Klatt, voyelle V_2, cf. texte pour les détails). Les barres rouges sur la première ligne symbolisent les instants clés pour lesquels les images échographiques correspondantes sont extraites (milieu de V_1, fin de V_1, le relâchement consonantique et enfin le milieu de V_2).

2.3.2. Données acoustiques

Afin de mettre à l'épreuve nos hypothèses de recherche concernant la précision des réalisations vocaliques, ainsi que la mesure de l'anticipation intra et extra-syllabique, nous avons effectué une série de mesures dans le domaine acoustique. Concernant les voyelles, la mesure des deux premiers formants, au centre de la voyelle isolée et de V_1, constitue à la fois un indice sur la précision des réalisations vocaliques et sur la mesure de l'anticipation extra-syllabique. La mesure des caractéristiques spectrales du *burst* consonantique et les équations de locus (exprimant les relations entre le F2 au début de la voyelle, tel que défini ci-dessus dans la section 2.3.1, et le F2 au milieu de V_2) permettent, quant à eux, une mesure de l'amplitude de la coarticulation intra-syllabique.

2.3.2.1. Indices mesurés

Une fois l'étiquetage du signal audio terminé, il reste à effectuer des mesures sur les moments d'intérêt du signal acoustique. L'estimation des valeurs paramètres qui nous intéressent ne sont pas extraites via le logiciel *Praat*, même si ce dernier le permet. L'estimation des valeurs de formants des voyelles produites par de jeunes enfants est en effet une tâche délicate (la fréquence fondamentale, F_0, très élevée chez les enfants complexifie grandement la séparation des composantes spectrales liées à la source de celles liées au filtre, ou conduit vocal), et la fiabilité de la détection des formants de *Praat*, si bonne soit-elle avec de la parole adulte, montre ses limites avec les production enfantines (notamment les changements d'estimation drastiques quand on change le paramètre du nombre de formants voulu dans la plage de fréquences définie). Les estimations des indices que nous avons choisis de mesurer sont effectuées sous MATLAB avec des programmes que nous avons écrits et paramétrés spécialement pour faire face à cette difficulté.

2.3.2.2. Indices mesurés pour les voyelles : estimations de formants

Toutes les informations concernant les moments définis par les bornes manuellement annotées sous *Praat* sont contenues dans des fichiers *.TextGrid*. Ces informations sont relues par des routines MATLAB afin de lire les morceaux choisis des signaux audio. Les estimations de formants pour les voyelles sont effectuées via la méthode de *Linear Predictive Coding* (LPC). Puisque la détection de formants pour les productions enfantines est délicate, nous avons combiné les résultats donnés par l'estimation des valeurs maximales dans le spectre LPC et ceux des valeurs estimées des pôles dans le filtre LPC. Pour chaque voyelle, une plage de valeurs autorisées a été définie dans le but d'éliminer les erreurs de détection et les *outliers*. Ces plages de valeur ont été définies spécifiquement pour les enfants, les adultes hommes et les adultes femmes, et sont données dans le tableau 2.1 ci-dessous.

Tableau 2.1. Valeurs en Hz des bornes inférieures et supérieures des intervalles de recherche pour F1 et F2, pour chaque voyelle, pour les enfants et les adultes hommes et femmes.

		/i/	/e/	/ɛ/	/a/	/u/
Enfants	F1	300-700	500-700	450-900	800-1500	300-850
	F2	3000-4200	2200-3300	2100-3400	1400-2500	600-1700
Adultes ♂	F1	210-400	500-700	350-700	500-850	210-400
	F2	1650-2450	2200-3300	1300-2200	1100-1800	500-900
Adultes ♀	F1	210-400	300-500	400-800	600-1100	210-450
	F2	2100-2800	1800-2600	1700-2600	1500-2200	600-1100

Les valeurs mesurées pour les voyelles sont les trois premiers formants, en Hz, mais nous ne nous préoccuperons que de la valeur des deux premiers formants, puisqu'ils suffisent à discriminer les voyelles entre-elles dans un espace en deux dimensions et que l'effet d'anticipation que nous souhaitons étudier devrait apparaître principalement dans ces deux dimensions. La mesure du troisième formant est utilisée seulement pour garantir la bonne mesure du second formant.

Les mesures que nous avons effectuées nous permettent l'estimation de **F1 et F2 au milieu des voyelles V_1 et V_2 et à l'initiale de V_2,** avec des fenêtres d'analyse glissantes de 10 ms se superposant de 40%. Le milieu de la voyelle est défini grâce aux bornes manuellement disposées via *Praat* ; cinq fenêtres autour du milieu de la voyelle sont analysées et la moyenne de ces cinq fenêtres nous donne les valeurs des estimations de F1 et F2.

2.3.2.3. Indices mesurés pour les consonnes : Equations de Locus

Concernant les consonnes, nous avons utilisé un indice acoustique de la coarticulation intra-syllabique de type plosive-voyelle : les **Equations de Locus**. Cet indice acoustique a été développé dans le but de mesurer la coarticulation entre plosives et voyelles au sein de syllabes CV (classiquement /b d g/ et /i a u/). Cet indice est basé sur la relation entre la valeur du second formant, F2, à la sortie de la consonne, lorsque l'écoulement de l'air devient laminaire (apparition d'une structure formantique appelée par la suite initiation

du F2) et au milieu de la voyelle. Si la production de la consonne est influencée par la voyelle suivante, alors cette influence devrait être observable dans les transitions formantiques dès l'initiation du F2. Pour le dire simplement, si la production de la consonne dépend de la voyelle suivante, le F2 de cette consonne doit varier en fonction de cette voyelle et adopter des valeurs différentes selon la voyelle suivante. Les transitions formantiques sont dans ce cas minimales, puisque les lieux d'articulation de la voyelle et de la consonne sont rapprochés au maximum. Dans le cas opposé, si la voyelle suivante n'a aucune influence sur la réalisation de la consonne, alors le F2 de cette consonne ne doit pas dépendre de la voyelle suivante, et les transitions formantiques doivent être maximales puisque les lieux d'occlusion de la voyelle et de la consonne ne sont pas maximalement rapprochés.

En représentant, dans un espace en 2 dimensions, dont l'abscisse est le F2 au milieu de la voyelle et l'ordonnée F2 à l'initiale, la régression entre ces 2 valeurs, on mesure bien d'une certaine façon la relation qu'entretient la consonne avec la voyelle suivante. La valeur de la pente de ces Equations de Locus est alors mesurée (régression entre les 2 valeurs susmentionnées), une pente de 0 indiquant l'absence de coarticulation (aucune influence de V sur la réalisation de C) et une pente de 1 indiquant un maximum de coarticulation (influence maximale de V sur la réalisation de C).

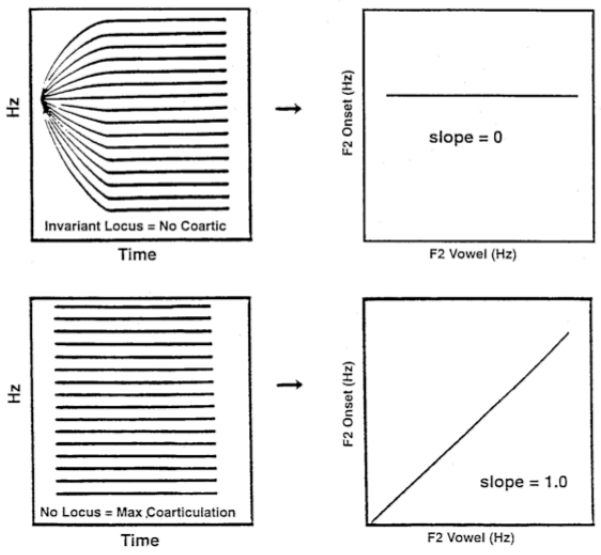


Figure 2.5. Illustration des transitions formantiques et des valeurs extrêmes des Equations de Locus correspondantes, d'après Sussman *et al.* (1999).

Cet indice, introduit par (Krull, 1987) a été extensivement utilisé notamment par Sussman et collègues pour mesurer la coarticulation entre plosives et voyelle au sein de syllabes CV. Cet indice, uniquement acoustique, est censé refléter la coarticulation, plus traditionnellement réservée aux mesures articulatoires. D'ailleurs, la validité de cet indice, en relation avec des mesures sur le plan articulatoire, établie par (Krull, 1987), est vivement contestée, notamment par (Löfqvist, 1999), qui ne retrouve aucune correspondance entre divers indices articulatoires et les Equations de Locus comme mesure de la coarticulation. Par ailleurs, il ne semble que ces mesures ne soient valables pour les plosives, car elles ne semble pas être pertinentes pour les fricatives (Tabain, 2000) et ne sont valables que pour les syllabes CV et pas pour les syllabes VC.

D'après Sussman (Sussman et al., 1992), les valeurs du formant F2 de V_2 à son tout début (fin du VOT de Klatt, cf. section 2.3.1) sont de précieux indices sur les consonnes, puisqu'elles donnent une indication, en relation avec la valeur de F2 au milieu de V_2, de l'amplitude de la coarticulation CV_2. Ces valeurs de F2 au début et au milieu de V_2 serviront à mesurer les **Equations de Locus** au sein de la syllabe CV_2, afin de tirer une information sur l'amplitude de la coarticulation au sein de la syllabe. Pour une opinion critique des Equations de Locus, cf. section 1.4.3)

2.3.2.4. Durée des segments et des séquences

En plus des mesures de ces indices sur les voyelles et les consonnes, les durées des segments ont été mesurées, ainsi que la durée totale de chaque séquence.

2.3.2.5. Normalisation des valeurs de formants

Puisque les Espaces Vocaliques Maximaux des hommes adultes, des femmes adultes, et ceux des enfants sont très différents (Peterson & Barney, 1952 ; Fant, 1960 ; Fant, 1975 ; Mol, 1970 ; Ménard et al., 2004 ; Vorperian & Kent, 2007) - la variabilité interindividuelle étant également importante - les données brutes des valeurs de formants en Hz ne permettent pas une confrontation directe des données enfant et adulte, ni pour la mesure de la variabilité ni pour celle de l'anticipation. En effet, les ellipses de dispersion des catégories vocaliques des enfants sont forcément plus grandes que celles des adultes comparées en Hz, parce que leur conduit vocal est plus court que celui des adultes et

produit donc des résonances associées à des longueurs d'onde plus courtes, donc perçues plus aigües (Fant, 1975 ; Mol, 1970). Afin de démêler ce qui relève de la longueur du conduit vocal de ce qui relève de la précision et de la stabilité du contrôle, une normalisation des données s'impose, afin de permettre une comparaison dans un référentiel unique, sans pour autant effacer les différences interindividuelles de stratégies de contrôle.

Parmi les différentes méthodes de normalisation, nous avons choisi le **Z-score**. Cette méthode consiste, pour chaque participant pris individuellement, à soustraire à chaque valeur la moyenne de l'échantillon, puis à diviser le tout par l'écart-type de ce même échantillon. Nous avons appliqué cette méthode pour les voyelles isolées mais également pour les différentes voyelles V_1 afin de mesurer l'anticipation de V_2 présente dans celles-ci. Pour les voyelles isolées, pour chaque voyelle, on a soustrait aux valeurs de F1 la moyenne de F1 pour toutes les voyelles (ceci afin de centrer l'Espace Vocalique Maximal sur une origine (0,0)), puis cette valeur est divisée par l'écart-type de l'ensemble des valeurs de F1, afin de ramener l'Espace vocalique Maximal à une taille unique pour tous les participants. La même opération a été effectuée pour les valeurs de F2. Enfin, pour les voyelles V1, /a/ et /ɛ/, la même opération (avec les valeurs moyennes et les écart-types calculés sur ces deux voyelles uniquement) a été effectuée pour F1 et F2, afin d'observer l'anticipation dans chacune des voyelles dans un espace normalisé par rapport aux amplitudes de variation.

2.3.3. Données articulatoires

L'enjeu, en ce qui concerne les images échographiques, est d'en d'extraire l'information pertinente. En effet, ces images ne sont composées que de niveaux de gris. L'information qui nous intéresse est le contour supérieur de la langue, visible à l'écran, puisque la densité des tissus de la langue est proche de celle de l'eau, et que juste au dessus du contour de la langue se trouve de l'air, ayant une densité très différente de celle des tissus mous. Une forte proportion des ondes échographiques est réfléchie à la frontière entre ces milieux de densités différentes. Cela se traduit visuellement sur ces images par un contour très noir situé juste en dessous d'une bande nette de contour blanc, qui correspond au passage du changement de densité, donc à l'air situé au dessus du contour

supérieur de la langue. La tâche consiste donc à repérer et définir le contour supérieur de la langue, identifiable grâce à ce changement brutal de gradient dans l'image échographique.

2.3.3.1. Méthodes d'extraction du contour supérieur de la langue

En ce qui concerne les prétraitements des données articulatoires, la première étape est d'identifier, à partir des repères posés via *Praat*, les instants clés pour lesquels nous souhaitons extraire et analyser des images échographiques. Nous avons défini 4 instants clés, exprimés en temps relatif par rapport à l'écoulement temporel des séquences V_1-C-V_2. Ces instants sont (1) le milieu de V_1, (2) la disparition du F2 de V_1, considéré comme la marque du passage de la prédominance de la configuration vocalique à la prédominance de la configuration consonantique (3) le relâchement de la consonne, correspondant au début du *burst* acoustique, et (4) le milieu de V_2. Pour les voyelles isolées, une seule image, correspondant au milieu de la voyelle, est extraite et analysée.

Une fois ces images identifiées, il est nécessaire d'en extraire l'information pertinente. Dans notre cas, l'information qui nous intéresse est la localisation du contour supérieur de la langue. Plusieurs méthodes semi-automatiques d'extraction du contour de langue existent à ce jour. Nous citerons les outils les plus utilisés, notamment *EdgeTrak* (Li et al., 2005), *TongueTrack* (Tang et al., 2012) ou encore *AutoTrace* (e.g. Berry et al., 2012). *Edgetrak* fonctionne grâce à un algorithme basé sur les contours actifs (*snakes*) : il suffit d'initialiser le premier contour (grâce à quelques ancres) pour que ce contour soit adapté aux 20 images suivantes. La limite de cette méthode - outre les erreurs de détection, qui peuvent être corrigées manuellement ou via la redéfinition du contour initial de l'image d'intérêt - concerne le nombre d'images très réduit que l'on peut traiter en une seule itération. Le formatage des données est donc très contraint et demande de se soumettre aux spécificités du logiciel. *TongueTrack* fonctionne selon les principes de l'apprentissage automatique (*Machine Learning*). L'objectif est de minimiser une énergie au sein de l'image grâce à une optimisation fondée sur les champs de Markov, afin de détecter automatiquement le contour de la langue. De telles méthodes peuvent être très efficaces, à condition de disposer d'une base de donnée suffisante pour lesquelles les images ont préalablement été traitées manuellement. Enfin, *AutoTrace* repose également sur les principes de l'apprentissage automatique en utilisant un Réseau de Neurones

Artificiel comportant un nombre important de couches cachées sans connexions entre les éléments au sein de chaque couche (Deep Belief Networks). Le point de départ de l'apprentissage est une large base de données pour lesquelles la détection manuelle des contours a déjà été effectuée. Son taux d'erreur dépend donc directement de la proximité entre sa base d'apprentissage et la base test. Dans notre cas, nous souhaitons détecter des contours de langues d'enfants, très différents de ceux des adultes, ce qui pourrait entrainer une erreur de détection importante avec ce logiciel.

Enfin, chaque logiciel proposé a des atouts et des faiblesses, mais une faiblesse commune à tous les outils de détection semi-automatique ou automatique est leur importante marge d'erreur. Ces logiciels présentent en effet en moyenne, dans les meilleurs cas, des erreurs de l'ordre de 2 mm, et dans les pires cas, des erreurs de l'ordre de 5 à 9 mm (Gàbor Csapò & Lulich, 2015). Etant donné que nous souhaitons effectuer des mesures quantitatives, afin de mettre en évidence un effet dont la taille devrait être de l'ordre de 7-8 mm ou 1 cm dans le meilleur des cas (la stabilité des voyelles isolées devrait également être de l'ordre de quelques millimètres), nous préférons en pas introduire d'erreur de mesure supplémentaire.

2.3.3.2. Notre choix : *GetContours*

Face à une telle situation, nous avons utilisé le logiciel *GetContours*, développé par Mark K. Tiede des *Haskins Laboratories* (New Haven, Connecticut) dans un environnement MATLAB, dans le but de permettre l'extraction manuelle des contours de langue sur un grand nombre d'images contenues dans des formats vidéo (à l'image de Articulate Assistant Advanced software, AAA, Articulate Instruments Ltd., 2010). Ce logiciel permet de lire une vidéo complète, de naviguer d'image en image sur cette vidéo en disposant du numéro de l'image et de disposer des ancres permettant de définir le contour supérieur de la langue. Pour chaque image d'intérêt, quelques ancres (entre 5 et 10 selon la complexité de la forme du contour de langue) sont disposées afin de repérer le contour supérieur de la langue. Ces ancres sont utilisées pour définir une fonction *spline*, qui approxime le contour de la langue, et dont 100 points seront extraits pour chaque image. Le format exporté par *GetContours* est une structure comprenant pour chaque image le numéro de l'image, les coordonnées x et y des ancres et les coordonnées x et y de 100 points répartis uniformément sur la fonction *spline* approximant le contour supérieur de

la langue. L'avantage de ce logiciel est qu'il permet de naviguer sur une vidéo complète, de repérer manuellement le contour de la langue et de pouvoir sans cesse modifier la position des ancres si une correction ultérieure est nécessaire. Le format de sortie est également pratique d'utilisation, contrairement à *EdgeTrak* qui écrit un fichier individuel pour chaque contour de langue traité.

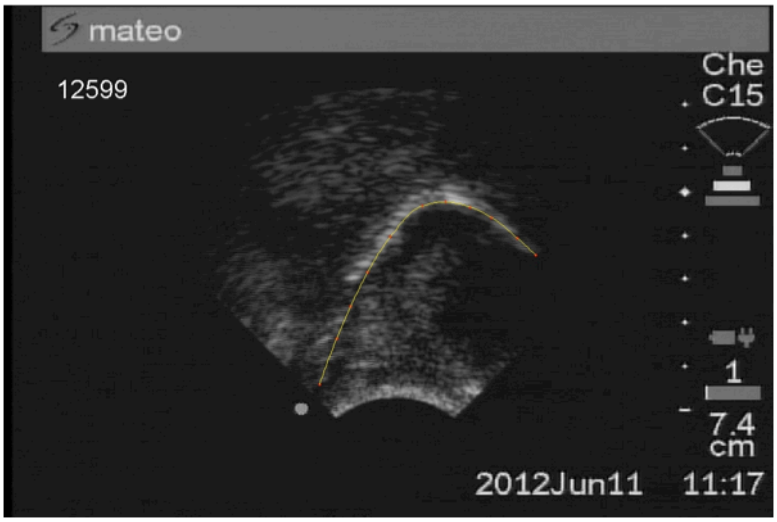


Figure 2.6. Exemple d'extraction du contour supérieur de la langue via l'interface *GetContours*. Les ancres définies manuellement apparaissent en rouge et le contour supérieur de la langue, optimalement approché par une fonction *spline*, apparait en jaune.

Ainsi, les contours supérieurs de la langue pour nos images échographiques d'intérêt ont été extraits manuellement via le logiciel *GetContours*. Nous nous affranchissons des erreurs de mesure liées à la détection du contour de la langue, comme nous nous sommes affranchis des erreurs de détection de repères sous *Praat*, en effectuant manuellement l'étiquetage. La fiabilité de ces données traitées manuellement dépend en revanche en grande partie du niveau d'expertise de l'utilisateur.

Le facteur de mise à l'échelle, ou facteur pixel/mm, varie selon les participants, car nous avons cherché à minimiser, autant que faire se peut, la profondeur de champ lors de l'acquisition, ceci afin que la langue occupe le maximum de la taille de l'image, afin de ne pas perdre en précision lors de l'extraction du contour. Pour chaque profondeur de champ, ce facteur doit donc être calculé afin d'obtenir, en millimètres, les coordonnées x et y des points constituant le contour de langue. Ce facteur est calculé automatiquement, pour chaque participant, grâce à une échelle présente sur chaque image échographique,

constituée de marqueurs blancs situés sur le côté droit de l'image. L'espace entre chacun de ces marqueurs est toujours de 1 cm, quelle que soit la profondeur de champ. Un programme de traitement d'images vient automatiquement repérer la position de ces marqueurs et calcule le nombre de pixels séparant chacun des marqueurs.

Les données articulatoires ont été analysées pour un sous-ensemble de nos participants : 6 enfants et 2 adultes. Ces participants ont été choisis uniquement pour la qualité de leurs images échographiques. La totalité des données articulatoires n'a pas pu être traitée compte tenu du caractère chronophage de l'extraction manuelle des contours de langue.

2.3.3.3. Réalignement des données dans un référentiel 3D unique

Une fois les données articulatoires prétraitées, il faut les réaligner dans un référentiel 3D unique, ayant pour origine les repères anatomiques du participant. Pour ce faire, il faut d'abord que les données échographiques et les données de l'Optotrak soient synchronisées temporellement. Cette synchronisation s'effectue via la **l'inter-corrélation des deux signaux audio** (*cross-correlation*) qui ont été enregistrés séparément, et de manière synchrone avec chacun des types de données (échographiques ou Optotrak). L'instant d'occurrence du maximum de cette inter-corrélation correspond au décalage temporel existant entre les deux signaux enregistrés (cf. figure 2.7 ci-dessous). Ce décalage est ensuite pris en compte afin de replacer toutes ces données dans un référentiel temporel unique.

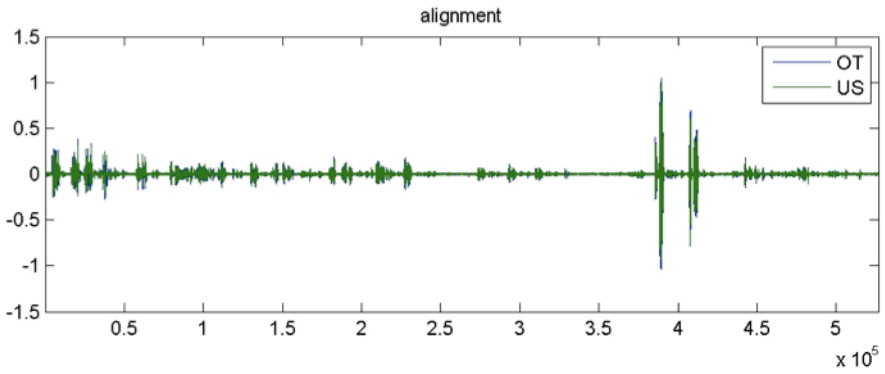


Figure 2.7. Résultat de l'inter-corrélation entre les deux signaux audio, respectivement celui issu de l'échographie (en vert) et celui issu de l'Optotrak (en bleu), représenté en amplitude en fonction du temps. Cette opération permet le réalignement des données dans un référentiel temporel unique.

La seconde étape est le réalignement de toutes ces données dans un seul et unique référentiel spatial. Cette étape, propre au dispositif HOCUS, consiste à réaligner les données échographiques dans le référentiel défini par les repères anatomiques de chaque participant, en prenant en compte les mouvements de la sonde relativement aux mouvements de la tête du participant.

La première manipulation consiste à définir un plan fixe, grâce à une séquence pendant laquelle le participant doit maintenir la sonde sous son menton et minimiser, autant que faire se peut, les mouvements de sa tête. Ce plan servira par la suite de référence pour réaligner les données. Nous avons en général effectué deux ou trois plans fixes, parmi lesquels nous choisissons celui dont l'immobilité du participant a été la plus effective. La première étape consiste en la visualisation des données lors de ce plan fixe, notamment la visualisation des diodes disposées sur le visage du participant et sur la sonde échographique, afin de vérifier la stabilité de la position des diodes (et donc du participant et de la sonde) au cours de la séquence, mais aussi l'alignement de l'image par rapport au plan sagittal médian (cf. Figure 2.8).

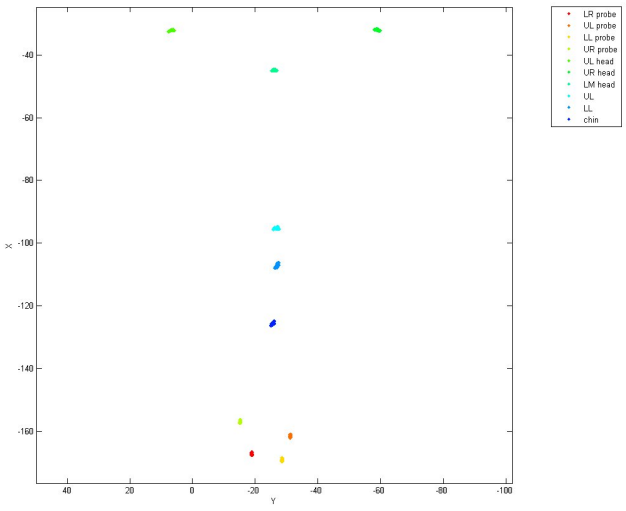


Figure 2.8. Séquence fixe servant au réalignement des données. On observe de face que les données sont bien prises dans le plan sagittal médian, puisque les diodes disposées sur le visage du sujet (en bleu, UL (*Upper Lip*), LL (*Lower Lip*) et *Chin* (menton)), que nous avons disposées dans ce plan, sont centrées par rapport aux diodes disposées sur la sonde (en bas de l'image, les 4 diodes nommées UR, UL, LL et LR légendées « *probe* » (sonde)). Les diodes « *head* » sont les trois diodes nécessaires au réalignement des données via le dispositif HOCUS.

La seconde étape consiste à vérifier et à choisir une séquence d'acquisition du plan occlusal, grâce au triangle (cf. Figure 2.8). De la même façon que pour l'étape précédente, il s'agit ici de visualiser les données et de choisir la séquence pour laquelle l'alignement est le meilleur et pour laquelle le mouvement du sujet durant la séquence est minimal.

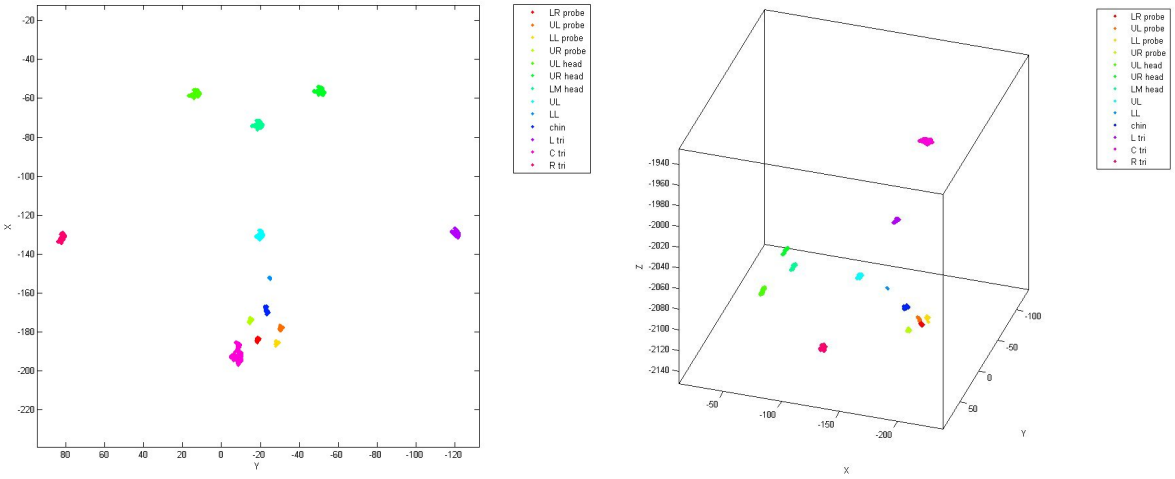


Figure 2.9. Séquence d'acquisition du plan occlusal, effectuée grâce à un triangle en plastique sur lequel sont disposées trois diodes (en violet, L, C et R légendés « tri »). Le participant doit serrer fermement ce triangle entre ses dents, et la position des 3 diodes définit le plan occlusal.

La troisième étape - la plus importante - vise à réaligner spatialement les données recueillies. Pour ce faire, nous avons maintenant à notre disposition (grâce au recalage temporel permis par l'inter-corrélation des signaux audio) la position des diodes sur la sonde et sur le visage du participant pour chacune des images échographiques dont le contour a été extrait. Nous sommes donc en mesure de connaitre, dans le référentiel 3D dont le plan occlusal définit la dimension horizontale, le plan à partir duquel chaque image échographique a été obtenue.

Durant cette dernière étape, les données sont réalignées en utilisant uniquement des transformations rigides (rotations et translations), ce qui respecte les propriétés des données. On replace, par rapport aux 3 diodes disposées sur le front du participant et par rapport aux 4 diodes disposées sur la sonde, toutes les données dans le plan sagittal médian, comme si le sujet n'avait pas bougé durant l'expérience et comme si la sonde n'avait subi que des mouvements verticaux exclusivement (cf. Figure 2.10).

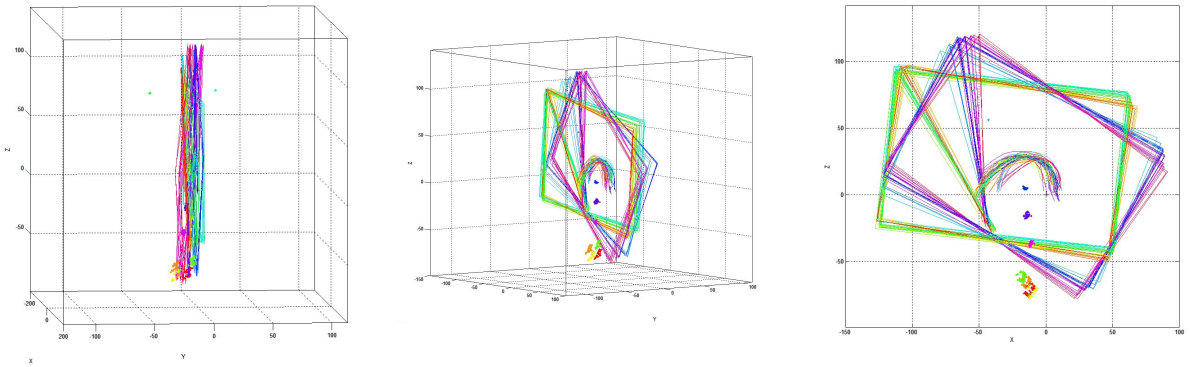


Figure 2.10. Données réalignées dans un référentiel unique 3D ayant pour origine l'incisive supérieure, grâce au système HOCUS. (a) Vue de face permettant de vérifier que les données ont bien été acquises dans le plan sagittal médian et que les mouvements, au cours de la séquence, ne sont pas trop importants (b) Vue de 3/4 et (c) Vue de profil mettant en évidence les différents plans dans lesquels les images échographiques 2D ont été acquises.

2.4. Traitement des données et analyses statistiques

2.4.1. Données acoustiques

2.4.1.1. Voyelles isolées

Afin de mesurer la stabilité des voyelles isolées, l'étude du seul écart-type autour de la moyenne (c'est-à-dire, visuellement, la taille des ellipses de dispersion pour chaque phonème vocalique, permettant une analyse uniquement descriptive) suffit à nous renseigner sur la stabilité du contrôle, ce que nous cherchions à mesurer via cet indice de variabilité entre répétitions. **Pour l'ensemble des valeurs F1 et F2 obtenues, nous avons donc mesuré les moyennes et écart-types, pour chaque participant et chaque voyelle produite**. Ces mesures ont été effectuées sur les données Z-scorées, puisque des mesures effectuées sur les données brutes en Hz n'auraient pas de sens pour comparer enfant et adultes (cf. section 2.3.2.5).

2.4.1.2. Séquences V_1-C-V_2

En ce qui concerne les séquences V_1-C-V_2, afin de quantifier les effets d'anticipation, que ceux-ci soient au sein de la syllabe ou au-delà de la frontière syllabique, la nécessité d'effectuer des tests statistiques s'impose. En effet, la simple observation des données

permet de pressentir la présence d'un effet ou non, mais l'amplitude et le degré de significativité de ces effets sont les seules informations pertinentes.

Afin de mesurer l'effet que V_2 produit sur les valeurs F1 et F2 de V_1, (anticipation au-delà de la frontière syllabique), **nous avons conduit des analyses statistiques basées sur des Modèles Linéaires à Effets Mixtes**, sous environnement R (R Foundation for Statistical Computing, 2013). Ces modèles fonctionnent globalement d'une manière équivalente à une analyse de variance (ANOVA), mais ont pour spécificité de pouvoir prendre en compte des facteurs aléatoires. Dans les modèles que nous avons défini, la valeur formantique est la variable dépendante (nous utilisons donc deux modèles, le premier pour F1 et le second pour F2, puisque ces modèles ne permettent de travailler que sur une seule variable dépendante à la fois), tandis que V_1, V_2 et le groupe d'âge (enfant ou adulte) sont des facteurs indépendants et que le participant est un facteur aléatoire. Puisque ces modèles à effet mixtes ne permettent pas le calcul exact de *p-values*, nous avons fait appel à la fonction *pamer.fnc*, (Tremblay & Ransijn, 2013), nous permettant de calculer des intervalles au sein desquels la *p-value* exacte est comprise. Le seuil de significativité que nous avons choisi pour cette étude est $p = 0.01$. Des comparaisons post-hoc par paires ont également été effectuées via la fonction *glht* (Hothorn et al., 2008).

2.4.1.3. Corrélation des indices de stabilité et de maturité

Afin d'observer s'il existe ou non une relation entre nos deux indices principaux (stabilité du contrôle et maturité du contrôle), nous avons testé la corrélation de ces deux indices entre eux. Ainsi, pour chaque participant, la stabilité du contrôle est reflétée par l'écart-type des valeurs de formants F1 et F2 (Z-scorés) des différentes voyelles isolées, et la maturité du contrôle est reflétée par l'amplitude moyenne de l'effet d'anticipation, mesurée ici par la distance euclidienne (dans l'espace F1 - F2 normalisé) entre les deux barycentres (pour chacune des deux voyelles) dans les deux contextes où V_2 diffère. Pour ces deux indices, nous avons effectué une simple régression linéaire, pour laquelle nous donnerons la pente *a* et l'indice *R*, permettant d'étudier la relation entre ces deux indices.

2.4.1.4. Coarticulation intra-syllabique

Afin d'étudier l'amplitude de la coarticulation au sein de la syllabe dans le signal acoustique, nous avons mesuré pour chaque syllabe CV_2 (pour rappel, dans notre

expérience, C est soit /b/ soit /d/ soit /g/ et V_2 est soit /i/ soit /ɛ/ soit /a/) les Équations de Locus correspondantes. Ces Équations de Locus sont une régression linéaire entre les valeurs de F2 à l'initiale de V_2 et au milieu de V_2. L'acquisition des mesures de F2 pour calculer ces équations de locus ont été décrites dans la section 2.3.1.2.

2.4.2. Données articulatoires

Concernant les données échographiques, une fois les contours supérieurs de la langue extraits et réalignés dans un référentiel unique, nous avons eu à choisir parmi de nombreuses méthodes permettant l'analyse statistique et la comparaison des contours de langue. Nous souhaitons effectuer des mesures quantitatives de l'anticipation et mesurer la significativité des résultats. Nous avons donc besoin de méthodes permettant des analyses quantitatives et statistiques.

2.4.2.1. Choix de la méthode d'analyse des contours linguaux

Il existe de nombreuses méthodes pour comparer deux courbes. Rien que pour les mesures de distance, la distance Euclidienne moyenne entre chaque point de deux courbes, les intégrales ou la distance de Hausdorff semblent être de bons candidats. En ce qui concerne les analyses statistiques, il existe également de multiples possibilités. Nous ne présenterons pas de manière exhaustive toutes les méthodes de comparaison de courbes, et ne présenterons que quelques méthodes qui semblent adaptées à l'analyses de données échographiques linguales.

Une première méthode consiste à calculer le point le plus haut de la langue (dans le référentiel de l'image échographique elle-même), et de comparer les contours de langue entre eux uniquement grâce aux coordonnées de ce point. Cette méthode, utilisée par exemple par Noiray et al. (2013), ne tire pas parti de toute l'information rendue disponible par l'imagerie échographique, puisque le contour entier (ou presque entier) est généralement visible. Puisque nous souhaitons tirer parti de toute l'information que nous avons extraite, nous ne privilégierons pas cette méthode.

Une seconde méthode consiste à extraire 3 points des contours linguaux, le point le plus avancé, le point le plus reculé, et le point le plus haut (Aubin, 2006). À partir des coordonnées de ces 3 points, la longueur de la base de la langue ainsi que la hauteur de la

langue sont calculées. Ces informations donnent des indications sur la position avant - arrière de la langue et son degré de courbure (Ménard et al., 2011).

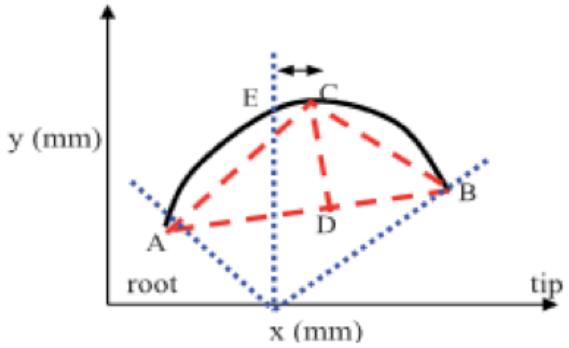


Figure 2.11. Points de mesure proposés par le logiciel *Lingua* (Aubin, 2006). A partir de 3 points A, B et C, il est possible de mesurer la longueur de la base de la langue (distance AB), la hauteur de la langue (distance CD) ainsi que la courbure de la langue (rapport distance AB/ distance CD).

Ces informations articulatoires permettent de discriminer les principales voyelles du français (Aubin, 2006). Cette méthode, comme la précédente, ne tire pas parti de toute l'information disponible, mais réussit à extraire, grâce à quelques variables seulement, des informations pertinentes concernant la déformation linguale. Puisque nous souhaitons tirer parti de toute l'information disponible et effectuer des mesures quantitatives de l'anticipation, nous ne privilégierons pas cette méthode.

Une autre méthode consiste à mesurer une distance moyenne entre deux contours. Il existe différentes façons de calculer une distance entre deux contours. On peut par exemple simplement calculer la distance euclidienne entre chaque point de deux courbes, un à un, et calculer la moyenne. Cet indice donne une idée de la distance moyenne entre deux contours, mais est tributaire de l'endroit où commencent et finissent les contours. Pour palier ce problème, une seconde méthode de calcul de distance moyenne entre deux contours a été proposée. Zharkova et al. (2011) ont en effet proposé la méthode dite du « plus proche voisin ». On recherche, pour chaque point du contour de langue, son plus proche voisin (en termes de distance euclidienne) sur le second contour de langue (celui que l'on veut comparer au premier). Cette opération est répétée pour tous les points du contour, et une distance moyenne est calculée.

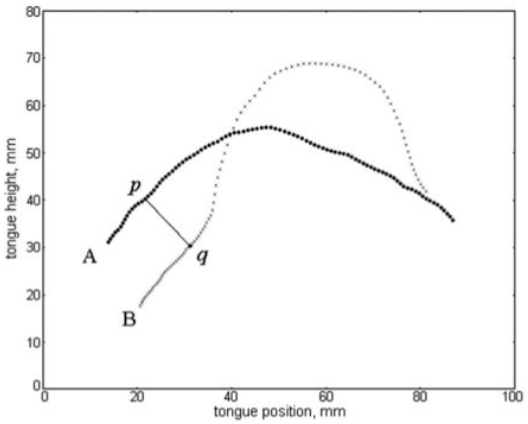


Figure 2.12. Calcul de la distance moyenne séparant deux courbes en utilisant la méthode dite du « plus proche voisin ». Tiré de Zharkova et al. (2011).

Cette méthode permet une mesure quantitative de la distance séparant deux contours de langue ou deux groupes de contours de langue, et peut donc s'avérer utile pour l'analyse de nos données. Cependant, la seule mesure quantitative ne nous suffit pas, et nous souhaitons effectuer des analyses statistiques de façon complémentaire aux mesures de distance. Il est évidemment possible d'effectuer des analyses statistiques à partir des valeurs extraites grâce à ces mesures de distance, mais de telles analyses, point par point, ne sont pas satisfaisantes d'un point de vue statistique, et nous verrons ensemble pourquoi dans les paragraphes qui suivent. Il est indispensable de notre point de vue d'effectuer une analyse statistique globale et robuste de nos contours de langue.

Une autre méthode, employée par Bressmann et al. (2005), consiste à disposer une grille semi-polaire fixe permettant de tracer l'intersection de cette grille avec le contour supérieur de la langue. On obtient alors des points également espacés sur le contour de la langue, et leur disposition reste la même relativement à la grille pour tous les contours (les lignes de cette grille sont les mêmes pour toutes les images). Des *t-test* (ou *test de Student*) sont alors effectués entre les points d'une même ligne de la grille pour différents contours de langue. Il est alors possible de comparer les contours entiers, en multipliant les tests tout en corrigeant les seuils avec une correction de Bonferroni, puisque les données ne sont pas indépendantes. Le défaut majeur de cette méthode est que sa sensibilité diminue drastiquement avec la taille de l'échantillon. Ainsi, si l'on veut représenter les données fidèlement, en 100 points par contour par exemple, il faudrait une différence énorme entre deux contours de langue pour que celle-ci soit considérée

significative par la méthode, puisque du fait de la correction de Bonferroni, sa sensibilité diminue avec le nombre de points utilisés pour définir chaque courbe.

Une autre méthode statistique répond à ce problème en traitant l'information de tous les points des contours de langue à la fois, traitant donc le contour complet de la langue, à partir d'un ensemble limité de valeurs. Cette méthode, appelée *Smoothing Spline Analysis of Variance*, ou SS ANOVA, consiste en une analyse fonctionnelle des données échographiques : elle est issue de méthodes statistiques générales (non spécifiques aux problèmes de parole), proposée par Gu (2002), et a déjà été utilisée en parole pour analyser des contours de langue issus de l'échographie (Davidson, 2006 ; Mielke, 2015). Selon cette méthode, chaque contour de langue est estimé par une fonction *spline*, d'où le nom de la méthode. L'analyse statistique porte alors sur les paramètres des fonctions *spline* et non sur les points des contours de langue. L'étude de la divergence des paramètres des deux fonctions *spline* moyennes (si on compare deux groupes de données), via l'étude de la variance des données autour de cette fonction moyenne, indique si les deux groupes de données diffèrent de façon significative. Cette méthode permet également de construire (mathématiquement et visuellement) des intervalles de confiance autour de la moyenne (cf. figure 2.13 ci-dessous). Il est alors possible d'observer si des différences significatives entre deux groupes de données existent, en observant si les intervalles de confiance se recoupent ou non, et il est également possible de repérer visuellement le lieu où les différences significatives, si elles existent, se situent.

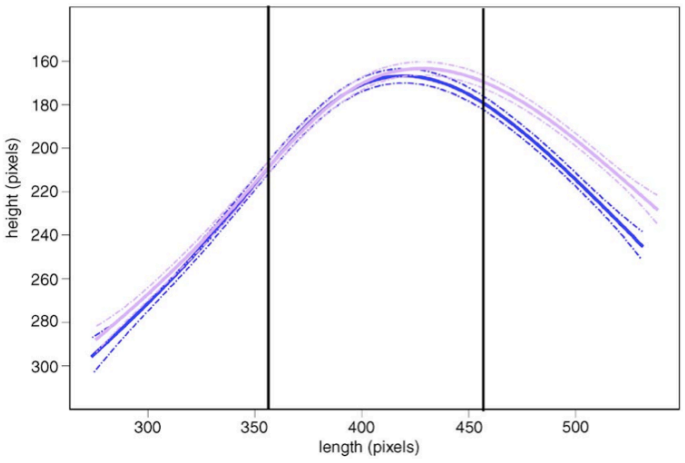


Figure 2.13. Illustration de la méthode SS ANOVA appliquée à des contours de langue extraits d'images échographiques, pour comparer deux groupes de données. Tiré de Davidson (2006).

2.4.2.1. Méthodes retenues pour l'analyse des contours linguaux

En ce qui concerne l'analyse des contours linguaux de notre étude, nous retiendrons deux méthodes complémentaires : les mesures quantitatives (millimétriques) de distance (que nous appliquerons pour mesurer la variabilité en répétition des voyelles isolées) et les analyses statistiques globales des SS ANOVA, effectuées sur le contour complet de la langue (que nous appliquerons afin d'observer les effets d'anticipation de V2 dans la réalisation de V1 et dans la réalisation de la consonne).

Pour les mesures de distance, nous retiendrons la méthode du « plus proche voisin » proposée par Zharkova et al. (2011). Cette méthode est utilisée dans le but d'offrir une mesure quantitative de la distance moyenne (distance Euclidienne) séparant deux groupes de contours de langue, dans deux contextes ou conditions différentes. Cette méthode nous servira également à mesurer la variabilité au sein d'un seul groupe de données, également représentée dans la méthode SS ANOVA par l'épaisseur des Intervalles de Confiance.

Pour ce qui est des analyses statistiques, nous retiendrons la méthode des SS ANOVA proposée par Gu (2002) (implémentée sous environnement R dans le package *gss*, Gu, 2012). Cette méthode permet une analyse statistique fondée sur la variance, comme l'ANOVA, permet de traiter toute l'information extraite des contours de langue, et permet de calculer des intervalles de confiance, rendant visible le recoupement des données, mais rendant également visible le lieu où les différences significatives existent entre les deux jeux de données.

Pour notre étude, nous ne retiendrons pas l'implémentation proposée par Davidson (2006), pour deux raisons principales. La première est que chez Davidson (2006), les données sont représentées dans des coordonnées cartésiennes. Cela pose problème pour comparer deux groupes de données issues d'images échographiques. En effet, les images échographiques sont recueillies originellement dans des coordonnées polaires. Il est donc plus naturel et plus juste de représenter et de comparer ces données entre elles dans des coordonnées polaires. En accord avec la proposition de Mielke (2015), reprise par Heyne & Derrick (2015), nos données seront donc transformées en coordonnées polaires avant toute analyse statistique. La seconde raison pour laquelle nous n'avons pas

adopté le code (sous environnement R) proposé par Davidson (2006), est que les intervalles de confiance qui sont construits le sont à partir de l'erreur-type et non pas de l'écart-type. Or les intervalles de confiance construits à partir de l'erreur-type, comme implémentés dans la version de base de Gu (2002), permettent de donner une indication sur la qualité de l'estimation de la moyenne, et ne renseignent pas directement sur la distribution des données autour de la moyenne (même si le calcul de l'écart-type à partir de l'erreur-type est possible). Les intervalles ainsi construits sont très fins et ne reflètent pas la variabilité des données.

Pour notre étude, nous nous appuierons sur le code (sous environnement R) proposé par Mielke (Mielke, 2015) qui permet d'effectuer des SS ANOVA en coordonnées polaires. Nos données étant enregistrées, lors de l'extraction des contours de langue, en coordonnées cartésiennes, il est nécessaire de les convertir en coordonnées polaires avant l'analyse statistique. L'origine est définie, afin de transformer les données cartésiennes en coordonnées polaires, comme la moyenne des X et le minimum des Y. Comme nos données sont déjà alignées dans un seul et même référentiel, cette méthode garantit que l'origine est proche de l'origine virtuelle de l'acquisition des images échographiques (Heyne & Derrick, 2015), et qu'elle permet une juste représentation et une juste comparaison des données.

Nous travaillerons donc sur des données converties en coordonnées polaires et appliquerons les analyses statistiques en coordonnées polaires, et construirons des intervalles de confiance à 95 %, correspondant à une valeur $p = 0.05$, à partir des valeurs de l'écart type, qui représente la variabilité des données. Puisqu'il n'existe pas, à notre connaissance, de méthode permettant de comparer les contours linguaux entre différents individus, présentant des morphologies linguales très différentes, nous comparerons les contours de langue sujet par sujet et contexte par contexte. Les figures sont préparées avec le package graphique *ggplot2*.

Afin d'illustrer cette méthode et de faciliter la lecture proposée par cette représentation statistique des contours de langue, nous proposons d'illustrer la méthode retenue à l'aide d'un exemple. Les figures 2.14. représentent les contours de langue bruts extraits pour le participant Enfant 8 dans le contexte V1 = /ɛ/ et C = /b/, au relâchement de la consonne

(panneau de gauche) et au milieu de V2 (panneau de droite), dans les contextes où V2 = /a/ (rouge) et V2 = /i/ (bleu).

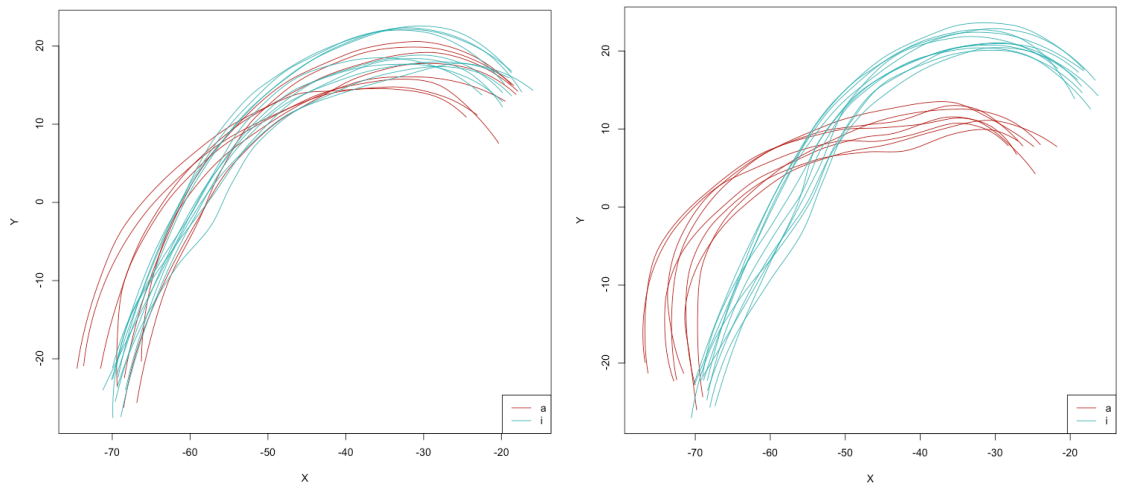


Figure 2.14. Exemples de données articulatoires brutes pour le participant Enfant 8 dans le contexte V1 = /ɛ/ et C = /b/, au relâchement de la consonne (panneau de gauche) et au milieu de V2 (panneau de droite), dans les contextes où V2 = /a/ (rouge) et V2 = /i/ (bleu).

Les figures 2.15. offrent une représentation statistique de ces données articulatoires, en représentant l'optimisation moyenne calculée par une fonction *spline* pour chacun de ces groupes de données (différenciées dans notre cas par le contexte de V2) ainsi que les intervalles de confiance à 95 % autour de ces moyennes. La méthode SS ANOVA est ici appliquée en coordonnées polaires (cf. la pointe de la langue dans les figures 2.15.).

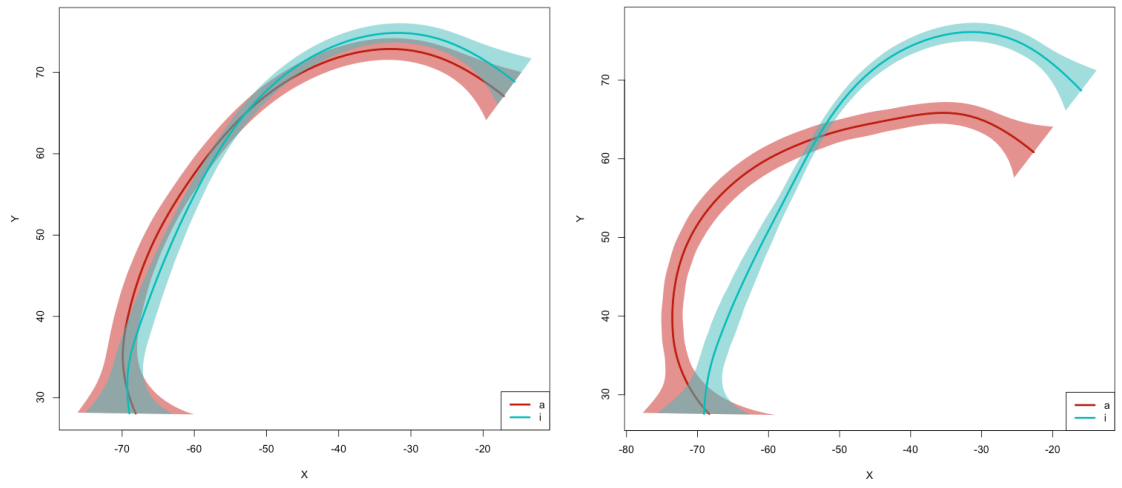


Figure 2.15. Représentations statistiques des données articulatoires pour le participant Enfant 8 dans le contexte V1 = /ɛ/ et C = /b/, au relâchement de la consonne (panneau de gauche) et au milieu de V2 (panneau de droite), dans les contextes où V2 = /a/ (rouge) et V2 = /i/ (bleu).

Dans le premier cas (figure 2.15., panneau de gauche, au relâchement de la consonne), les intervalles de confiance à 95 % autour des moyennes se recoupent, et ce tout le long des contours linguaux. Aucune différence significative entre les deux groupes de données n'est observée.

Dans le second cas (figure 2.15., panneau de droite, au milieu de V2), les intervalles de confiance à 95 % autour des moyennes ne se recoupent pas tout au long des contours linguaux, un espace est observé entre ces intervalles de confiance. Une différence significative est observée, au niveau de la pointe de la langue et du dos de la langue, car en ces lieux, les intervalles de confiance ne se recoupent pas.

Chapitre 3

Résultats

Dans ce chapitre, nous présentons les résultats de l'étude expérimentale dont nous avons décrit les outils et méthodes au chapitre 2. Pour chaque indice mesuré, nous présentons les résultats pour les adultes, puis pour les enfants, et enfin la confrontation des résultats en fonction des groupes d'âge. Nous commencerons par présenter les résultats portant sur la variabilité en répétition des voyelles en contexte isolé (section 3.1.). Ces résultats permettront de tester notre hypothèse 1, à savoir que l'on devrait observer davantage de variabilité liée à la répétition d'une même tâche dans les productions des enfants par rapport aux productions adultes. Nous présenterons ensuite les résultats portant sur la coarticulation intra-syllabique (section 3.2.). Ces résultats permettront de tester notre hypothèse 2, à savoir que nous devrions observer une coarticulation intra-syllabique d'amplitude similaire entre enfants et adultes. Nous présenterons ensuite les résultats concernant la coarticulation extra-syllabique (section 3.3.). Ces résultats permettront de tester notre hypothèse 3, à savoir que l'on devrait observer une moindre amplitude de la coarticulation extra-syllabique dans les productions enfantines par rapport aux productions adultes. Nous présenterons ensuite les résultats portant sur les durées des segments et des séquences V_1-C-V_2 (section 3.4.). Ces résultats permettront de tester notre hypothèse 4, à savoir que nous devrions observer des gestes plus lents chez les enfants que chez les adultes. Pour finir, nous présenterons les résultats concernant les relations entre nos principaux indices (section 3.5.), afin d'observer s'il existe une relation entre indices de stabilité et d'anticipation et d'observer s'il existe une relation entre indice d'anticipation et débit d'élocution.

3.1. Stabilité de la production des voyelles en contexte isolé

Les premiers résultats que nous présentons concernent la stabilité du contrôle des productions vocaliques, via l'étude la variabilité liée à la répétition d'une même tâche, ici la production de voyelles en contexte isolé. Ces résultats permettront de tester notre hypothèse 1, à savoir que l'on devrait observer davantage de variabilité liée à la répétition d'une même tâche dans les productions des enfants par rapport aux productions adultes.

3.1.1. Résultats individuels

3.1.1.1. Résultats individuels concernant les adultes

Nous présentons tout d'abord les résultats concernant la variabilité en répétition des voyelles isolées pour les 10 participants adultes. La figure 3.1 ci-dessous représente les 8 à 10 répétitions des 5 voyelles en contexte isolé /i e ɛ a u/ dans l'espace acoustique F1 - F2 et les ellipses de dispersion (à 95 %) représentent la variabilité liée à cette répétition. On observe que les productions des participants adultes sont relativement stables d'une répétition à l'autre, mais qu'il existe une certaine variabilité interindividuelle. De plus, toutes les catégories vocaliques ne montrent pas la même variabilité chez tous les participants. Par exemple, le participant Adulte 2 montre une grande variabilité en répétition pour le /ɛ/ et le participant Adulte 9 montre une grande variabilité pour le /e/. Le participant Adulte 10 montre une grande consistance des productions, pour chacune des catégories vocaliques.

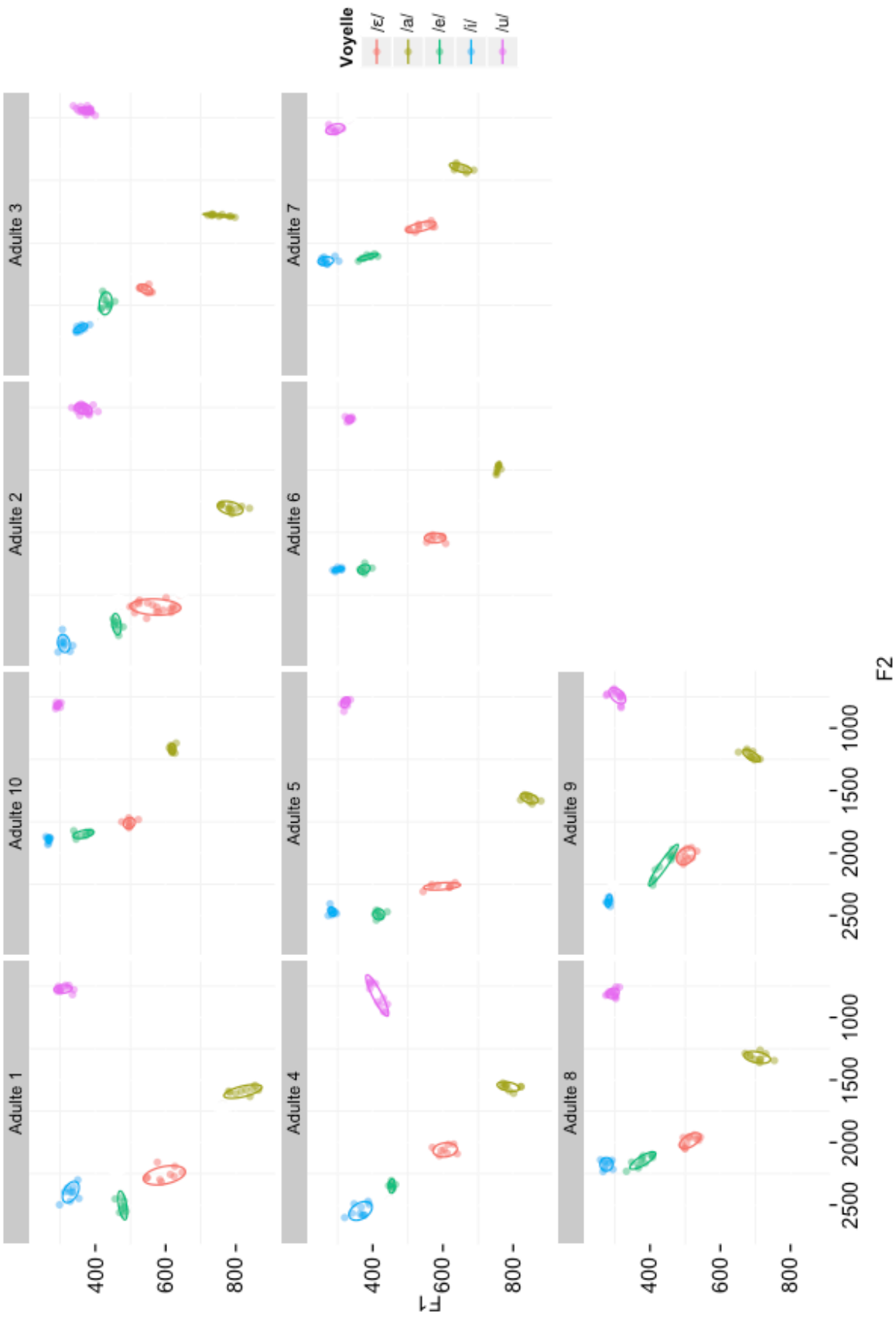


Figure 3.1. Stabilité des productions vocaliques chez les adultes dans l'espace F1 - F2 en Hertz.

Afin de quantifier la stabilité des productions, via la mesure de la variabilité en répétition, nous proposons de mesurer pour chaque participant adulte et chaque catégorie vocalique, l'écart type en F1 et en F2, dans l'espace acoustique normalisé, afin de pouvoir effectuer une comparaison avec les mêmes mesures chez les enfants.

Tableau 3.1. Mesures des écarts type en F1 et en F2 pour chaque participant adulte et chaque catégorie vocalique séparément et indice individuel de stabilité du contrôle.

	Stabilité des voyelles isolées (écarts type des valeurs Z-scorées)										
	/i/		**/e/**		**/ɛ/**		**/a/**		**/u/**		**Indice individuel de stabilité**
Participant	**F1**	**F2**	**F1**	**F2**	**F1**	**F2**	**F1**	**F2**	**F1**	**F2**	
Adulte 1	0,086	0,087	0,050	0,141	0,184	0,076	0,183	0,062	0,087	0,033	**0,099**
Adulte 2	0,066	0,074	0,046	0,092	0,144	0,018	0,133	0,142	0,067	0,022	**0,080**
Adulte 3	0,087	0,031	0,086	0,085	0,091	0,034	0,192	0,012	0,105	0,026	**0,075**
Adulte 4	0,136	0,071	0,037	0,066	0,143	0,052	0,128	0,038	0,122	0,146	**0,094**
Adulte 5	0,034	0,042	0,052	0,041	0,155	0,032	0,079	0,038	0,036	0,047	**0,056**
Adulte 6	0,054	0,024	0,064	0,054	0,112	0,053	0,026	0,064	0,042	0,035	**0,053**
Adulte 7	0,109	0,054	0,125	0,044	0,179	0,066	0,135	0,060	0,130	0,079	**0,098**
Adulte 8	0,074	0,061	0,147	0,078	0,115	0,068	0,151	0,056	0,071	0,051	**0,087**
Adulte 9	0,043	0,073	0,169	0,159	0,117	0,069	0,116	0,049	0,106	0,070	**0,097**
Adulte 10	0,027	0,048	0,138	0,052	0,095	0,062	0,048	0,065	0,043	0,042	**0,062**
Moyenne	**0,072**	**0,056**	**0,092**	**0,081**	**0,133**	**0,053**	**0,119**	**0,058**	**0,081**	**0,055**	**0,080**

Le tableau 3.1. rapporte les mesures de dispersion (variabilité en F1 et en F2 mesurés par les écarts type dans l'espace F1 - F2 normalisé) pour chaque participant adulte et chaque catégorie vocalique. Un indice individuel de stabilité du contrôle (la moyenne des dispersions pour chaque catégorie vocalique pour un participant donné) est calculé à partir de ces valeurs. L'indice de stabilité moyen des participants adultes est de **0.08**.

La stabilité de la production de ces voyelles en contexte isolé a également été étudiée dans le domaine articulatoire, pour 3 participants adultes. Les données articulatoires concernant les productions de ces voyelles isolées sont présentées dans les figures 3.2, 3.3 et 3.4 ci-dessous. Par convention, la pointe de la langue est toujours à droite.

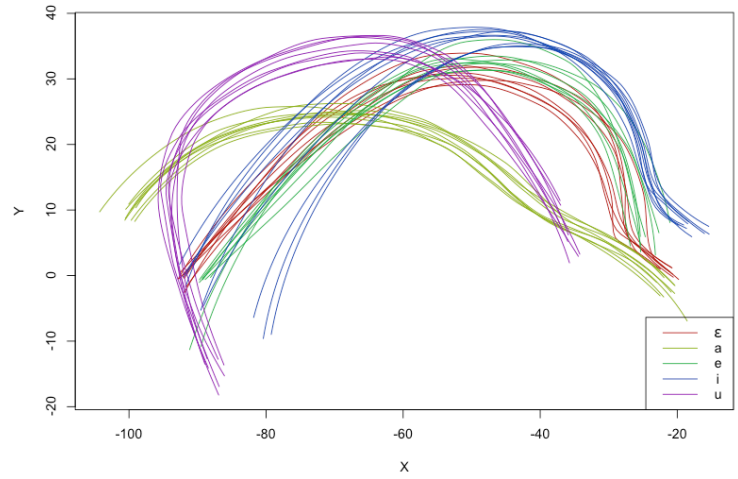


Figure 3.2. Données articulatoires des voyelles isolées pour le participant Adulte 5, axes en mm.

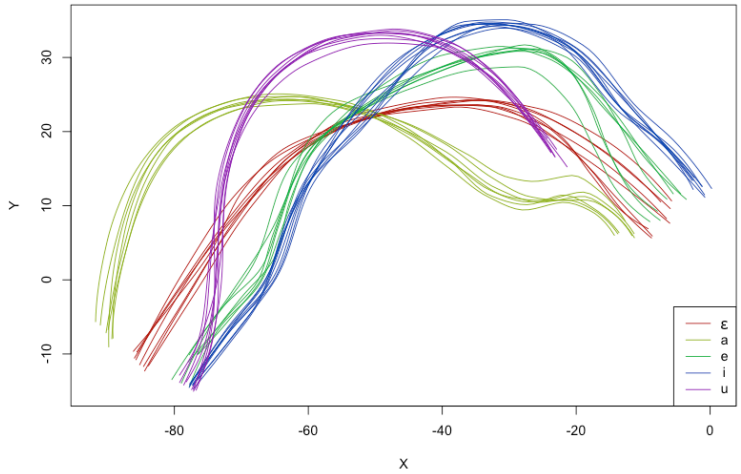


Figure 3.3. Données articulatoires des voyelles isolées pour le participant Adulte 6, axes en mm.

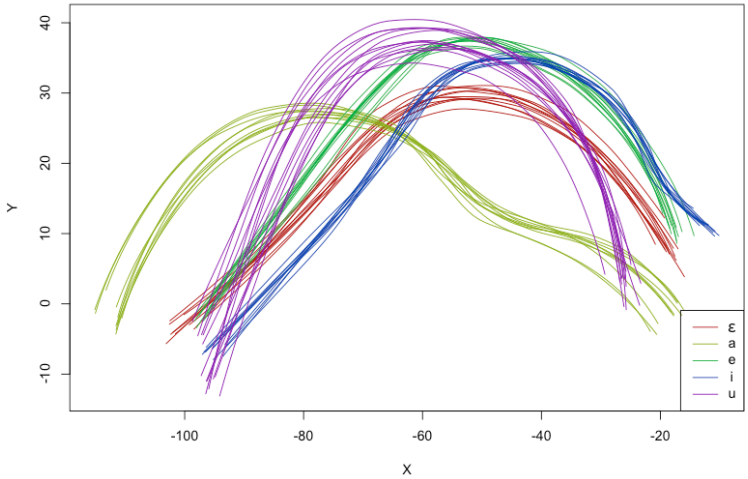


Figure 3.4. Données articulatoires des voyelles isolées pour le participant Adulte 8, axes en mm.

Afin de quantifier, dans le domaine articulatoire, la stabilité des productions, nous proposons de mesurer pour chaque participant et chaque catégorie vocalique, la variabilité intra-catégorielle, mesurée par la distance moyenne entre chaque contour de langue au sein d'une même catégorie vocalique.

Tableau 3.2. Mesures de stabilité dans le domaine articulatoire (distance moyenne en mm entre chaque contour de langue au sein de chaque catégorie vocalique et pour chaque participant adulte).

	Distance moyenne au sein de chaque catégorie vocalique (mm)					
Participant	**/i/**	**/e/**	**/ɛ/**	**/a/**	**/u/**	**Moyenne**
Adulte 5	3,9	3,6	1,9	1,9	3,6	3,0
Adulte 6	1,3	2,7	1,5	1,7	1,5	1,7
Adulte 8	1,7	2,4	2,8	2,3	3,6	2,6
Moyenne	2,3	2,7	2,1	2,0	2,9	**2,4**

Le tableau 3.2. rapporte les distances moyennes entre les contours de langue (en mm) au sein de chaque catégorie vocalique et pour chaque participant adulte. La variabilité moyenne de ces mesures articulatoires de stabilité pour ces participants adultes est de **2.4 mm**.

3.1.1.2. Résultats individuels concernant les enfants

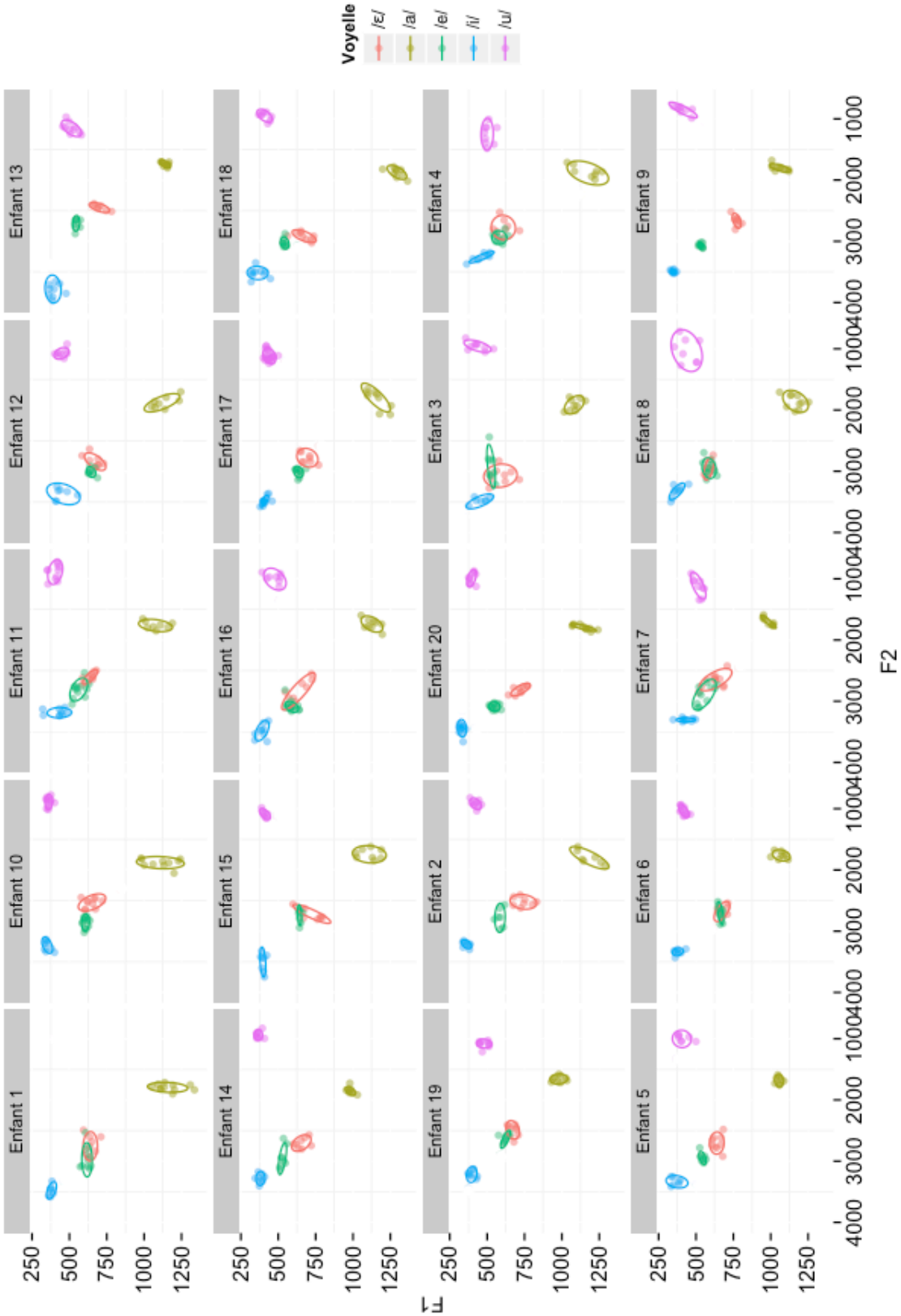


Figure 3.5. Stabilité des productions vocaliques chez les enfants dans l'espace F1 - F2 en Hertz.

On observe visuellement une plus grande variabilité en répétition chez les enfants (figure 3.5 ci-dessus) par rapport aux adultes (figure 3.1.). On observe également une grande variabilité interindividuelle, un moins grand contraste /e/ ~ /ɛ/ et l'on observe que la variabilité liée à la répétition varie considérablement d'une catégorie vocalique à une autre et d'un participant à un autre.

Afin de quantifier la stabilité des productions, via la mesure de la variabilité en répétition, nous proposons de mesurer pour chaque participant enfant et chaque catégorie vocalique, l'écart type en F1 et en F2, dans l'espace acoustique normalisé.

Tableau 3.3. Mesures des écarts type en F1 et en F2 pour chaque participant enfant et chaque catégorie vocalique séparément et indice individuel de stabilité du contrôle.

	Stabilité des voyelles isolées (écarts-type des valeurs Z-scorées)										
	/i/		**/e/**		**/ɛ/**		**/a/**		**/u/**		**Indice individuel de stabilité**
Participant	**F1**	**F2**	**F1**	**F2**	**F1**	**F2**	**F1**	**F2**	**F1**	**F2**	
Enfant 1	0,099	0,152	0,075	0,303	0,112	0,240	0,290	0,083	-	-	**0,169**
Enfant 2	0,070	0,055	0,078	0,173	0,188	0,086	0,258	0,129	0,083	0,073	**0,119**
Enfant 3	0,240	0,105	0,068	0,255	0,274	0,134	0,168	0,099	0,239	0,071	**0,165**
Enfant 4	0,204	0,080	0,124	0,097	0,205	0,170	0,319	0,203	0,112	0,211	**0,172**
Enfant 5	0,227	0,073	0,062	0,087	0,122	0,147	0,076	0,070	0,182	0,146	**0,119**
Enfant 6	0,109	0,054	0,053	0,142	0,148	0,140	0,163	0,069	0,092	0,082	**0,105**
Enfant 7	0,226	0,028	0,252	0,191	0,388	0,159	0,162	0,082	0,160	0,181	**0,183**
Enfant 8	0,128	0,115	0,103	0,140	0,075	0,143	0,194	0,133	0,220	0,246	**0,150**
Enfant 9	0,047	0,026	0,044	0,037	0,077	0,091	0,165	0,055	0,242	0,093	**0,088**
Enfant 10	0,085	0,098	0,058	0,104	0,266	0,120	0,364	0,088	0,061	0,093	**0,134**
Enfant 11	0,246	0,063	0,163	0,156	0,138	0,088	0,300	0,075	0,134	0,145	**0,151**
Enfant 12	0,294	0,131	0,084	0,071	0,179	0,112	0,293	0,110	0,120	0,080	**0,147**
Enfant 13	0,143	0,170	0,058	0,088	0,174	0,060	0,080	0,040	0,153	0,095	**0,106**
Enfant 14	0,084	0,083	0,074	0,203	0,177	0,094	0,097	0,067	0,085	0,069	**0,103**
Enfant 15	0,050	0,172	0,036	0,124	0,281	0,108	0,252	0,096	0,119	0,065	**0,130**
Enfant 16	0,126	0,127	0,120	0,095	0,276	0,199	0,192	0,097	0,184	0,155	**0,157**
Enfant 17	0,084	0,067	0,083	0,066	0,186	0,109	0,237	0,162	0,130	0,086	**0,121**
Enfant 18	0,131	0,089	0,085	0,075	0,162	0,068	0,136	0,081	0,089	0,073	**0,099**
Enfant 19	0,111	0,122	0,121	0,141	0,128	0,127	0,175	0,064	0,143	0,071	**0,120**

Enfant 20	0,090	0,129	0,082	0,052	0,140	0,071	0,185	0,046	0,065	0,100	**0,096**
Moyenne	**0,140**	**0,097**	**0,091**	**0,130**	**0,185**	**0,123**	**0,205**	**0,092**	**0,137**	**0,112**	**0,132**

Le tableau 3.3. rapporte les mesures de dispersion (variabilité en F1 et en F2 mesurés par les écarts type dans l'espace F1 - F2 normalisé) pour chaque participant enfant et chaque catégorie vocalique. Un indice individuel de stabilité du contrôle (la moyenne des dispersions pour chaque catégorie vocalique pour un participant donné) est calculé à partir de ces valeurs. L'indice de stabilité moyen des participants enfants est de **0.13**.

La stabilité de la production de ces voyelles en contexte isolé a également été étudiée dans le domaine articulatoire, pour 5 participants enfants. Les données articulatoires concernant ces productions de voyelles isolées sont présentées dans les figures 3.6 à 3.10. ci-dessous. Par convention, la pointe de la langue est toujours à droite.

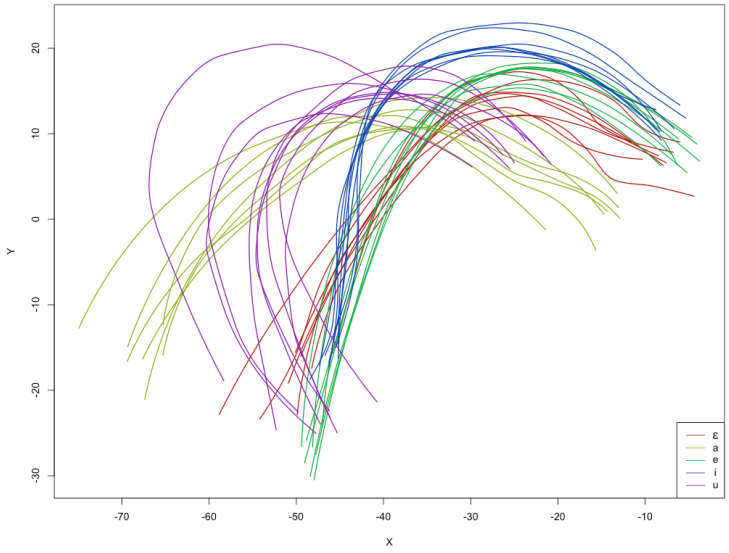


Figure 3.6. Données articulatoires des voyelles isolées pour le participant Enfant 2, axes en mm.

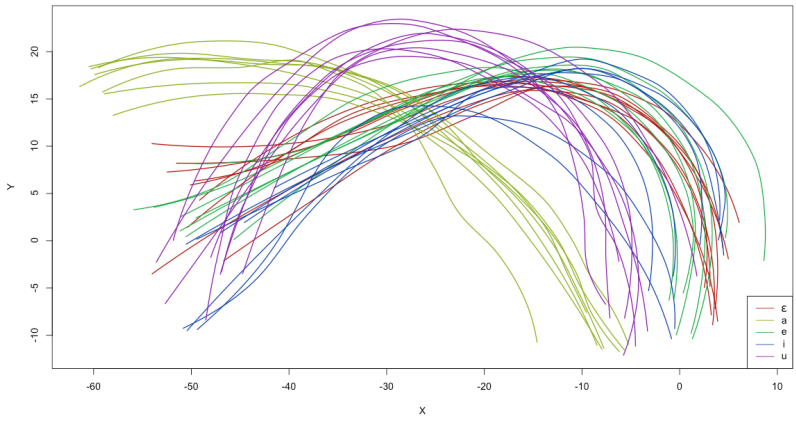


Figure 3.7. Données articulatoires des voyelles isolées pour le participant Enfant 5, axes en mm.

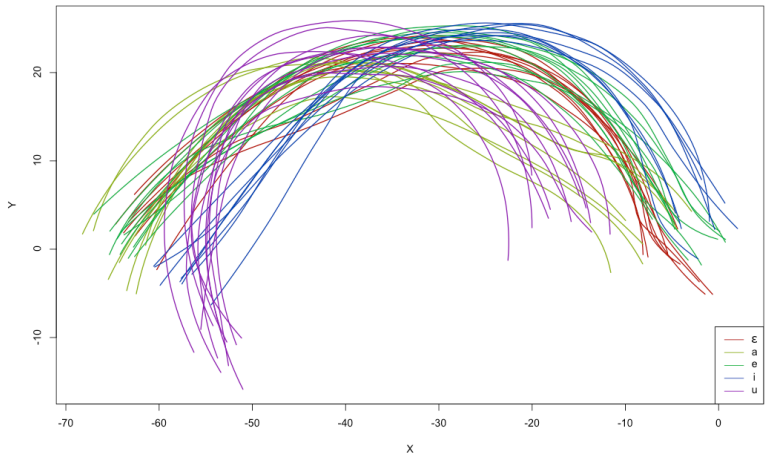


Figure 3.8. Données articulatoires des voyelles isolées pour le participant Enfant 6, axes en mm.

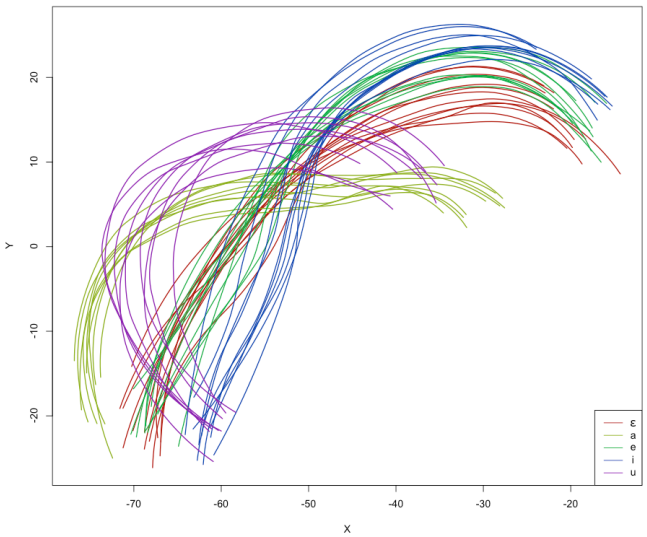


Figure 3.9. Données articulatoires des voyelles isolées pour le participant Enfant 7, axes en mm.

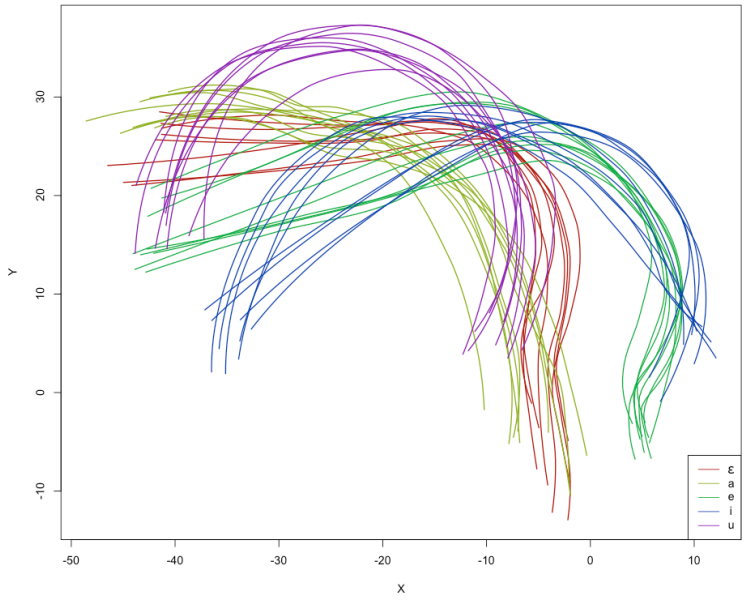


Figure 3.10. Données articulatoires des voyelles isolées pour le participant Enfant 9, axes en mm.

Afin de quantifier, dans le domaine articulatoire, la stabilité des productions, nous proposons de mesurer pour chaque participant et chaque catégorie vocalique, la variabilité intra-catégorielle, mesurée par la distance moyenne entre chaque contour de langue au sein d'une même catégorie vocalique.

Tableau 3.4. Mesures de stabilité dans le domaine articulatoire (distance moyenne en mm entre chaque contour de langue au sein de chaque catégorie vocalique et pour chaque enfant).

	Distance moyenne au sein de chaque catégorie vocalique (mm)					
Participant	**/i/**	**/e/**	**/ɛ/**	**/a/**	**/u/**	**Moyenne**
Enfant 2	2,2	2,6	4,6	4,2	6,7	4,1
Enfant 5	6,3	4,3	3,7	2,7	4,7	4,4
Enfant 6	2,6	3,3	3,0	3,4	4,3	3,3
Enfant 7	3,7	3,0	3,6	3,8	4,8	3,8
Enfant 9	4,7	3,3	3,9	4,2	3,2	3,9
Moyenne	3,9	3,3	3,8	3,7	4,7	**3,9**

Le tableau 3.4. rapporte les distances moyennes entre les contours de langue (en mm) au sein de chaque catégorie vocalique et pour chaque participant enfant. La variabilité moyenne de ces mesures articulatoires de stabilité pour ces participants enfants est de **3.9 mm.**

3.1.2. Résultats par groupe d'âge : comparaison entre enfants et adultes

Afin de comparer les valeurs de dispersion (donc de stabilité) entre enfants et adultes, nous présentons graphiquement les indices numériques individuels de stabilité (calculés dans le domaine acoustique, tableaux 3.1. et 3.3.) en fonction de l'âge. Nous présentons ensuite les productions vocaliques de tous les enfants regroupés et de tous les adultes regroupés, dans l'espace acoustique normalisé, afin de permettre une comparaison visuelle de la stabilité des productions enfantines et adultes.

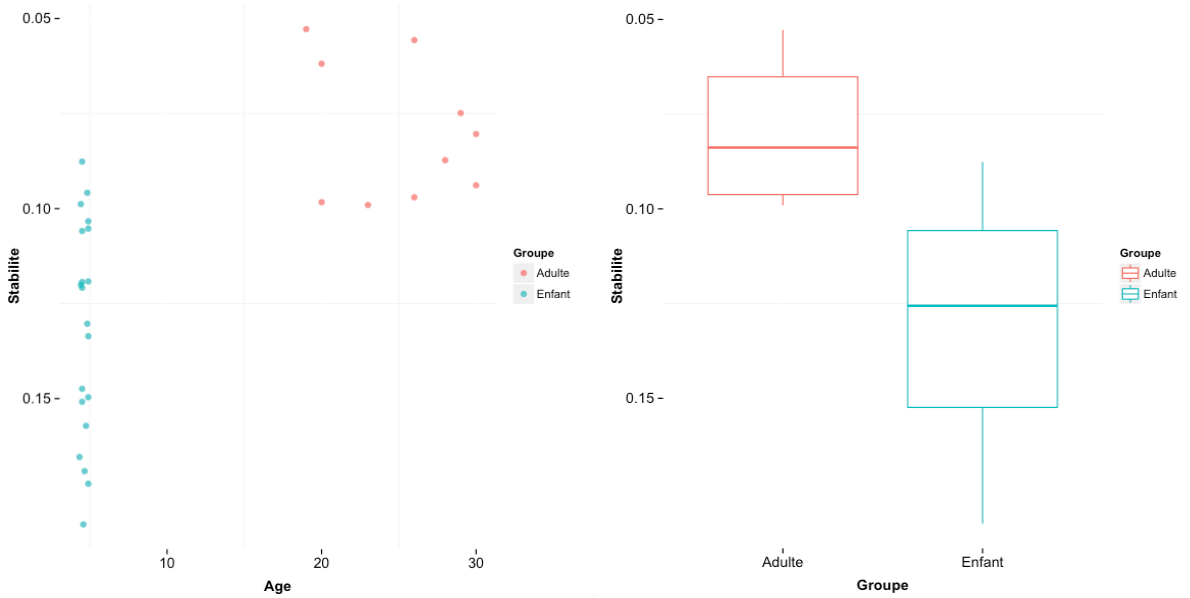


Figure 3.11. Indices numériques individuels de stabilité (calculés dans le domaine acoustique) en fonction de l'âge, pour chacun des participants, groupe Adulte en rouge, groupe Enfant en bleu.

La figure 3.11. ci-dessus représente les indices numériques individuels de stabilité pour chacun des participants. On observe que les adultes montrent globalement une grande stabilité et qu'il existe une grande variabilité interindividuelle chez les participants enfants. Seuls 6 participants enfants montrent une stabilité en répétition d'un ordre de grandeur semblable aux scores adultes les plus faibles. Ces indices numériques individuels de stabilité, présentés par groupe d'âge (figure 3.11., panneau de droite), permettent une comparaison de la mesure de stabilité entre ces deux groupes d'âge. Une analyse statistique (modèle linéaire, prenant comme variable dépendante les indices individuels et comme facteur le groupe d'âge) révèle que la différence entre le groupe Enfant et le groupe Adulte est significative $F(1, 28) = 27{,}09$; $p < 1.59\ e^{-05}$.

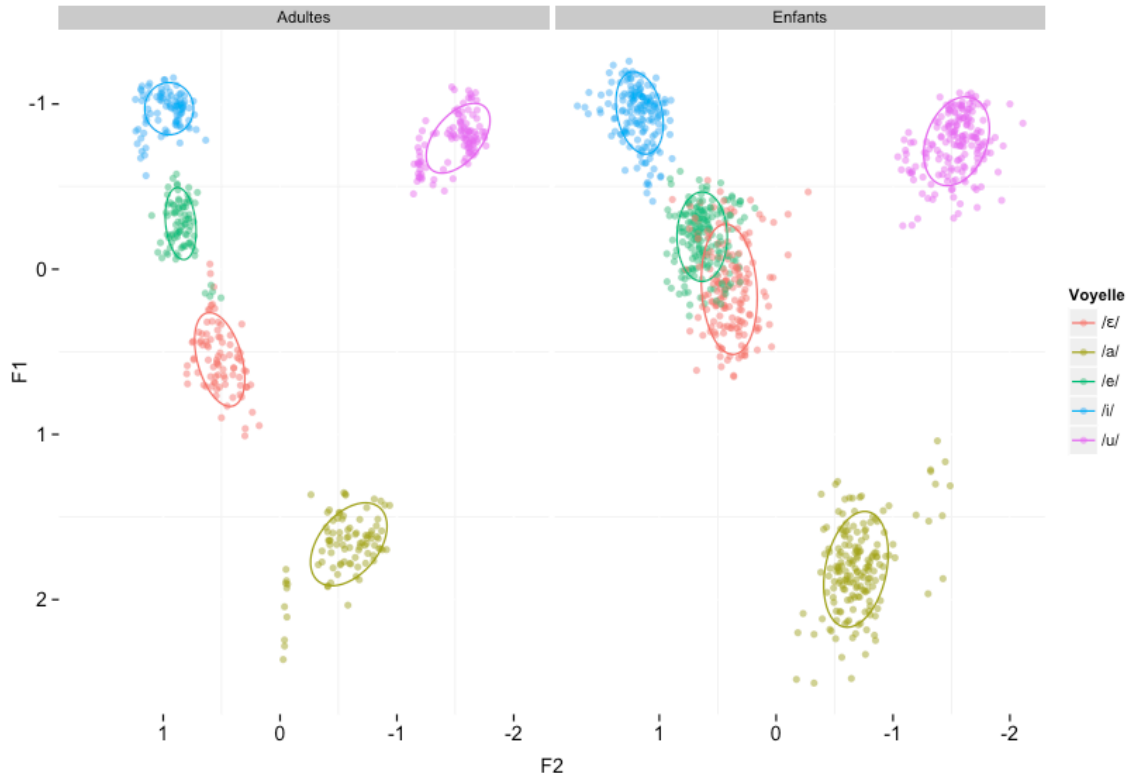


Figure 3.12. Données acoustiques, dans l'espace F1 - F2 normalisé, pour tous les participants adultes (panneau de gauche) et tous les participants enfants (panneau de droite).

La figure 3.12. ci-dessus représente les productions acoustiques des voyelles isolées, pour tous les participants adultes regroupés (panneau de gauche) et tous les participants enfants regroupés (panneau de droite), dans l'espace F1 - F2 normalisé.

L'indice moyen de stabilité (calculé à partir des indices individuels) dans le domaine acoustique pour les adultes est de **0,08** ; pour les enfants, il est de **0,13**. Le rapport entre ces deux indices est de **1.64**. Cela indique qu'en moyenne, on observe **64** % de variabilité supplémentaire dans les productions des enfants par rapport à celles des adultes.

L'indice moyen de stabilité (calculé à partir des indices individuels) dans le domaine articulatoire pour les adultes est de **2,4 mm** ; pour les enfants, il est de **3,9 mm.** Le rapport entre ces deux indices est de **1.63**. Cela indique qu'en moyenne, on observe **63** % de variabilité supplémentaire dans les productions des enfants par rapport à celles des adultes.

On observe donc effectivement davantage de variabilité liée à la répétition d'une même tâche dans les productions des enfants par rapport à celles des adultes.

3.2. Coarticulation intra-syllabique

Nous présentons maintenant les résultats concernant la coarticulation intra-syllabique. Ces résultats permettront de tester notre hypothèse 2 (nous devrions observer une coarticulation intra-syllabique d'amplitude similaire entre enfants et adultes).

3.2.1. Mesures acoustiques : équations de locus

Un indice acoustique de l'amplitude de la coarticulation intra-syllabique, la pente des équations de locus, a été utilisé dans le but de comparer l'amplitude de la coarticulation chez les enfants et chez les adultes. Les résultats sont présentés dans les tableaux 3.5 concernant les adultes et 3.6 concernant les enfants.

Tableau 3.5. Pentes des équations de locus pour les participants adultes.

Adultes	**/b/**	**/d/**	**/g/**
Adulte 1	0,82	0,66	0,72
Adulte 2	0,94	0,63	0,99
Adulte 3	0,99	0,51	0,86
Adulte 4	0,91	0,65	0,92
Adulte 5	0,97	0,68	0,73
Adulte 6	0,96	0,66	0,89
Adulte 7	0,93	0,91	0,93
Adulte 8	1,04	0,97	0,85
Adulte 9	0,87	0,72	0,89
Adulte 10	0,99	0,82	0,86
Moyenne	**0,94**	**0,72**	**0,86**

Tableau 3.6. Pentes des équations de locus pour les participants enfants.

Enfants	**/b/**	**/d/**	**/g/**
Enfant 1	1,08	0,80	0,94
Enfant 2	0,99	0,97	0,87
Enfant 3	0,95	0,98	0,95
Enfant 4	0,95	0,86	0,90
Enfant 5	0,97	0,70	0,83

Enfant 6	0,97	0,94	0,97
Enfant 7	1,02	0,78	0,94
Enfant 8	1,00	0,90	0,91
Enfant 9	1,02	0,87	0,87
Enfant 10	0,94	0,81	0,94
Enfant 11	0,96	0,87	0,85
Enfant 12	0,96	0,83	0,94
Enfant 13	0,96	0,92	0,88
Enfant 14	0,98	0,62	0,75
Enfant 15	0,92	0,90	0,95
Enfant 16	0,87	0,73	0,93
Enfant 17	0,92	0,81	1,03
Enfant 18	0,94	0,85	0,93
Enfant 19	1,05	0,56	0,85
Enfant 20	0,93	0,86	0,77
Moyenne	**0,97**	**0,83**	**0,90**

Les tableaux 3.5 et 3.6 rapportent les pentes des équations de locus pour les participants adultes et enfants. Ces résultats indiquent que les pentes des équations de locus pour les adultes sont de **0.94 pour /b/** ; **0,72 pour /d/** et **0.86 pour /g/.** Les pentes des équations de locus pour les enfants sont de **0.97 pour /b/** ; **0.83 pour /d/** et **0.90 pour /g/.** La hiérarchie dans les valeurs des pentes des équations de locus est la même chez les enfants et les adultes : bilabiales > vélaires > alvéolaires. Cette hiérarchie est conforme à ce qui est habituellement décrit dans la littérature.

Des analyses statistiques (ANOVA à mesures répétées prenant les consonnes /b/ /d/ /g/ comme variables intra sujet et prenant le facteur groupe comme variable inter sujet) révèlent un effet de la consonne ($F(2,56) = 32.419$, $p < 0.00001$), pas d'effet du groupe ($F(1,28) = 7.2080$, $p < 0.01206$) et, de manière plus importante, pas d'effet d'interaction entre consonne et groupe ($F(2,56) = 1.88$, $p < 0.162$).

Des comparaisons par paires (modèles linéaires, prenant les valeurs de pentes des équations de locus des consonnes /b/, /d/ et /g/ comme variables dépendantes et prenant le groupe d'âge comme facteur fixe) révèlent que les différences entre enfants et adultes

ne sont pas significatives : F (1,28) = 1.154 et $p < 0.292$ pour /b/ ; F (1,28) = 4.892 et $p < 0.0353$ pour /d/ ; F (1,28) = 1.791 et $p < 0.192$ pour /g/. Ces résultats indiquent qu'il ne semble pas exister de différence significative dans l'amplitude de la coarticulation intra-syllabique entre enfants et adultes.

Afin de permettre la comparaison entre mesures acoustiques (équations de locus) et mesures articulatoires de la coarticulation intra-syllabique, nous avons calculé les pentes des équations de locus pour les participants enfants en séparant les contextes V1 = /ɛ/ et V1 = /a/, puisque les mesures articulatoires ne peuvent être effectuées que contexte par contexte. Les tableaux 3.7. et 3.8. rapportent respectivement les pentes des équations de locus pour les enfants dans les contextes V1 = /ɛ/ et V1 = /a/.

Tableau 3.7. Pentes des équations de locus pour les participants enfants dans le contexte V1 = /ɛ/.

Pentes des équations de locus pour V1 = /ɛ/			
Enfants	**/b/**	**/d/**	**/g/**
Enfant 1	1,04	1,04	0,85
Enfant 2	0,84	0,74	0,79
Enfant 3	0,98	0,32	0,69
Enfant 4	0,74	1,26	0,91
Enfant 5	0,90	0,78	1,17
Enfant 6	0,81	0,29	0,31
Enfant 7	0,98	0,65	0,90
Enfant 8	0,86	0,50	0,99
Enfant 9	1,08	0,68	0,90
Enfant 10	0,92	0,69	0,52
Enfant 11	0,95	0,62	0,71
Enfant 12	1,06	0,43	1,04
Enfant 13	0,89	0,87	0,92
Enfant 14	0,81	0,42	0,49
Enfant 15	0,45	0,53	0,96
Enfant 16	0,77	0,47	0,94
Enfant 17	0,82	0,55	0,95
Enfant 18	0,85	0,66	0,60
Enfant 19	1,02	0,27	1,27

Enfant 20	0,77	0,47	1,14

Tableau 3.8. Pentes des équations de locus pour les participants enfants dans le contexte V1 = /a/.

Pentes des équations de locus pour V1 = /a/			
Enfants	**/b/**	**/d/**	**/g/**
Enfant 1	0,99	0,62	0,59
Enfant 2	0,96	0,37	0,81
Enfant 3	0,37	0,98	1,03
Enfant 4	0,78	1,05	1,03
Enfant 5	0,94	0,16	1,29
Enfant 6	0,92	0,78	1,04
Enfant 7	1,11	0,20	0,50
Enfant 8	0,78	0,32	1,05
Enfant 9	1,11	0,64	0,77
Enfant 10	0,84	0,70	1,04
Enfant 11	0,62	0,69	1,14
Enfant 12	1,13	0,29	0,94
Enfant 13	0,97	0,86	0,98
Enfant 14	1,04	0,56	1,12
Enfant 15	0,33	0,59	0,82
Enfant 16	0,81	1,07	0,86
Enfant 17	1,08	0,95	0,92
Enfant 18	0,96	0,62	1,12
Enfant 19	1,09	0,36	0,94
Enfant 20	0,91	0,62	1,38

Les tableaux 3.7. et 3.8. permettent de mettre en lumière le fait que l'amplitude de la coarticulation intra-syllabique peut dépendre du contexte de V1 : les mesures entre les deux contextes de V1 ne sont pas toujours équivalentes chez tous les participants (eg. le participant Enfant 1). Une analyse statistique (ANOVA à mesures répétées prenant les consonnes comme facteur intra sujet et V1 comme variable inter sujet) révèle un effet de la consonne ($F (2,38) = 15.11$, $p < 0.0001$), pas d'effet de V1 ($F (1,18) = 1.24$, $p < 0.279$) et pas d'effet d'interaction entre consonne et V1 ($F (2,38) = 0.92$, $p < 0.405$).

Dans l'ensemble, la coarticulation dans la syllabe CV chez les participants enfants est de même amplitude dans les contextes V1 = /a/ et V1 = /ɛ/. Cependant, les résultats individuels suggèrent que ce n'est pas toujours le cas : certains participants n'anticipent pas de la même façon la voyelle V2 dans la réalisation de C en fonction de V1. Nous n'entrerons pas dans la lecture détaillée de ces résultats, qui sont avant tout destinés à confronter les mesures des pentes des équations de locus aux données articulatoires (cf. section 4.2. de la discussion).

Afin d'illustrer la coarticulation intra-syllabique chez l'enfant d'un point de vue articulatoire, nous présentons les données articulatoires de 6 participants enfants concernant les mesures de contour linguaux effectuées au relâchement de la consonne et au milieu de V2 (figures 3.13. à 3.48.). Les données concernant les sujets 6 et 10 ne sont pas orientées par rapport au plan occlusal réel, car les participants n'ont pas mordu suffisamment fort le triangle occlusal permettant le réalignement des données par rapport à ce plan, il faut donc les faire pivoter mentalement pour mieux les lire. Nous n'avons pas effectué de mesures articulatoires de la coarticulation intra-syllabique pour les participants adultes. Les données sont toujours présentées dans l'ordre qui suit : /ɛba/ versus /ɛbi/, /aba/ versus /abi/, /ɛda/ versus /ɛdi/, /ada/ versus /adi/, /ɛga/ versus /ɛgi/, /aga/ versus /agi/.

Les différences entre les configurations articulatoires sont évaluées sur la base du traitement statistique des données par la méthode SS-ANOVA, à partir de l'observation des intervalles de confiance à 95 % autour des tracés linguaux moyens (cf. Chapitre 2, section 2.4.2.1.). Si les intervalles de confiance se recoupent sur toute la longueur des contours linguaux, aucune différence statistiquement significative n'est observée. Si les intervalles de confiance ne se recoupent pas en au moins un lieu, le long des contours linguaux, alors une différence statistiquement significative est observée en ce lieu précis où les intervalles de confiance ne se recoupent pas.

Enfant 1

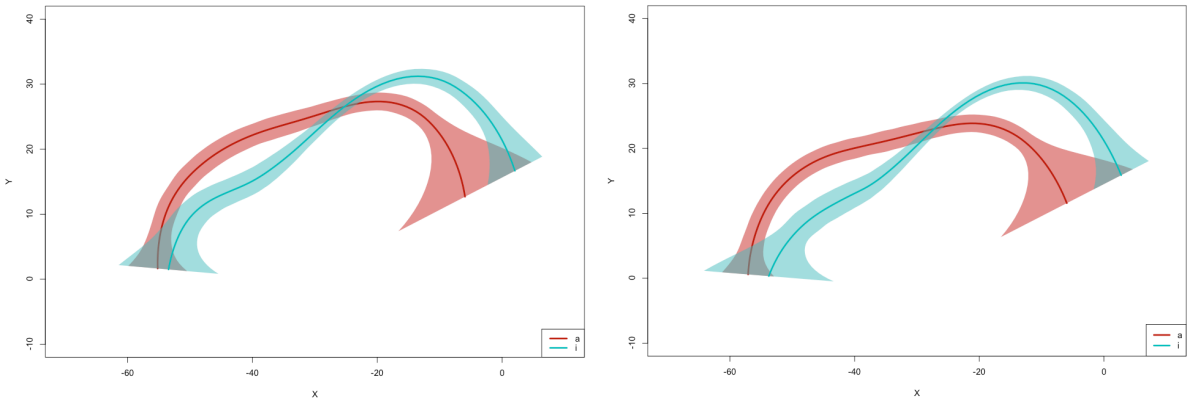


Figure 3.13. Représentation statistique des contours linguaux pour le participant Enfant 1 au relâchement de la consonne (panneau gauche) et au milieu de V2 (panneau droit) pour C = /b/, V1 = /ɛ/, en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

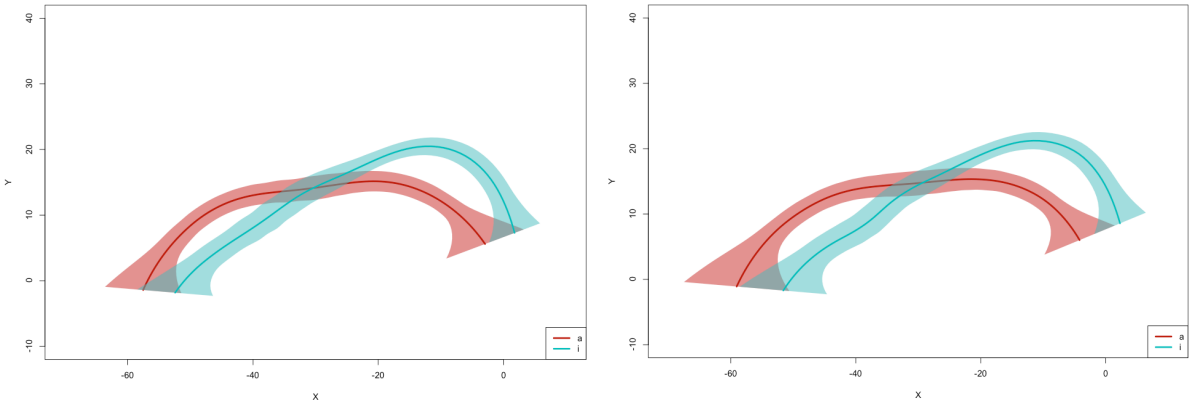


Figure 3.14. Représentation statistique des contours linguaux pour le participant Enfant 1 au relâchement de la consonne (panneau gauche) et au milieu de V2 (panneau droit) pour C = /b/, V1 = /a/, en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

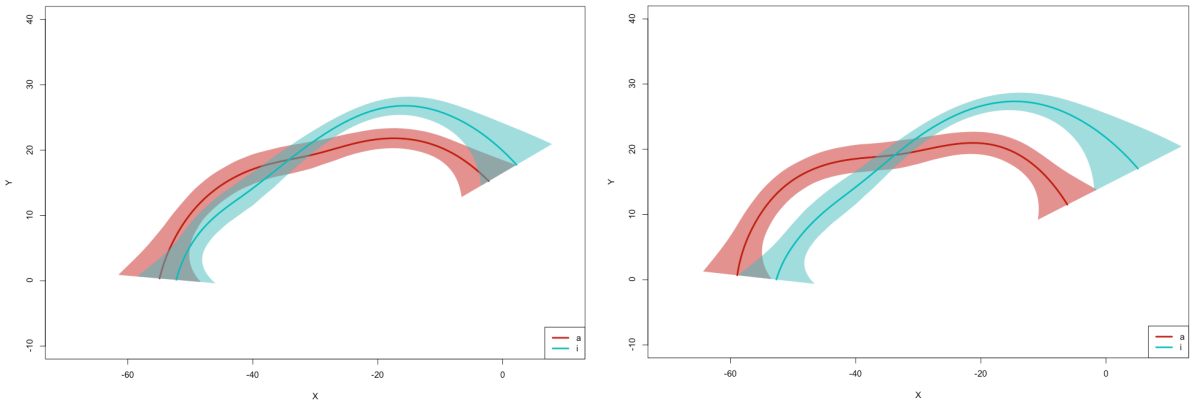


Figure 3.15. Représentation statistique des contours linguaux pour le participant Enfant 1 au relâchement de la consonne (panneau gauche) et au milieu de V2 (panneau droit) pour C = /d/, V1 = /ɛ/, en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

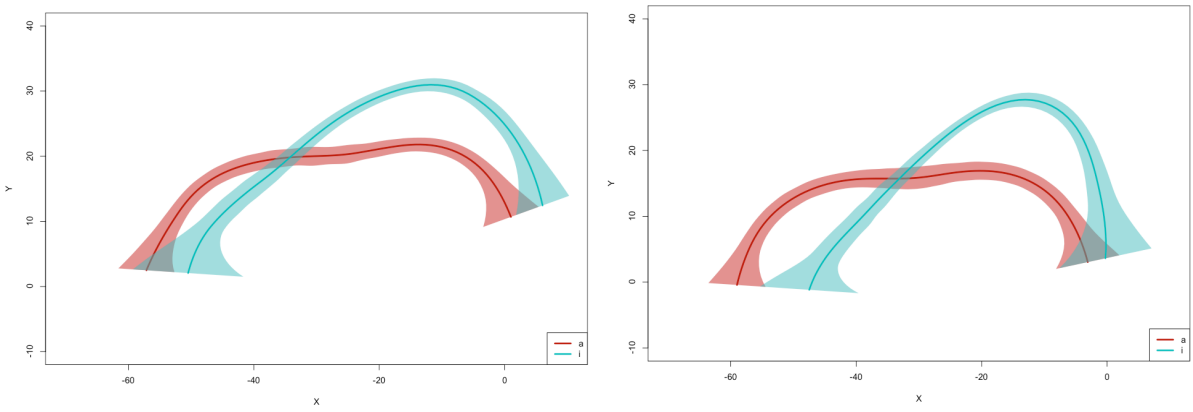


Figure 3.16. Représentation statistique des contours linguaux pour le participant Enfant 1 au relâchement de la consonne (panneau gauche) et au milieu de V2 (panneau droit) pour C = /d/, V1 = /a/, en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

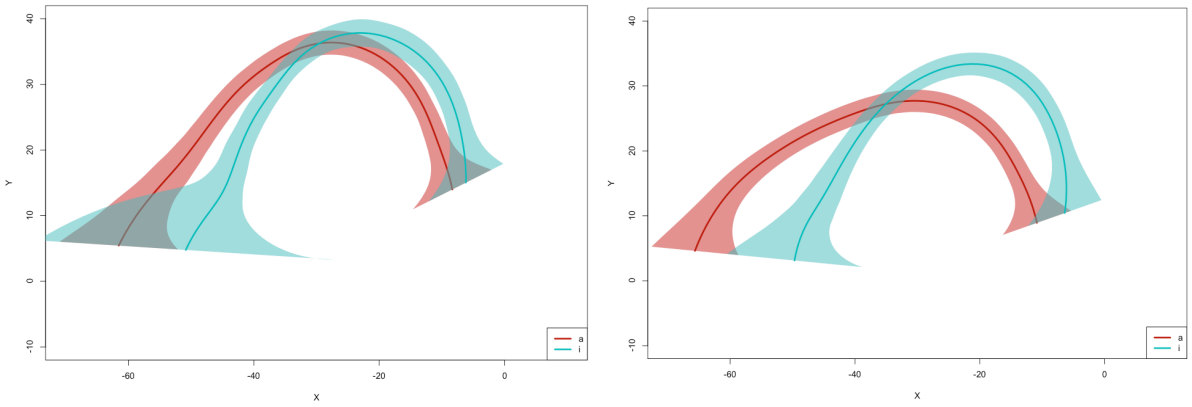


Figure 3.17. Représentation statistique des contours linguaux pour le participant Enfant 1 au relâchement de la consonne (panneau gauche) et au milieu de V2 (panneau droit) pour C = /g/, V1 = /ɛ/, en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

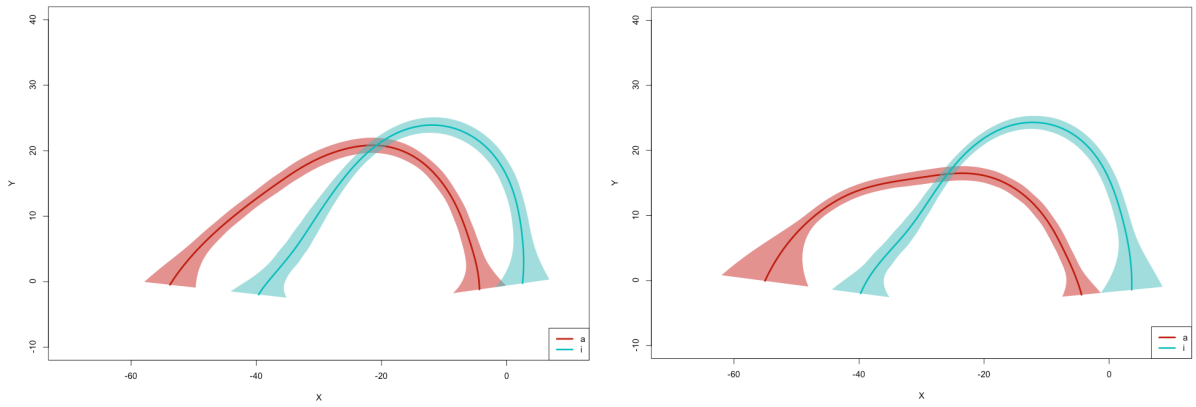


Figure 3.18. Représentation statistique des contours linguaux pour le participant Enfant 1 au relâchement de la consonne (panneau gauche) et au milieu de V2 (panneau droit) pour C = /g/, V1 = /a/, en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

On observe, chez ce participant, un effet significatif de V2 sur la réalisation de la consonne (au relâchement de la consonne) dans tous les contextes étudiés.

Enfant 2

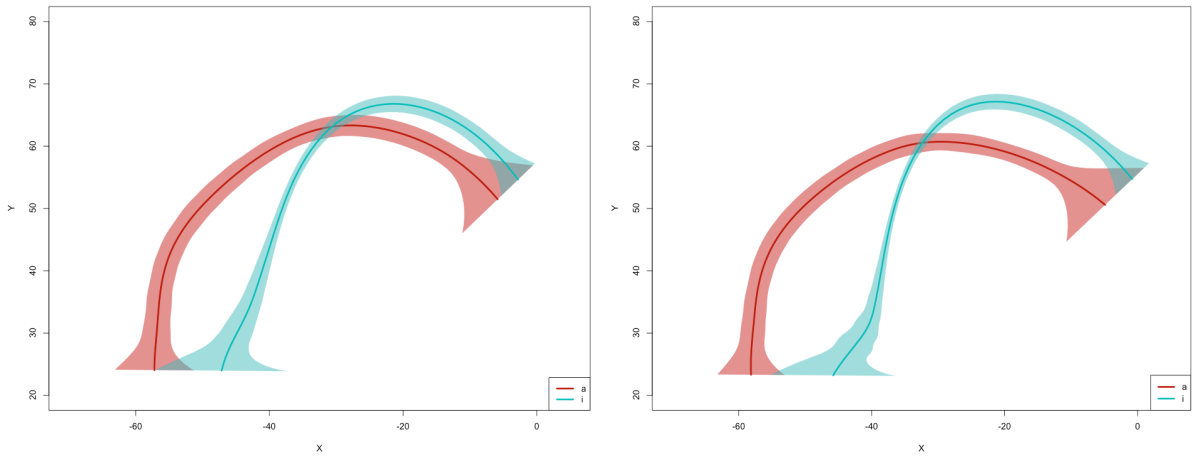


Figure 3.19. Représentation statistique des contours linguaux pour le participant Enfant 2 au relâchement de la consonne (panneau gauche) et au milieu de V2 (panneau droit) pour C = /b/, V1 = /ɛ/, en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

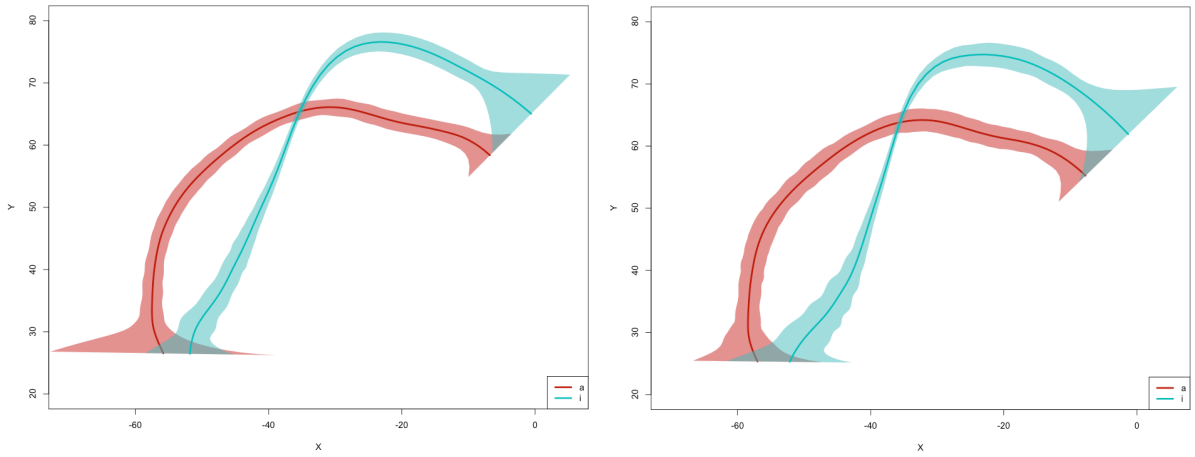


Figure 3.20. Représentation statistique des contours linguaux pour le participant Enfant 2 au relâchement de la consonne (panneau gauche) et au milieu de V2 (panneau droit) pour C = /b/, V1 = /a/, en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

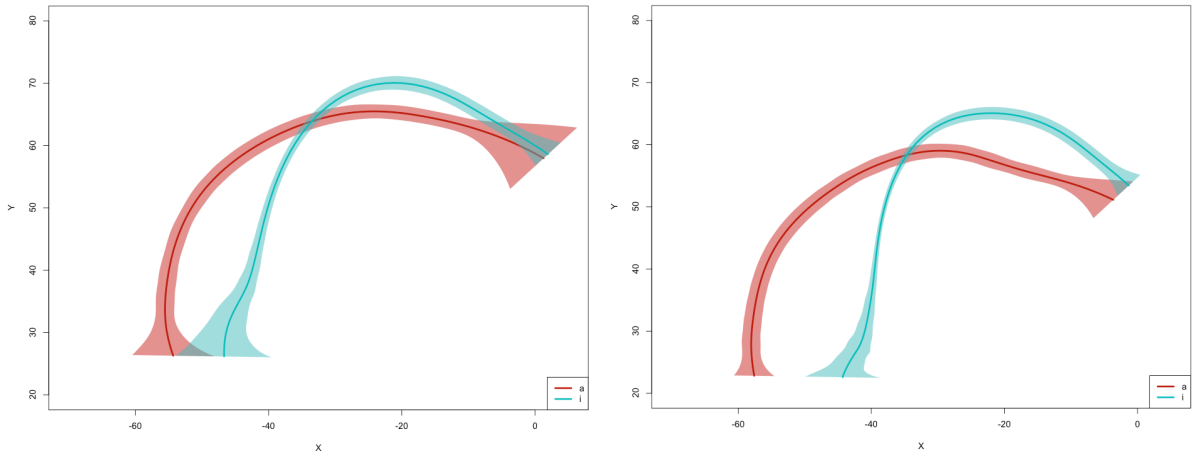


Figure 3.21. Représentation statistique des contours linguaux pour le participant Enfant 2 au relâchement de la consonne (panneau gauche) et au milieu de V2 (panneau droit) pour C = /d/, V1 = /ɛ/, en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

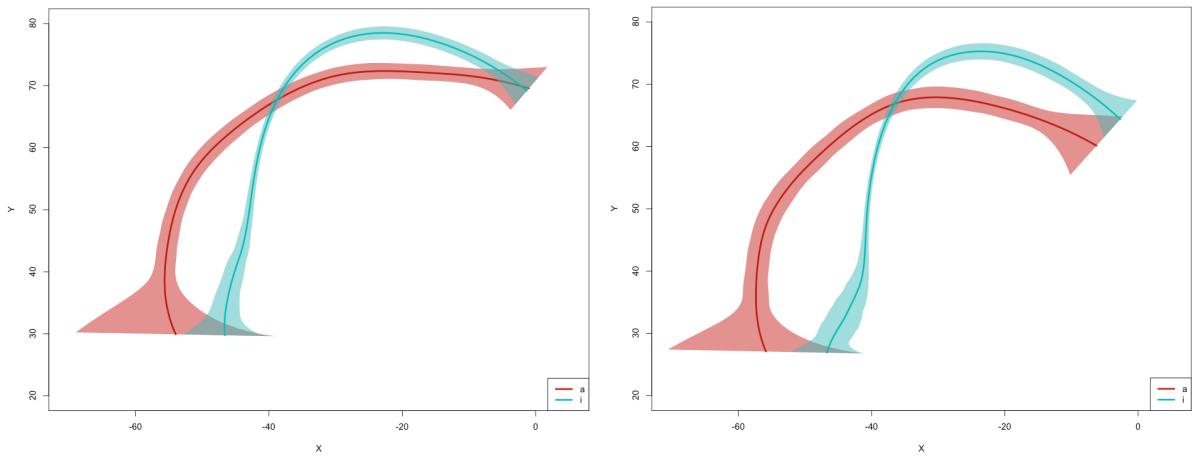


Figure 3.22. Représentation statistique des contours linguaux pour le participant Enfant 2 au relâchement de la consonne (panneau gauche) et au milieu de V2 (panneau droit) pour C = /d/, V1 = /a/, en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

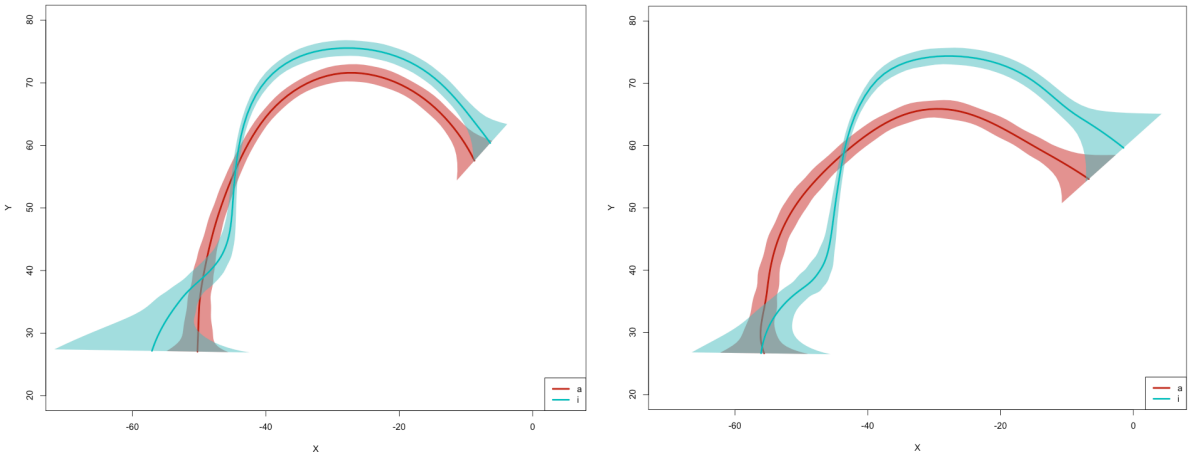


Figure 3.23. Représentation statistique des contours linguaux pour le participant Enfant 2 au relâchement de la consonne (panneau gauche) et au milieu de V2 (panneau droit) pour C = /g/, V1 = /ɛ/, en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

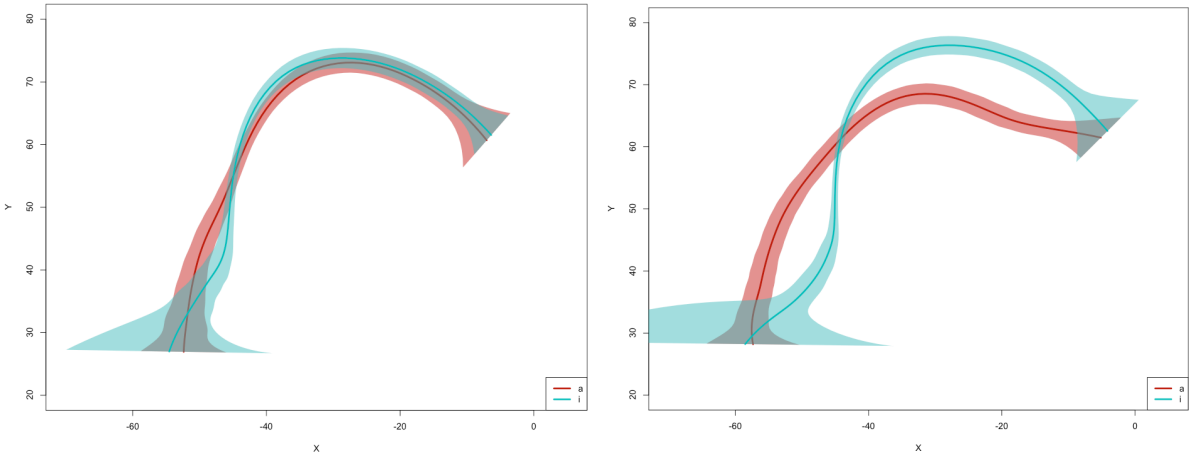


Figure 3.24. Représentation statistique des contours linguaux pour le participant Enfant 2 au relâchement de la consonne (panneau gauche) et au milieu de V2 (panneau droit) pour C = /g/, V1 = /a/, en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

On observe, chez ce participant, un effet significatif de V2 sur la réalisation des consonnes /b/ et /d/ (au relâchement de la consonne) que ce soit pour V1 = /a/ ou V1 = /ɛ/, mais pas sur la consonne /g/ dans le contexte où V1 = /a/ (soit dans 5 des 6 contextes étudiés).

Enfant 5

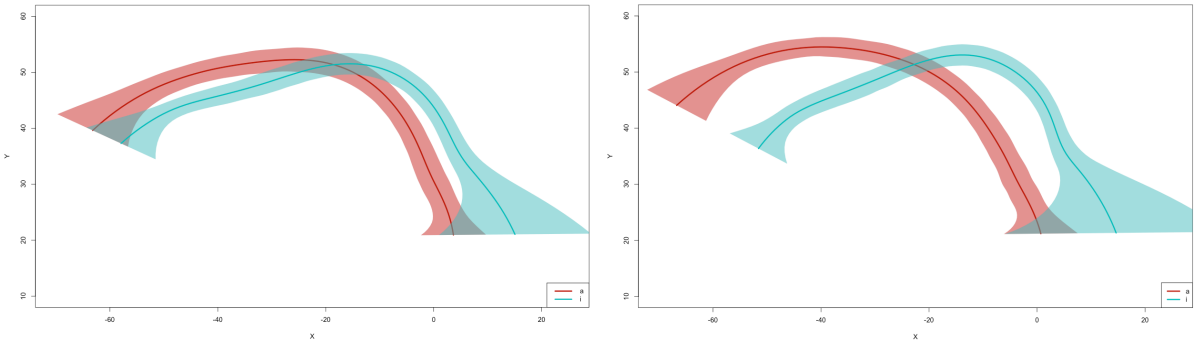


Figure 3.25. Représentation statistique des contours linguaux pour le participant Enfant 5 au relâchement de la consonne (panneau gauche) et au milieu de V2 (panneau droit) pour C = /b/, V1 = /ɛ/, en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

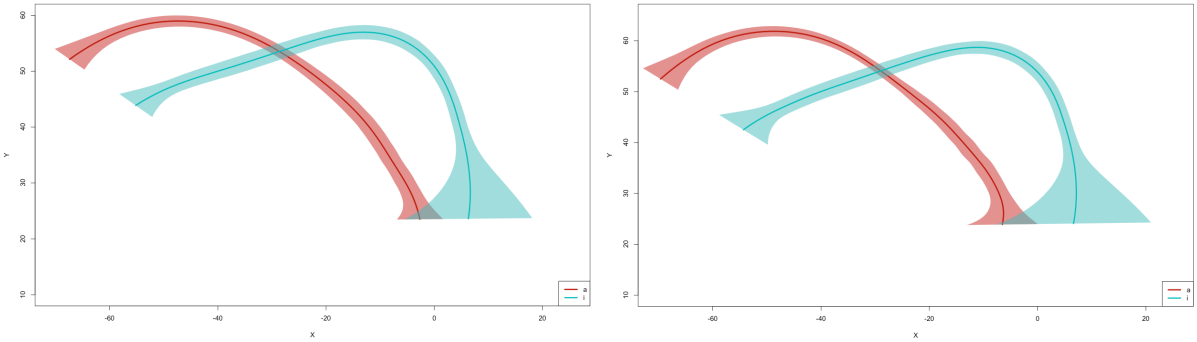


Figure 3.26. Représentation statistique des contours linguaux pour le participant Enfant 5 au relâchement de la consonne (panneau gauche) et au milieu de V2 (panneau droit) pour C = /b/, V1 = /a/, en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

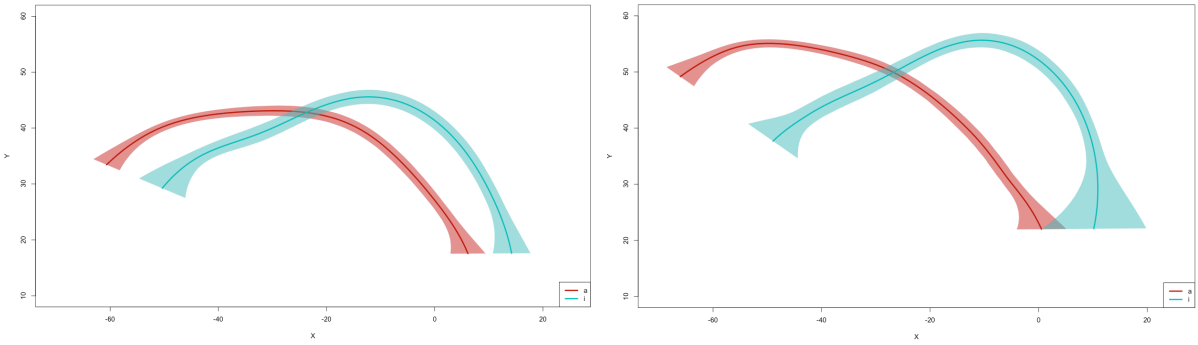


Figure 3.27. Représentation statistique des contours linguaux pour le participant Enfant 5 au relâchement de la consonne (panneau gauche) et au milieu de V2 (panneau droit) pour C = /d/, V1 = /ɛ/, en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

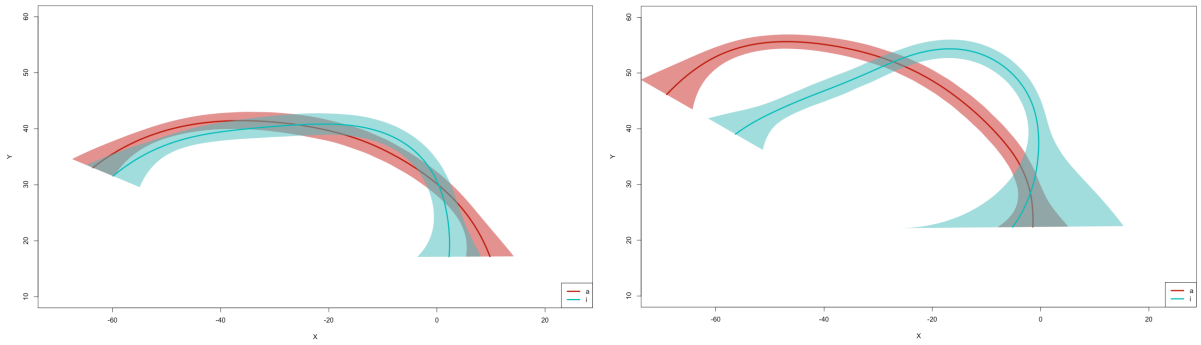


Figure 3.28. Représentation statistique des contours linguaux pour le participant Enfant 5 au relâchement de la consonne (panneau gauche) et au milieu de V2 (panneau droit) pour C = /d/, V1 = /a/, en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

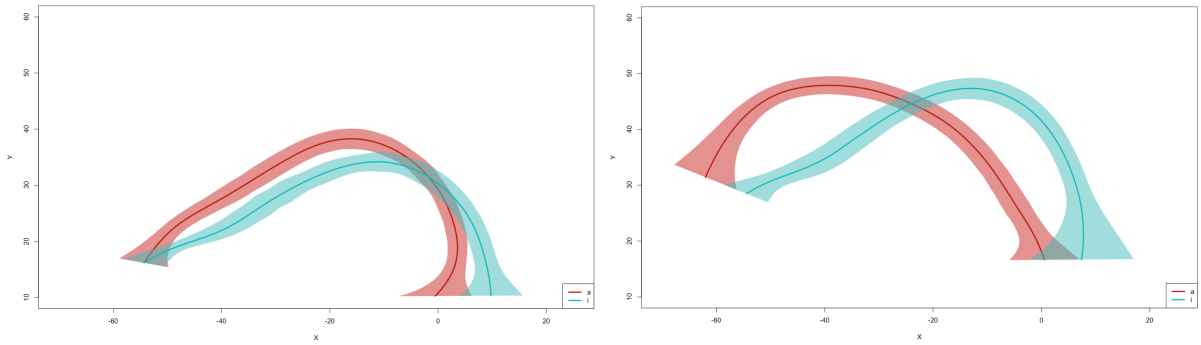


Figure 3.29. Représentation statistique des contours linguaux pour le participant Enfant 5 au relâchement de la consonne (panneau gauche) et au milieu de V2 (panneau droit) pour C = /g/, V1 = /ɛ/, en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

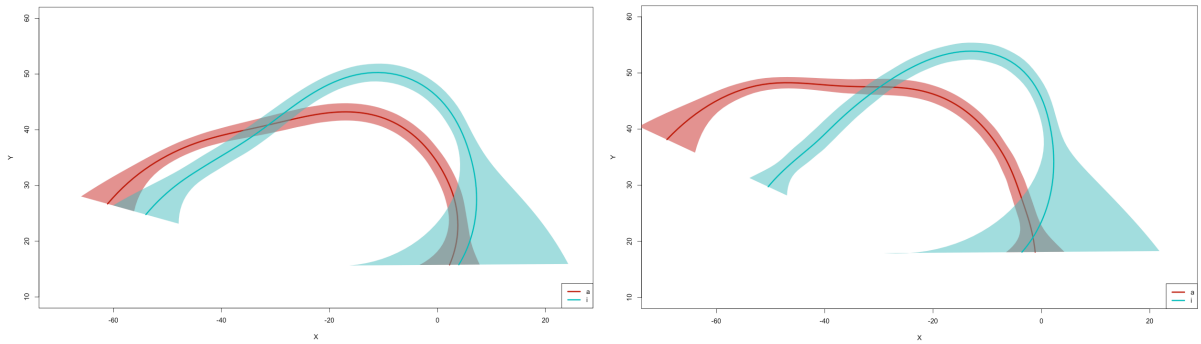


Figure 3.30. Représentation statistique des contours linguaux pour le participant Enfant 5 au relâchement de la consonne (panneau gauche) et au milieu de V2 (panneau droit) pour C = /g/, V1 = /a/, en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

On observe, chez ce participant, un effet significatif de V2 sur la réalisation des consonnes /b/ et /g/ (au relâchement de la consonne), dans les contextes V1 = /a/ et V1 = /ɛ/, et sur la consonne /d/ quand V1 = /ɛ/, mais pas sur la consonne /d/ dans le contexte où V1 = /a/ (soit dans 5 des 6 contextes étudiés).

Enfant 6

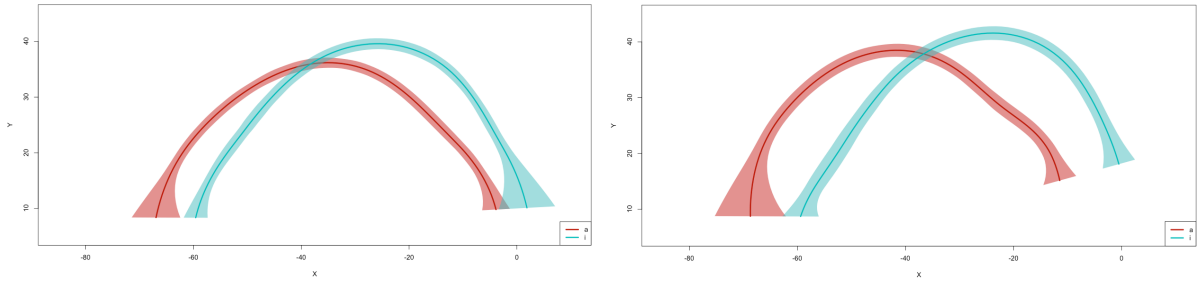


Figure 3.31. Représentation statistique des contours linguaux pour le participant Enfant 6 au relâchement de la consonne (panneau gauche) et au milieu de V2 (panneau droit) pour C = /b/, V1 = /ɛ/, en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

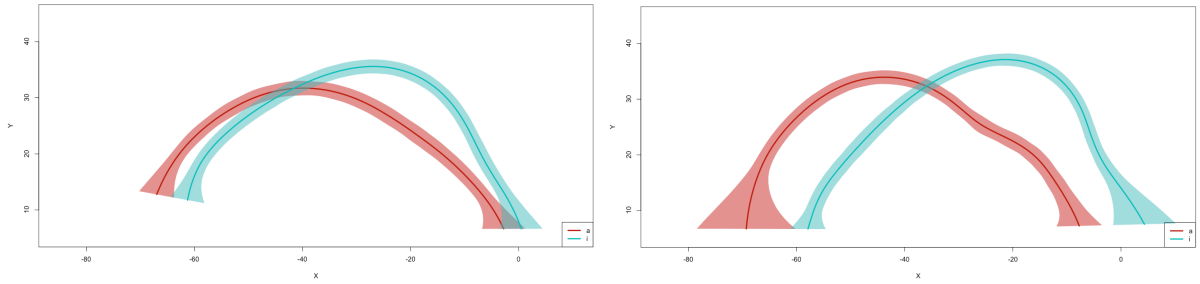


Figure 3.32. Représentation statistique des contours linguaux pour le participant Enfant 6 au relâchement de la consonne (panneau gauche) et au milieu de V2 (panneau droit) pour C = /b/, V1 = /a/, en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

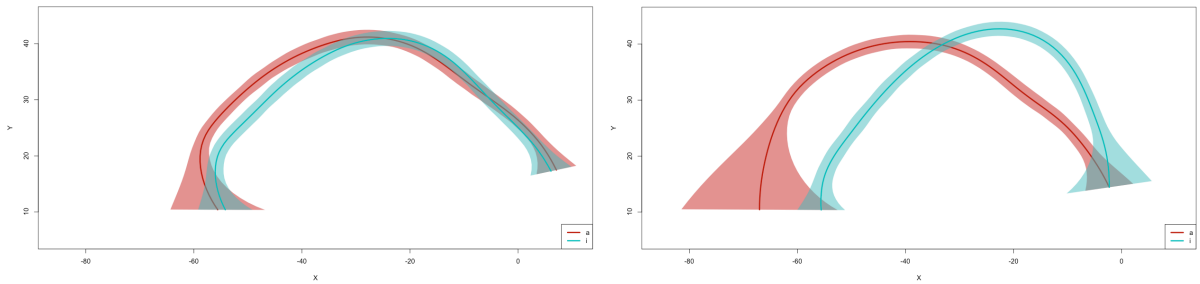


Figure 3.33. Représentation statistique des contours linguaux pour le participant Enfant 6 au relâchement de la consonne (panneau gauche) et au milieu de V2 (panneau droit) pour C = /d/, V1 = /ɛ/, en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

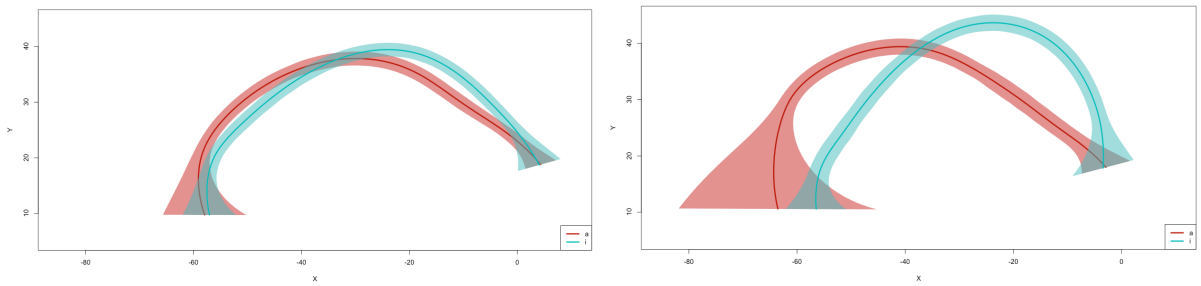


Figure 3.34. Représentation statistique des contours linguaux pour le participant Enfant 6 au relâchement de la consonne (panneau gauche) et au milieu de V2 (panneau droit) pour C = /d/, V1 = /a/, en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

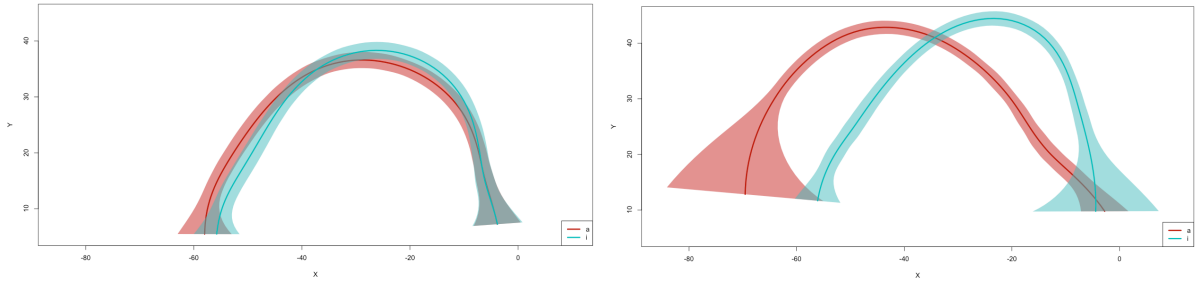


Figure 3.35. Représentation statistique des contours linguaux pour le participant Enfant 6 au relâchement de la consonne (panneau gauche) et au milieu de V2 (panneau droit) pour C = /g/, V1 = /ɛ/, en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

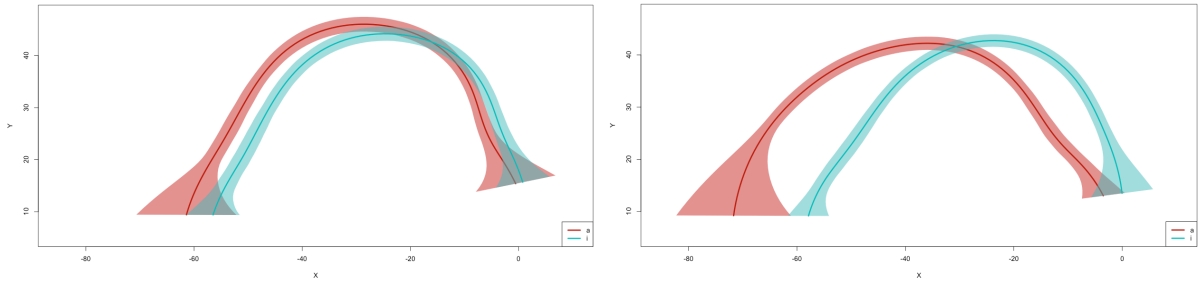


Figure 3.36. Représentation statistique des contours linguaux pour le participant Enfant 6 au relâchement de la consonne (panneau gauche) et au milieu de V2 (panneau droit) pour C = /g/, V1 = /a/, en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

On observe, chez ce participant, un effet significatif de V2 sur la réalisation des consonnes /b/ et /d/ (au relâchement de la consonne) dans les contextes où V1 = /a/ et V1 = /ɛ/, et sur la consonne /g/ quand V1 = /a/, mais pas dans le contexte où C = /g/ et V1 = /ɛ/ (soit dans 5 des 6 contextes étudiés).

Enfant 7

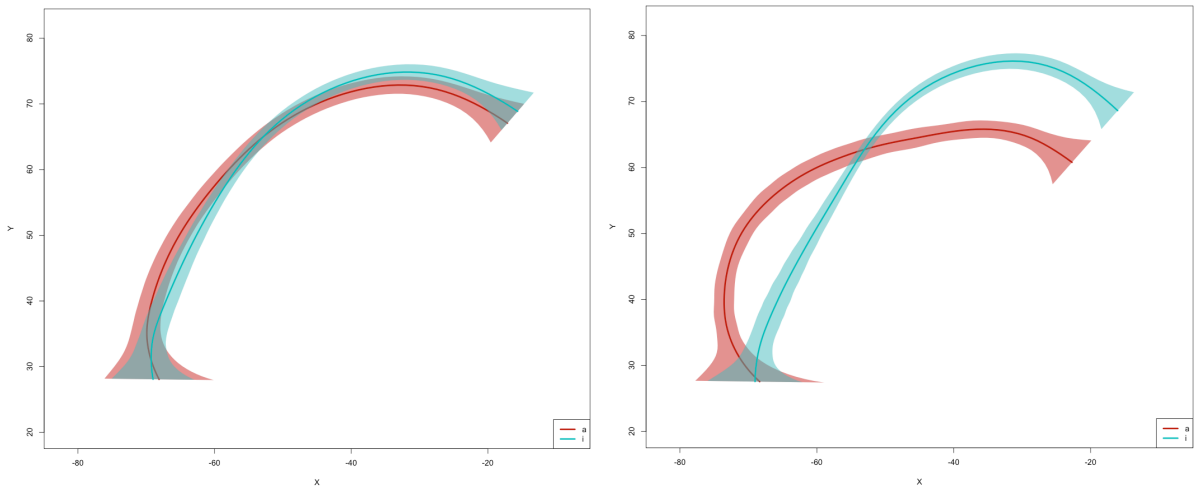


Figure 3.37. Représentation statistique des contours linguaux pour le participant Enfant 7 au relâchement de la consonne (panneau gauche) et au milieu de V2 (panneau droit) pour C = /b/, V1 = /ɛ/, en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

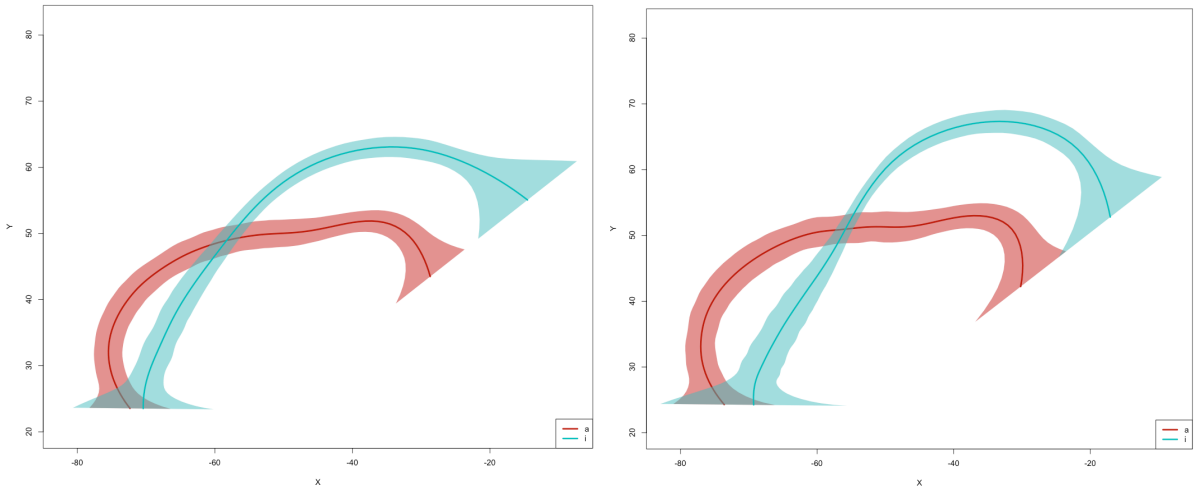


Figure 3.38. Représentation statistique des contours linguaux pour le participant Enfant 7 au relâchement de la consonne (panneau gauche) et au milieu de V2 (panneau droit) pour C = /b/, V1 = /a/, en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

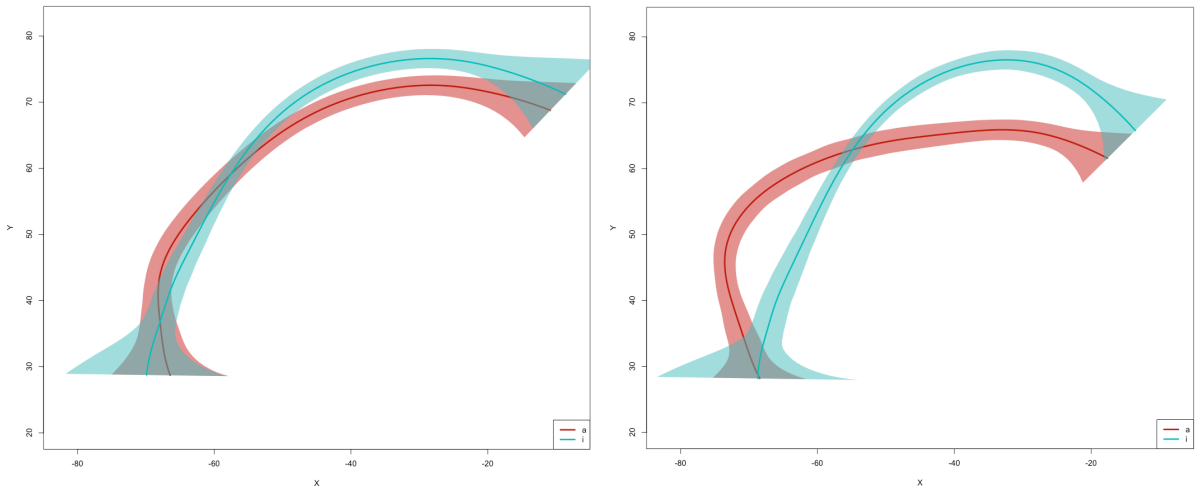


Figure 3.39. Représentation statistique des contours linguaux pour le participant Enfant 7 au relâchement de la consonne (panneau gauche) et au milieu de V2 (panneau droit) pour C = /d/, V1 = /ɛ/, en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

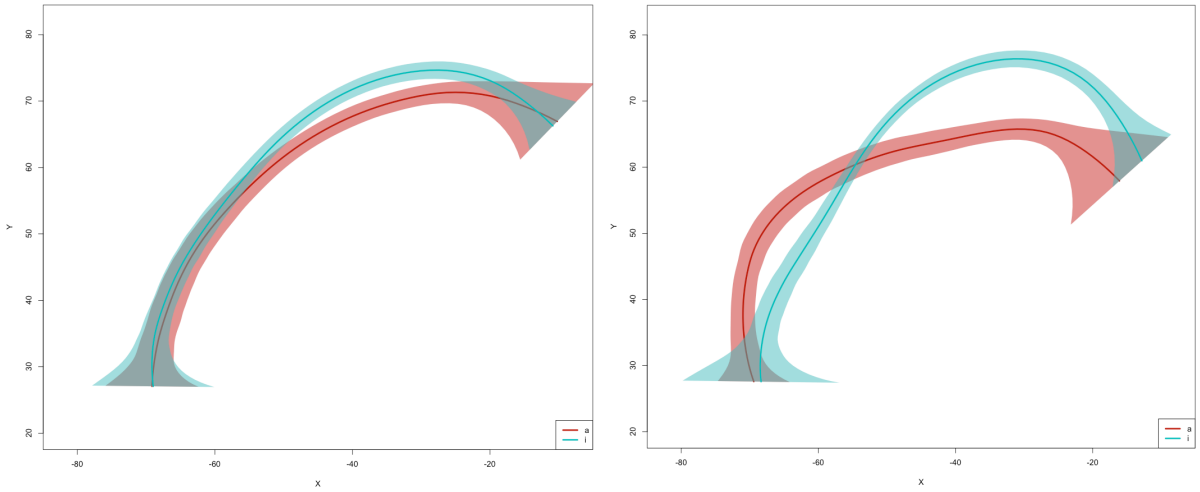


Figure 3.40. Représentation statistique des contours linguaux pour le participant Enfant 7 au relâchement de la consonne (panneau gauche) et au milieu de V2 (panneau droit) pour C = /d/, V1 = /a/, en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

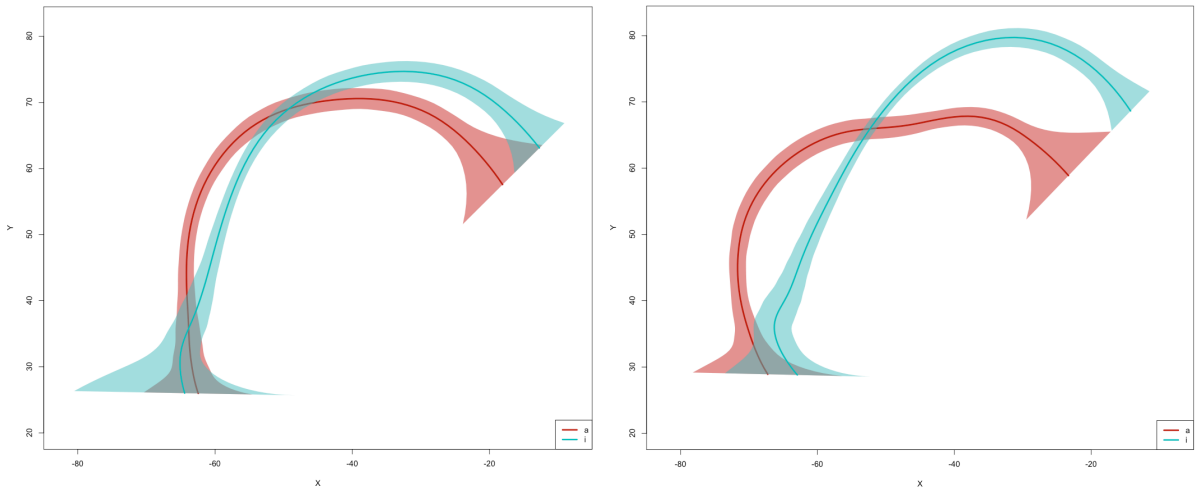


Figure 3.41. Représentation statistique des contours linguaux pour le participant Enfant 7 au relâchement de la consonne (panneau gauche) et au milieu de V2 (panneau droit) pour C = /g/, V1 = /ɛ/, en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

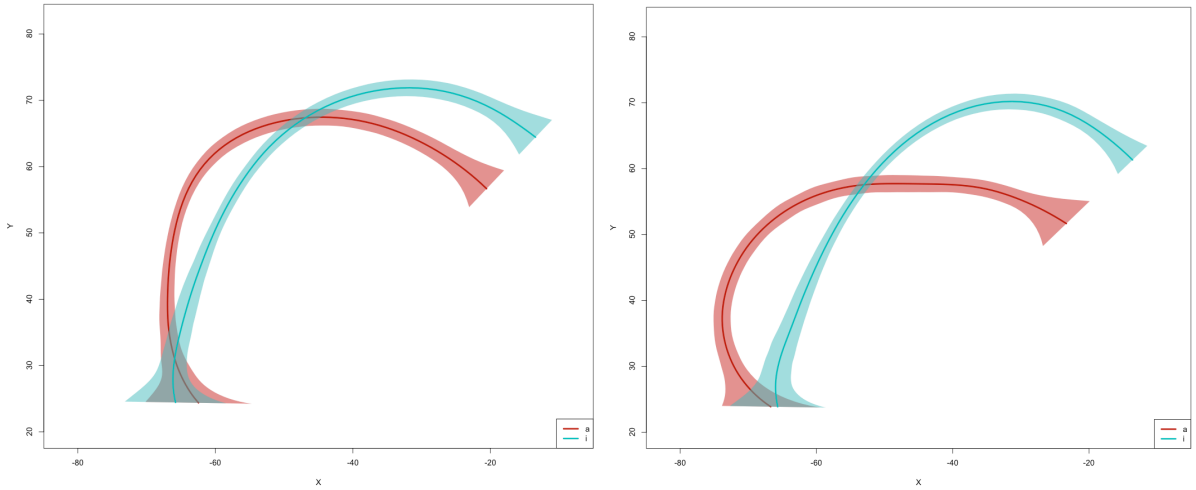


Figure 3.42. Représentation statistique des contours linguaux pour le participant Enfant 7 au relâchement de la consonne (panneau gauche) et au milieu de V2 (panneau droit) pour C = /g/, V1 = /a/, en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

On observe, chez ce participant, un effet significatif de V2 sur la réalisation des consonnes /d/ et /g/ (au relâchement de la consonne) dans les contextes V1 = /a/ et V1 = /ɛ/, et sur la réalisation de la consonne /b/ quand V1 = /a/, mais pas dans le contexte où C = /b/ et V1 = /ɛ/ (soit dans 5 des 6 contextes étudiés).

Enfant 9

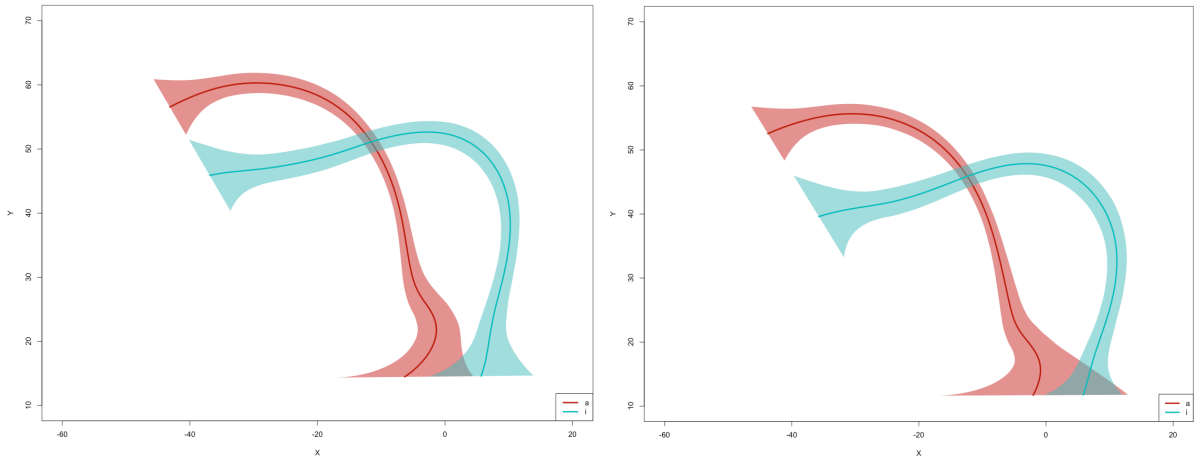


Figure 3.43. Représentation statistique des contours linguaux pour le participant Enfant 9 au relâchement de la consonne (panneau gauche) et au milieu de V2 (panneau droit) pour C = /b/, V1 = /ɛ/, en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

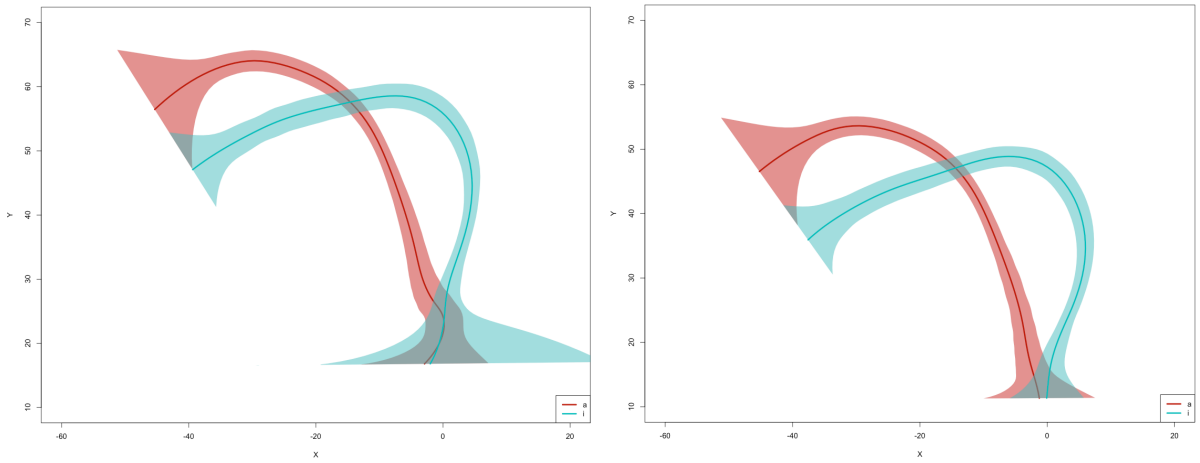


Figure 3.44. Représentation statistique des contours linguaux pour le participant Enfant 9 au relâchement de la consonne (panneau gauche) et au milieu de V2 (panneau droit) pour C = /b/, V1 = /a/, en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

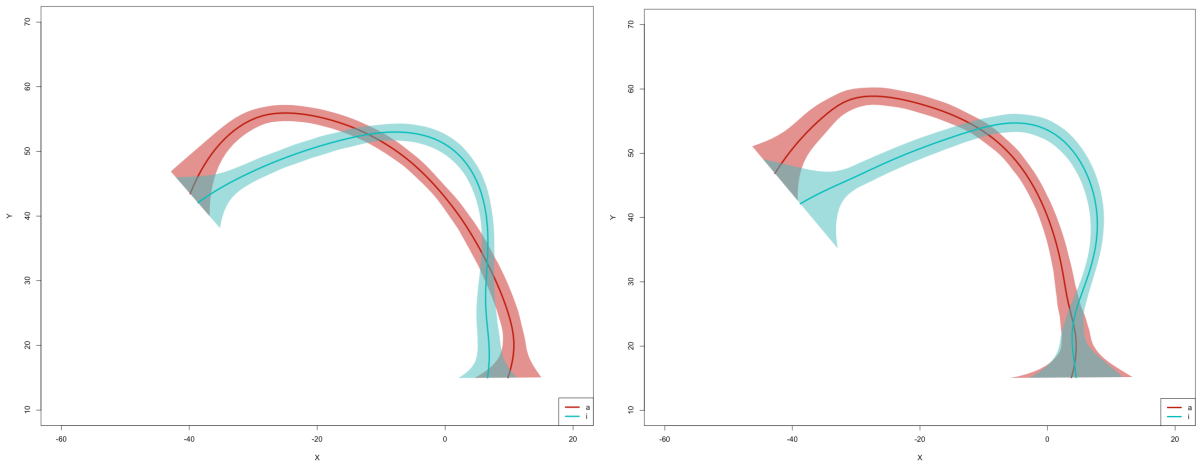


Figure 3.45. Représentation statistique des contours linguaux pour le participant Enfant 9 au relâchement de la consonne (panneau gauche) et au milieu de V2 (panneau droit) pour C = /d/, V1 = /ɛ/, en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

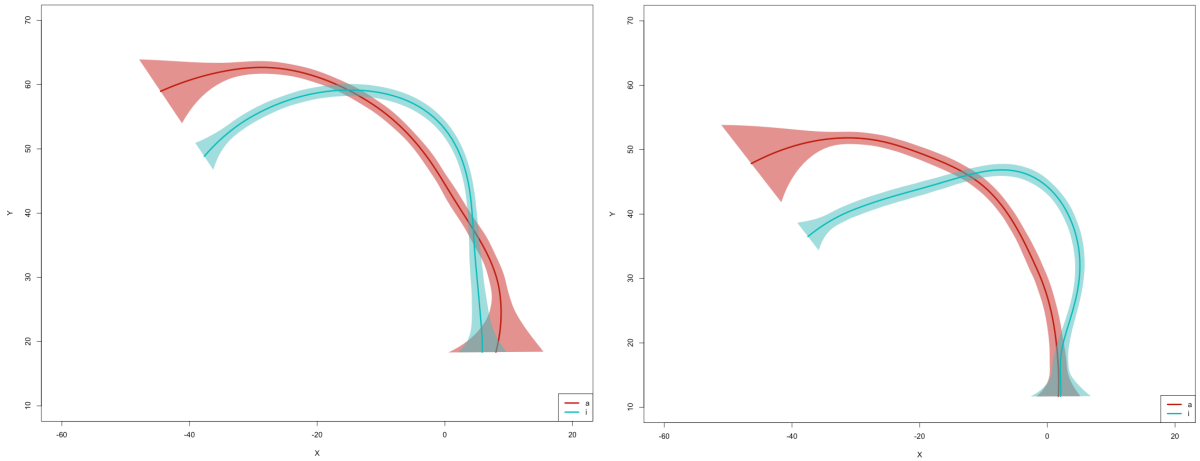


Figure 3.46. Représentation statistique des contours linguaux pour le participant Enfant 9 au relâchement de la consonne (panneau gauche) et au milieu de V2 (panneau droit) pour C = /d/, V1 = /a/, en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

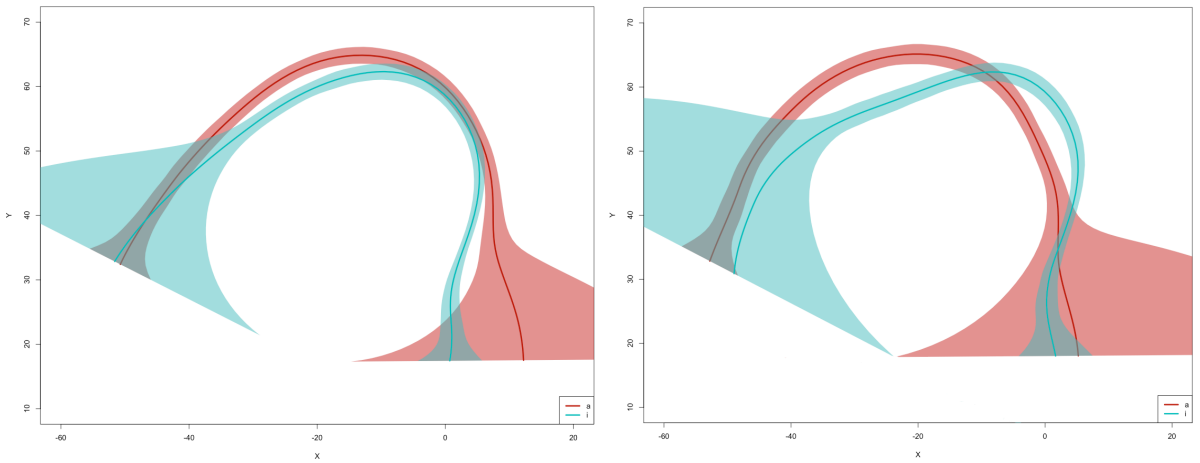


Figure 3.47. Représentation statistique des contours linguaux pour le participant Enfant 9 au relâchement de la consonne (panneau gauche) et au milieu de V2 (panneau droit) pour C = /g/, V1 = /ɛ/, en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

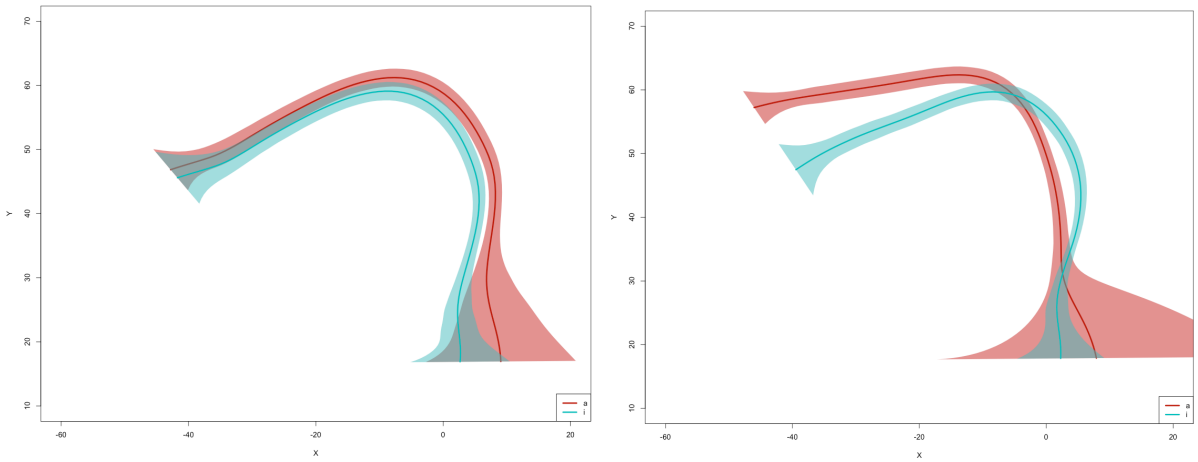


Figure 3.48. Représentation statistique des contours linguaux pour le participant Enfant 9 au relâchement de la consonne (panneau gauche) et au milieu de V2 (panneau droit) pour C = /g/, V1 = /a/, en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

On observe, chez ce participant, un effet significatif de V2 sur la réalisation des consonnes /b/ et /d/ (au relâchement de la consonne) dans les contextes V1 = /a/ et V1 = /ɛ/, et sur la réalisation de /g/ dans le contexte où V1 = /ɛ/. Dans le contexte C = /g/ et V1 = /a/, un effet significatif de V2 sur la réalisation de C est observé, mais cet effet va dans la direction opposée à l'anticipation. On observe donc chez ce participant un effet significatif d'anticipation dans 5 des 6 contextes étudiés.

Dans l'ensemble, les enfants anticipent V2 dans la réalisation de la consonne, dans la majorité des contextes étudiés. Cependant, cette anticipation n'est pas systématique puisqu'il existe des contextes pour lesquels cette anticipation n'a pas été observée ou des contextes dans lesquels l'amplitude de cette anticipation n'est pas suffisante pour que les différences observées soient statistiquement significatives.

3.3. Coarticulation extra-syllabique

Nous présentons maintenant les résultats principaux de cette étude, ceux concernant la coarticulation extra-syllabique. Rappelons que nous étudions l'anticipation de V2 dans la réalisation de V1 dans des séquences de type V_1-C-V_2. Ces résultats permettront de tester notre hypothèse 3 (nous devrions observer une moindre amplitude de la coarticulation anticipatoire au-delà de la frontière syllabique chez l'enfant par rapport à l'adulte).

3.3.1. Résultats individuels concernant les adultes

Nous présentons dans la figure 3.49 ci-dessous, les réalisations de V1 = /a/ pour les participants adultes, dans l'espace acoustique F1 - F2 normalisé, dans les contextes où V2 = /i/ (bleu), et V2 = /a/ (rouge), tous contextes consonantiques confondus.

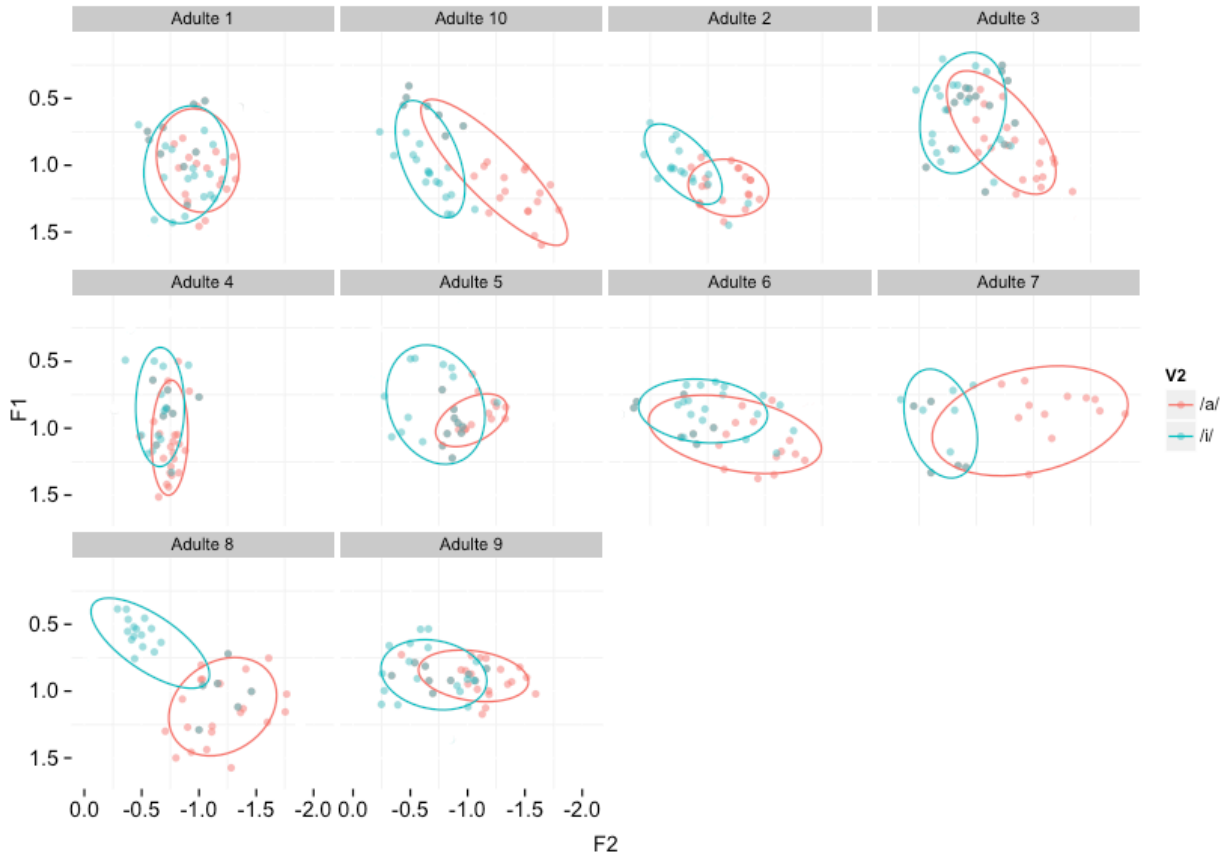


Figure 3.49. Anticipation de V2 dans la réalisation de V1 = /a/ dans l'espace acoustique F1 -F2 normalisé, pour les participants adultes, tous contextes consonantiques confondus.

Nous présentons dans la figure 3.50 ci-dessous, les réalisations de V1 = /ɛ/ pour les participants adultes, dans l'espace acoustique F1 - F2 normalisé, dans les contextes où V2 = /i/ (bleu), et V2 = /a/ (rouge), tous contextes consonantiques confondus.

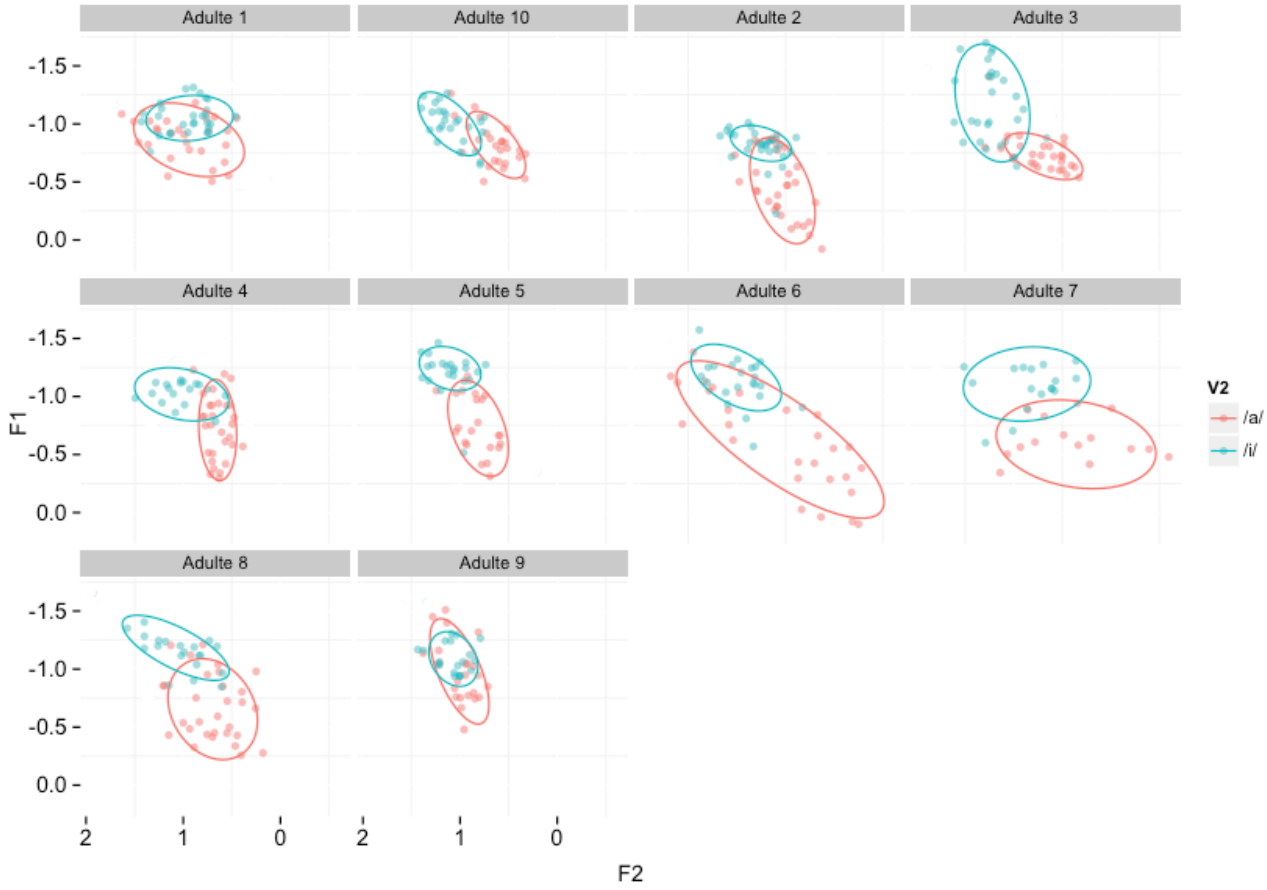


Figure 3.50. Anticipation de V2 dans la réalisation de V1 = /ɛ/ dans l'espace acoustique F1 -F2 normalisé, pour les participants adultes, tous contextes consonantiques confondus.

Les figures 3.49. et 3.50. permettent d'observer que l'effet d'anticipation est clair et important pour tous les participants adultes, sauf pour le participant Adulte 1 dans le contexte V1 = /a/ et pour le participant Adulte 9 dans le contexte V1 = / ɛ /.

Afin de quantifier, dans le domaine acoustique, la taille de l'effet d'anticipation, pour chaque participant adulte et chaque contexte de V1 séparément, nous proposons dans le tableau 3.9. ci-dessous, le calcul d'un indice individuel d'anticipation. Cet indice est calculé comme la moyenne des indices globaux pour V1 = /a/ et V1 = /ɛ/, indices représentant les distances euclidiennes séparant les centres des ellipses de dispersion des réalisations de V1 en fonction des deux contextes de V2.

Tableau 3.9. Mesures de la taille de l'effet en F1, en F2 et de l'effet global et indice individuel d'anticipation pour les participants adultes, tous contextes consonantiques confondus.

Indice de la taille de l'effet d'anticipation							
	V1 = /ɛ/			V1 = /a/			Indice individuel d'anticipation
Participant	Effet F1	Effet F2	Effet global	Effet F1	Effet F2	Effet global	
Adulte 1	0,20	0,04	**0,20**	-0,05	0,10	**0,12**	**0,16**
Adulte 2	0,38	0,24	**0,45**	0,14	0,36	**0,39**	**0,42**
Adulte 3	0,37	0,46	**0,59**	0,13	0,33	**0,36**	**0,47**
Adulte 4	0,40	0,05	**0,41**	0,26	0,10	**0,27**	**0,34**
Adulte 5	0,51	0,28	**0,58**	0,01	0,54	**0,54**	**0,56**
Adulte 6	0,47	0,44	**0,64**	0,17	0,28	**0,33**	**0,49**
Adulte 7	0,35	0,59	**0,69**	0,05	0,70	**0,70**	**0,69**
Adulte 8	0,50	0,39	**0,63**	0,36	0,55	**0,65**	**0,64**
Adulte 9	0,06	0,05	**0,08**	0,00	0,37	**0,37**	**0,22**
Adulte 10	0,18	0,46	**0,49**	0,12	0,53	**0,55**	**0,52**
Moyenne	0,34	0,30	**0,48**	0,12	0,39	**0,43**	**0,45**

Le tableau 3.9. rapporte les tailles d'effet en F1 et en F2 (mesurées par la distance entre les moyennes des productions en F1 et en F2 entre les 2 contextes de V2 étudiés (V2 =/i/ ou V2 = /a/) et la taille de l'effet global sur V1 (mesuré par la distance euclidienne entre les centres des ellipses de dispersion). Un signe a été ajouté à cet effet global, positif si l'effet va dans le sens de l'anticipation, négatif s'il va dans le sens opposé à l'anticipation de V2. Dans ce tableau, tous les effets globaux sont positifs, ce qui indique qu'on observe un effet d'anticipation de V2 dans la réalisation de V1 pour tous les participants adultes. Un indice individuel d'anticipation est proposé, calculé comme la moyenne de l'effet global sur V1 = /ɛ/ et sur V1 = /a/. L'indice moyen d'anticipation pour les participants adultes est de **0.45**. Cet indice est moins grand pour les participants Adulte 1 et 9.

Afin de tester statistiquement si les effets d'anticipation observés sont significatifs, nous avons conduit des analyses statistiques individuelles : des modèles linéaires prenant comme variables dépendantes F1 et F2 de chaque individu en fonction du contexte de V2. Les résultats de ces analyses statistiques sont reportées dans le tableau 3.10. ci-dessous. Nous avons choisi une valeur de $p = 0.01$, les résultats significatifs sont indiqués en gras.

Tableau 3.10. Analyses statistiques des effets d'anticipation en F1 et en F2 pour chaque participant adulte, tous contextes consonantiques confondus.

		Analyses statistiques des effets d'anticipation			
		V1 = /ɛ/		**V1 = /a/**	
Participant	**Valeurs statistiques**	**F1**	**F2**	**F1**	**F2**
Adulte 1	**F**	**18.46**	0.215	0.207	2.725
	p	**6.69e-05**	0.645	0.651	0.104
Adulte 2	**F**	**41.9**	**18.49**	**11.69**	**26.9**
	p	**1.85e-08**	**6.27e-05**	**0.00149**	**6.95e-06**
Adulte 3	**F**	**39.88**	**47.41**	3.162	**24.08**
	p	**4.36e-08**	**4.94e-09**	0.0797	**5.67e-06**
Adulte 4	**F**	**19.25**	**28.36**	**7.502**	4.636
	p	**5.37e-05**	**2,00E-06**	**0.00842**	0.036
Adulte 5	**F**	**56.8**	**20.31**	3.657	**21.39**
	p	**1.44e-09**	**4.5e-05**	0.0625	**3.42e-05**
Adulte 6	**F**	**28.94**	**10.21**	**9.911**	4.049
	p	**1.98e-06**	**0.00243**	**0.00282**	0.0498
Adulte 7	**F**	**37.58**	**15.37**	0.013	**17.42**
	p	**5.21e-07**	**0.000392**	0.911	**0.000279**
Adulte 8	**F**	**57.88**	**16.75**	**36.49**	**36.67**
	p	**5.32e-10**	**0.000149**	**2.17e-07**	**2.06e-07**
Adulte 9	**F**	**36.67**	1.623	0.019	**16.96**
	p	**2.06e-07**	0.209	0.89	**0.000129**
Adulte 10	**F**	**8.463**	**55.09**	1.879	**34.67**
	p	**0.00532**	**1.05e-09**	0.177	**3.27e-07**

On observe, d'après le tableau 3.10., que les effets d'anticipation sont statistiquement significatifs pour la majorité des participants adultes et des contextes étudiés.

Afin d'illustrer l'effet de l'anticipation de V2 dans la réalisation de V1 dans le domaine articulatoire, et de vérifier d'un point de vue articulatoire ce que nous observons à travers les mesures acoustiques, nous présentons dans les figures 3.51. à 3.56. ci-dessous les représentations statistiques des contours de langue correspondant au milieu de V1 pour deux participants adultes.

Adulte 6

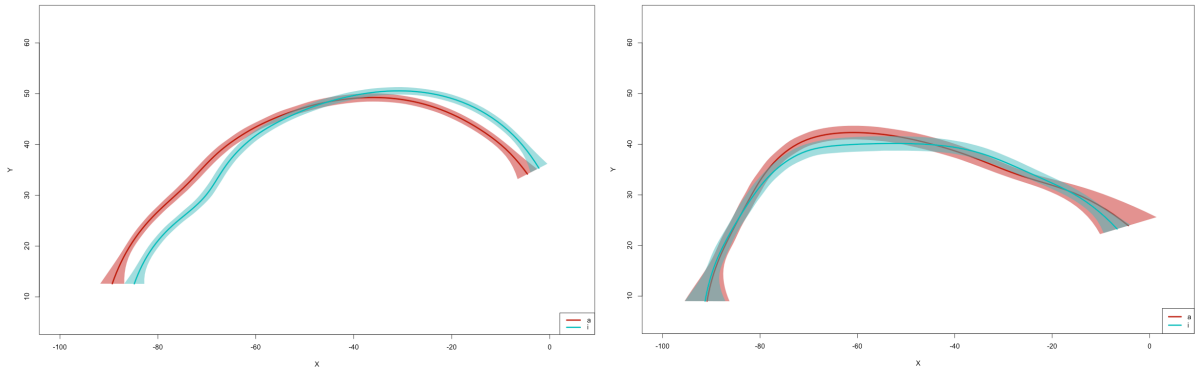


Figure 3.51. Représentation statistique des contours linguaux pour le participant Adulte 6 au milieu de V1 pour C = /b/ et V1 = /ɛ/ (panneau gauche), et pour C = /b/ et V1 = /a/ (panneau droit), en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

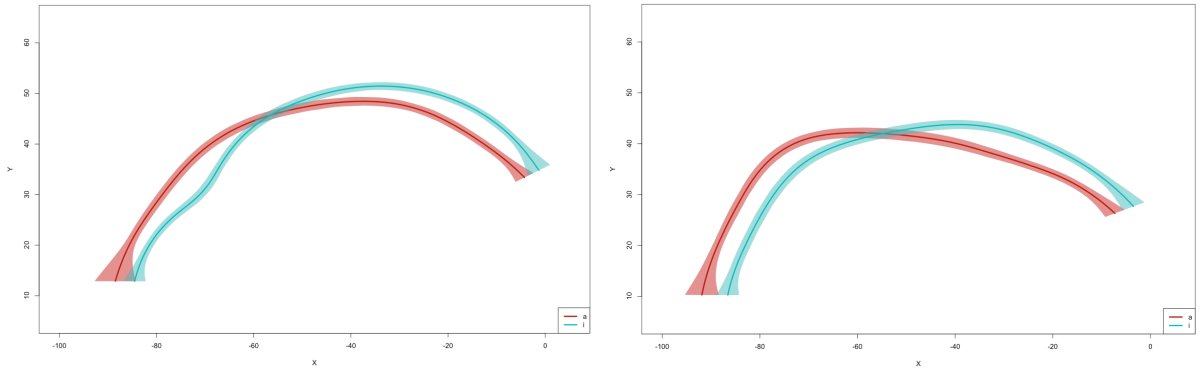


Figure 3.52. Représentation statistique des contours linguaux pour le participant Adulte 6 au milieu de V1 pour C = /d/ et V1 = /ɛ/ (panneau gauche), et pour C = /d/ et V1 = /a/ (panneau droit), en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

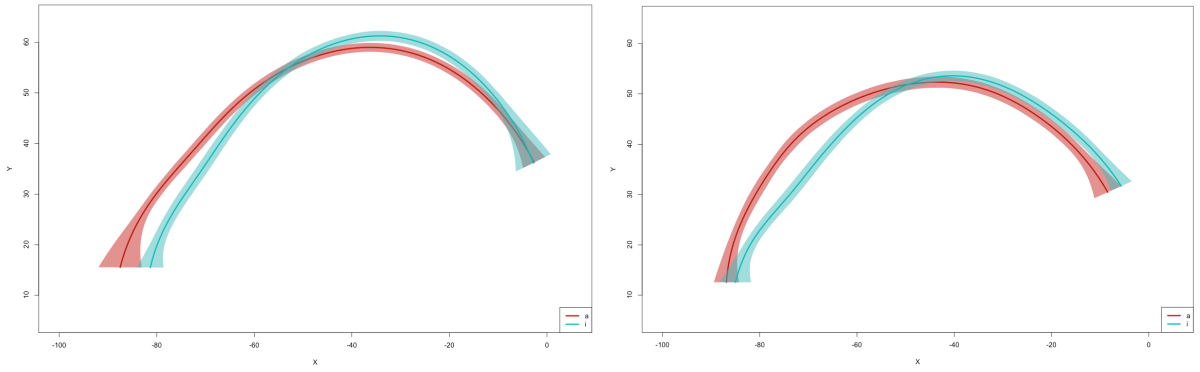


Figure 3.53. Représentation statistique des contours linguaux pour le participant Adulte 6 au milieu de V1 pour C = /g/ et V1 = /ɛ/ (panneau gauche), et pour C = /g/ et V1 = /a/ (panneau droit), en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

On observe, chez ce participant, un effet significatif de V2 sur la réalisation de V1 (au milieu de V1) dans tous les contextes, sauf dans le contexte où C = /b/ et V1 = /a/. Notons que pour ce contexte, une tendance qui va dans la direction de l'anticipation est observée, mais l'amplitude de cette anticipation n'est pas suffisante pour que la différence entre les deux contextes soit statistiquement significative.

Adulte 8

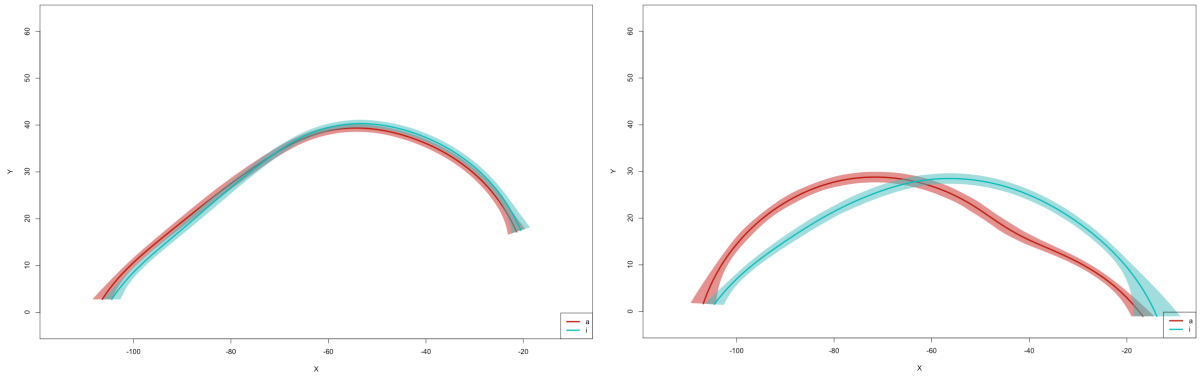


Figure 3.54. Représentation statistique des contours linguaux pour le participant Adulte 8 au milieu de V1 pour C = /b/ et V1 = /ɛ/ (panneau gauche), et pour C = /b/ et V1 = /a/ (panneau droit), en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

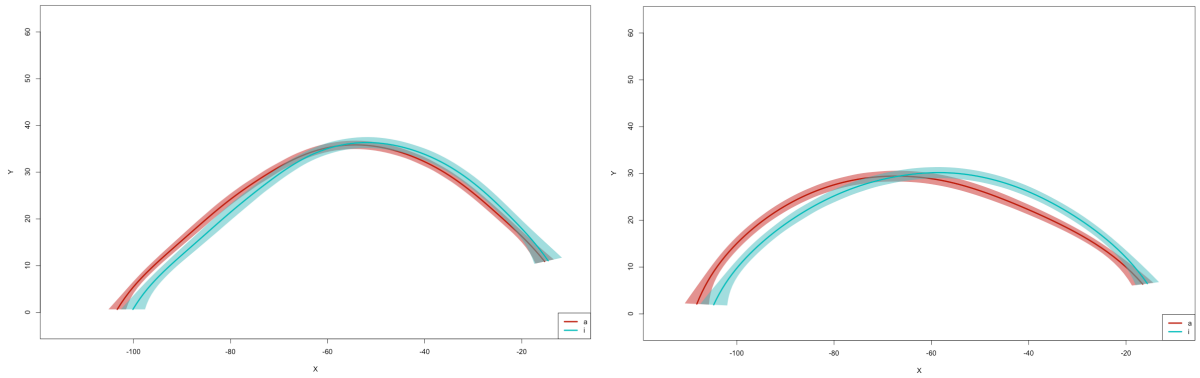


Figure 3.55. Représentation statistique des contours linguaux pour le participant Adulte 8 au milieu de V1 pour C = /d/ et V1 = /ɛ/ (panneau gauche), et pour C = /d/ et V1 = /a/ (panneau droit), en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

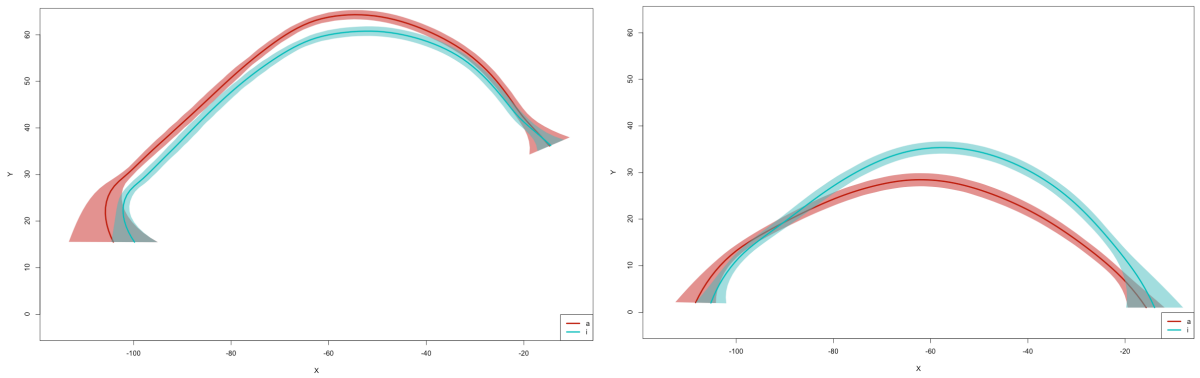


Figure 3.56. Représentation statistique des contours linguaux pour le participant Adulte 8 au milieu de V1 pour C = /g/ et V1 = /ɛ/ (panneau gauche), et pour C = /g/ et V1 = /a/ (panneau droit), en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

On observe, chez ce participant, un effet significatif de V2 sur la réalisation de V1 (au milieu de V1) dans tous les contextes, sauf dans le contexte où C = /b/ et V1 = /ɛ/. Notons que pour ce contexte, une tendance qui va dans la direction de l'anticipation est observée, mais l'amplitude de cette anticipation n'est pas suffisante pour que la différence entre les deux contextes soit statistiquement significative.

Dans l'ensemble, on observe une anticipation de V2 dans la réalisation de V1, puisque les 2 participants adultes étudiés montrent une anticipation statistiquement significative dans 5 des 6 contextes étudiés, et pour le contexte dans lequel cet effet d'anticipation n'est pas significatif, une tendance allant dans la direction de l'anticipation est observée.

3.3.2. Résultats individuels concernant les enfants

Nous présentons dans la figure 3.57. ci-dessous, les réalisations de V1 = /a/ pour les participants enfants, dans l'espace acoustique F1 - F2 normalisé, dans les contextes où V2 = /i/ (bleu), et V2 = /a/ (rouge), tous contextes consonantiques confondus.

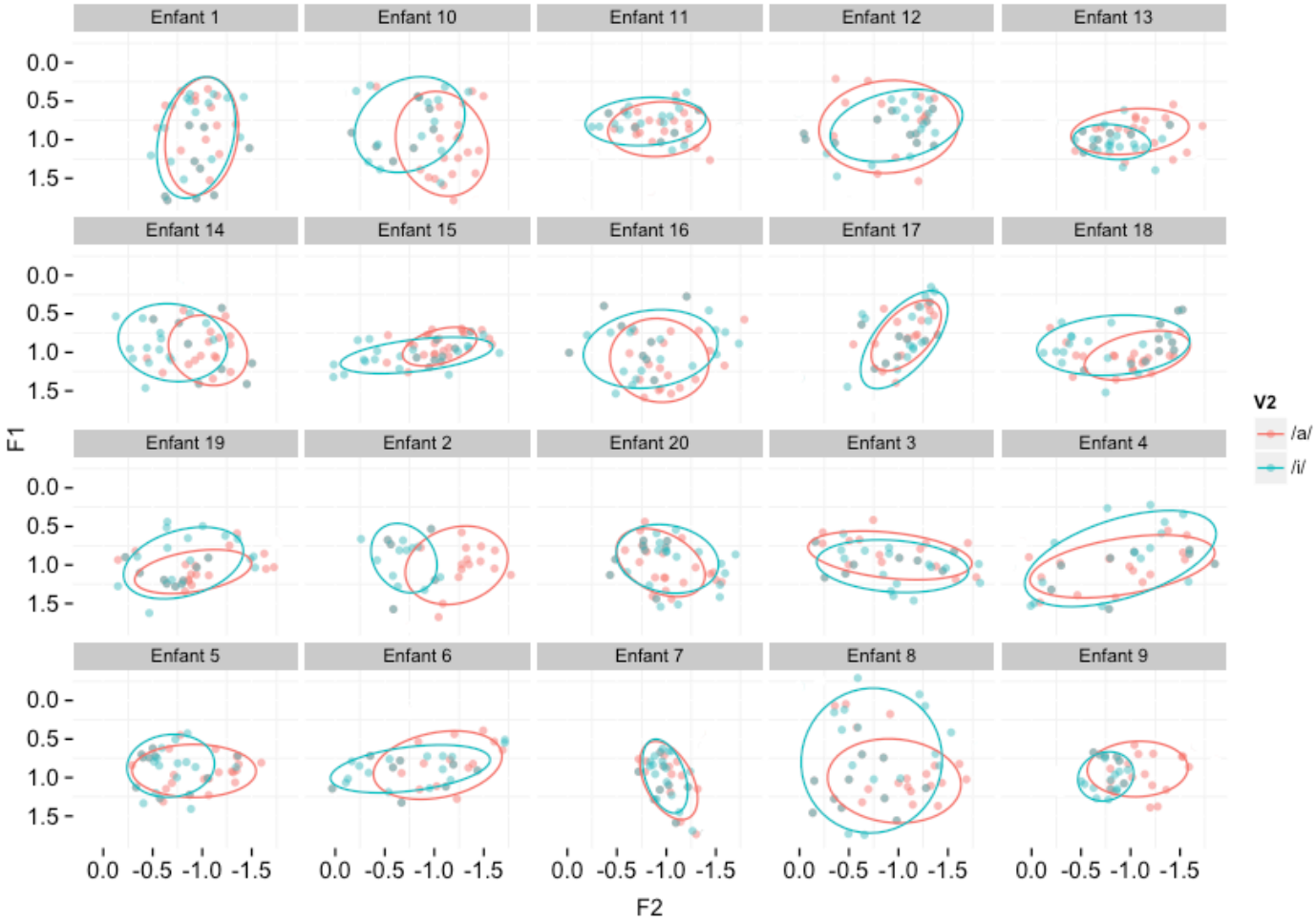


Figure 3.57. Anticipation de V2 dans la réalisation de V1 = /a/ dans l'espace acoustique F1 -F2 normalisé, pour les participants enfants, tous contextes consonantiques confondus.

Nous présentons dans la figure 3.58. ci-dessous, les réalisations de V1 = /ɛ/ pour les participants enfants, dans l'espace acoustique F1 - F2 normalisé, dans les contextes où V2 = /i/ (bleu), et V2 = /a/ (rouge), tous contextes consonantiques confondus.

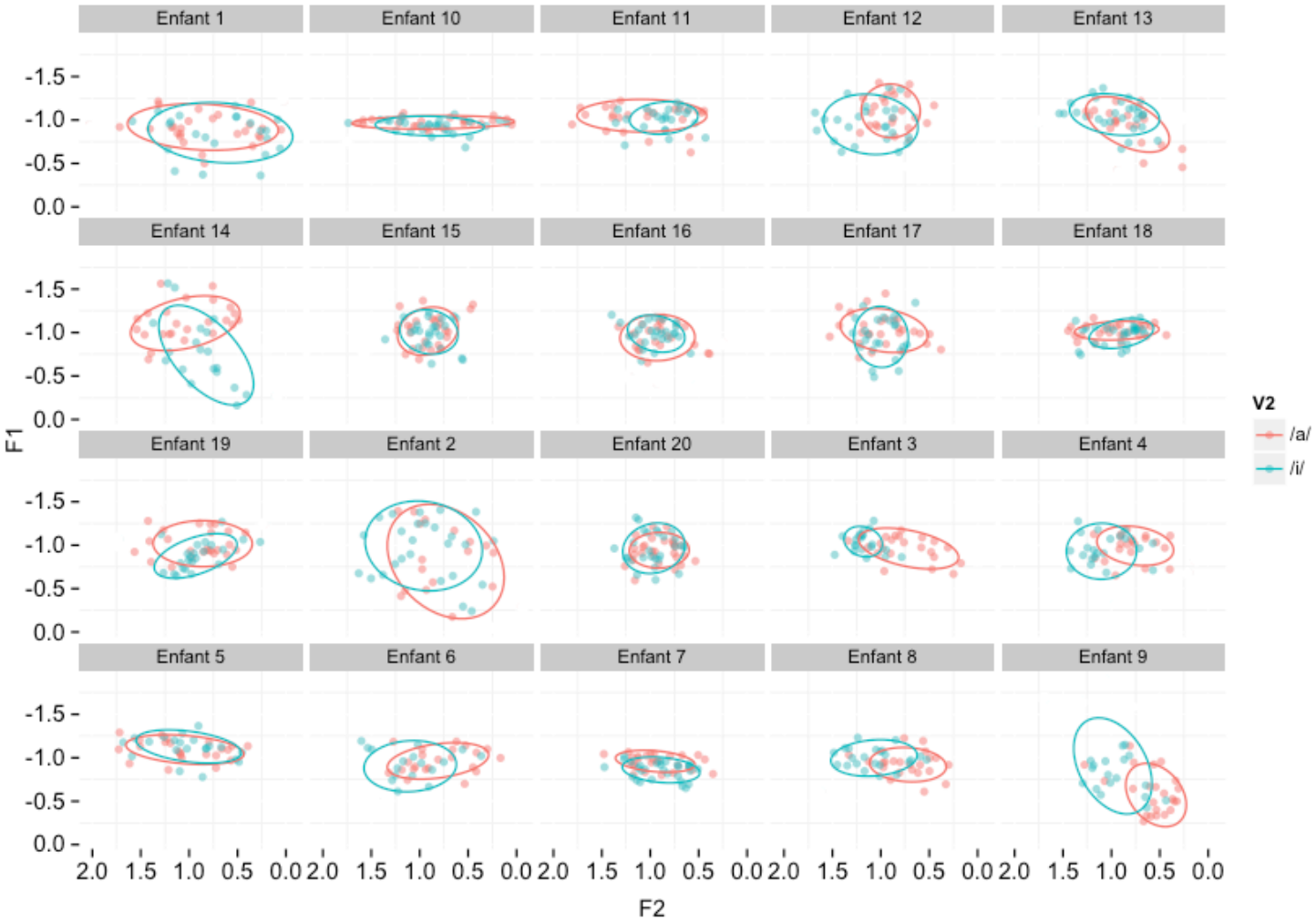


Figure 3.58. Anticipation de V2 dans la réalisation de V1 = /ɛ/ dans l'espace acoustique F1 -F2 normalisé, pour les participants enfants, tous contextes consonantiques confondus.

Les figures 3.57. et 3.58. permettent d'observer que l'effet d'anticipation, pour les participants enfants, n'est pas aussi clair que pour les participants adultes. Seuls quelques participants enfants montrent un effet clair d'anticipation : les participants Enfant 10, 14 et 2 pour le contexte V1 = /a/ et les participants Enfant 2, 4, 6, 8, et 9 pour le contexte V1 = /ɛ/ (d'après l'observation de ces figures uniquement).

Afin de quantifier, dans le domaine acoustique, la taille de l'effet d'anticipation, pour chaque participant enfant et chaque contexte de V1 séparément, nous proposons dans le tableau 3.11. ci-dessous, le calcul d'un indice individuel d'anticipation. Cet indice est calculé comme la moyenne des indices globaux pour V1 = /a/ et V1 = /ɛ/, indices représentant les distances euclidiennes séparant les centres des ellipses de dispersion des réalisations de V1 en fonction des deux contextes de V2.

Tableau 3.11. Mesures de la taille de l'effet en F1, en F2 et de l'effet global et indice individuel d'anticipation pour les participants enfants, tous contextes consonantiques confondus.

Indice de la taille de l'effet d'anticipation							
	V1 = /ɛ/			V1 = /a/			Indice individuel d'anticipation
Participant	Effet F1	Effet F2	Effet global	Effet F1	Effet F2	Effet global	
Enfant 1	-0,15	-0,17	**-0,22**	-0,02	0,04	**0,05**	**-0,09**
Enfant 2	0,18	0,28	**0,33**	0,07	0,37	**0,38**	**0,35**
Enfant 3	0,09	0,38	**0,39**	-0,19	0,02	**-0,19**	**0,10**
Enfant 4	-0,04	0,25	**0,25**	0,05	0,02	**0,06**	**0,15**
Enfant 5	0,04	-0,04	**0,06**	0,02	0,15	**0,15**	**0,10**
Enfant 6	0,00	0,29	**0,29**	-0,19	0,26	**0,32**	**0,31**
Enfant 7	-0,12	-0,01	**-0,12**	0,07	0,10	**0,12**	**0,00**
Enfant 8	0,06	0,35	**0,35**	0,25	0,21	**0,32**	**0,34**
Enfant 9	0,16	0,49	**0,52**	-0,07	0,37	**0,38**	**0,45**
Enfant 10	0,00	0,03	**0,03**	0,18	0,23	**0,29**	**0,16**
Enfant 11	0,01	-0,17	**-0,17**	0,00	0,07	**0,07**	**-0,05**
Enfant 12	-0,15	0,22	**0,27**	0,02	-0,07	**0,07**	**0,17**
Enfant 13	0,17	0,25	**0,30**	-0,13	0,18	**0,22**	**0,26**
Enfant 14	-0,30	-0,16	**-0,34**	0,09	0,25	**0,27**	**-0,03**
Enfant 15	-0,06	0,01	**0,06**	-0,12	0,25	**0,27**	**0,17**
Enfant 16	0,06	0,04	**0,07**	0,12	0,08	**0,14**	**0,11**
Enfant 17	-0,11	0,01	**-0,11**	-0,06	-0,05	**-0,08**	**-0,10**
Enfant 18	-0,04	-0,02	**-0,04**	0,13	0,18	**0,22**	**0,09**
Enfant 19	-0,12	0,05	**-0,13**	0,11	0,02	**0,12**	**-0,01**
Enfant 20	0,01	0,04	**0,04**	-0,01	-0,06	**-0,06**	**-0,01**
Moyenne	-0,02	0,12	**0,20**	0,02	0,13	**0,19**	**0,19**

Le tableau 3.11. rapporte les tailles d'effet en F1 et en F2 (mesurées par la distance entre les moyennes des productions en F1 et en F2 entre les 2 contextes de V2 étudiés (V2 =/i/ ou V2 = /a/) et la taille de l'effet global sur V1 (mesuré par la distance euclidienne entre les centres des ellipses de dispersion). Un signe a été ajouté à cet effet global, positif si l'effet va dans le sens de l'anticipation, négatif s'il va dans le sens opposé à l'anticipation de V2. Un indice individuel d'anticipation est proposé, calculé comme la moyenne de

l'effet global sur V1 = /ɛ/ et sur V1 = /a/. L'indice moyen d'anticipation pour les participants enfants est de **0.19**. Comparé au groupe Adulte, on observe une moindre anticipation, voire une absence d'anticipation de V2 dans la réalisation de V1 pour presque tous les participants enfants. Seuls cinq participants enfants semblent montrer un effet d'anticipation de V2 dans V1, les participants Enfant 2, Enfant 6, Enfant 8, Enfant 9 et Enfant 13.

Afin de tester statistiquement si les effets d'anticipation observés sont significatifs, nous avons conduit des analyses statistiques individuelles : des modèles linéaires prenant comme variables dépendantes F1 et F2 de chaque individu en fonction du contexte de V2. Les résultats de ces analyses statistiques sont reportées dans le tableau 3.12. ci-dessous. Nous avons choisi une valeur de $p = 0.01$, les résultats significatifs sont indiqués en gras.

Tableau 3.12. Analyses statistiques des effets d'anticipation en F1 et en F2 pour chaque participant enfant, tous contextes consonantiques confondus.

		Analyses statistiques des effets d'anticipation			
		V1 = /ɛ/		**V1 = /a/**	
Participant		**F1**	**F2**	**F1**	**F2**
Enfant 1	**F**	2.736	1.436	0.029	0.377
	p	0.105	0.236	0.866	0.564
Enfant 2	**F**	2.297	3.748	0.66	**21.05**
	p	0.137	0.0595	0.422	**5.53e-05**
Enfant 3	**F**	4.709	**36.11**	5.822	0.011
	p	0.0356	**3.55e-07**	0.02	0.915
Enfant 4	**F**	0.824	**19.12**	0.908	0
	p	0.369	**7.41e-05**	0.346	0.995
Enfant 5	**F**	0.342	0.166	0.433	3.228
	p	0.562	0.686	0.514	0.0782
Enfant 6	**F**	1.066	**8.347**	0.88	2.61
	p	0.308	**0.00635**	0.354	0.114
Enfant 7	**F**	**16.48**	0.186	0.179	0.242
	p	**0.000157**	0.668	0.674	0.625
Enfant 8	**F**	2.779	**17.64**	3.198	1.775
	p	0.102	**0.000124**	0.0793	0.188

Enfant 9	**F**	**12.72**	**28.12**	1.855	**21.74**
	p	**0.000903**	**3.74e-06**	0.18	**2.91e-05**
Enfant 10	**F**	6.154	0.005	3.743	**10.86**
	p	0.0164	0.946	0.0588	**0.00183**
Enfant 11	**F**	0.227	3.651	1.743	1.151
	p	0.636	0.0623	0.193	0.289
Enfant 12	**F**	6.344	**8.893**	0.004	0.127
	p	0.0152	**0.00452**	0.952	0.723
Enfant 13	**F**	6.202	5.338	4.038	4.065
	p	0.0163	0.0252	0.0499	0.0492
Enfant 14	**F**	**16.14**	4.699	0.804	**9.898**
	p	**0.000233**	0.0358	0.375	**0.0029**
Enfant 15	**F**	0.354	0.024	**7.319**	5.692
	p	0.554	0.877	**0.00911**	0.0206
Enfant 16	**F**	1.38	0.404	1.528	0.559
	p	0.245	0.528	0.222	0.458
Enfant 17	**F**	1.983	0.038	0.416	0.458
	p	0.165	0.847	0.522	0.502
Enfant 18	**F**	1.987	0.078	2.336	3.03
	p	0.164	0.782	0.132	0.0876
Enfant 19	**F**	6.445	0.289	1.493	0.879
	p	0.0142	0.593	0.228	0.353
Enfant 20	**F**	0.589	0.692	0.036	0.72
	p	0.446	0.409	0.85	0.4

On observe, d'après le tableau 3.12., que les effets d'anticipation observés ne sont pas statistiquement significatifs pour la majorité des participants enfants et la majorité des contextes étudiés. Cependant, quelques effets observés sont significatifs, en particulier pour le participant Enfant 9, dont l'effet d'anticipation est statistiquement significatif dans trois contextes sur quatre.

Afin d'illustrer l'effet de l'anticipation de V2 dans la réalisation de V1 dans le domaine articulatoire, et de vérifier d'un point de vue articulatoire ce que nous observons à travers les mesures acoustiques, nous présentons dans les figures 3.59. à 3.94. ci-dessous les

représentations statistiques des contours de langue correspondant au milieu de V1 et à la fin de V1 pour 6 participants enfants.

Enfant 1

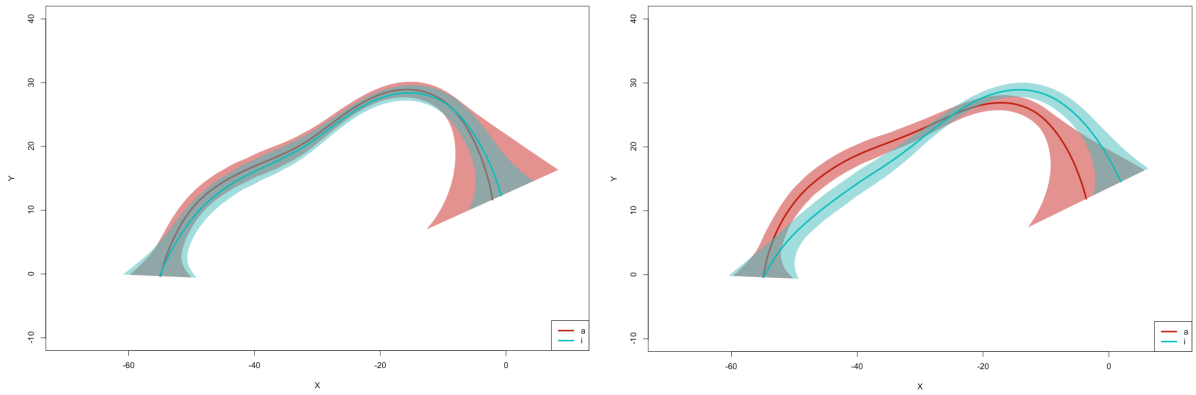


Figure 3.59. Représentation statistique des contours linguaux pour le participant Enfant 1 pour C = /b/ et V1 = /ɛ/, au milieu de V1 (panneau gauche) et à la fin de V1 (panneau droit), en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

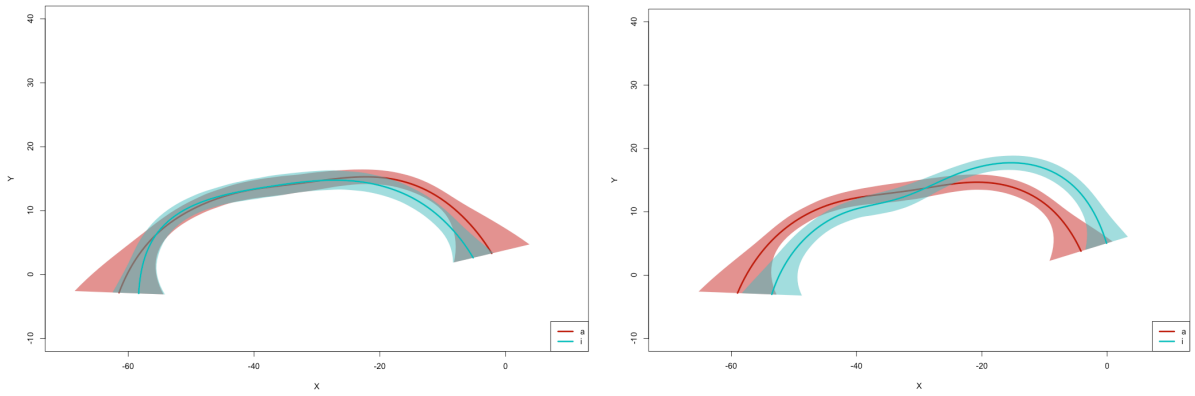


Figure 3.60. Représentation statistique des contours linguaux pour le participant Enfant 1 pour C = /b/ et V1 = /a/, au milieu de V1 (panneau gauche) et à la fin de V1 (panneau droit), en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

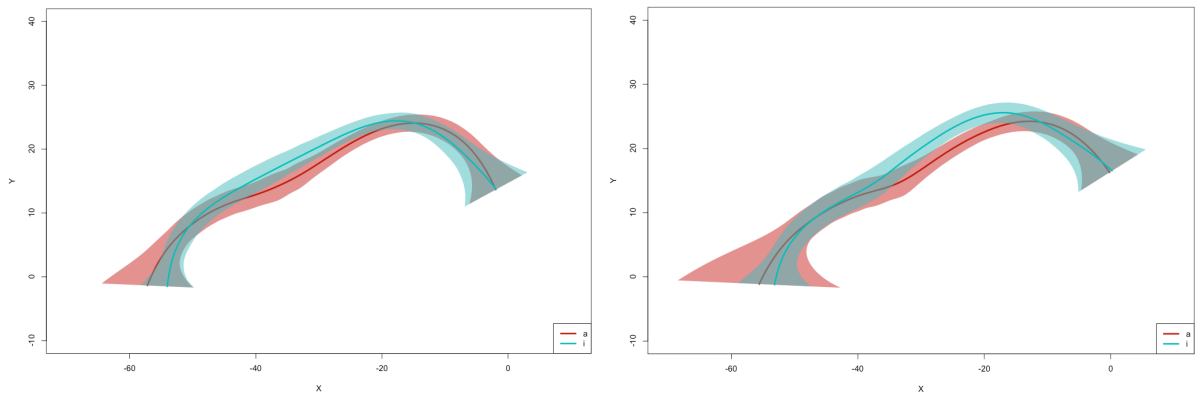


Figure 3.61. Représentation statistique des contours linguaux pour le participant Enfant 1 pour C = /d/ et V1 = /ɛ/, au milieu de V1 (panneau gauche) et à la fin de V1 (panneau droit), en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

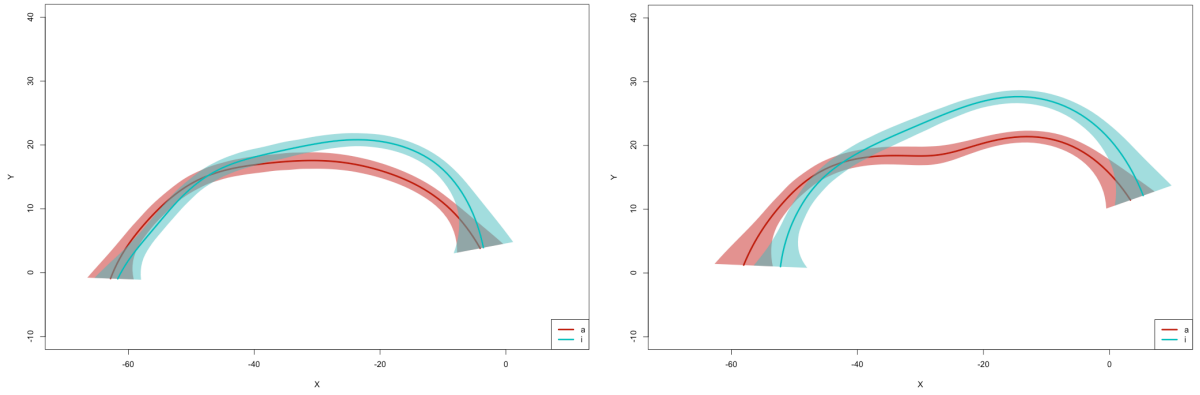


Figure 3.62. Représentation statistique des contours linguaux pour le participant Enfant 1 pour C = /d/ et V1 = /a/, au milieu de V1 (panneau gauche) et à la fin de V1 (panneau droit), en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

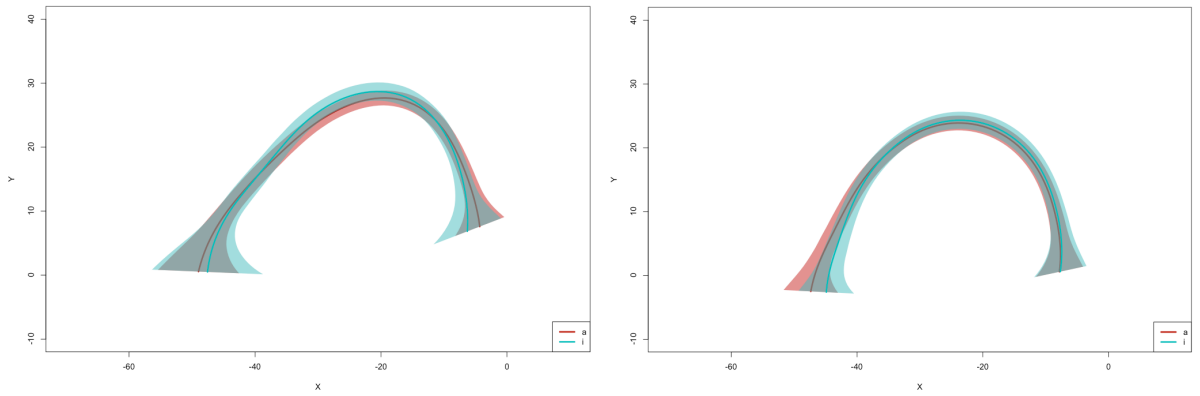


Figure 3.63. Représentation statistique des contours linguaux pour le participant Enfant 1 pour C = /g/ et V1 = /ɛ/, au milieu de V1 (panneau gauche) et à la fin de V1 (panneau droit), en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

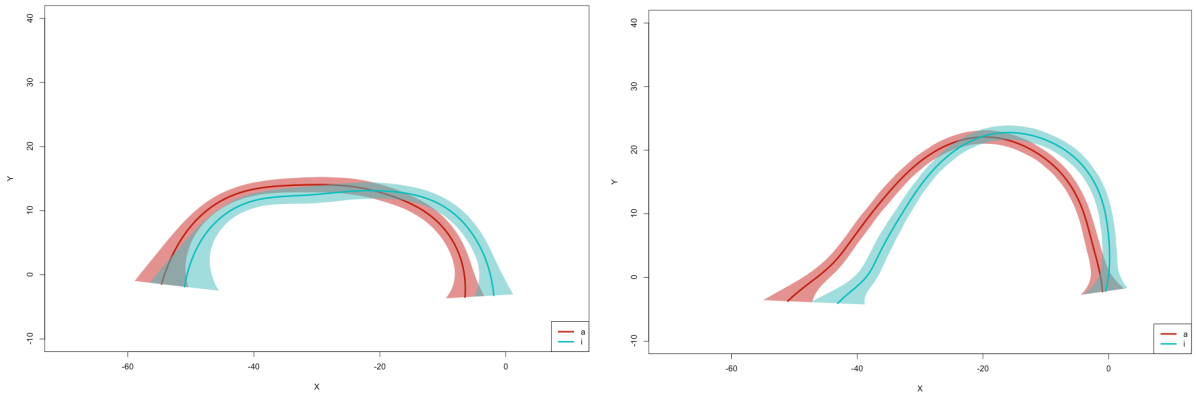


Figure 3.64. Représentation statistique des contours linguaux pour le participant Enfant 1 pour C = /g/ et V1 = /a/, au milieu de V1 (panneau gauche) et à la fin de V1 (panneau droit), en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

On observe, chez ce participant, un effet significatif de V2 sur la réalisation de V1 (au milieu de V1) dans les contextes où C = /d/ et V1 = /a/ et C = /g/ et V1 = /a/. L'effet d'anticipation est donc statistiquement significatif dans 2 contextes sur 6, et non significatif dans 4 contextes sur 6.

Enfant 2

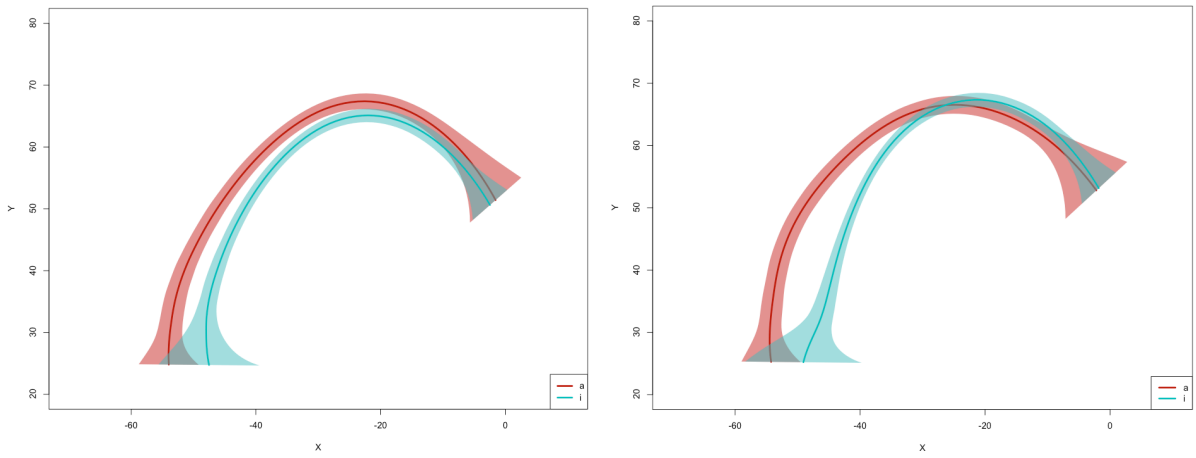


Figure 3.65. Représentation statistique des contours linguaux pour le participant Enfant 2 pour C = /b/ et V1 = /ɛ/, au milieu de V1 (panneau gauche) et à la fin de V1 (panneau droit), en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

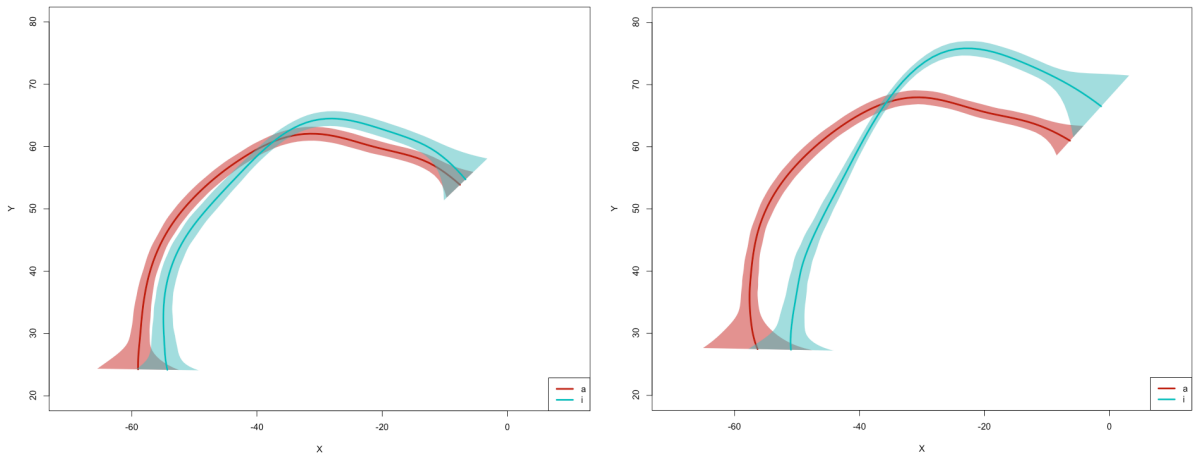


Figure 3.66. Représentation statistique des contours linguaux pour le participant Enfant 2 pour C = /b/ et V1 = /a/, au milieu de V1 (panneau gauche) et à la fin de V1 (panneau droit), en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

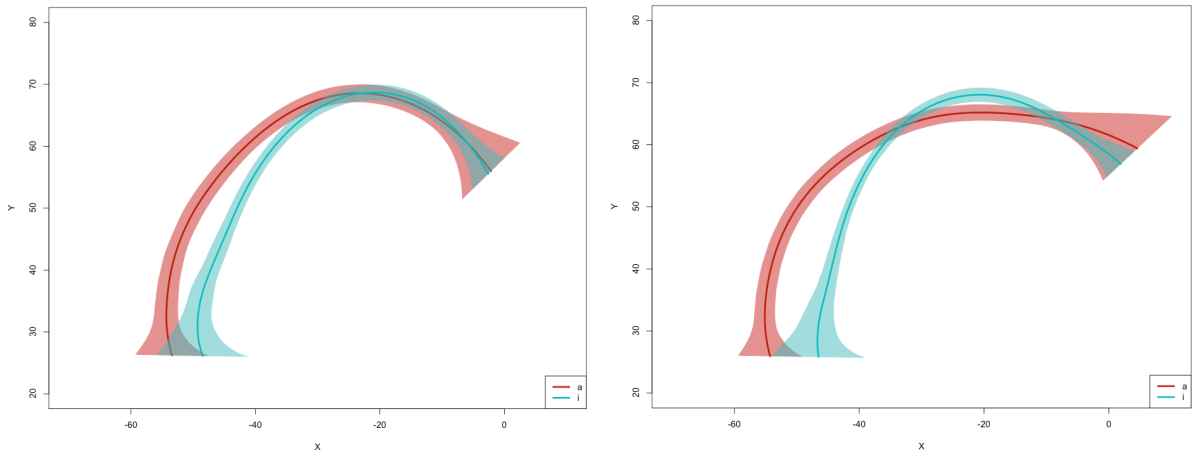


Figure 3.67. Représentation statistique des contours linguaux pour le participant Enfant 2 pour C = /d/ et V1 = /ɛ/, au milieu de V1 (panneau gauche) et à la fin de V1 (panneau droit), en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

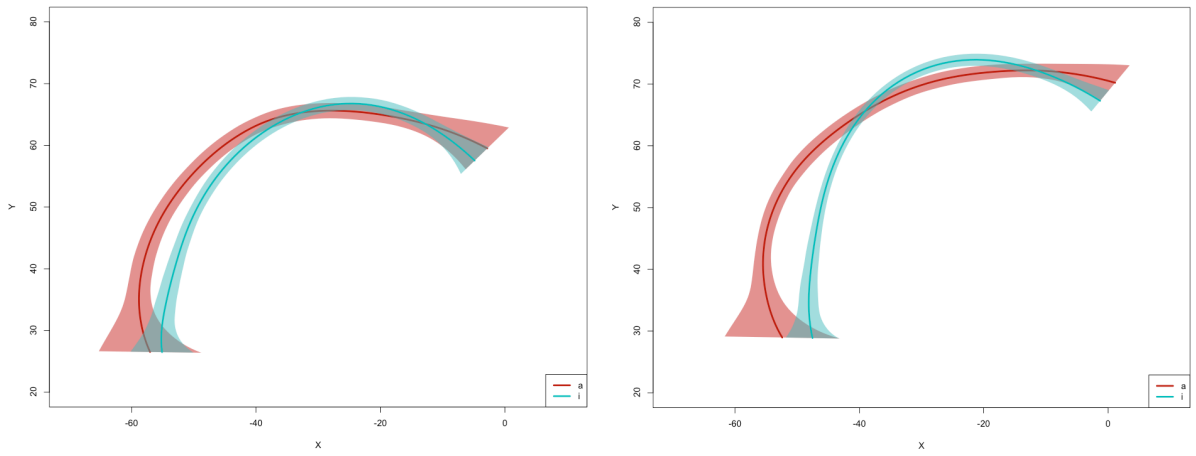


Figure 3.68. Représentation statistique des contours linguaux pour le participant Enfant 2 pour C = /d/ et V1 = /a/, au milieu de V1 (panneau gauche) et à la fin de V1 (panneau droit), en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

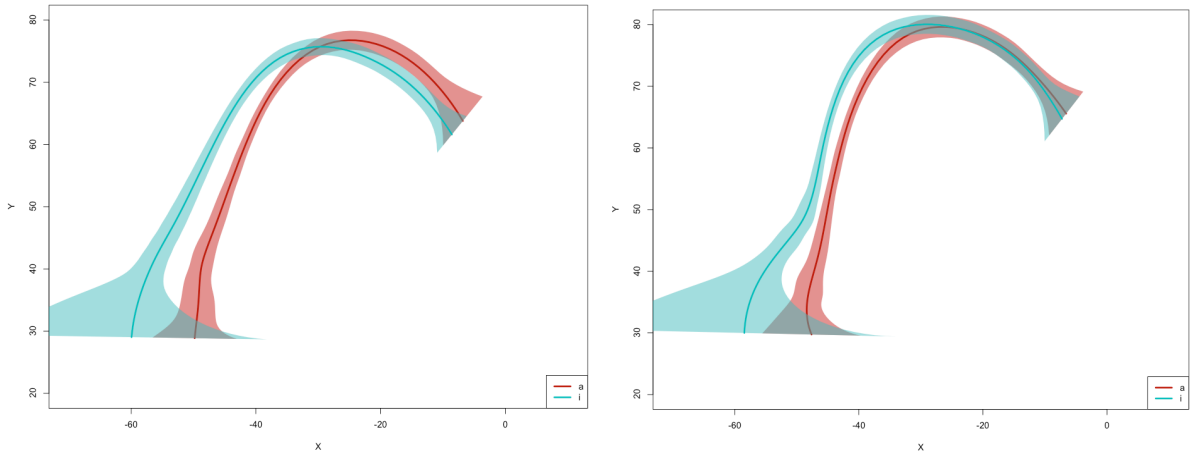


Figure 3.69. Représentation statistique des contours linguaux pour le participant Enfant 2 pour C = /g/ et V1 = /ɛ/, au milieu de V1 (panneau gauche) et à la fin de V1 (panneau droit), en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

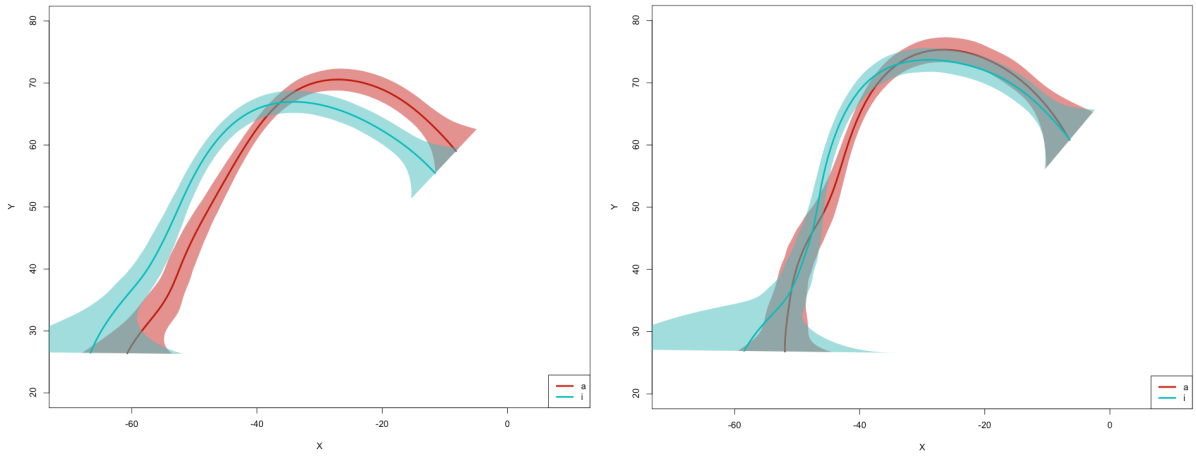


Figure 3.70. Représentation statistique des contours linguaux pour le participant Enfant 2 pour C = /g/ et V1 = /a/, au milieu de V1 (panneau gauche) et à la fin de V1 (panneau droit), en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

On observe, chez ce participant, un effet significatif de V2 sur la réalisation de V1 (au milieu de V1) dans les contextes où C = /b/ et V1 = /a/ ; et pour les contextes où C = /d/. On observe également un effet de V2 sur la réalisation de V1 dans le contexte où C = /g/ (et dans le contexte C = /b/ et V1 = /ɛ/), mais cet effet va dans la direction opposée à

l'anticipation. On observe donc, pour ce participant, un effet d'anticipation de V2 dans la réalisation de V1 dans 3 des 6 contextes étudiés.

Enfant 5

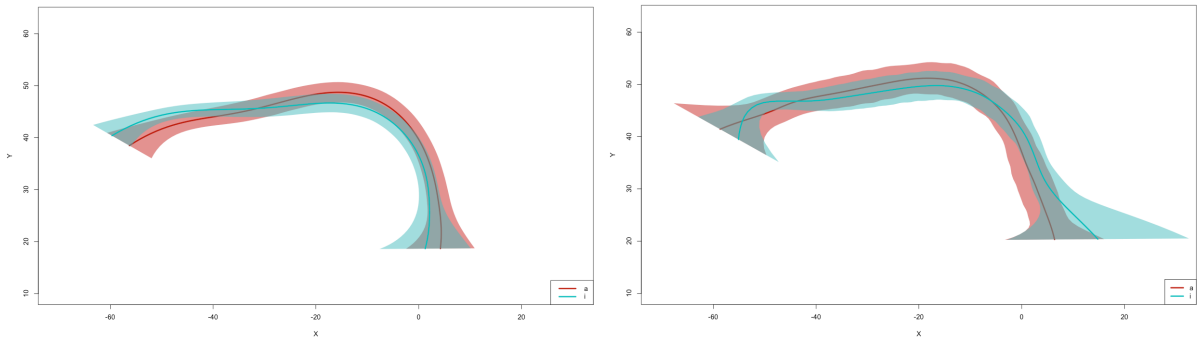


Figure 3.71. Représentation statistique des contours linguaux pour le participant Enfant 5 pour C = /b/ et V1 = /ɛ/, au milieu de V1 (panneau gauche) et à la fin de V1 (panneau droit), en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

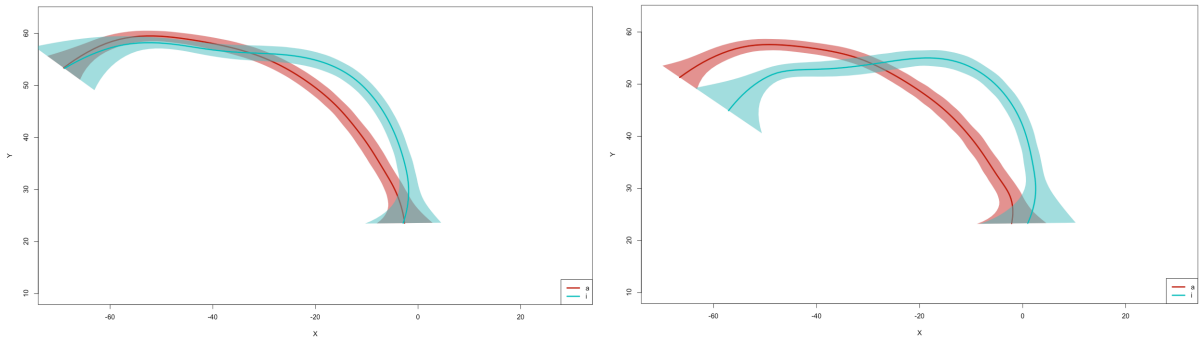


Figure 3.72. Représentation statistique des contours linguaux pour le participant Enfant 5 pour C = /b/ et V1 = /a/, au milieu de V1 (panneau gauche) et à la fin de V1 (panneau droit), en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

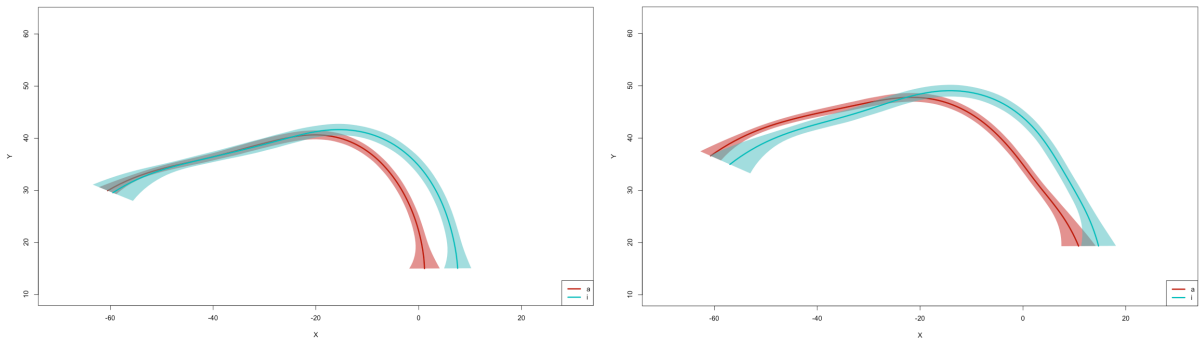


Figure 3.73. Représentation statistique des contours linguaux pour le participant Enfant 5 pour C = /d/ et V1 = /ɛ/, au milieu de V1 (panneau gauche) et à la fin de V1 (panneau droit), en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

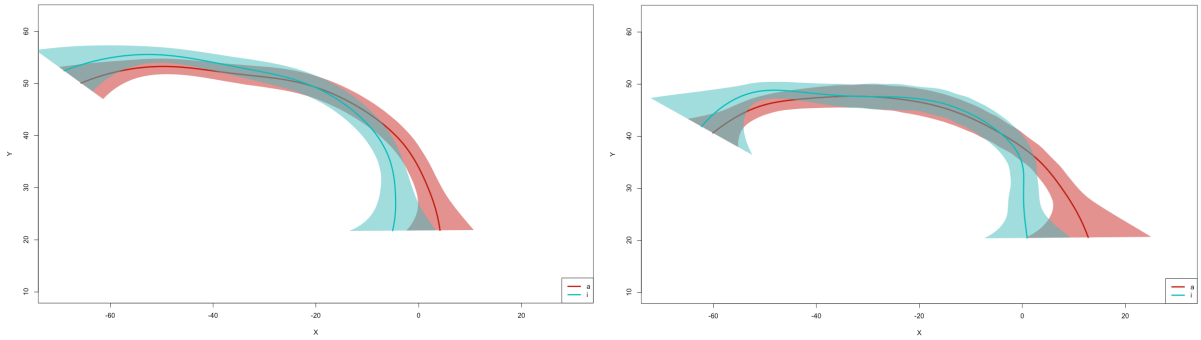


Figure 3.74. Représentation statistique des contours linguaux pour le participant Enfant 5 pour C = /d/ et V1 = /a/, au milieu de V1 (panneau gauche) et à la fin de V1 (panneau droit), en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

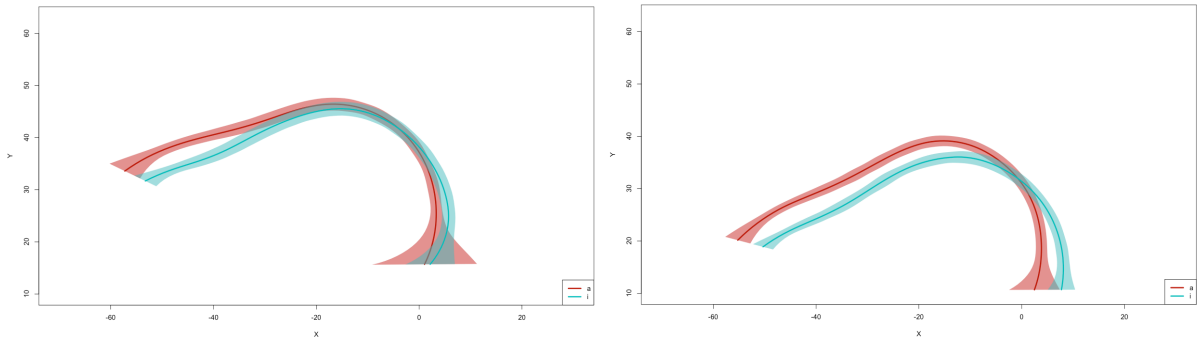


Figure 3.75. Représentation statistique des contours linguaux pour le participant Enfant 5 pour C = /g/ et V1 = /ɛ/, au milieu de V1 (panneau gauche) et à la fin de V1 (panneau droit), en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

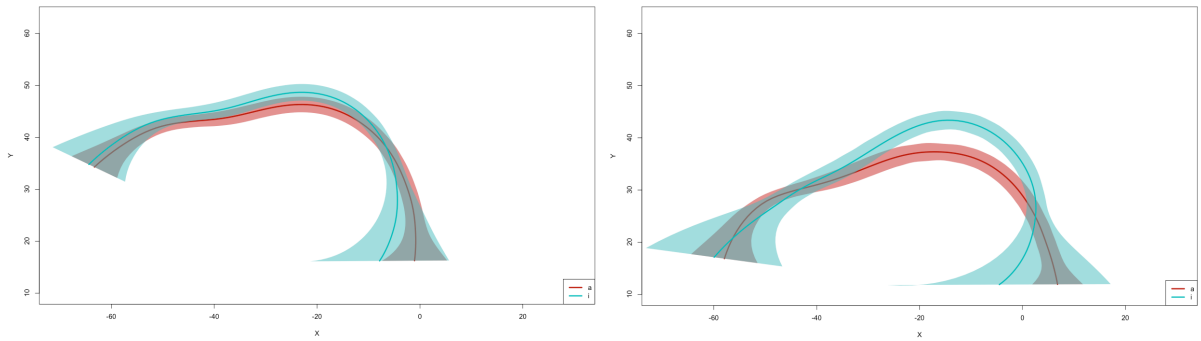


Figure 3.76. Représentation statistique des contours linguaux pour le participant Enfant 5 pour C = /g/ et V1 = /a/, au milieu de V1 (panneau gauche) et à la fin de V1 (panneau droit), en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

On observe, chez ce participant, un effet significatif de V2 sur la réalisation de V1 (au milieu de V1) dans les contextes où C = /b/ et V1 = /a/ ; et dans les contextes C = /d/ et V1 = /ɛ/ et dans le contexte C = /g/ et V1 = /ɛ/ (soit 3 des 6 contextes étudiés).

Enfant 6

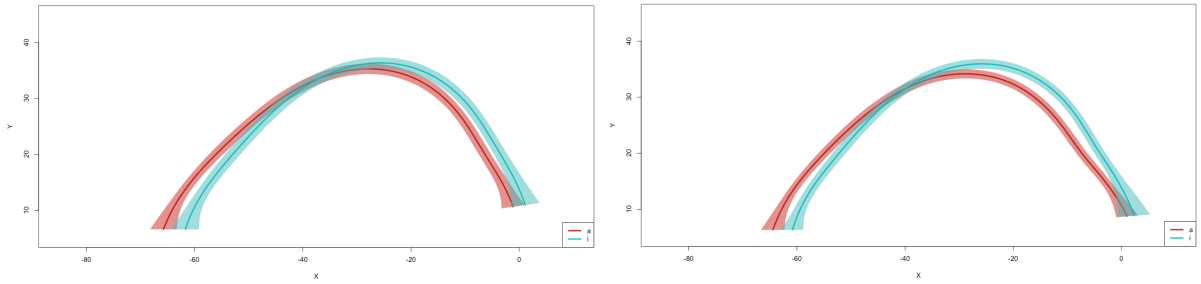


Figure 3.77. Représentation statistique des contours linguaux pour le participant Enfant 6 pour C = /b/ et V1 = /ɛ/, au milieu de V1 (panneau gauche) et à la fin de V1 (panneau droit), en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

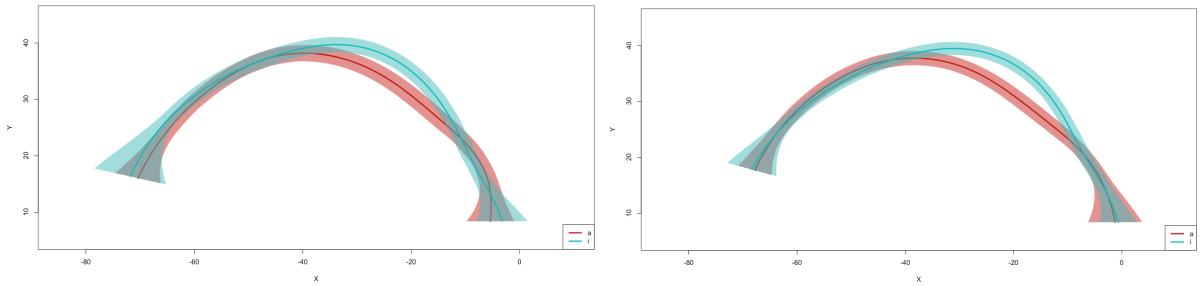


Figure 3.78. Représentation statistique des contours linguaux pour le participant Enfant 6 pour C = /b/ et V1 = /a/, au milieu de V1 (panneau gauche) et à la fin de V1 (panneau droit), en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

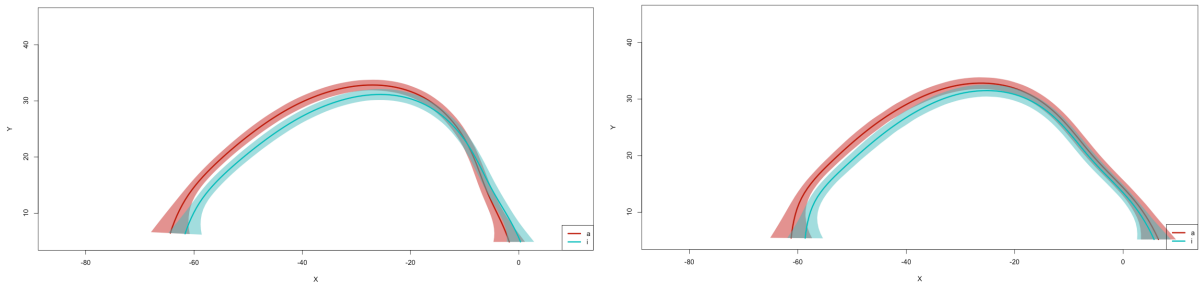


Figure 3.79. Représentation statistique des contours linguaux pour le participant Enfant 6 pour C = /d/ et V1 = /ɛ/, au milieu de V1 (panneau gauche) et à la fin de V1 (panneau droit), en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

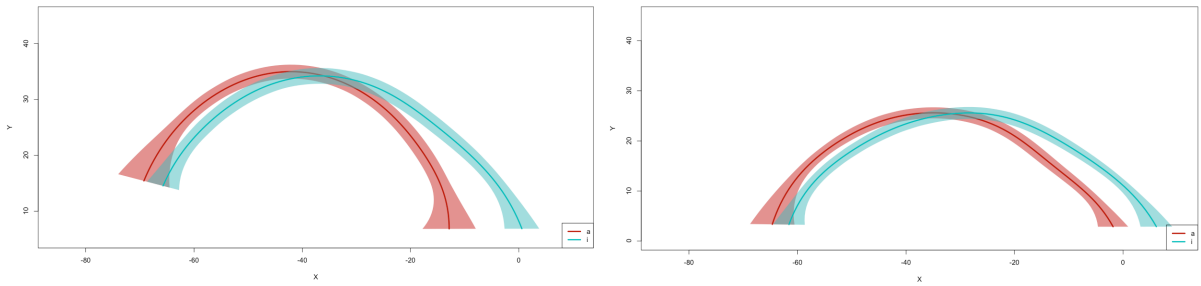


Figure 3.80. Représentation statistique des contours linguaux pour le participant Enfant 6 pour C = /d/ et V1 = /a/, au milieu de V1 (panneau gauche) et à la fin de V1 (panneau droit), en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

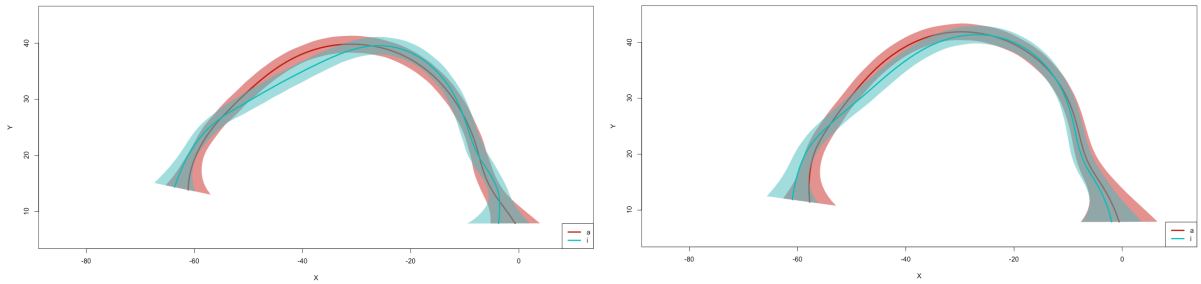


Figure 3.81. Représentation statistique des contours linguaux pour le participant Enfant 6 pour C = /g/ et V1 = /ɛ/, au milieu de V1 (panneau gauche) et à la fin de V1 (panneau droit), en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

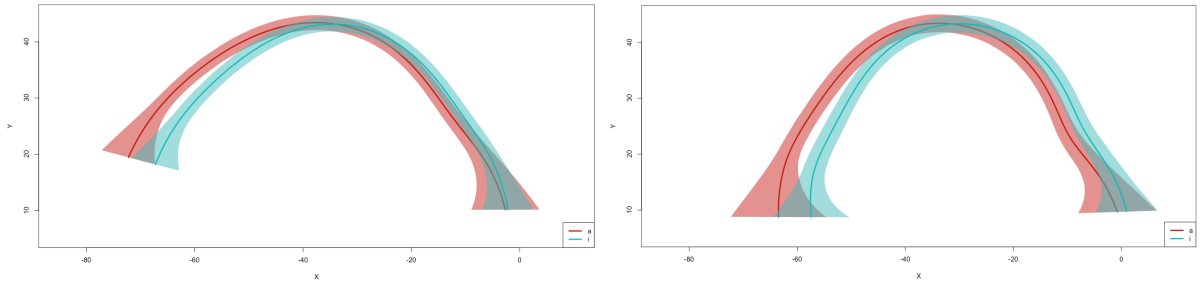


Figure 3.82. Représentation statistique des contours linguaux pour le participant Enfant 6 pour C = /g/ et V1 = /a/, au milieu de V1 (panneau gauche) et à la fin de V1 (panneau droit), en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

On observe, chez ce participant, un effet significatif de V2 sur la réalisation de V1 (au milieu de V1) dans tous les contextes sauf pour le contexte C = /g/ et V1 = /ɛ/ (soit 5 des 6 contextes étudiés).

Enfant 7

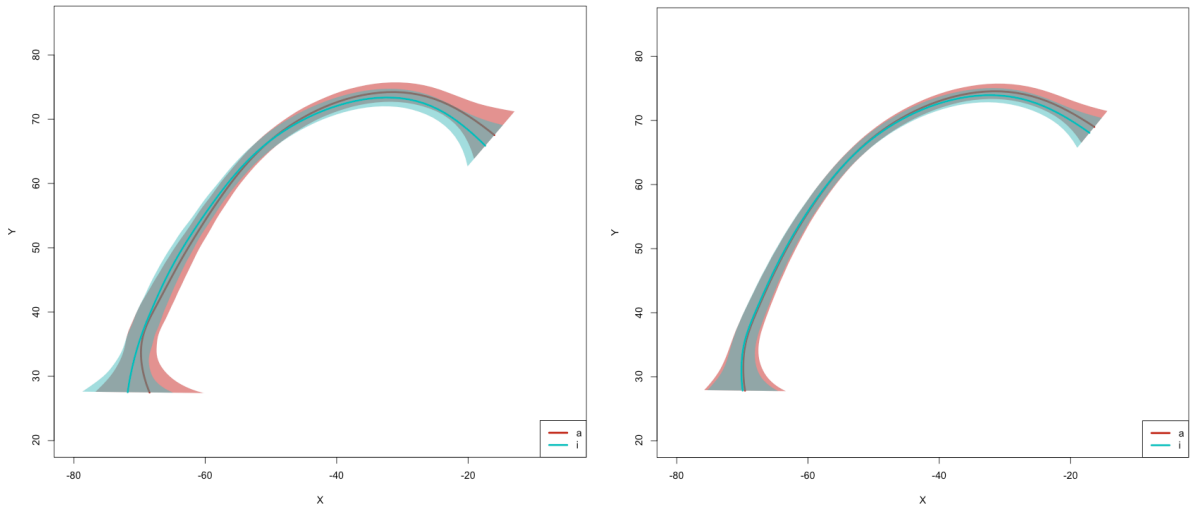


Figure 3.83. Représentation statistique des contours linguaux pour le participant Enfant 7 pour C = /b/ et V1 = /ɛ/, au milieu de V1 (panneau gauche) et à la fin de V1 (panneau droit), en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

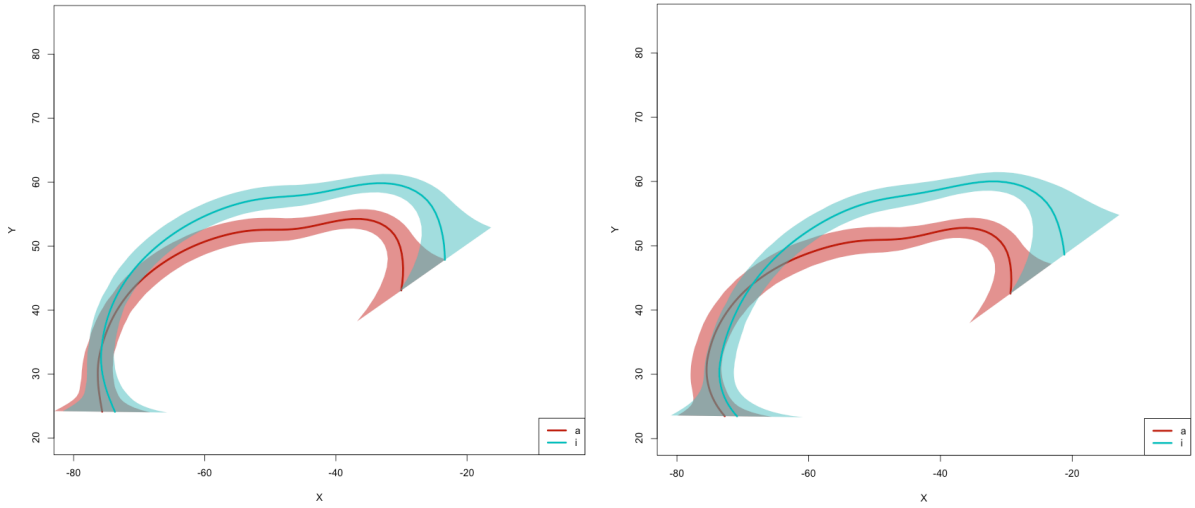


Figure 3.84. Représentation statistique des contours linguaux pour le participant Enfant 7 pour C = /b/ et V1 = /a/, au milieu de V1 (panneau gauche) et à la fin de V1 (panneau droit), en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

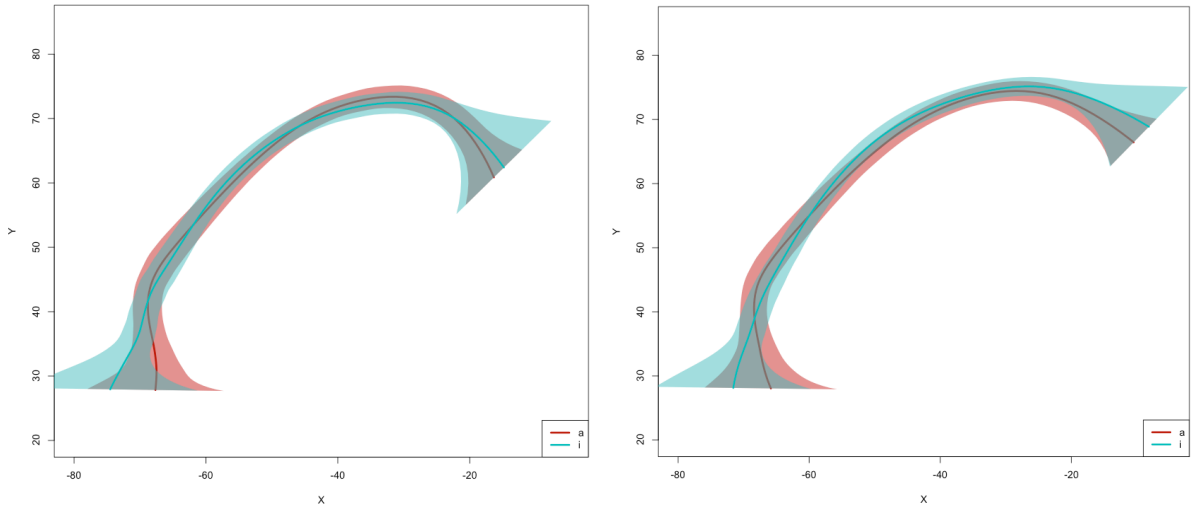


Figure 3.85. Représentation statistique des contours linguaux pour le participant Enfant 7 pour C = /d/ et V1 = /ɛ/, au milieu de V1 (panneau gauche) et à la fin de V1 (panneau droit), en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

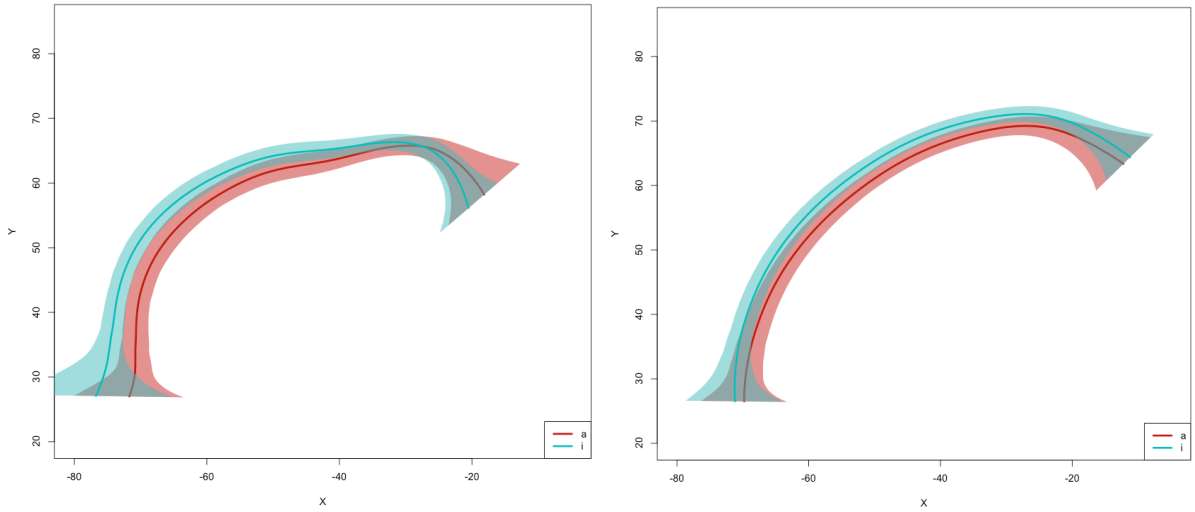


Figure 3.86. Représentation statistique des contours linguaux pour le participant Enfant 7 pour C = /d/ et V1 = /a/, au milieu de V1 (panneau gauche) et à la fin de V1 (panneau droit), en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

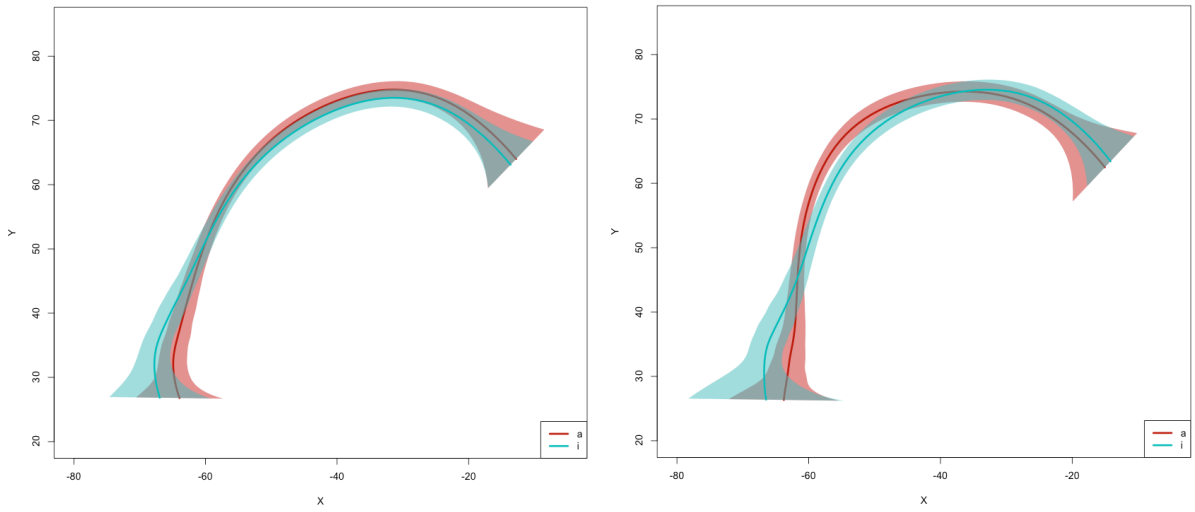


Figure 3.87. Représentation statistique des contours linguaux pour le participant Enfant 7 pour C = /g/ et V1 = /ɛ/, au milieu de V1 (panneau gauche) et à la fin de V1 (panneau droit), en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

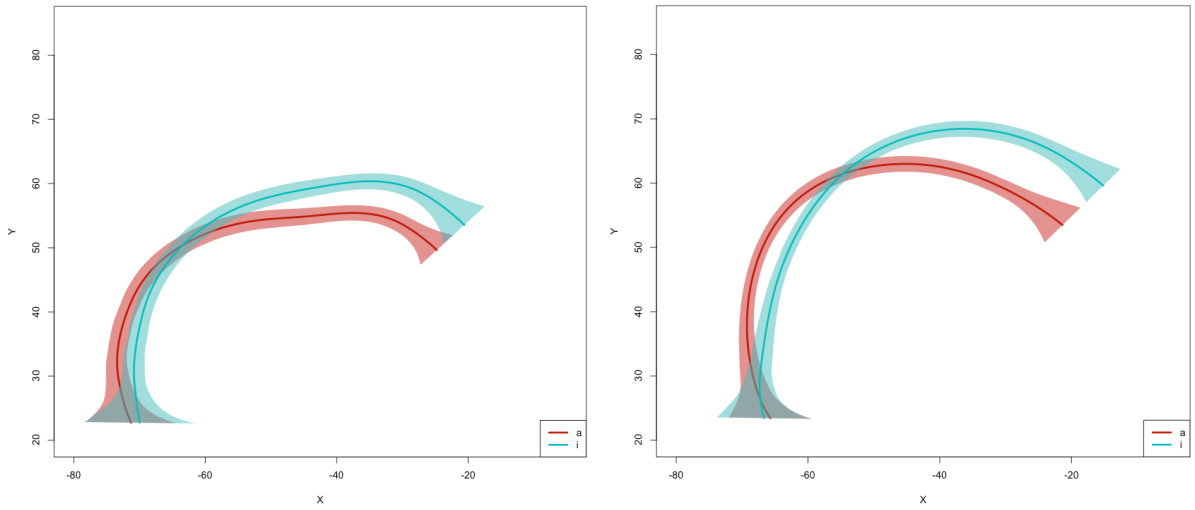


Figure 3.88. Représentation statistique des contours linguaux pour le participant Enfant 7 pour C = /g/ et V1 = /a/, au milieu de V1 (panneau gauche) et à la fin de V1 (panneau droit), en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

On observe, chez ce participant, un effet significatif de V2 sur la réalisation de V1 (au milieu de V1) pour les contextes C = /b/ et V1 = /a/ et C = /g/ et V1 = /a/ (soit 2 des 6 contextes étudiés).

Enfant 9

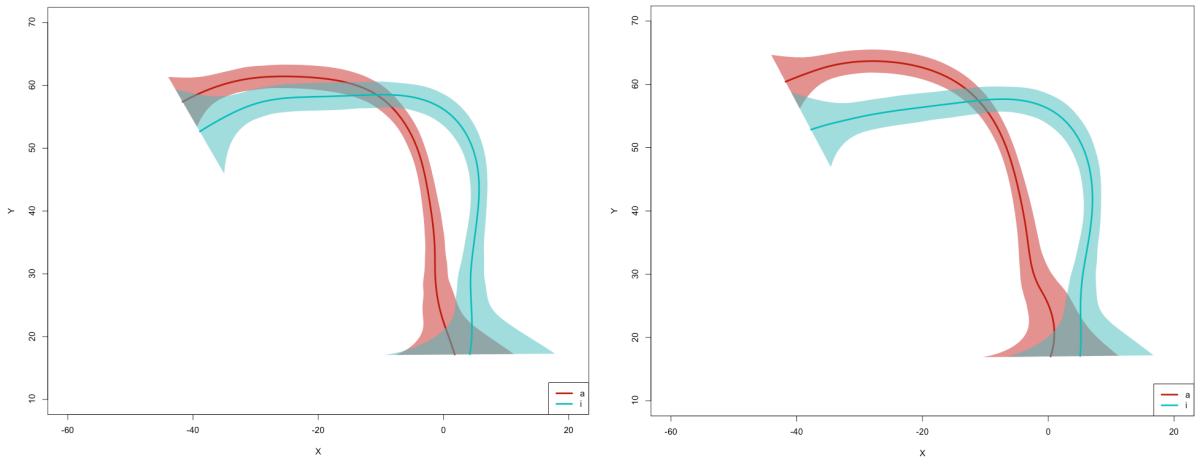


Figure 3.89. Représentation statistique des contours linguaux pour le participant Enfant 9 pour C = /b/ et V1 = /ɛ/, au milieu de V1 (panneau gauche) et à la fin de V1 (panneau droit), en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

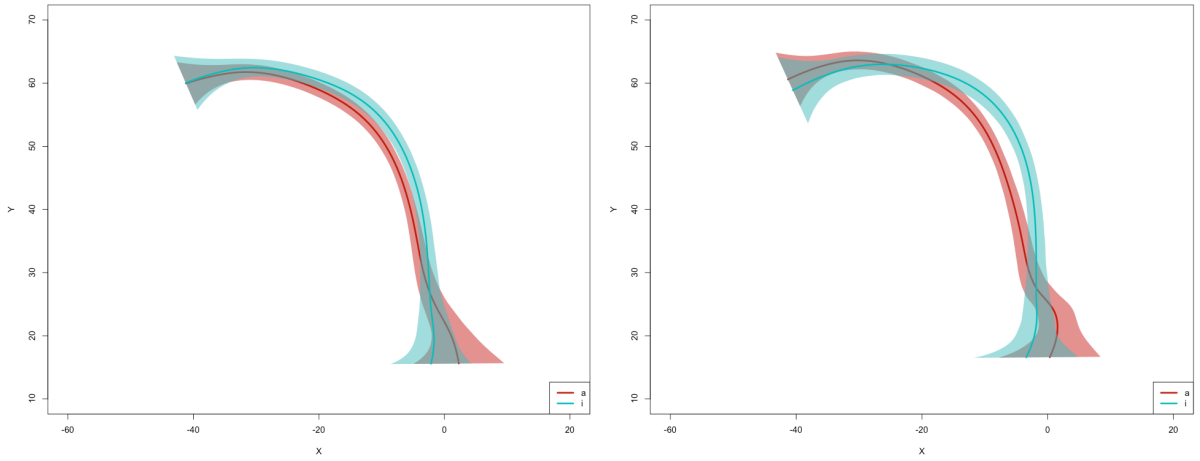


Figure 3.90. Représentation statistique des contours linguaux pour le participant Enfant 9 pour C = /b/ et V1 = /a/, au milieu de V1 (panneau gauche) et à la fin de V1 (panneau droit), en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

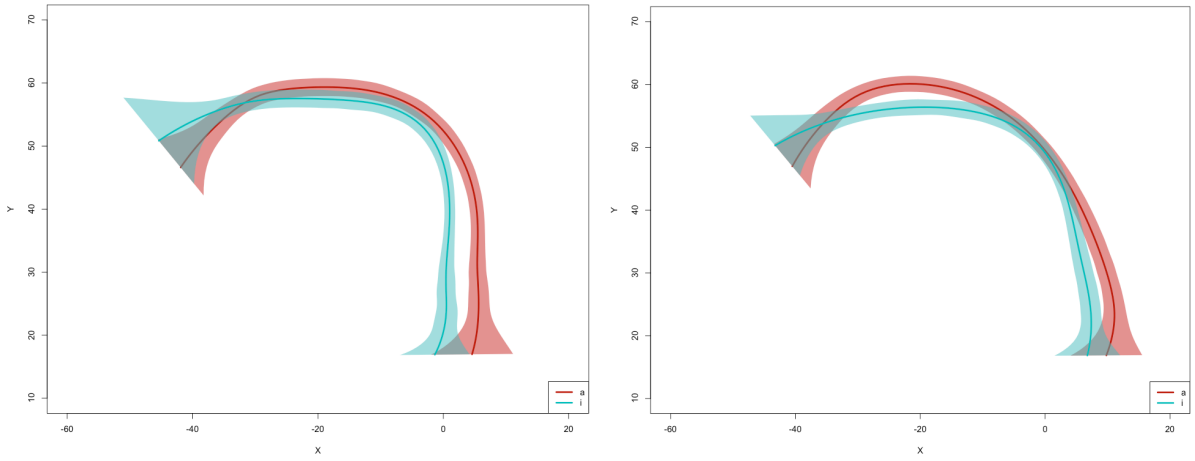


Figure 3.91. Représentation statistique des contours linguaux pour le participant Enfant 9 pour C = /d/ et V1 = /ɛ/, au milieu de V1 (panneau gauche) et à la fin de V1 (panneau droit), en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

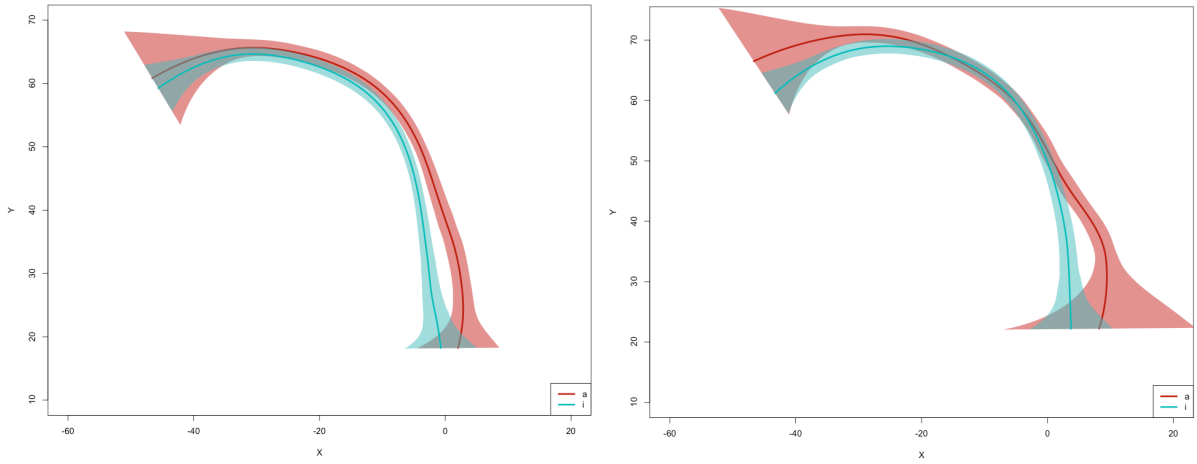


Figure 3.92. Représentation statistique des contours linguaux pour le participant Enfant 9 pour C = /d/ et V1 = /a/, au milieu de V1 (panneau gauche) et à la fin de V1 (panneau droit), en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

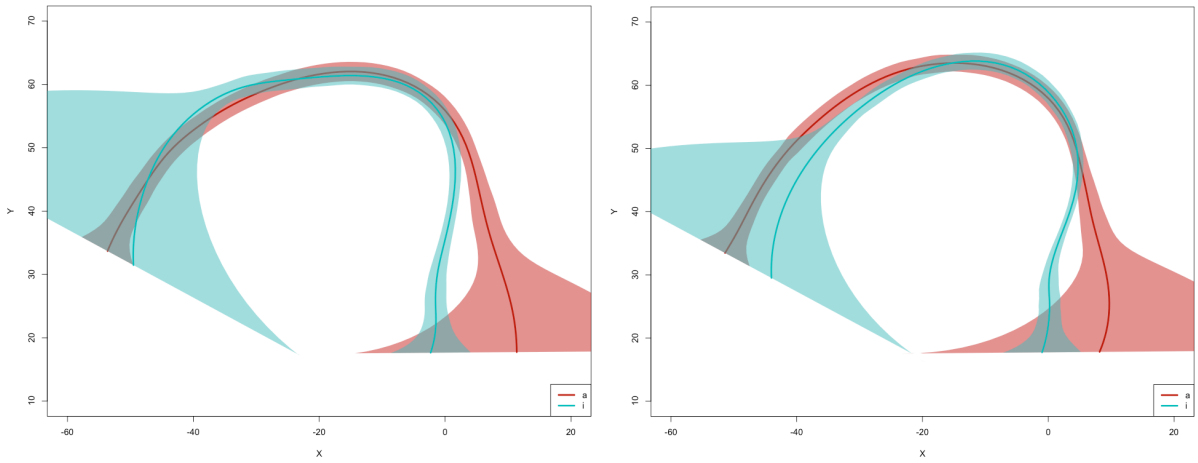


Figure 3.93. Représentation statistique des contours linguaux pour le participant Enfant 9 pour C = /g/ et V1 = /ɛ/, au milieu de V1 (panneau gauche) et à la fin de V1 (panneau droit), en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

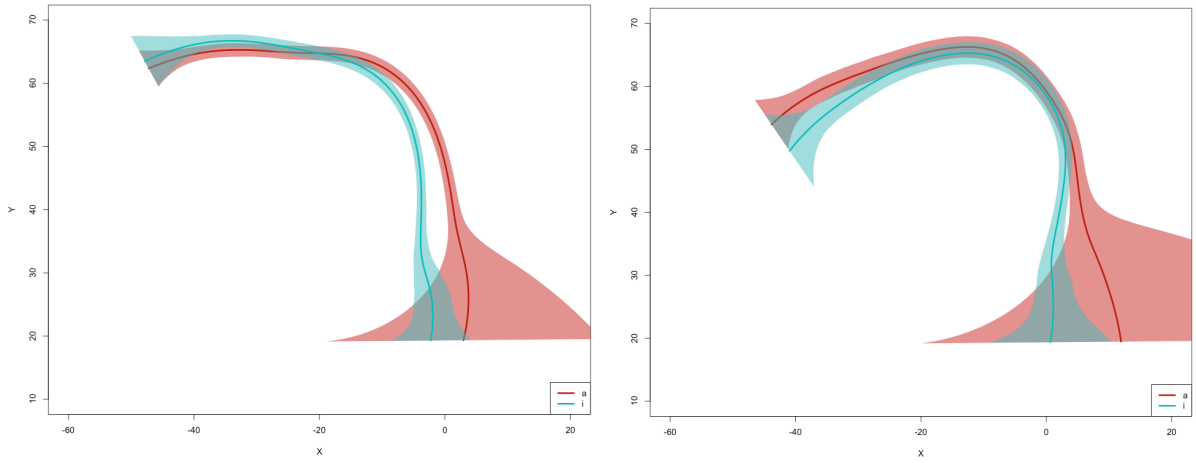


Figure 3.94. Représentation statistique des contours linguaux pour le participant Enfant 9 pour C = /g/ et V1 = /a/, au milieu de V1 (panneau gauche) et à la fin de V1 (panneau droit), en contextes V2 = /a/ (rouge) et V2 = /i/ (bleu).

On observe, chez ce participant, un effet significatif de V2 sur la réalisation de V1 (au milieu de V1) pour le contextes C = /b/ et V1 = /ɛ/. Pour les contextes C = /d/ et C = /g/, un effet est observé, mais cet effet va dans la direction opposée à l'anticipation. On observe donc, chez ce participant, un effet significatif de l'anticipation dans 1 seul des 6 contextes étudiés.

Dans l'ensemble, l'anticipation de V2 dans la réalisation de V1 n'est pas systématique chez les participants enfants. Certains participants, à l'image de l'Enfant 6, anticipent effectivement dans la majorité des contextes étudiés, mais la plupart des participants enfants n'anticipe que dans quelques contextes. L'absence de toute tendance allant dans la direction de l'anticipation (à l'image des enfants 1 et 7) semble être un caractère propre aux enfants, que nous n'avons pas observé dans les comportements adultes.

3.3.3. Comparaison entre adultes et enfants

Nous présentons, dans la figure 3.95. ci-dessous, les réalisations de V1 = /a/ dans l'espace acoustique F1 - F2 normalisé, dans les contextes où V2 = /i/ (bleu), et V2 = /a/ (rouge), pour l'ensemble des participants adultes (panneau de gauche) et l'ensemble des participants enfants (panneau de droite), tous contextes consonantiques confondus.

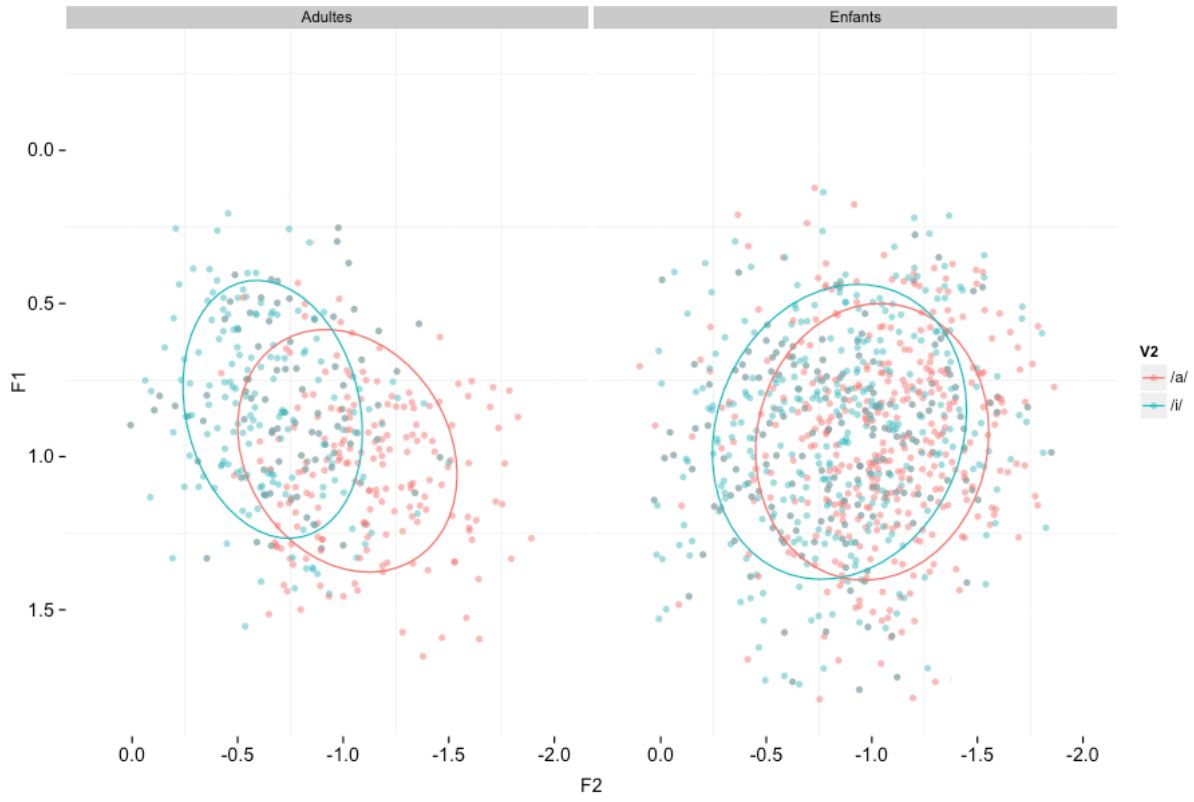


Figure 3.95. Anticipation de V2 dans la réalisation de V1 = /a/ dans l'espace acoustique F1 -F2 normalisé, pour tous les participants adultes (panneau de gauche) et tous les participants enfants (panneau de droite), tous contextes consonantiques confondus.

Nous présentons, dans la figure 3.96. ci-dessous, les réalisations de V1 = /ɛ/ dans l'espace acoustique F1 - F2 normalisé, dans les contextes où V2 = /i/ (bleu), et V2 = /a/ (rouge), pour l'ensemble des participants adultes (panneau de gauche) et l'ensemble des participants enfants (panneau de droite), tous contextes consonantiques confondus.

Figure 3.96. Anticipation de V2 dans la réalisation de V1 = /ɛ/ dans l'espace acoustique F1 -F2 normalisé, pour tous les participants adultes (panneau gauche) et tous les participants enfants (panneau droit), tous contextes consonantiques confondus.

On observe visuellement d'après les figures 3.95. et 3.96. un effet de V2 sur la réalisation de V1 pour le groupe Adulte, mais pas pour le groupe Enfant. L'effet sur V1 = /a/ semble être prédominant dans la dimension F2 et l'effet sur V1 = /ɛ/ semble être prédominant dans la dimension F2. Afin de mesurer la taille de ces effets et leur significativité, nous avons conduit des analyses statistiques à l'aide de modèles linéaires à effets mixtes et de modèles linéaires.

Les analyses statistiques conduites avec deux modèles linéaires à effets mixtes (un premier pour F1 et un second pour F2, pour lesquels V1, V2 et Groupe sont des facteurs indépendants à effet fixe et Participant est un facteur aléatoire) ont révélé que l'interaction entre V2 et le facteur Groupe est significative (pour F1, $F = 107.4371$, $p < 0.00001$; pour F2, $F = 85.8757$, $p < 0.00001$). Cela indique que l'effet d'anticipation varie de façon significative entre les deux groupes.

Les analyses conduites avec 4 modèles linéaires (pour F1 et F2 séparément et pour les groupes Adulte et Enfant séparément) révèlent que l'effet de V2 sur V1 est significatif en F1 (F (1, 1072) = 14,72 ; p < 0.0001323) et en F2 (F (1, 1072) = 35.99 ; p < 2.709e-09) pour le groupe Adulte, et n'est pas significatif en F1 (F (1, 2086) = 0.1084 ; p < 0.742) mais est significatif en F2 (F (1, 2086) = 8.463 ; p < 0.003663) pour le groupe Enfant.

Concernant la taille de ces effets, les tests post-hoc par paires ont révélé que quand V2 = /i/, les valeurs z-scorées de F1 diminuent en moyenne de 0.196 (comparées à V2 = /a/) et les valeurs de F2 augmentent en moyenne de 0.636. Pour les enfants, comparés aux adultes, la diminution des valeurs de F1 liée à l'effet de V2 est réduite de 0.168 ; l'augmentation de F2 est réduite de 0.405. La taille de l'effet est importante dans les deux dimensions pour le groupe Adulte et est non significative pour F1 et très réduite pour F2 pour le groupe Enfant.

Après cette analyse globale, nous proposons de comparer les indices numériques individuels d'anticipation, calculés dans le domaine acoustique. Nous présentons dans la figure 3.97. ci-dessous ces indices numériques individuels d'anticipation pour chacun des participants en fonction de l'âge.

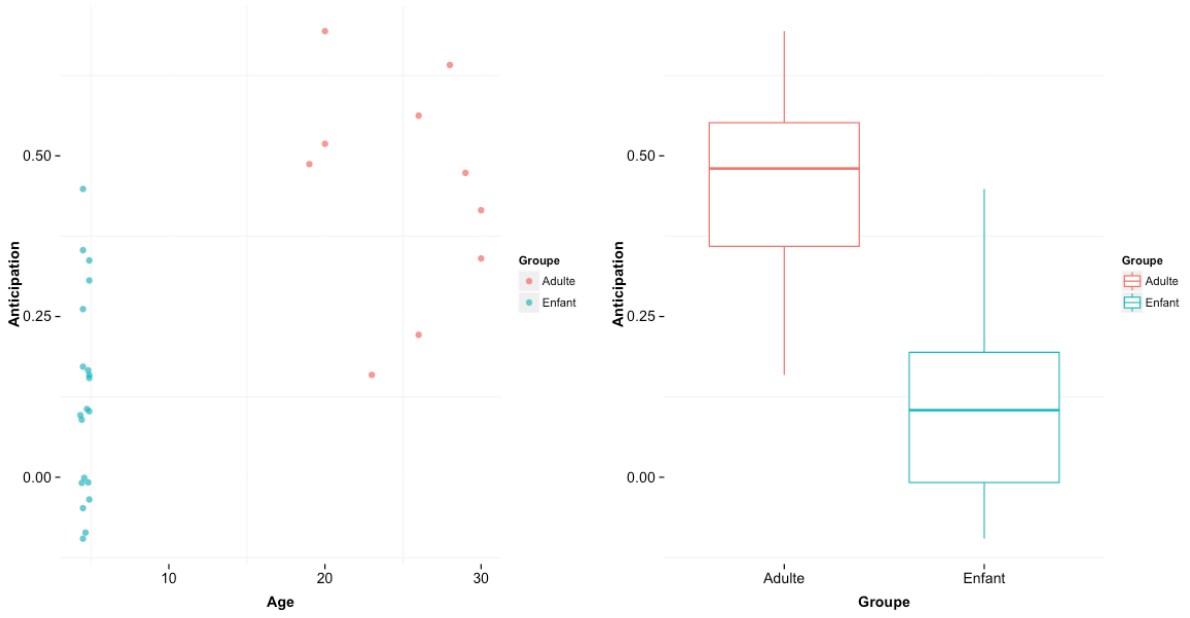


Figure 3.97. Indices numériques individuels d'anticipation (calculés dans le domaine acoustique) en fonction de l'âge, pour chacun des participants, groupe Adulte en rouge, groupe Enfant en bleu.

On observe une grande variabilité interindividuelle concernant cet indice d'anticipation, que ce soit au sein du groupe Adulte ou Enfant. Malgré cette grande variabilité, on observe que le groupe Adulte montre globalement une forte anticipation (indice moyen d'anticipation de 0,45), et que le groupe Enfant montre une anticipation plus faible, voire une absence d'anticipation (indice moyen d'anticipation de 0,19). On pourrait subdiviser le groupe Enfant en trois sous-groupes : un premier groupe montrant une absence d'anticipation, un second groupe montrant une faible anticipation, et un troisième groupe montrant une anticipation plus importante. En effet, au moins 7 participants enfants montrent une absence d'anticipation, puisqu'ils possèdent un indice d'anticipation inférieur à zéro, la majorité des participants enfants possède un faible indice d'anticipation (8 participants), et enfin 5 participants enfants possèdent un indice de stabilité d'un ordre de grandeur comparable aux scores adultes les plus faibles.

Ces indices individuels d'anticipation, présentés par groupe d'âge (figure 3.97., panneau de droite), permettent une comparaison de la mesure d'anticipation entre ces deux groupes d'âge. Une analyse statistique (modèle linéaire, prenant comme variable dépendante les indices individuels et comme facteur le groupe d'âge) révèle que la différence entre le groupe Enfant et le groupe Adulte est significative $F(1, 28) = 27{,}57$; $p < 1.398\ e^{-05}$.

Concernant les mesures articulatoires, on constate que les adultes anticipent bien V2 dès le milieu de V1, et ce pour presque tous les contextes étudiés (pour chaque adulte, dans un seul des contextes étudiés, l'anticipation n'est pas significative). Pour les enfants, certains participants (à l'image des participants Enfant 2 et 6) anticipent effectivement V2 dès le milieu de V1, mais pas dans tous les contextes étudiés. La très large majorité des enfants semble ne pas anticiper ou anticiper très peu V2 dans la réalisation de V1, à l'image des participants Enfant 1 et 7.

3.4. Durées des segments et des séquences

Nous présentons maintenant les résultats concernant les durées des segments et des séquences V_1-C-V_2. Ces résultats concernent notre hypothèse 4 (nous devrions observer des gestes légèrement plus lents chez l'enfant par rapport à l'adulte).

Le tableau 3.13. ci-dessous présente pour chaque participant adulte, les durées moyennes et les écarts types de la première voyelle, de la tenue consonantique, du relâchement consonantique et de la seconde voyelle.

Tableau 3.13. Durées des segments (Voyelle 1 (V1), tenue consonantique (tenue C), relâchement consonantique (relâchement C), Voyelle 2 (V2)) et des séquences V_1-C-V_2 pour les adultes.

	Durées moyennes ($\bar{x}$) et écarts types (σ) pour les adultes (en secondes)									
	V1		**tenue C**		**relâchement C**		**V2**		**Séquences V_1-C-V_2**	
Participant	**$\bar{x}$**	**σ**	**$\bar{x}$**	**σ**	**$\bar{x}$**	**σ**	**$\bar{x}$**	**σ**	**$\bar{x}$**	**σ**
Adulte 1	0,121	0,014	0,122	0,028	0,015	0,007	0,142	0,030	0,401	0,046
Adulte 2	0,138	0,022	0,098	0,015	0,022	0,010	0,130	0,041	0,387	0,051
Adulte 3	0,100	0,014	0,084	0,013	0,026	0,018	0,173	0,032	0,382	0,036
Adulte 4	0,176	0,029	0,095	0,014	0,025	0,021	0,183	0,027	0,479	0,047
Adulte 5	0,179	0,027	0,093	0,014	0,024	0,015	0,125	0,019	0,420	0,033
Adulte 6	0,074	0,017	0,068	0,013	0,022	0,010	0,092	0,032	0,256	0,037
Adulte 7	0,088	0,015	0,078	0,016	0,024	0,012	0,058	0,017	0,247	0,028
Adulte 8	0,102	0,017	0,078	0,020	0,021	0,006	0,135	0,031	0,336	0,037
Adulte 9	0,164	0,030	0,111	0,018	0,017	0,006	0,117	0,035	0,409	0,059
Adulte 10	0,126	0,016	0,077	0,014	0,027	0,010	0,077	0,023	0,307	0,030
Moyenne	**0,127**	0,020	**0,090**	0,016	**0,022**	0,011	**0,123**	0,029	**0,362**	0,040

Pour les participants adultes, la durée moyenne de V1 est de 127 ms (± 20 ms), la durée moyenne de la tenue consonantique est de 90 ms (± 16 ms), la durée moyenne du relâchement consonantique est de 22 ms (± 11 ms), et la durée moyenne de V2 est de 123 ms (± 29 ms). La durée moyenne des séquences V_1-C-V_2 pour les participants adultes est de 362 ms (± 40 ms).

Le tableau 3.14. ci-dessous présente pour chaque participant enfant, les durées moyennes et les écarts types de la première voyelle, de la tenue consonantique, du relâchement consonantique et de la seconde voyelle.

Tableau 3.14. Durées des segments (Voyelle 1 (V1), tenue consonantique (tenue C), relâchement consonantique (relâchement C), Voyelle 2 (V2)) et des séquences V_1-C-V_2 pour les enfants.

	Durées moyennes ($\bar{x}$) et écarts types (σ) pour les enfants (en secondes)									
	V1		**tenue C**		**relâchement C**		**V2**		**Séquences V_1-C-V_2**	
Participant	**$\bar{x}$**	**σ**	**$\bar{x}$**	**σ**	**$\bar{x}$**	**σ**	**$\bar{x}$**	**σ**	**$\bar{x}$**	**σ**
Enfant 1	0,329	0,104	0,149	0,044	0,029	0,018	0,178	0,074	0,685	0,116
Enfant 2	0,145	0,043	0,079	0,031	0,031	0,020	0,163	0,053	0,418	0,076
Enfant 3	0,178	0,054	0,133	0,051	0,044	0,037	0,301	0,119	0,656	0,182
Enfant 4	0,228	0,064	0,176	0,045	0,030	0,020	0,214	0,086	0,649	0,153
Enfant 5	0,114	0,039	0,087	0,026	0,022	0,018	0,123	0,044	0,346	0,094
Enfant 6	0,131	0,037	0,102	0,026	0,032	0,023	0,188	0,035	0,452	0,057
Enfant 7	0,176	0,042	0,120	0,034	0,045	0,036	0,200	0,037	0,541	0,087
Enfant 8	0,102	0,035	0,082	0,031	0,037	0,028	0,124	0,024	0,345	0,065
Enfant 9	0,158	0,038	0,123	0,040	0,030	0,023	0,161	0,030	0,472	0,083
Enfant 10	0,124	0,032	0,118	0,022	0,039	0,041	0,203	0,033	0,484	0,520
Enfant 11	0,141	0,049	0,104	0,031	0,045	0,031	0,172	0,045	0,462	0,098
Enfant 12	0,183	0,035	0,120	0,030	0,029	0,025	0,275	0,069	0,607	0,077
Enfant 13	0,186	0,032	0,137	0,034	0,040	0,032	0,281	0,049	0,643	0,081
Enfant 14	0,148	0,044	0,098	0,040	0,026	0,019	0,156	0,041	0,427	0,073
Enfant 15	0,150	0,042	0,114	0,040	0,051	0,050	0,255	0,055	0,570	0,091
Enfant 16	0,196	0,049	0,147	0,058	0,047	0,047	0,226	0,058	0,615	0,106
Enfant 17	0,132	0,073	0,100	0,030	0,035	0,023	0,094	0,040	0,360	0,106
Enfant 18	0,254	0,047	0,172	0,045	0,063	0,062	0,276	0,089	0,764	0,097
Enfant 19	0,174	0,026	0,122	0,041	0,037	0,015	0,255	0,037	0,587	0,068
Enfant 20	0,202	0,050	0,134	0,031	0,031	0,036	0,276	0,055	0,642	0,083
Moyenne	**0,173**	0,047	**0,121**	0,037	**0,037**	0,030	**0,206**	0,054	**0,536**	0,116

Pour les participants enfants, la durée moyenne de V1 est de 173 ms (± 47 ms), la durée moyenne de la tenue consonantique est de 121 ms (± 37 ms), la durée moyenne du relâchement consonantique est de 37 ms (± 30 ms), et la durée moyenne de V2 est de 206

ms (± 54 ms). La durée moyenne des séquences V_1-C-V_2 pour les participants enfants est de 536 ms (± 116 ms).

Nous observons que les gestes des enfants sont plus lents que ceux des adultes et présentent une plus grande variabilité temporelle. Une analyse statistique effectuée sur la durée globale des séquences V_1-C-V_2 (modèle linéaire, prenant comme variable dépendante la durée des séquences V_1-C-V_2 et comme facteur le groupe d'âge) révèle que la différence entre le groupe Enfant et le groupe Adulte est significative $F(1, 28) = 16{,}69$; $p < 0{,}0003346$.

3.5. Relations entre nos principaux indices

Dans cette dernière section, nous présentons les résultats concernant les relations entre nos principaux indices. Nous avons élaboré dans les sections précédentes un indice de la stabilité du contrôle et un indice de la taille de l'effet d'anticipation. Nous voulons maintenant observer s'il existe une relation entre ces deux indices, à savoir si les participants qui anticipent le plus sont également ceux qui montrent une meilleure stabilité du contrôle.

Nous présentons, dans la figure 3.98. ci-dessous, la relation entre indice individuel d'anticipation et indice individuel de stabilité, via la régression linéaire observée entre ces deux indices, séparément pour les groupes Enfant et Adulte.

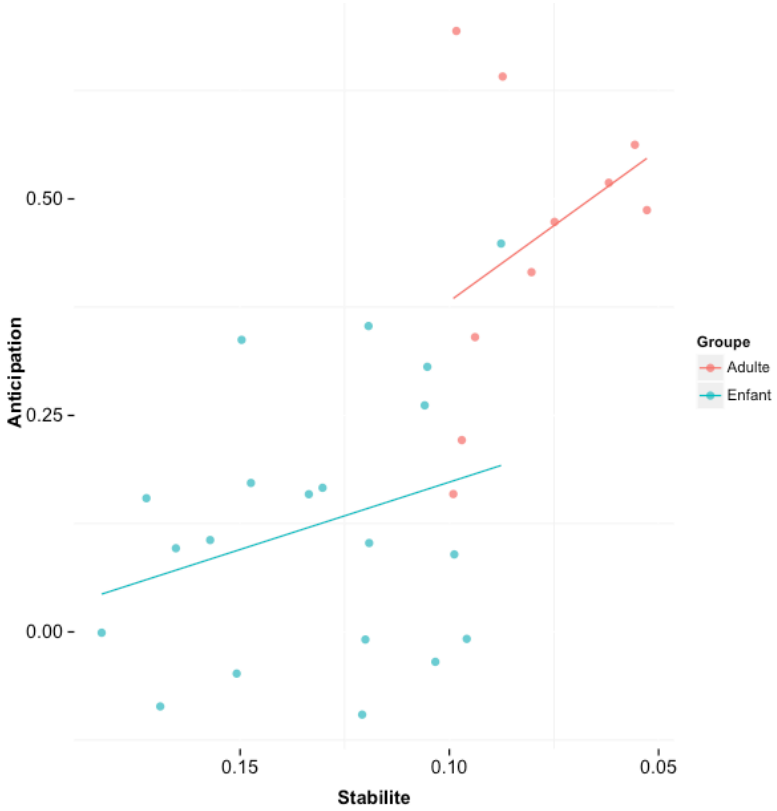


Figure 3.98. Relations entre indices numériques de stabilité et d'anticipation, séparément pour les groupes Enfant et Adulte.

La pente de la régression linéaire entre indice de stabilité et d'anticipation pour le groupe Adulte est de 0.5469. Le coefficient de corrélation R est de 0.3670. La valeur de F est 1.245 et $p < 0.2969$. Cette corrélation n'est pas significative.

La pente de la régression linéaire entre indice de stabilité et d'anticipation pour le groupe Enfant est de 0.1922. Le coefficient de corrélation R est de 0.2847. La valeur de F est de 1.587 et $p < 0.2238$. Cette corrélation n'est pas significative. Les participants qui montrent une plus grande amplitude d'anticipation ont tendance à être ceux qui ont la meilleure stabilité du contrôle, et ceci semble être vrai au sein de chaque groupe. Cette observation est une tendance, puisque la corrélation entre ces deux indices n'est pas significative.

Nous souhaitons également observer s'il existe une relation entre le débit d'élocution et l'amplitude de l'anticipation. Nous présentons, dans la figure 3.99. ci-dessous, la relation entre indice individuel d'anticipation et débit d'élocution, via la régression linéaire observée entre ces deux indices, séparément pour les groupes Enfant et Adulte.

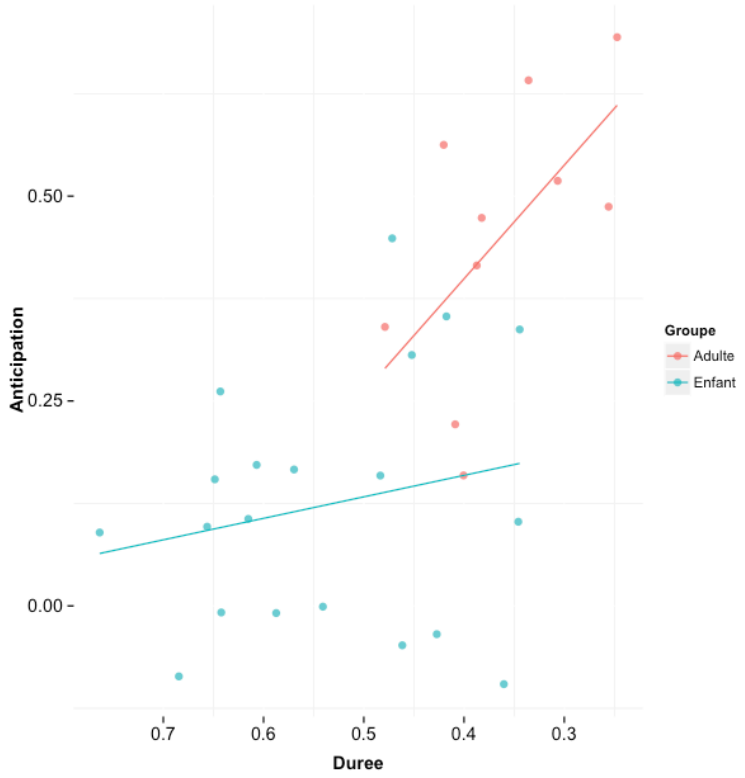


Figure 3.99. Relations entre indices numériques d'anticipation et débit d'élocution, séparément pour les groupes Enfant et Adulte.

La pente de la régression linéaire entre indice d'anticipation et débit d'élocution pour le groupe Adulte est de 0.6108. Le coefficient de corrélation R est de 0.6016. La valeur de F est de 4.538 et $p < 0.06576$. Cette corrélation n'est pas significative.

La pente de la régression linéaire entre indice d'anticipation et débit d'élocution pour le groupe Enfant est de 0.1738. Le coefficient de corrélation R est de 0.2070. La valeur de F est de 0.8058 et $p < 0.3812$. Cette corrélation n'est pas significative. Les participants qui montrent une plus grande amplitude d'anticipation ont tendance à être aussi ceux qui ont le débit d'élocution le plus rapide, et ceci semble être vrai au sein de chaque groupe. Cette observation est une tendance, puisque la corrélation entre ces deux indices n'est pas significative.

Chapitre 4

Discussion

Nous avions formulé quatre hypothèses de travail *a priori*. Ces hypothèses étaient que (1) nous devrions observer plus de variabilité en répétition dans les productions enfantines par rapport à celles des adultes ; (2) nous devrions observer une amplitude similaire de coarticulation intra-syllabique entre enfants et adultes ; (3) nous devrions observer une moindre amplitude de coarticulation extra-syllabique chez les enfants par rapport aux adultes et (4) nous devrions observer des gestes de parole plus lents chez les enfants que chez les adultes. Nous venons de présenter les résultats expérimentaux qui nous permettent de tester ces hypothèses. Dans le même ordre que celui proposé au chapitre 3, nous proposons dans ce chapitre une lecture de ces résultats, nos interprétations de ces résultats et une confrontation aux données de la littérature, pour chacun des résultats présentés au chapitre 3.

4.1. Stabilité de la production des voyelles en contexte isolé

Nos résultats concernant la stabilité de la production des voyelles en contexte isolé mettent en évidence une plus grande variabilité liée à la répétition d'une même tâche dans les productions enfantines que dans celles des adultes.

Dans le domaine acoustique, les figures 3.1. et 3.5. montrent la variabilité liée à la répétition (8 à 10 répétitions) de ces voyelles en contexte isolé, pour chacun des participants. Les valeurs des écarts types pour chaque participant et chaque catégorie vocalique, reportés dans les tableaux 3.1. et 3.3. révèlent que l'on observe en moyenne 1.64 fois plus de variabilité dans les productions enfantines que dans celles des adultes. Les différences observées sont statistiquement significatives.

Du point de vue articulatoire, les figures 3.2. à 3.4. et 3.6. à 3.10. montrent la variabilité liée à la répétition de ces voyelles en contexte isolé pour quelques participants adultes et

enfants. Les tableaux 3.2. et 3.4., qui reportent les distances moyennes entre les contours de langue au sein de chaque catégorie vocalique (mesurées par la distance dite du « plus proche voisin »), révèlent que l'on observe en moyenne 1.63 fois plus de variabilité dans les productions enfantines que dans celles des adultes.

Les figures 3.6. à 3.10. montrent également davantage de recoupement entre catégories vocaliques chez l'enfant dans le domaine articulatoire. En effet, chez l'adulte (figures 3.2. à 3.4.), la variabilité au sein de chaque catégorie vocalique est faible, et les contrastes entre catégories vocaliques sont clairs et importants : le recoupement entre catégorie est très faible. Chez les enfants, la variabilité au sein de chaque catégorie est plus importante et les contrastes entre catégories sont moins marqués, de sorte que le recoupement entre catégories vocaliques est plus important. Ces résultats semblent suggérer (en s'appuyant sur les contrastes en production seulement) que l'intelligibilité de leurs productions vocaliques est probablement moindre que chez l'adulte, en accord avec Eguchi & Hirsh (1969).

Les figures 3.1., 3.5. et 3.12. montrent également que les catégories vocaliques /i/, /e/, /ɛ/ et /a/ sont bien contrastées en F1 chez l'adulte, et dispersées de façon homogène sur la dimension F1 (en accord avec la théorie de la dispersion maximale (Liljencrants & Lindblom, 1972 ; Schwartz et al., 1997). Chez les enfants en revanche (figure 3.12. panneau droit), ces catégories vocaliques ne sont pas dispersées de façon homogène et optimale sur la dimension F1, et l'on observe une confusion entre les catégories /e/ et /ɛ/. Cette observation peut être interprétée comme le fait que les enfants de 4 ans n'ont pas encore bien acquis le contraste /e/ ~ /ɛ/.

Une plus grande variabilité est donc observée dans les productions des enfants par rapport à celles des adultes (cf. figure 3.11.). Selon notre point de vue, ces résultats indiquent que, par rapport aux adultes, les enfants disposent d'une **moindre stabilité du contrôle** pour la production de gestes vocaliques. Selon notre cadre théorique, l'apprentissage de la production de la parole repose sur l'apprentissage de deux modèles internes, un modèle inverse et un modèle direct. Une fois qu'un apprentissage suffisant est effectué, la production repose presque exclusivement sur le modèle inverse, selon un schéma *feedforward* de contrôle de la production de parole. Chez les participants adultes,

on observe une très faible variabilité liée à la répétition d'une même tâche. Cela est en cohérence avec l'hypothèse d'un modèle inverse efficace et la définition précise de prototypes auditifs représentant les buts de la production. La faible variabilité observée chez l'adulte reflèterait presque uniquement le bruit dans les commandes motrices, c'est-à-dire dans l'exécution, la réalisation de la tâche. La variabilité plus importante observée chez l'enfant reflèterait l'immaturité du modèle inverse, mais aussi l'imprécision de la définition du but (ou prototype) auditif, ainsi qu'un possible plus grand bruit biologique dans la réalisation même de la tâche. Notons que les différentes contributions de ces possibles facteurs de variabilité restent difficiles à évaluer. Quoi qu'il en soit, nos résultats concernant la variabilité en répétition des voyelles en contexte isolé révèlent une moindre stabilité du contrôle chez l'enfant par rapport à l'adulte.

Les estimations des valeurs de formants pour les production vocaliques des enfants de 4 ans peuvent être directement comparées avec d'autres mesures similaires effectuées dans la littérature, par exemple Ménard (2002) (p. 233) et MacGowan et al. (2014). Les valeurs reportées dans ces travaux sont tout à fait comparables à celles que nous avons obtenues. Concernant l'Espace Vocalique Maximal, les méta analyses de Vorperian et al. (2007) et Boë et al. (2013), reportant les résultats de plusieurs dizaines d'études développementales, rapportent également des valeurs tout à fait comparables à celles que nous avons obtenues.

Du point de vue articulatoire, il est possible de comparer la variabilité au sein des catégories vocaliques avec celles qui ont été reportées dans Zharkova et al. (2011). Nous observons une variabilité moyenne de l'ordre de 2,4 mm chez les adultes et de 3,9 mm chez les enfants. Zharkova et al. (2011) reportent une variabilité moyenne de l'ordre de 1,8 mm chez les enfants entre 6 et 9 ans et de 1,2 mm chez les adultes. La variabilité reportée par Zharkova et al. (2011) est inférieure à celle que nous reportons, peut-être parce que les voyelles étudiées par Zharkova et al. (2011) sont des voyelles en contexte, et en contexte /ʃV/, contexte particulièrement contraignant. Quoi qu'il en soit, la variabilité en répétition semble bien diminuer avec l'âge.

Ces résultats sont également à mettre en relation avec les résultats des études développementales portant sur le contrôle du bras et de la main (Lambert & Bard, 2005 ;

Hourcade et al., 2004 ; Forssberg et al., 1991) et celles portant sur le contrôle des lèvres et de la mandibule (e.g. Smith & Goffman, 1998). Ces études mettent toutes en évidence une diminution de la variabilité liée à la répétition d'une même tâche avec l'âge. Cela souligne que la diminution de la variabilité liée à la répétition reflète bien une amélioration de la stabilité du contrôle au cours du développement tardif.

Ces résultats confirment notre hypothèse 1, à savoir que l'on observe davantage de variabilité liée à la répétition d'une même tâche dans les productions des enfants par rapport à celles des adultes. Nous interprétons ces résultats comme l'indice d'une **moindre stabilité du contrôle** pour la production de gestes vocaliques chez l'enfant de 4 ans par rapport à l'adulte.

4.2. Coarticulation intra-syllabique

Nos résultats acoustiques concernant la coarticulation intra-syllabique (utilisant les équations de locus) suggèrent que l'amplitude de cette coarticulation est similaire entre enfants de 4 ans et adultes. En revanche, la lecture des données articulatoires concernant la coarticulation intra-syllabique révèle que l'anticipation de V2 dans la réalisation de la consonne n'est pas systématique chez les enfants.

Les tableaux 3.5. et 3.6. reportent les valeurs de pentes des équations de locus en contextes /b/, /d/ et /g/ pour les adultes et les enfants. Ces résultats suggèrent que l'amplitude de la coarticulation intra-syllabique est similaire entre les groupes d'âge. Les pentes des équations de locus pour les adultes sont de 0.94 pour /b/ ; 0,72 pour /d/ et 0.86 pour /g/. Les pentes des équations de locus pour les enfants sont de 0.97 pour /b/ ; 0.83 pour /d/ et 0.90 pour /g/. Ces résultats suggèrent que les enfants anticipent bien V2 dans la réalisation de C, et ce d'une manière relativement similaire aux adultes (les différences entre les groupes d'âge ne sont pas significatives).

Les données articulatoires (figures 3.13. à 3.48.) montrent, dans l'ensemble, que la majorité des enfants, dans la plupart des cas, anticipent effectivement V2 dans la réalisation de la consonne. Par exemple, l'enfant 1 anticipe bien V2 dans la réalisation de /b/, puisqu'au moment du relâchement de la consonne, on observe bien une différence

significative entre la réalisation de /b/ en contexte V2 = /i/ et la réalisation de /b/ en contexte V2 = /a/, qui va bien dans la direction de l'anticipation (figures 3.13. et 3.14.).

En revanche, les données articulatoires révèlent également que dans certains cas de figure, les enfants n'anticipent pas V2 dans la réalisation de la consonne. Par exemple, l'enfant 2 ne montre pas d'effet d'anticipation de V2 dans le contexte C = /g/ et V1 = /a/ (figure 3.24.) ; l'enfant 5 ne montre pas d'effet d'anticipation de V2 dans le contexte C = /d/ et V1 = /a/ (figure 3.28) ; l'enfant 7 ne montre pas d'effet d'anticipation de V2 dans le contexte C = /b/ et V1 = /ɛ/ (figure 3.37.) ; l'enfant 9 ne montre pas d'effet d'anticipation de V2 dans le contexte C = /g/ et V1 = /a/ (figure 3.48).

Si l'on compare les valeurs des pentes des équations de locus pour les contextes de V1 séparément (tableaux 3.7. et 3.8.) aux observations effectuées dans le domaine articulatoire, il ne semble pas exister de relation entre les deux mesures.

En effet, les observations articulatoires concernant l'Enfant 2 dans le contexte C = /g/ montrent que ce participant anticipe dans le contexte V1 = /ɛ/ mais n'anticipe pas dans le contexte V1 = /a/. Or les pentes des équations de locus pour ce participant et pour C = /g/ sont sensiblement équivalentes, de 0,79 pour V1 = /ɛ/ et de 0,81 pour V1 = /a/, indiquant une forte amplitude de coarticulation pour ces deux contextes.

Les observations articulatoires concernant l'Enfant 5 dans le contexte C = /d/ montrent que ce participant anticipe dans le contexte V1 = /ɛ/ mais n'anticipe pas dans le contexte V1 = /a/. Les pentes des équations de locus pour ce participant et pour C = /d/ sont de 0,78 pour V1 = /ɛ/ et de 0,16 pour V1 = /a/. Dans ce cas de figure, les pentes des équations de locus semblent en cohérence avec les données articulatoires.

Les observations articulatoires concernant l'Enfant 7 dans le contexte C = /b/ montrent que ce participant anticipe dans le contexte V1 = /a/ mais n'anticipe pas dans le contexte V1 = /ɛ/. Or les pentes des équations de locus pour ce participant et pour C = /b/ sont de 0,98 pour V1 = /ɛ/ et de 1,11 pour V1 = /a/, soit relativement fortes, indiquant une grande amplitude de coarticulation, ce qui ne semble pas être le cas pour le contexte V1 = /ɛ/.

Les observations articulatoires concernant l'Enfant 9 dans le contexte C = /g/ montrent que ce participant anticipe dans le contexte V1 = /ɛ/ mais n'anticipe pas dans le contexte

V1 = /a/. Or les pentes des équations de locus pour ce participant et pour C = /g/ sont de 0,90 pour V1 = /ɛ/ et de 0,77 pour V1 = /a/. Une pente de 0,77 indique une amplitude relativement forte de coarticulation, or dans ce cas précis, ce participant montre un effet qui va dans la direction opposée à l'anticipation.

La comparaison entre données articulatoires et pentes des équations de locus suggère que les pentes des équations de locus ne traduisent pas toujours, de manière sensible, l'amplitude de la coarticulation. Souvent, une valeur de pente d'équation de locus relativement forte est observée alors que les données articulatoires suggèrent que l'amplitude de la coarticulation est en réalité relativement faible.

La lecture des données articulatoires révèle que les participants enfants n'anticipent pas de manière systématique V2 dans la réalisation de C. Nos données articulatoires concernant la coarticulation intra-syllabique chez l'enfant, confrontées aux mesures des pentes des équations de locus, suggèrent donc qu'il n'existe pas de relation entre les valeurs des équations de locus et la réalité articulatoire de l'anticipation. Il semblerait donc, comme l'avait suggéré Löfqvist (1999), que les équations de locus ne représentent pas un indice fiable et sensible de l'amplitude de la coarticulation intra-syllabique.

Les mesures effectuées à travers les équations de locus suggèrent que l'amplitude de la coarticulation intra-syllabique est similaire entre enfants et adultes. Mais la lecture des données articulatoires révèle que l'anticipation de V2 dans la réalisation de la consonne n'est pas systématique chez les participants enfants. Il existe un certain nombre de cas de figure où les enfants ne montrent aucun signe d'anticipation de V2 dans la réalisation de C, et un certain nombre de cas de figure où l'anticipation est faible voire très faible. Nous n'avons pas analysé les données articulatoires concernant la coarticulation intra-syllabique chez l'adulte. En revanche, nous observons une forte anticipation de V2 dès la réalisation de V1. Il est raisonnable de penser que cette forte anticipation de V2 dès le milieu de V1 soit également présente lors du relâchement consonantique. De plus, les observations cinéradiographiques chez l'adulte (Perkell, 1969 ; Bothorel et al., 1986) montrent que les adultes anticipent fortement la voyelle suivante dans la réalisation de la consonne. Nos données articulatoires mettant en évidence une absence de coarticulation intra-syllabique ou une faible amplitude de coarticulation intra-syllabique chez l'enfant

révèlent donc très probablement que l'amplitude de la coarticulation intra-syllabique est plus faible chez les enfants que chez les adultes.

Une anticipation non systématique de V2 dans la réalisation de la consonne suggère une moindre amplitude de la coarticulation intra-syllabique chez l'enfant, par rapport aux adultes. Selon notre point de vue, ces résultats indiquent une **moindre maturité du contrôle**, en particulier pour ce qui est de la prise en compte de la variabilité allouée à chaque son de parole et de l'exploitation de cette variabilité dans la production d'une séquence de sons de parole dans le but de minimiser un effort articulatoire. Dans notre cadre théorique, celui de la coarticulation anticipatoire expliquée à travers l'utilisation de la variabilité sub-phonémique (ou intra-catégorielle) et le contrôle optimal, visant à minimiser un coût le long de la séquence de sons de parole, ces résultats peuvent être expliqués par l'immaturité du contrôle moteur. En effet, si l'on fait l'hypothèse que l'organisation de la parole est de type segmentale, nous devrions observer une amplitude croissante de la coarticulation avec l'âge, reflétant une amélioration de la capacité à planifier efficacement une séquence de buts moteurs. Une faible amplitude de coarticulation reflète, dans ce cadre, une moindre connaissance de la variabilité allouée à chaque son de parole et une moindre exploitation de cette variabilité dans la planification et l'exécution d'une séquence de sons de parole ; une forte amplitude de coarticulation reflète au contraire une certaine maturité motrice permettant l'exploitation de cette variabilité dans le but de minimiser l'effort articulatoire le long de la séquence.

Nos résultats concernant la coarticulation intra-syllabique peuvent être mis en relation avec de nombreuses études de la littérature. A la lecture de ces études, nous nous attendions à observer autant de coarticulation chez l'enfant que chez l'adulte pour ce type particulier de coarticulation. En effet, la majorité des études de la littérature conclut que l'amplitude de la coarticulation intra-syllabique de type plosive-voyelle chez l'enfant est similaire à celle de l'adulte (Sussman et al., 1999 ; Repp, 1986 ; Nittrouer, 1993 ; Turnbaugh et al., 1985 ; Sereno & Lieberman, 1987 ; Sereno et al., 1987 ; Noiray et al., 2013). Or ces études reposent uniquement sur l'indice des équations de locus, et comme nous venons de le voir, cet indice ne semble pas refléter l'amplitude de la coarticulation intra-syllabique de façon sensible. Nos résultats sont, dans une certaine mesure, cohérents avec ces études, puisque nos équations de locus semblent indiquer une

amplitude de coarticulation similaire entre enfants et adultes. Mais la lecture des données articulatoires nous amène à relativiser cette conclusion, et révèle que les enfants n'anticipent pas de manière systématique le geste vocalique dans le geste consonantique précédent. Cela souligne la nécessité de collecter et d'analyser des données articulatoires des productions enfantines.

Ces résultats permettent également de discuter de la nature des unités de la production au cours du développement. Etant donnée l'acquisition précoce des syllabes de type plosive-voyelle au cours du développement (dès le babillage canonique, autour de 7 mois) et la supposée forte cohésion entre gestes consonantiques et vocaliques au sein de la syllabe CV chez l'enfant mise en évidence par des études acoustiques, nous nous attendions à une forte anticipation de V2 dans la réalisation de C chez les enfants. De nombreux auteurs, dont Nittrouer, ont interprété ces résultats comme l'indice de l'organisation syllabique de la production de la parole chez les jeunes enfants. En effet, si les enfants rapprochent gestes consonantiques et vocaliques au sein de la syllabe, cela pourrait s'expliquer par le fait que leurs unités de production sont d'ordre syllabique plutôt que phonémique (les enfants apprendraient et stockeraient /ba/ et /bi/ comme des unités de base de la production, ces deux réalisations de /b/ seraient donc fortement contextualisées, donc différentes). Or nos résultats semblent indiquer le contraire : nous n'observons pas de forte amplitude systématique de la coarticulation intra-syllabique chez l'enfant. Ces résultats semblent donc indiquer que la syllabe n'est pas ou n'est plus l'unité de base de la production de parole chez l'enfant de 4 ans (si elle l'était, nous devrions observer une forte amplitude de coarticulation intra-syllabique, et ce pour tous les contextes étudiés) et semblent également indiquer que le phonème serait l'unité de base de la production de la parole chez l'enfant. En effet, le fait par exemple que les gestes /g/ en contexte V2 = /i/ et V2 = /a/ de l'enfant 10 ne soient pas différenciés indique que les représentations sensori-motrices de cet enfant sont bien organisées autour des segments, des phonèmes, et qu'il ne connait pas encore toutes les possibilités de production de ces phonèmes pour exploiter cette variabilité lors de la production de séquences de sons de parole. De plus, l'émergence du phonème comme unité de base de la perception (cf. notamment Cutler & Mehler, 1993) et de la production (cf. notamment Hallé et Christia, 2012) semble relativement précoce, puisque l'enfant est attiré très tôt

par les unités segmentales, au cours de la première année de vie, et semble organiser sa production autour des phonèmes dès la seconde année de vie. Il semblerait donc qu'il faille étudier des productions très précoces, c'est-à-dire avant les premiers mots, pour remettre en question l'organisation segmentale de la parole et réellement mettre en difficulté les modèles de production centrés autour du phonème, tel que le modèle de contrôle optimal à travers lequel nous interprétons nos résultats.

Ces résultats infirment notre hypothèse 2, à savoir que l'on devrait observer des patrons de coarticulation intra-syllabique d'amplitudes similaires entre enfants et adultes. Nous interprétons ces résultats comme un indice de **l'immaturité du contrôle moteur** pour la production de séquences de sons de parole et comme un indice de **l'organisation segmentale de la parole enfantine**. Ces résultats soulignent également la nécessité de collecter et d'analyser des données articulatoires des productions enfantines, particulièrement dans le cas où les indices acoustiques développés ne reflètent pas fidèlement la réalité articulatoire.

4.3. Coarticulation extra-syllabique

Nos résultats concernant la coarticulation extra-syllabique mettent en évidence une moindre amplitude de la coarticulation extra-syllabique chez les enfants de 4 ans par rapport aux adultes.

Dans le domaine acoustique, les figures 3.49., 3.50., 3.57. et 3.58. montrent l'effet de V2 sur la réalisation de V1, pour chacun des participants. On observe que l'effet d'anticipation de V2 dès le milieu de V1 est important pour tous les participants adultes (seuls le participant Adulte 1 pour le contexte V1 = /a/ et le participant Adulte 9 pour le contexte où V1 = /ɛ/ ne semblent pas montrer d'effet d'anticipation). On observe que l'effet d'anticipation de V2 dès le milieu de V1 ne semble pas être évident pour la majorité des participants enfants (seuls les participants Enfant 2, 6, 8, 9 et 13 montrent un effet d'anticipation dans certains des contextes). Les valeurs de la taille de l'effet d'anticipation rapportées dans les tableaux 3.10. et 3.11. révèlent que tous les participants adultes anticipent V2 dès la réalisation de V1, et que seuls 5 enfants parmi les 20 enfants étudiés anticipent V2 dès la réalisation de V1 (les enfants 2, 6, 8, 9 et 13 montrent un effet d'anticipation. Les figures par groupe d'âge (figures 3.95. et 3.96.) montrent que l'effet

d'anticipation est clair pour les adultes, dans les deux dimensions, F1 et F2, et que cet effet ne semble exister que dans la dimension F2 pour le contexte V1 = /a/ pour les enfants. Les analyses statistiques révèlent que l'effet de V2 sur V1 ne s'exerce pas de la même façon entre les groupes d'âge, et que cet effet est significatif dans les deux dimensions pour les adultes et n'est pas significatif dans la dimension F1 pour les enfants, mais est significatif dans la dimension F2. Les analyses statistiques visant à mesurer les tailles de ces effets révèlent que la taille de l'effet est importante dans les deux dimensions pour le groupe Adulte et est négligeable en F1 et très réduite en F2 pour le groupe Enfant. L'amplitude de la coarticulation extra-syllabique est donc moindre au sein du groupe Enfant.

Dans le domaine articulatoire, les figures 3.51. à 3.56. montrent que l'effet de V2 sur la réalisation de V1 est significatif dans presque tous les contextes pour les 2 adultes étudiés, à l'exception du contexte C = /b/ et V1 = /a/ pour le participant Adulte 6 et des contextes C = /b/ et V1 = /ɛ/ et C = /g/ et V1 = /ɛ/ pour le participant Adulte 8.

Les figures 3.59. à 3.94. montrent que l'effet de V2 sur la réalisation de V1 n'est pas significatif, pour environ la moitié des contextes étudiés. Une anticipation significative de V2 dans la réalisation de V1 est tout de même observée pour l'enfant 2 dans le contexte où C = /b/ et V1 = /a/ et dans les contextes où C = /d/ ; pour l'enfant 6 dans les contextes où C = /b/ et V1 = /a/ et dans les contextes C = /d/ et C = /g/ et V1 = /ɛ/ ; pour l'enfant 7 dans tous les contextes sauf pour le contexte où C = /g/ et V1 = /ɛ/ ; pour l'enfant 8 dans les contextes où C = /b/ et V1 = /a/ et C = /d/ et V1 = /a/ ; pour l'enfant 10 dans les contextes où C = /b/. Pour les participants Enfant 1 et 8, l'anticipation n'est pas significative dans la majorité des contextes étudiés. En somme, l'enfant 7 anticipe bien V2 dans la réalisation de V1 dans presque tous les contextes, mais ce n'est pas le cas de l'enfant 2 qui montrait pourtant des signes d'anticipation dans le domaine acoustique. Cependant, dans le domaine acoustique, nous n'avons pas séparé les contextes consonantiques. C'est pourquoi un effet d'anticipation visible dans le domaine acoustique peut être dû à une anticipation dans quelques contextes seulement, comme cela semble être le cas pour l'Enfant 2.

Une moindre amplitude de la coarticulation extra-syllabique est donc observée chez l'enfant par rapport à l'adulte. Nous interprétons ces résultats comme un indice de l'**immaturité du contrôle moteur**, en particulier pour ce qui est de la prise en compte de la variabilité allouée à chaque son de parole et de l'exploitation de cette variabilité dans la production d'une séquence de sons de parole dans le but de minimiser un effort articulatoire. Dans le cadre théorique qui est le notre, une faible amplitude de coarticulation extra-syllabique reflète une immaturité du contrôle moteur. Nous observons bien une amplitude plus faible de coarticulation extra-syllabique chez l'enfant par rapport à l'adulte, que ce soit dans le domaine acoustique ou articulatoire. Les données articulatoires sont cohérentes avec les données acoustiques (les enfants qui anticipent peu dans le domaine acoustique sont aussi ceux qui anticipent peu dans les données articulatoires), et offrent une lecture plus détaillé de cet effet d'anticipation. Dans la majorité des contextes et pour la majorité des enfants, l'effet d'anticipation de V2 dans V1 n'est pas significatif. L'anticipation de V2 est souvent visible à la fin de V1, mais pas toujours, et comme nous l'avons vu dans la section précédente sur la coarticulation intra-syllabique, l'effet de V2 n'est quelques fois même pas visible dans le relâchement de la consonne, qui s'effectue pourtant en moyenne 30 ms avant le début de la réalisation de V2. Pris ensemble, ces résultats indiquent que l'anticipation de V2 est observable dès le milieu de V1 pour les adultes, et est observable dès le milieu de V1 pour quelques enfants (5 d'après les données acoustiques), mais n'est observable que dès le fin de V1 pour la majorité des participants enfants, et n'est toujours pas observable dans le relâchement de la consonne précédent V2 pour certains enfants et dans certains contextes.

Nos résultats concernant la coarticulation extra-syllabique peuvent être mis en relation avec les résultats des études de Repp (1986) et Hodge (1989). L'étude de Repp (1986) mettait en évidence l'absence de coarticulation extra-syllabique chez un enfant de 4 ans, et la présence d'un effet d'anticipation chez un enfant de 9 ans et un adulte. L'étude de Hodge (1989) mettait en évidence que les enfants de 3 et 5 ans ne montraient pas de signe d'anticipation extra-syllabique, alors que les enfants de 9 ans et les adultes montraient un effet d'anticipation. Nos résultats sont en cohérence avec ces deux études. Notons que les résultats de ces deux études n'ont pas été interprétés à la lumière des modèles théoriques de coarticulation basés sur l'exploitation de la variabilité sub-

phonémique (Keating, 1990), ou à travers la notion de planification (Henke, 1966 ; Whalen, 1990), ou encore de contrôle optimal (Jordan & Rumelhart, 1992), puisque ces idées ont été développées après la parution de ces études.

Notons également que nous nous sommes volontairement placés dans un cadre visant à maximiser la possibilité d'anticiper au delà de la frontière syllabique : nous avons utilisé des séquences V_1-C-V_2 et des voyelles V1 pour lesquelles une large variabilité est tolérée. Dans les études de Repp (1986) et Hodge (1989), les séquences étudiées sont de type /ə.stV/, au sein desquelles l'anticipation de V dans le schwa est mesurée. Dans ce cadre, l'anticipation de la voyelle dans le schwa est contrainte par le cluster consonantique /st/, qui présente des contraintes articulatoires fortes. L'anticipation dans ce cadre est donc logiquement moindre que dans le cadre que nous avons proposé. Le fait que nous observons une anticipation extra-syllabique pour 5 de nos participants enfants n'est donc pas en contradiction avec les précédentes études, mais souligne simplement que dans ce cadre qui maximise la possibilité d'anticiper V2 dans la réalisation de V1, certains participants enfants démontrent une certaine maturité motrice, en particulier en ce qui concerne la prise en compte de la variabilité allouée à chaque son de parole.

Nous évoquions dans le premier chapitre de cette thèse les liens entre les expériences de coarticulation et celles de perturbation. De notre point de vue, ces expériences mettent en jeu les différentes stratégies articulatoires effectuées dans un seul et même but, et de ce point de vue, partagent des mécanismes communs. Elles révèlent en effet les capacités des locuteurs à utiliser différentes stratégies motrices pour atteindre un but acoustique et somatosensoriel, et renseignent de ce point de vue sur les représentations sensori-motrices des buts de la production de parole. Mais dans le cadre de perturbations, ce sont les stratégies d'équivalence motrice qui sont étudiées. Dans le cadre de perturbations articulatoires, les participants sont conscients de la perturbation et tentent consciemment de compenser cette perturbation. Dans le cadre de la coarticulation, le processus étudié est inconscient, et l'utilisation des différentes stratégies motrices n'est pas imposée, mais fortement suggérée par le contexte. L'étude de la coarticulation constitue, de ce point de vue, un cadre plus écologique que celui de la perturbation de l'articulation, afin d'étudier la maturité des représentations sensori-motrices des buts de la production de la parole.

Une interprétation alternative de nos résultats concernant la coarticulation extra-syllabique aurait pu être formulée : celle du poids de la syllabe comme unité de la production. Si l'on avait observé une faible amplitude de coarticulation extra-syllabique et une forte amplitude de coarticulation intra-syllabique, alors notre interprétation de ces résultats comme étant un indice d'une immaturité motrice aurait été mise en concurrence avec l'interprétation selon laquelle les enfants organisent leur production autour d'unités plus larges que le phonème, par exemple la syllabe. Or nos résultats concernant la coarticulation intra-syllabique viennent mettre à mal cette interprétation alternative de nos résultats : la coarticulation intra-syllabique chez l'enfant semble être d'amplitude moindre que chez l'adulte. L'explication alternative du poids de la syllabe dans la production est donc écartée. Nos résultats semblent conforter l'hypothèse selon laquelle la parole enfantine est organisée autour des mêmes unités que la parole adulte, à savoir les phonèmes, et semblent conforter l'hypothèse selon laquelle la coarticulation résulte d'un processus de planification motrice basée sur la minimisation d'un coût. Nous tenons à préciser que nous ne sommes pas partisans du phonème comme point de départ de la production et que nous défendons le point de vue selon lequel les enfants passent progressivement de représentations globales, holistiques, passant par une étape de forte domination syllabique, vers une organisation spécifiée en termes de segments.

Ces résultats confirment notre hypothèse 3, à savoir que l'on devrait observer une moindre amplitude de coarticulation extra-syllabique chez les enfants par rapport aux adultes. Nous interprétons ces résultats comme un indice de **l'immaturité du contrôle moteur** pour la production de séquences de sons de parole. La moindre amplitude de coarticulation intra-syllabique et la moindre amplitude de coarticulation extra-syllabique chez l'enfant par rapport à l'adulte soulignent une moindre maturité du contrôle, en particulier en ce qui concerne la capacité à tenir compte de la variabilité allouée à chaque son de parole et à exploiter cette variabilité dans le but de minimiser l'effort articulatoire. Ces données sont en cohérence avec le cadre théorique dans lequel ce travail de recherche s'inscrit, et semblent conforter les hypothèses selon lesquelles la parole enfantine serait organisée très tôt autour des unités de parole adulte, à savoir les phonèmes, et semble conforter l'hypothèse selon laquelle la coarticulation résulterait d'un processus de planification motrice reposant sur un contrôle moteur optimal.

4.4. Durées des segments et des séquences

Nos résultats concernant les durées des segments et des séquences mettent en évidence que les gestes de parole des enfants de 4 ans sont plus lents que ceux des adultes.

Les tableaux 3.13. et 3.14. reportent les durées moyennes pour chacun des segments (V1, tenue et relâchement de C et V2) et la durée moyenne des séquences V_1-C-V_2, pour les participants adultes et enfants respectivement. Ces résultats montrent que les gestes des enfants sont plus lents que ceux des adultes, puisque la durée moyenne des séquences V_1-C-V_2 pour les adultes est de 326 ms alors qu'elle est de 536 ms pour les enfants. Plus en détail, la durée moyenne de V1 pour les adultes est de 127 ms contre 173 ms pour les enfants ; la durée moyenne de la tenue de C est de 90 ms pour les adultes contre 121 ms pour les enfants ; la durée moyenne du relâchement de C est de 22 ms pour les adultes contre 37 ms pour les enfants ; la durée moyenne de V2 est de 123 ms pour les adultes contre 206 ms pour les enfants. On constate qu'une grande partie de l'allongement de la durée de la séquence V_1-C-V_2 chez les enfants par rapport aux adultes est dû à l'allongement de V2. Mais les gestes vocaliques (V1) et consonantiques (C) sont tout de même plus lents que ceux des adultes.

Nous ne pouvons donc pas rejeter l'hypothèse selon laquelle les enfants coarticulent moins car leur débit d'élocution est plus faible. Néanmoins, ces deux questions sont de notre point de vue intimement liées. Il est possible d'interpréter ces résultats à travers deux cadres : soit on considère que les gestes des enfants sont plus lents, et de ce fait, ils ne se trouvent pas devant la nécessité d'anticiper, soit on considère que les enfants n'anticipent pas et parlent lentement car ils ne possèdent pas la maturité motrice nécessaire pour le faire. Dans le premier cadre, la question de la cause de la lenteur des gestes enfantins reste à définir. Le second cadre, que nous privilégions, offre l'avantage de proposer des explications cohérentes pour tous les comportements observés : imprécision motrice, manque d'anticipation et lenteur des gestes. En effet, selon notre cadre théorique, parler avec un débit d'élocution rapide suppose un contrôle moteur mature, en particulier une bonne définition des cibles et un contrôle exclusivement de type prédictif (*feedforward*), car à débit d'élocution rapide le contrôle de type *feedback* est trop lent pour apporter des informations utiles au contrôle. Nos résultats concernant

les durées des segments et des séquences vont bien dans ce sens, puisque les adultes ont un débit d'élocution rapide (indiquant un contrôle moteur mature) et les enfants ont un débit d'élocution plus lent (indiquant que le contrôle moteur est encore en phase d'apprentissage).

Les gestes de parole des enfants sont donc plus lents que les gestes de parole des adultes. Nous interprétons ces résultats comme un indice de **l'immaturité du contrôle moteur** pour la production de parole chez l'enfant, en particulier concernant l'immaturité du modèle direct. En effet, l'apprentissage du modèle direct suppose une première phase de production de gestes lents, afin de pouvoir effectuer la comparaison entre ce qui est prédit par le modèle et ce qui est effectivement réalisé. Le temps de calcul de cette divergence sensorielle ainsi que la rétro propagation du gradient de l'erreur au sein du modèle direct, qui constituent les deux étapes clés de l'apprentissage du modèle direct, supposent un temps de calcul supplémentaire par rapport au cas de figure des adultes, pour lesquels l'apprentissage du modèle direct n'est plus en cours et pour lesquels la production repose essentiellement sur le modèle inverse. Nous supposons que les gestes des enfants sont plus lents que ceux des adultes, justement pour permettre aux enfants l'apprentissage du modèle direct. Leurs gestes plus lents reflèteraient donc de ce point de vue le fait que leur modèle direct est encore en construction, en phase d'apprentissage.

Nos résultats concernant la durée des voyelles peuvent être mis en relation avec ceux de Lyasko & Grigor'ev (2015), qui ont étudié entre autres les durées des segments vocaliques produits par 115 enfants entre 3 mois et 7 ans. A l'âge de 4 ans, leurs résultats suggèrent que la durée moyenne des voyelles est de l'ordre de 200 ms. Ces résultats sont cohérents avec les nôtres, puisque nous observons chez les enfants de 4 ans une durée moyenne de V1 de 173 ms et une durée moyenne de V2 de 206 ms. Notons que Lyasko & Grigor'ev (2015) relèvent une évolution importante de la durée des voyelles entre 4 et 5 ans, puisque cette dernière passe de 200 ms environ à 4 ans à environ 150 ms à l'âge de 5 ans, et reste relativement stable entre 5 et 7 ans. Cela suggère qu'il s'opère une maturation du contrôle moteur relativement importante entre l'âge de 4 ans et celui de 5 ans.

Ces résultats confirment notre hypothèse 4, à savoir que l'on devrait observer des gestes légèrement plus lents chez les enfants que chez les adultes. Nous interprétons ces résultats comme un indice de **l'immaturité du contrôle moteur** pour la production de parole, en particulier en ce qui concerne l'immaturité du modèle direct.

4.5. Relations entre nos principaux indices

Dans la dernière partie de nos résultats, nous avons cherché à savoir s'il existait une relation entre nos principaux indices. Plus précisément, nous avons cherché à savoir s'il existait une relation entre les indices numériques de stabilité et d'anticipation, et à savoir s'il existait une relation entre indice d'anticipation et débit d'élocution.

Nos résultats concernant la relation entre indices numériques de stabilité et d'anticipation montrent qu'il existe une relation entre ces deux indices. Les participants qui anticipent le plus V2 dans la réalisation de V1 sont aussi ceux qui montrent la meilleure stabilité du contrôle (c'est-à-dire la plus petite variabilité liée à la répétition des voyelles en contexte isolé). Ceci semble être vrai au sein de chacun des groupes d'âge (régression linéaire de pente 0.5469 pour le groupe Adulte ; régression linéaire de pente 0.1922 pour le groupe Enfant). Les participants du groupe Enfant qui anticipent le plus ont aussi tendance à être ceux qui montrent la meilleure stabilité du contrôle. Le fait que cette relation existe au sein du groupe Adulte montre également qu'il existe une certaine variabilité au sein du groupe Adulte concernant stabilité du contrôle et anticipation, et le fait que cette relation soit forte indique que l'indice d'anticipation est un indice robuste la de maturité motrice, puisque ce sont bien les participants qui montrent la meilleure stabilité qui montrent également le plus d'anticipation. Cependant, le fait que cette régression linéaire ne soit pas significative indique que la relation entre anticipation et stabilité du contrôle est plus complexe qu'une simple relation linéaire ne prenant en compte que ces deux facteurs.

L'indice de stabilité du contrôle est un indice couramment utilisé et constitue sans conteste un bon indice de maturité motrice. L'indice d'anticipation, quant à lui, n'a pas été extensivement utilisé et sa validité mérite encore d'être démontrée. Nos résultats concernant la relation entre indices numériques de stabilité et d'anticipation soulignent,

de notre point de vue, que l'indice d'anticipation est un indice robuste de la maturité motrice.

Nos résultats concernant la relation entre indices numériques d'anticipation et débit d'élocution montrent qu'il existe une relation entre ces deux indices. Les participants qui anticipent le plus V2 dans la réalisation de V1 sont aussi ceux qui ont un débit d'élocution rapide. Ceci semble être vrai au sein de chacun des groupes d'âge (régression linéaire de pente 0.6108 pour le groupe Adulte ; régression linéaire de pente 0.1738 pour le groupe Enfant). Les participants du groupe Enfant qui anticipent le plus sont donc aussi ceux qui ont un débit d'élocution plus rapide. Le fait que cette relation existe au sein du groupe Adulte montre également qu'il existe une certaine variabilité au sein du groupe Adulte concernant l'anticipation et le débit d'élocution, et le fait que cette relation soit forte indique que l'indice d'anticipation est fortement lié à l'indice du débit d'élocution, puisque ce sont bien les participants qui montrent la plus grande anticipation qui ont également un débit d'élocution rapide. Cependant, le fait que cette régression linéaire ne soit pas significative indique que la relation entre anticipation et débit d'élocution est plus complexe qu'une simple relation linéaire ne prenant en compte que ces deux facteurs.

Le fait qu'il existe une relation entre débit d'élocution et coarticulation est relativement bien admis. Le facteur débit d'élocution est probablement le facteur ayant le plus d'influence sur l'amplitude de la coarticulation. Une parole lente est peu coarticulée et une parole avec un débit d'élocution rapide est nécessairement très coarticulée. Cet élément ne remet pas en cause, selon nous, nos résultats concernant la coarticulation intra et extra syllabique. Parler avec un débit d'élocution rapide nécessite d'avoir une bonne connaissance de la variabilité associée à chaque son de parole, afin de rapprocher les cibles et de minimiser l'effort articulatoire. Sans connaissance suffisante de la variabilité associée à chaque son de parole et sans stratégie efficace de minimisation de l'effort articulatoire, il n'est pas possible de parler avec un débit d'élocution rapide. Ces résultats concernant le débit d'élocution et les résultats concernant la coarticulation, pris ensemble, mettent en évidence une moindre maturité du contrôle moteur pour la production de parole chez les enfants de 4 ans par rapport aux adultes.

4.6. Limites et perspectives

4.6.1. Limites de l'étude

Quelles que soient les qualités de l'étude expérimentale présentée dans ce travail de thèse, cette étude souffre de nombreuses limites. La première limite que nous identifions est celle liée au caractère non écologique de la situation expérimentale. La parole que nous enregistrons est en effet une parole de laboratoire, dans une situation d'interaction peu naturelle, puisque les participants n'interagissent qu'avec des marionnettes. La parole de laboratoire a souvent été décrite comme lente, sur articulée et sur planifiée par rapport à la parole naturelle, mais ces caractéristiques ne sont pas inhérentes à la parole de laboratoire, mais dépendent avant tout du design expérimental, et peuvent facilement être contrôlées (cf. Xu, 2010 pour un plaidoyer en faveur de la parole de laboratoire). Néanmoins, la parole de laboratoire reste de la parole et conserve nombre de ses caractéristiques naturelles, et à ce titre, la plupart des observations formulées en étudiant de la parole de laboratoire peuvent être généralisées à de la parole naturelle. Enfin, recueillir de la parole dans un cadre expérimental contrôlé est souvent une nécessité. Dans notre cas, nous devions recueillir de la parole d'adultes et d'enfants dans les mêmes conditions expérimentales, afin d'effectuer des comparaisons entre les deux groupes de participants.

La deuxième limite que nous identifions, qui relève davantage de la critique qui aurait pu être formulée à l'encontre de nos travaux que de la limite en soi, est celle liée à la langue. Nous étudions un phénomène dans une langue donnée, le français québécois, or nombre de phénomènes sont spécifiques à une langue. La coarticulation extra-syllabique de voyelle à voyelle peut effectivement être tributaire de plusieurs aspects spécifiques à une langue donnée, par exemple la taille du système vocalique (Manuel & Krakow, 1984). En effet, plus le système vocalique comporte de voyelles, plus la variabilité de ces voyelles semble être limitée. Le français québécois comportant un système vocalique riche, nous ne nous plaçons pas dans un cadre extrême où nous maximisons la possibilité d'observer ce type de coarticulation. De plus, la coarticulation extra-syllabique de voyelle à voyelle est un phénomène qui a été observé dans de nombreuses langues, par exemple en anglais (Beddor et al., 2002), en suédois (Öhman, 1966), ou en catalan (Recasens, 1984),

et nos résultats peuvent donc être étendus à toutes les langues pour lesquelles ce phénomène a été observé. Cependant, ce phénomène n'est pas universel, puisque dans certaines langues où la frontière syllabique joue un rôle plus important, par exemple en mandarin (Ma et al., 2015) ou en japonais (Chiu et al., 2015), ce phénomène n'est pas ou très peu observé. La généralisation de nos résultats ne peut donc pas s'appliquer à ces langues.

La troisième limite que nous identifions concerne la mesure de la coarticulation intra-syllabique et en particulier la mesure des équations de locus. L'absence du /u/ comme V_2 dans notre corpus, donc l'absence de contraste maximal en F2, peut limiter la bonne mesure des équations de locus, qui reposent sur la transition du F2. C'est peut-être une des raisons qui expliquent le manque de cohérence, dans nos résultats, entre pente des équations de locus et réalité articulatoire de la coarticulation.

Enfin, la quatrième limite que nous identifions, la plus importante, est le fait que notre étude expérimentale, seule, ne permet pas d'identifier clairement une composante unique de la parole enfantine qui serait immature. En effet, comme nous l'avons vu, il existe plusieurs possibles sources de variabilité qui amènent à l'observation d'un comportement immature (par exemple, la définition des buts de la production, le bruit biologique résiduel du système périphérique, ou encore l'immaturité des relations sensori-motrices). Nous défendons l'hypothèse selon laquelle l'absence d'anticipation relèverait avant tout de l'immaturité des relations sensori-motrices, mais d'autres hypothèses alternatives peuvent être formulées, comme celle du poids de la syllabe dans l'organisation de la parole des jeunes enfants. Afin de clarifier le rôle de chacune des sources de variabilité, plusieurs perspectives de travail peuvent être envisagées.

4.6.2. Perspectives

Afin de mieux comprendre le rôle, au cours du développement, de chacune des possibles sources de variabilité de la production de la parole, plusieurs pistes de recherche peuvent être envisagées. Tout d'abord, la mise en place d'études développementales portant sur la production de la parole, à différents stades ontogéniques, et comportant différents types de tâches, mettant en jeu de façon différente chacune des composantes (stabilité des cibles, anticipation, vitesse d'exécution), permettrait de mieux mesurer la maturité de

chacune de ces composantes au cours du développement, en particulier la maturation des représentations sensori-motrices en termes de précision, de stabilité, et de capacité d'anticipation.

La mise en place d'études développementales portant sur le contrôle lingual en dehors du cadre de la parole permettrait d'évaluer le rôle du bruit biologique résiduel seul, et de constituer une mesure de référence du contrôle lingual. La mise en place d'études développementales portant sur l'anticipation en dehors du cadre de la parole permettrait de comprendre si nos résultats portant sur l'anticipation, observés dans le cadre de la parole, sont spécifiques à ce cadre ou non. Si un manque d'anticipation est observé dans et hors du cadre de la parole, alors nos résultats portant sur l'anticipation seraient dus à une immaturité générale des représentations sensori-motrices, alors que si ce manque d'anticipation n'est observé que dans le cadre de la parole, cette immaturité des représentations sensori-motrices pourrait être spécifique aux cibles phonémiques.

L'implémentation d'un modèle de contrôle de la production de la parole, couplé à un modèle biomécanique de langue, et confronté aux résultats expérimentaux, permettrait de tester indépendamment les conséquences théoriques de chacune des possibles sources de variabilité. Le modèle actuel repose sur l'idée de planification motrice optimale utilisant un modèle interne décrivant les classes phonologiques, les liens entre commandes motrices et ces classes phonologiques, incluant les degrés de liberté et la variabilité autorisée autour de chacune de ces classes. La limitation d'un de ces aspects indépendamment des autres permettrait une évaluation quantitative de ces effets, et la comparaison avec les données expérimentales permettrait d'étayer les différentes hypothèses. En particulier, cela permettrait d'évaluer le rôle théorique de la précision de la définition des buts de la production, composante très difficilement accessible expérimentalement.

Il serait également intéressant de conduire des études portant sur la perception de la parole au cours du développement, car ce domaine a encore été très peu exploré, or on sait combien la perception influe sur la production et vice-versa. Par exemple, une perception catégorielle immature pourrait conduire à une moins bonne définition des

buts de la production et de l'exploitation de la variabilité autorisée autour de chaque son de parole.

Enfin, des études de perturbation de l'articulation et de perturbation des retours sensoriels permettraient de mieux caractériser la maturité des relations sensori-motrices au cours du développement et permettraient de mieux comprendre le rôle de chacune des modalités (auditive, somatosensorielle) dans la définition des buts de la production.

Conclusion

Ce travail de thèse a proposé de s'intéresser à la production de la parole à l'âge de 4 ans, selon une perspective de contrôle moteur, à travers la mesure de différents indices et la comparaison avec la mesure de ces mêmes indices chez une population de jeunes adultes. L'objectif était de contribuer à qualifier et à quantifier la maturité du contrôle moteur pour la production de parole à l'âge de 4 ans, à travers la proposition d'un cadre théorique pour le contrôle moteur de la parole et d'indices adaptés à la mesure de la maturité de ce contrôle. La motivation de ce travail de thèse vient du constat qu'il existe un manque cruel de données, et particulièrement de données articulatoires relatives au développement tardif de la production de la parole. Elle vient également du fait que la littérature portant sur la coarticulation linguale anticipatoire a produit des résultats contradictoires et essentiellement basés sur des analyses acoustiques. Un des objectifs de ce travail était de contribuer à clarifier ces résultats, en particulier via l'apport de données articulatoires.

Nous avons recueilli des données acoustiques et articulatoires concernant 20 participants enfants âgés de 4 ans et 10 jeunes adultes. Le corpus étudié était composé de voyelles isolées et de séquences V_1-C-V_2. Les données articulatoires ont été obtenues via l'échographie linguale, couplée à un dispositif HOCUS.

Ces données ont été interprétées à la lumière du cadre théorique qui a été présenté dans le premier chapitre de cette thèse (section 1.3.). Les principales hypothèses de ce cadre théorique sont que les unités de base de la production de la parole sont les phonèmes ; que la définition des buts de la production s'effectue dans l'espace acoustique et somatosensoriel ; que le processus d'apprentissage repose sur deux phases distinctes, une première phase d'exploration libre (dont font partie les premières productions du nouveau-né et le babillage), permettant l'apprentissage d'un modèle direct, et une seconde phase d'apprentissage, mettant en jeu le modèle inverse, permettant de spécifier les commandes motrices pour une séquence de buts moteurs donnée ; que les trajectoires des gestes de parole, dans le cadre de la production d'une séquence de sons de parole (dont le phénomène de coarticulation fait partie), résultent d'un processus de

planification motrice optimale, reposant sur l'exploitation de la variabilité allouée à chaque son de parole.

En accord avec ce cadre théorique et les données de la littérature, nous avions formulé 3 hypothèses théoriques *a priori* (engendrant 4 hypothèses ou attentes sur les observations) sur les caractéristiques du contrôle moteur de la parole chez les enfants de 4 ans : (1) ils ne disposent pas encore de la capacité à mémoriser avec précision les patrons moteurs associés à l'exécution d'une tâche motrice liée au phonème ; (2) le caractère répétitif du babillage a engendré chez eux une forte cohésion de la structure syllabique dont on devrait encore observer les traces à 4 ans ; (3) les enfants de 4 ans n'ont pas encore acquis des représentations sensori-motrices internes précises de leur système de production de la parole dans leur système nerveux central. Dans ce contexte, nous nous attendions aux observations suivantes :

(1) davantage de variabilité liée à la répétition d'une même tâche chez les enfants par rapport aux adultes ;

(2) des patrons de coarticulation intra-syllabique d'amplitude similaire voire plus importante chez les enfants comparé aux adultes ;

(3) une moindre amplitude de coarticulation extra-syllabique chez les enfants par rapport aux adultes ;

(4) des gestes légèrement plus lents chez les enfants que chez les adultes.

Nos résultats concernant la stabilité de la production de voyelles isolées ont permis de confirmer l'hypothèse (1) : nous observons effectivement davantage de variabilité liée à la répétition dans les productions enfantines. Ceci est cohérent avec l'hypothèse d'une moindre stabilité du contrôle chez l'enfant.

Nos résultats concernant la coarticulation intra-syllabique tendent à infirmer notre hypothèse (2) : l'amplitude de la coarticulation intra-syllabique semble moins importante chez l'enfant que chez l'adulte. Ceci laisse penser que la forte cohésion au sein de la syllabe initiée par le babillage a déjà laissé la place à un contrôle sériel décomposant la syllabe en phonèmes. La coarticulation anticipatoire intra-syllabique serait alors selon notre cadre théorique, tout comme la coarticulation extra-syllabique, le résultat d'un

contrôle optimal qui n'est pas encore maîtrisé par l'enfant (cf. ci-dessous notre analyse de la coarticulation anticipatoire extra-syllabique). Les mesures acoustiques semblaient indiquer une amplitude similaire de coarticulation intra-syllabique chez les enfants et les adultes, alors que la lecture des données articulatoires a suggéré au contraire que l'amplitude de la coarticulation intra-syllabique était moindre chez l'enfant. Ce résultat met en avant la force des données articulatoires pour l'étude fine du contrôle moteur. Dans le cas précis de la coarticulation intra-syllabique, le calcul des équations du locus nécessite que les formants soient mesurables : temporellement des mesures plus tardives sont effectuées par rapport aux mesures articulatoires, c'est-à-dire des mesures effectuées plus proches de la voyelle. C'est probablement cette différence qui explique le désaccord entre les conclusions tirées par les deux types de données.

Nos résultats concernant la coarticulation extra-syllabique et les mesures de durées ont permis de confirmer les hypothèses (3) et (4) : nous observons une amplitude plus faible de la coarticulation extra-syllabique et un débit d'élocution plus faible chez l'enfant par rapport à l'adulte. Ces deux résultats sont cohérents avec l'idée que chez les enfants, les modèles internes, s'ils existent, sont moins développés, moins complets, moins généraux que chez les adultes. En effet, les enfants semblent disposer d'une moindre capacité que les adultes à simuler mentalement les conséquences acoustiques et somatosensorielles des commandes motrices, afin d'optimiser la séquence des commandes motrices pour atteindre les buts de parole donnés. Leurs modèles internes étant moins efficaces, les enfants n'ont pas encore la capacité ou disposent d'une capacité moindre à utiliser un contrôle de type *feedforward*, permettant de générer une séquence de parole efficacement, en l'absence de perturbation inattendue, sans avoir à attendre les retours sensoriels biologiques externes (*feedback*). Le fait que les enfants se reposent davantage sur le feedback biologique externe, qui nécessite d'attendre ce retour sensoriel, dont le délai peut être long, peut expliquer, dans notre cadre d'interprétation, le débit d'élocution plus lent des enfants.

La figure 5.1. ci-dessous résume à travers une vue d'artiste les principaux résultats de cette étude expérimentale, à savoir la moindre stabilité du contrôle et la moindre anticipation des gestes suivants dans la production d'une séquence de sons de parole chez l'enfant.

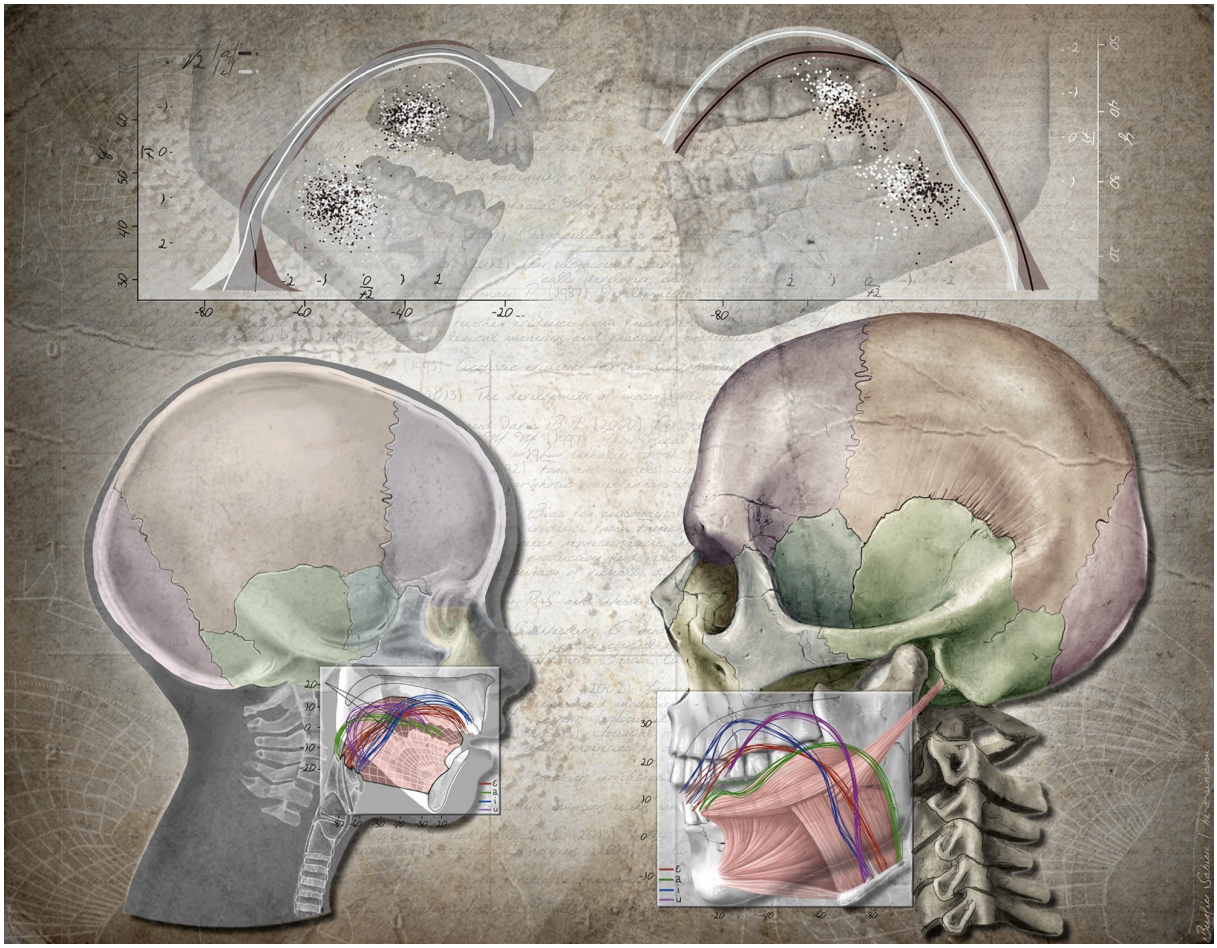


Figure 5.1. Vue d'artiste de nos principaux résultats. Dans les panneaux du haut, illustration de l'absence d'anticipation chez les enfants de 4 ans (à gauche) et de la forte anticipation chez les adultes (à droite). On observe les tracés linguaux de la voyelle V1 selon que V2 soit /a/ (tracé noir) ou /i/ (tracé blanc), et, en superposition, les nuages de points correspondants dans le plan F1 - F2. Dans les panneaux du bas, on observe la variabilité des articulations d'une même voyelle répétée (enfant à gauche, adulte à droite), illustrant la moindre stabilité du contrôle chez les enfants et de la grande stabilité du contrôle chez les adultes. Illustration de Sabine Burfin.

L'objectif principal de cette thèse n'était pas de démontrer que le système de production de parole chez l'enfant de 4 ans est immature. De nombreuses études expérimentales (souvent acoustiques), voire simplement l'observation, ont en effet déjà permis de mettre en évidence cet élément. L'objectif de cette thèse était d'évaluer en quoi la parole des enfants de 4 ans est immature par rapport à celle des adultes. De ce point de vue, nous soutenons le fait que la variabilité en répétition est un bon indice concernant une mesure de la stabilité du contrôle moteur. Nous avons également, et c'est plus original, mis en évidence que l'anticipation extra-syllabique est un bon indice de maturité du

contrôle moteur. Nous défendons le point de vue selon lequel, pris ensemble, les indices de stabilité, de débit d'élocution et d'anticipation fournissent des mesures quantitatives permettant d'inférer le degré de maturation du système moteur pour la production de parole à un stade ontogénique donné.

Concernant le débat sur le développement de la coarticulation linguale, notre étude a également apporté une contribution en participant à clarifier les résultats contradictoires de la littérature. Notre revue de la littérature (section 1.4.4) tentait déjà de clarifier les résultats des différentes études, en séparant ce qui relève du développement précoce de ce qui relève du développement tardif, en séparant les domaines d'organisation étudiés, via la séparation entre coarticulation intra et extra-syllabique, et en séparant les types de sons de parole étudiés via la séparation entre coarticulation de type fricative-voyelle et de type plosive-voyelle. Nos résultats apportent des éléments de réponse quant à la coarticulation intra-syllabique de type plosive-voyelle et à la coarticulation extra-syllabique, en indiquant que l'amplitude de ce type de coarticulation est moindre chez l'enfant de 4 ans par rapport à l'adulte. Nos résultats indiquant une moindre coarticulation intra-syllabique chez l'enfant revêtent un intérêt tout particulier puisqu'ils mettent en évidence un élément nouveau, qui n'avait pas été mis en évidence par les études acoustiques. Ces résultats participent à dresser un schéma développemental typique cohérent des patrons de coarticulation, en indiquant qu'il semble bien que l'amplitude de la coarticulation croisse avec l'âge. Ces résultats sont par ailleurs cohérents avec le cadre théorique dans lequel ce travail s'inscrit, et apportent des éléments en défaveur de l'hypothèse selon laquelle la parole enfantine serait organisée autour d'unités plus larges que le phonème.

D'un point de vue théorique, les résultats expérimentaux de cette thèse sont en effet cohérents avec le cadre théorique selon lequel (1) la parole est organisée très tôt autour d'unités segmentales, les phonèmes, et (2) la coarticulation repose sur un processus de planification motrice optimale au sein de laquelle les cibles sont phonémiques. Nos résultats soutiennent donc l'hypothèse selon laquelle la parole des enfants de 4 ans est organisée, comme l'est celle de l'adulte, autour du phonème. Pour réellement mettre en difficulté ces modèles théoriques de production de parole, il semblerait qu'il faille étudier

des stades relativement précoces, pour lesquels la parole, si on peut la nommer ainsi, n'est pas encore organisée comme le semble être la parole adulte.

Pris ensemble, ces résultats soulignent l'immaturité du contrôle moteur pour la production de la parole chez les enfants de 4 ans par rapport aux adultes, et ce sous différents aspects. Si l'on reprend le modèle théorique de production de parole que nous avons adopté pour interpréter nos résultats, ces derniers soulignent l'immaturité du système moteur à plusieurs niveaux : immaturité dans la **définition des buts de la production** (mis en évidence dans nos données par la stabilité des voyelles isolées, mais aussi et surtout par les données de la littérature concernant le développement tardif de la perception catégorielle), possible immaturité dans le **bruit biologique résiduel du système périphérique** (bruit dans les commandes motrices, mis en évidence par la plus grande variabilité des productions enfantines en répétition), **immaturité des modèles internes, inverse et direct** (mise en évidence par la faible anticipation de V2 dans V1 et également par la moindre anticipation intra-syllabique et par la lenteur des gestes articulatoires enfantins).

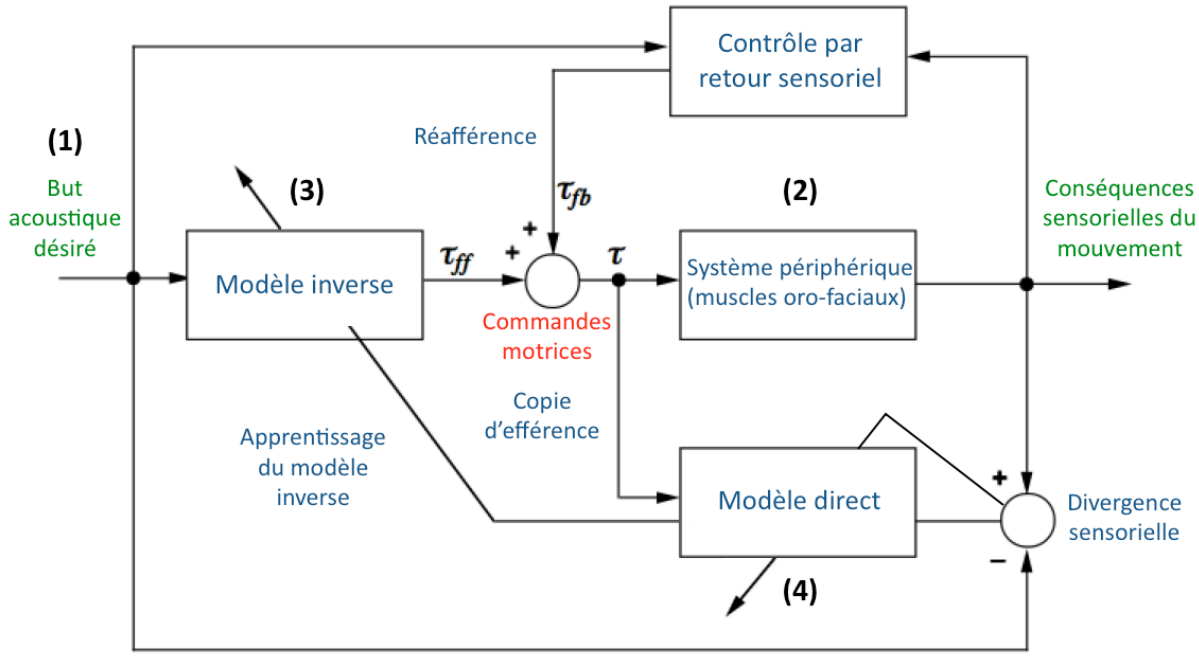


Figure 5.2. Modèle théorique de la production de la parole et possibles sources de variabilité et d'immaturité : (1) Définition des buts de la production, (2) Bruit biologique résiduel, (3) Immaturité du modèle inverse et (4) Immaturité du modèle direct. Les flèches traversant les modèles indiquent un apprentissage. Adapté de Jordan et Rumelhart (1992).

Plusieurs facteurs potentiels de variabilité entrent donc en jeu à différents niveaux, et les contributions individuelles de chacun de ces facteurs de variabilité restent difficiles à évaluer. Nous pensons que l'outil de modélisation pourrait contribuer à désambigüiser le rôle de chacun de ces facteurs de variabilité, via l'implémentation d'un modèle de contrôle moteur couplé à un modèle biomécanique de langue (à l'instar de Ma, 2008) au sein duquel il serait possible d'évaluer la contribution individuelle de chacun de ces facteurs de variabilité, en modifiant séparément les paramètres de chacun de ces facteurs et en observant les conséquences de ces modifications sur le comportement du modèle, en les comparant aux comportements observés via l'expérimentation. Une telle démarche constitue, de notre point de vue, une perspective de travail intéressante concernant le volet développemental. Mais le premier effort à fournir vers une meilleure compréhension du processus d'apprentissage de la production de parole semble rester le développement d'études expérimentales, portant à la fois sur la perception de la parole et sur sa production, à différents stades ontogéniques, dans le but de mieux comprendre les contributions de chacun de ces facteurs de variabilité.

Pour conclure, le chemin vers l'élaboration du profil ontogénétique du développement typique de la production de la parole est encore long et les efforts à fournir dans cette direction sont encore nombreux. Nous restons convaincus que le développement de la production de la parole, à l'instar du développement précoce de la perception, dispose encore de nombreux éléments de réflexion à offrir quant à la compréhension de la nature même de cette faculté et des processus mis en jeu dans les mécanismes de perception et de production de la parole. Nous espérons que ce travail de thèse aura modestement contribué à l'effort qu'il reste à fournir dans cette direction.

Références bibliographiques

Abd-El-Malek, S. (1939). Observations on the morphology of the human tongue. *Journal of Anatomy*, 73 (2), 201.

Abercrombie, D. (1967). *Elements of general phonetics* (Vol. 203). Edinburgh: Edinburgh University Press.

Abramson, A. S. & Lisker, L. (1970). Discriminability along the voicing continuum: Cross-language tests. In *Proceedings of the sixth international congress of phonetic sciences*, 196 (7), Academia Prague, 569-573.

Abry, C. & Lallouache, T. (1995). Le MEM: un modèle d'anticipation paramétrable par locuteur: Données sur l'arrondissement en français. *Les Cahiers de l'ICP, Bulletin de la Communication Parlée*, (3), 85-99.

Acher, A. (2014). *Corrélats cérébraux de l'adaptation de la parole après exérèse de la cavité orale*. Thèse de Doctorat. Université de Grenoble.

Ackermann, H., Wildgruber, D., Daum, I. & Grodd, W. (1998). Does the cerebellum contribute to cognitive aspects of speech production? A functional magnetic resonance imaging (fMRI) study in humans. *Neuroscience letters*, 247 (2), 187-190.

Articulate Instruments, L. T. D. (2010). *Articulate assistant user guide: version: 2.11*. Edinburgh: Articulate Instruments.

Assmann, P. F. & Katz, W. F. (2000). Time-varying spectral change in the vowels of children and adults. *The Journal of the Acoustical Society of America*, 108 (4), 1856-1866.

Aubin, J. (2006). *Effets acoustiques et articulatoires de perturbations labiales sur la parole des enfants et des adultes*. Mémoire de Maitrise, Université du Québec à Montréal.

Baer, T., Alfonso, P. J. & Honda, K. (1988). Electromyography of the tongue muscles during vowels in /gpVp/ environment. *Ann. Bull. RILP*, (22), 7-19.

Bailly, G. (1997). Learning to speak. Sensori-motor control of speech movements. *Speech Communication*, 22 (2), 251-267.

Barbier, G. (2010). *Croissance du conduit vocal de la naissance à l'âge adulte Étude radiographique longitudinale*. Mémoire de Master. Sciences du langage, Université Stendhal, Grenoble.

Barbier, G. (2011). *Production de la parole chez l'enfant de 4 ans : modélisation biomécanique de la langue, échographie linguale et contrôle moteur*. Master en Sciences Cognitives, Université de Grenoble.

Barbier, G., Boë, L. J. & Captier, G. (2012). La croissance du conduit vocal du fœtus à l'adulte: une étude longitudinale. *Biométrie Humaine et Anthropologie*, 30 (1-2), 11-22.

Barbier, G., Boë, L. J., Captier, G. & Laboissière, R. (2015). Human vocal tract growth: A longitudinal study of the development of various anatomical structures. In *Proceedings of Interspeech 2015*, Dresden, Germany.

Bard, C., Hay, L. & Fleury, M. (1990). Timing and accuracy of visually directed movements in children: Control of direction and amplitude components. *Journal of Experimental Child Psychology*, 50 (1), 102-118.

Baum, S. R. & Katz, W. F. (1988). Acoustic analysis of compensatory articulation in children. *The Journal of the Acoustical Society of America*, 84 (5), 1662-1668.

Beddor, P. S., Harnsberger, J. D. & Lindemann, S. (2002). Language-specific patterns of vowel-to-vowel coarticulation: acoustic structures and their perceptual correlates. *Journal of Phonetics*, 30 (4), 591-627.

Berry, J., Fasel, I., Fadiga, L. & Archangeli, D. (2012) Training Deep Networks with Imbalanced and Unlabeled Data. In *Proceedings of Interspeech 2012*, Portland, OR, USA, 1756–1759.

Bizzi, E., Kalil, R. E. & Tagliasco, V. (1971). Eye-head coordination in monkeys: evidence for centrally patterned organization. *Science*, 173 (3995), 452-454.

Bizzi, E., Mussa-Ivaldi, F. A. & Giszter, S. (1991). Computations underlying the execution of movement: a biological perspective. *Science*, 253 (5017), 287-291.

Bizzi, E., Polit, A. & Morasso, P. (1976). Mechanisms underlying achievement of final head position. *Journal of Neurophysiology*, 39 (2), 435-444.

Boë, L. J. (1989). Maximal vowel space. In *Proceedings of Eurospeech 89*, 281-284.

Boë, L. J. & Maeda, S. (1998). Modélisation de la croissance du conduit vocal. *Journées d'Études Linguistiques, La voyelle dans tous ses états*, 98-105.

Boë, L. J., Badin, P., Ménard, L., Captier, G., Davis, B., MacNeilage, P. ... & Schwartz, J. L. (2013). Anatomy and control of the developing human vocal tract: A response to Lieberman. *Journal of Phonetics*, 41 (5), 379-392.

Boë, L. J., Heim, J. L., Honda, K. & Maeda, S. (2002). The potential Neandertal vowel space was as large as that of modern humans. *Journal of Phonetics*, 30 (3), 465-484.

Boë, L. J., Heim, J. L., Honda, K., Maeda, S., Badin, P. & Abry, C. (2007). The vocal tract of newborn humans and Neanderthals: Acoustic capabilities and consequences for the debate on the origin of language. A reply to Lieberman (2007a). *Journal of Phonetics*, 35 (4), 564-581.

Boë, L.-J., Perrier, P., Guérin, B. & Schwartz, J.-L. (1989). Maximal Vowel Space, *First European Conference on Speech Communication and Technology*, 2, 281-284.

Boersma, P. & Weenink, D. (1996). Praat, a system for doing phonetics by computer, version 3.4, *Report No. 132, Institute of Phonetic Sciences of the University of Amsterdam*, 1–182.

Bothorel, A., Simon, P., Wioland, F. & Zerling, J.-P. (1986). *Cinéradiographies des voyelles et consonnes du Français* [Cineradiographs of French vowels and consonants]. Strasbourg, Institut de Phonétique.

Bouchet A. & Cuilleret J. (1991). *Anatomie : Tome I*. Editions Masson.

Bourguignon, N. J., Baum, S. R. & Shiller, D. M. (2014). Lexical-perceptual integration influences sensorimotor adaptation in speech. *Frontiers in human neuroscience*, 8, article 208.

Bressman, T., Thind, P., Bollig, C. M., Uy, C., Gilbert, R. W., & Irish, J. C. (2005). Quantitative three- dimensional ultrasound analysis of tongue protrusion and grooving: Data from 12 normal speakers and a partial glossectomee. *Clinical Linguistics and Phonetics*, 19, 573–588.

Browman, C. & Goldstein, L. (2000). Competing constraints on intergestural coordination and self-organization of phonological structures. *Les Cahiers de l'ICP, Bulletin de la Communication Parlée*, vol. 5, 25-34.

Browman, C. P. & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology*, 6 (02), 201-251.

Browman, C. P. & Goldstein, L. M. (1986). Towards an articulatory phonology. *Phonology*, 3 (01), 219-252.

Bruderer, A. G., Danielson, D. K., Kandhadai, P. & Werker, J. F. (2015). Sensorimotor influences on speech perception in infancy. *Proceedings of the National Academy of Sciences*, 112 (44), 13531-13536.

Buchaillard, S. (2007). *Activations musculaires et mouvements linguaux : modélisation en parole naturelle et pathologique*. Thèse de Doctorat Université Joseph-Fourier Grenoble I.

Buchaillard, S., Perrier, P. & Payan, Y. (2009). A biomechanical model of cardinal vowel production: Muscle activations and the impact of gravity on tongue positioning. *The Journal of the Acoustical Society of America*, 126 (4), 2033-2051.

Buhr, R. D. (1980). The emergence of vowels in an infant. *Journal of Speech, Language, and Hearing Research*, 23 (1), 73-94.

Busby, P. A. & Plant, G. L. (1995). Formant frequency values of vowels produced by preadolescent boys and girls. *The Journal of the Acoustical Society of America*, 97 (4), 2603-2606.

Byrd, D. (1995). C-centers revisited. *Phonetica*, 52(4), 285-306.

Callan, D. E., Kent, R. D., Guenther, F. H. & Vorperian, H. K. (2000). An auditory-feedback-based neural network model of speech production that is robust to developmental changes in the size and shape of the articulatory system. *Journal of Speech, Language, and Hearing Research*, 43 (3), 721-736.

Canault, M. (2007). *L'émergence du contrôle articulatoire au stade du babillage. Une étude acoustique et cinématique*. Thèse de Doctorat, Université Marc Bloch-Strasbourg II.

Captier, G., Boë, L. J. & Barbier, G. (2010). Anatomie et croissance du conduit vocal du fœtus à l'enfant de 5 ans. *Biométrie humaine et anthropologie*, 28 (3-4), 65-73.

Casal, C., Dominguez, C., Fernandez, A., Sarget, R. O. N., Martínez-Celdrán, E., Sentís-Vilalta, J. & Gay-Escoda, C. (2002). Spectrographic measures of the speech of young children with cleft lip and cleft palate. *Folia phoniatrica et logopaedica*, 54 (5), 247-257.

Cheng, H. Y., Murdoch, B. E., Goozee, J. V. & Scott, D. (2007). Physiologic development of tongue–jaw coordination from childhood to adulthood. *Journal of Speech, Language, and Hearing Research*, 50 (2), 352-360.

Childers, D. G. & Wu, K. (1991). Gender recognition from speech. Part II: Fine analysis. *The Journal of the Acoustical society of America*, 90 (4), 1841-1856.

Chiu, F., Fromont, L., Lee, A. & Xu, Y. (2015). Long-distance anticipatory vowel-to-vowel assimilatory effects in French and Japanese. In *Proceedings of the 18th International Congress of Phonetic Sciences, ICPhS* 2015.

Chomsky, N. (2000). *New horizons in the study of language and mind*. Cambridge University Press.

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

Cutler, A. & Mehler, J. (1993). The periodicity bias. *Journal of Phonetics*, 21 (1-2), 103-108.

Damon, F., Bayet, L., Quinn, P. C., Hillairet de Boisferon, A., Méary, D., Dupierrix, E. ... & Pascalis, O. (2015). Can human eyes prevent perceptual narrowing for monkey faces in human infants? *Developmental Psychobiology*, 57 (5), 637-642.

Dang, J. & Honda, K. (2002). Estimation of vocal tract shapes from speech sounds with a physiological articulatory model. Journal of Phonetics, 30 (3), 511-532.

Dang, J. & Honda, K. (2004). Construction and control of a physiological articulatory model. *The Journal of the Acoustical Society of America*, 115 (2), 853-870.

Darwin, C. (1877). A biographical sketch of an infant. *Mind*, (7), 285-294.

Davidson, L. (2006). Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance, *Journal of the Acoustical Society of America*, 120, 407-415.

De Boer, B. & Fitch, W. T. (2010). Computer models of vocal tract evolution: An overview and critique. *Adaptive Behavior*, 18 (1), 36-47.

de Boysson-Bardies, B., & Vihman, M. M. (1991). Adaptation to language: Evidence from babbling and first words in four languages. *Language*, 297-319.

de Boysson-Bardies, B., Hallé, P., Sagart, L. & Durand, C. (1989). A crosslinguistic investigation of vowel formants in babbling. *Journal of Child Language*, 16 (01), 1-17.

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

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

Diehl, R. L. & Kluender, K. R. (1989). On the objects of speech perception. *Ecological Psychology*, 1 (2), 121-144.

Edwards, J. (1992). Compensatory speech motor abilities in normal and phonologically disordered children. *Journal of Phonetics*, 20 (2), 189-207.

Edwards, J., Beckman, M. E. & Munson, B. (2015). Frequency effects in phonological acquisition. *Journal of Child Language*, 42 (2), 306-311.

Eguchi, S. & Hirsh, I. J. (1969). Development of speech sounds in children. *Acta oto-laryngologica.* Supplementum, 257, 1-51.

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

Eimas, P. D. & Miller, J. L. (1980). Discrimination of information for manner of articulation. *Infant Behavior and Development*, 3, 367-375.

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

Enard, W., Gehre, S., Hammerschmidt, K., Hölter, S. M., Blass, T., Somel, M., ... & Becker, L. (2009). A humanized version of Foxp2 affects cortico-basal ganglia circuits in mice. *Cell*, 137 (5), 961-971.

Fant, G. (1960). *Acoustic theory of speech production*. The Hague: Mouton.

Fant, G. (1975). A note on vocal tract size factors and non-uniform F-pattern scalings. *Speech Transmission Laboratory Quarterly Progress & Status Reports*, 4, 22–30.

Fant, G. (1975). *Non-uniform vowel normalization*. STL-QPSR, 16 (2-3), 1-19.

Ferrante, C., Van Borsel, J. & Pereira, M. M. D. B. (2008). Phonological acquisition in socio-economical high-class children. *Revista CEFAC*, 10(4), 452-460.

Fitch, W. T. & Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *The Journal of the Acoustical Society of America*, 106 (3), 1511-1522.

Forssberg, H., Eliasson, A. C., Kinoshita, H., Johansson, R. S. & Westling, G. (1991). Development of human precision grip I: basic coordination of force. *Experimental Brain Research*, 85 (2), 451-457.

Fowler, C. A. & Turvey, M. T. (1980). Immediate compensation in bite-block speech. *Phonetica*, 37 (5-6), 306-326.

Gàbor Csapò, T. & Lulich, S. M. (2015) Error analysis of extracted tongue contours from 2D ultrasound images. In *Proceeding of Interspeech 2015*, Dresden, Germany, 1156-1186.

Gangji, N., Pascoe, M. & Smouse, M. (2015). Swahili speech development: preliminary normative data from typically developing pre-school children in Tanzania. *International Journal of Language & Communication Disorders*, 50 (2), 151-164.

Gay, T., Lindblom, B. & Lubker, J. (1981). Production of bite-block vowels: acoustic equivalence by selective compensation. *The Journal of the Acoustical Society of America*, 69 (3), 802-810.

Ghosh, S. S., Matthies, M. L., Maas, E., Hanson, A., Tiede, M., Ménard, L., ... & Perkell, J. S. (2010). An investigation of the relation between sibilant production and somatosensory and auditory acuity. *The Journal of the Acoustical Society of America*, 128 (5), 3079-3087.

Gibbon, F. & Hardcastle, W. (1989). Deviant articulation in a cleft palate child following late repair of the hard palate: A description and remediation procedure using electropalatography (EPG). . *Clinical Linguistics and Phonetics*, 3, 93-110.

Gibson, A. & McPhearson, L. (1980). Production of bite-block vowels by children. *Phonetic Exp. Res. Inst. Linguistics* Univ. Stockholm II, 26-43.

Gilbert, H. R., Robb, M. P. & Chen, Y. (1997). Formant frequency development: 15 to 36 months. *Journal of Voice*, 11 (3), 260-266.

Goffman, L. & Smith, A. (1999). Development and phonetic differentiation of speech movement patterns. *Journal of Experimental Psychology: Human Perception and Performance*, 25 (3), 649.

Goldstein, U. G. (1980). *An articulatory model for the vocal tracts of growing children*. Doctoral dissertation, Massachusetts Institute of Technology.

Gomi, H., Shidara, M., Takemura, A., Inoue, Y., Kawano, K. & Kawato, M. (1998). Temporal firing patterns of Purkinje cells in the cerebellar ventral paraflocculus during ocular following responses in monkeys I. Simple spikes. *Journal of Neurophysiology*, 80 (2), 818-831.

Goodell, E. W. & Studdert-Kennedy, M. (1993). Acoustic Evidence for the Development of Gestural Coordination in the Speech of 2-Year-Olds: A Longitudinal Study. *Journal of Speech, Language, and Hearing Research*, 36 (4), 707-727.

Grabski, K. (2012). *Les cartes sensorimotrices de la parole: Corrélats neurocognitifs et couplage fonctionnel des systèmes de perception et de production des voyelles du Français*. Thèse de Doctorat. Université de Grenoble.

Green, J. R., Moore, C. A. & Reilly, K. J. (2002). The sequential development of jaw and lip control for speech. *Journal of Speech, Language, and Hearing Research*, 45 (1), 66-79.

Gregoire, A. (1937). *L'apprentissage du langage. (1). Les deux premières années*. Faculté de Philosophie et Lettres.

Grieser, D. & Kuhl, P. K. (1989). Categorization of speech by infants: Support for speech-sound prototypes. *Developmental Psychology*, 25 (4), 577.

Grosvald, M. (2009). Interspeaker variation in the extent and perception of long-distance vowel-to-vowel coarticulation. *Journal of Phonetics*, 37 (2), 173-188.

Gu, C. (2002). *Smoothing Spline ANOVA Models*. Springer, New York.

Gu, C. (2012). *GSS: General smoothing splines*. R package version 2.0-9, http://cran.r-project.org/web/packages/gss/index.html.

Guenther, F. H. (1995). Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psychological review*, 102 (3), 594.

Guenther, F. H. (2006). Cortical interactions underlying the production of speech sounds. *Journal of Communication Disorders*, 39 (5), 350-365.

Guenther, F. H., Hampson, M., & Johnson, D. (1998). A theoretical investigation of reference frames for the planning of speech movements. *Psychological Review*, 105 (4), 611.

Guenther, F. H., Perkell, J. S., Maassen, B., Kent, R. D., Peters, H. F. M., van Lieshout, P. H. H. M., & Hulstijn, W. (2004). A neural model of speech production and its application to studies of the role of auditory feedback in speech. In *Speech motor control in normal and disordered speech*, 29-49.

Hagiwara, R. (1997). Dialect variation and formant frequency: The American English vowels revisited. *The Journal of the Acoustical Society of America*, 102 (1), 655-658.

Hallé, P. & Cristia, A. (2012). Global and detailed speech representations in early language acquisition. In *Speech planning and dynamics*. Peter Lang, 11-38.

Hazan, V. & Barrett, S. (2000). The development of phonemic categorization in children aged 6–12. *Journal of Phonetics*, 28 (4), 377-396.

Henke, W. L. (1966). *Dynamic articulatory model of speech production using computer simulation*. Doctoral dissertation, Massachusetts Institute of Technology.

Heyne, M. & Derrick, D. (2015). Using a radial ultrasound probe's virtual origin to compute midsagittal smoothing splines in polar coordinates. *The Journal of the Acoustical Society of America*, 138 (6), EL509-EL514.

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*, 13 (2), 135-145.

Hillenbrand, J., Getty, L. A., Clark, M. J. & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical society of America*, 97 (5), 3099-3111.

Hillenbrand, J., Minifie, F. D. & Edwards, T. J. (1979). Tempo of spectrum change as a cue in speech-sound discrimination by infants. *Journal of Speech, Language, and Hearing Research*, 22(1), 147-165.

Hintzman, D. L. (1984). MINERVA 2: A simulation model of human memory. *Behavior Research Methods, Instruments & Computers*, 16 (2), 96-101.

Hodge, M. M. (1989). *A Comparison of Spectral Temporal Measures Across Speaker Age: Implications for an Acoustic Characterization of Speech Maturation*. Doctoral Dissertation. Ann Arbor, Michigan: University Microfilms International.

Hogan, N. (1984). An organizing principle for a class of voluntary movements. *The Journal of Neuroscience*, 4 (11), 2745-2754.

Honda, K., Hirai, H. & Kusakawa, N. (1993). Modeling vocal tract organs based on MRI and EMG observations and its implication on brain function. *Research Institute of Logopedics and Phoniatrics Annual Bulletin*, 27, 37-49.

Hothorn, T., Bretz, F. and Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, 50, 3, 346–363.

Houde, J. F. & Jordan, M. I. (1998). Sensorimotor adaptation in speech production. *Science*, 279 (5354), 1213-1216.

Houde, J. F. & Jordan, M. I. (2002). Sensorimotor Adaptation of Speech ICompensation and Adaptation. *Journal of Speech, Language, and Hearing Research*, 45 (2), 295-310.

Hourcade, J. P., Bederson, B. B., Druin, A. & Guimbretière, F. (2004). Differences in pointing task performance between preschool children and adults using mice. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 11 (4), 357-386.

Howard, I. S. & Messum, P. (2011). Modeling the development of pronunciation in infant speech acquisition. *Motor Control*, 15 (1), 85-117.

Howard, I. S. & Messum, P. (2014). Learning to pronounce first words in three languages: An investigation of caregiver and infant behavior using a computational model of an infant. *PloS one*, 9 (10), e110334.

Imamizu, H., Miyauchi, S., Tamada, T., Sasaki, Y., Takino, R., Pütz, B., ... & Kawato, M. (2000). Human cerebellar activity reflecting an acquired internal model of a new tool. *Nature*, 403 (6766), 192-195.

Ito, T., Tiede, M. & Ostry, D. J. (2009). Somatosensory function in speech perception. *Proceedings of the National Academy of Sciences*, 106 (4), 1245-1248.

Jansen-Osmann, P., Richter, S., Konczak, J. & Kalveram, K. T. (2002). Force adaptation transfers to untrained workspace regions in children. *Experimental Brain Research*, 143 (2), 212-220.

Jones, J. A., & Munhall, K. G. (2005). Remapping auditory-motor representations in voice production. *Current Biology*, 15 (19), 1768-1772.

Jordan, M. I. & Rumelhart, D. E. (1992). Forward models: Supervised learning with a distal teacher. *Cognitive science*, 16 (3), 307-354.

Junge, C., Cutler, A. & Hagoort, P. (2014). Successful word recognition by 10-month-olds given continuous speech both at initial exposure and test. *Infancy*, 19 (2), 179-193.

Jusczyk, P. W. (1977). Perception of syllable-final stop consonants by 2-month-old infants. *Perception & Psychophysics*, 21 (5), 450-454.

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

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

Jusczyk, P. W., Rosner, B. S., Cutting, J. E., Foard, C. F. & Smith, L. B. (1977). Categorical perception of nonspeech sounds by 2-month-old infants. *Perception & Psychophysics*, 21 (1), 50-54.

Kakita,Y., Fujimura, O. & Honda, K. (1985) Computation of mapping from muscular contraction patterns to formant patterns in vowel space. In *Phonetic Linguistics*, Orlando: Academic Press, 133 – 144.

Kamina, P. (2009). *Anatomie Clinique, Tome 2 : Tête, cou, dos*. 3e édition, Maloine, Paris.

Katz, W. F., Kripke, C. & Tallal, P. (1991). Anticipatory Coarticulation in the Speech of Adults and Young Children: Acoustic, Perceptual, and Video Data. *Journal of Speech, Language, and Hearing Research*, 34 (6), 1222-1232.

Kawato, M., Furukawa, K. & Suzuki, R. (1987). A hierarchical neural-network model for control and learning of voluntary movement. *Biological Cybernetics*, 57 (3), 169-185.

Kawato, M., Kuroda, T., Imamizu, H., Nakano, E., Miyauchi, S. & Yoshioka, T. (2003). Internal forward models in the cerebellum: fMRI study on grip force and load force coupling. *Progress in Brain Research*, 142, 171-188.

Kawato, M., Maeda, Y., Uno, Y. & Suzuki, R. (1990). Trajectory formation of arm movement by cascade neural network model based on minimum torque-change criterion. *Biological Cybernetics*, 62 (4), 275-288.

Keating, P. A. (1990). The window model of coarticulation: articulatory evidence. *Papers in laboratory phonology* I, 26, 451-470.

Keller, E. & Ostry D. J. (1983) Computerized measurement of tongue dorsum movements with pulsed-echo ultrasound. *Journal of the Acoustical Society of America*, 73, 1309-1315.

Kelso, J. S. & Tuller, B. (1983). Compensatory Articulation Under Conditions of Reduced Afferent Information: A Dynamic Formulation. *Journal of Speech, Language, and Hearing Research*, 26 (2), 217-224.

Kent, R. D. (1984). Psychobiology of speech development: Coemergence of language and a movement system. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, 246 (6), R888-R894.

Kent, R. D. & Forner, L. L. (1979). Developmental study of vowel formant frequencies in an imitation task. *The Journal of the Acoustical Society of America*, 65 (1), 208-217.

Kent, R. D. & Murray, A. D. (1982). Acoustic features of infant vocalic utterances at 3, 6, and 9 months. *The Journal of the Acoustical Society of America*, 72 (2), 353-365.

Kiritani, S., Miyawaki, K., Fujimura, O. & Miller, J. E. (1975). Computational model of the tongue. *The Journal of the Acoustical Society of America*, 57 (S1), S3-S3.

Konczak, J., Jansen-Osmann, P. & Kalveram, K. T. (2003). Development of force adaptation during childhood. *Journal of Motor Behavior*, 35 (1), 41-52.

Kooijman, V., Hagoort, P. & Cutler, A. (2005). Electrophysiological evidence for prelinguistic infants' word recognition in continuous speech. *Cognitive Brain Research*, 24 (1), 109-116.

Koopmans-van Beinum, F. J. & van der Stelt, J. M. (1986). Early stages in the development of speech movements. In *Precursors of early speech* (pp. 37-50). Palgrave Macmillan UK.

Krull D. (1987). Second formant locus patterns as a measure of consonant-vowel coarticulation. *Phonetic Experimental Research at the Institute of Linguistics, University of Stockholm, PERILUS*, 5, 43–61.

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.

Kuhl, P. K. (1983). Perception of auditory equivalence classes for speech in early infancy. *Infant Behavior and Development*, 6 (2), 263-285.

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature reviews neuroscience*, 5 (11), 831-843.

Kuhl, P. K. & Meltzoff, A. N. (1982). The Bimodal Perception of Speech in Infancy. *Science*, 218, 1138-1141.

Kuhl, P. K. & Meltzoff, A. N. (1996). Infant vocalizations in response to speech: Vocal imitation and developmental change. *The Journal of the Acoustical Society of America*, 100 (4), 2425-2438.

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

Kuhl, P. K., 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.

Lalevée, C. & Vilain, A. (2004). Développement du contrôle des cadres de la parole: une étude longitudinale du contrôle oral/nasal et de la coordination oro-laryngée. In *Atelier de conjoncture de la SFP: Acquisition du langage: vers une approche pluridisciplinaire*.

Lalevée, C. & Vilain, A. (2008). From proto-syllabic frames to first words: syllable: an audio-visual survey of 2 children from 6 to 16 months. In *Emergence of Language Abilities: from gestures to grammar*, 45-60.

Lambert, J. & Bard, C. (2005). Acquisition of visuomanual skills and improvement of information processing capacities in 6-to 10-year-old children performing a 2D pointing task. *Neuroscience Letters*, 377 (1), 1-6.

Lametti, D. R., Nasir, S. M. & Ostry, D. J. (2012). Sensory preference in speech production revealed by simultaneous alteration of auditory and somatosensory feedback. *The Journal of Neuroscience*, 32 (27), 9351-9358.

Lee, S., Potamianos, A. & Narayanan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *The Journal of the Acoustical Society of America*, 105 (3), 1455-1468.

Levelt, W. J. (1993). *Speaking: From intention to articulation* (Vol. 1). MIT press, 413.

Lewis, M. M. (1936). *Infant speech*. Kegan Paul, Trench, Trubner & Company.

Li, F. (2012). Language-Specific Developmental Differences in Speech Production: A Cross-Language Acoustic Study. *Child development*, 83 (4), 1303-1315.

Li, M., Kambhamettu, C. & Stone, M. (2005). Automatic contour tracking in ultrasound images. *Clinical Linguistics & Phonetics*, 19 (6-7), 545–554.

Liberman, A. M. & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21 (1), 1-36.

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

Liberman, A. M., Delattre, P. C. & Cooper, F. S. (1958). Some cues for the distinction between voiced and voiceless stops in initial position. *Language and speech*, 1 (3), 153-167.

Lieberman, D. E., McCarthy, R. C., Hiiemae, K. M. & Palmer, J. B. (2001). Ontogeny of postnatal hyoid and larynx descent in humans. *Archives of oral biology*, 46 (2), 117-128.

Lieberman, P. (2007). The evolution of human speech. *Current Anthropology*, 48 (1), 39-66.

Lieberman, P. & Crelin, E. S. (1971). On the speech of Neanderthal man. *Linguistic Inquiry*, 2 (2), 203-222.

Liljencrants, J. & Lindblom, B. (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language*, 839-862.

Locke, J. L. (1983). *Phonological acquisition and change*. Academic Press.

Löfqvist, A. (1999). Interarticulator phasing, locus equations, and degree of coarticulation. *The Journal of the Acoustical Society of America*, 106 (4), 2022-2030.

Lyakso, E. E. & Grigor'ev, A. S. (2015). Dynamics of the Duration and Frequency Characteristics of Vowels during the First Seven Years of Life in Children. *Neuroscience and Behavioral Physiology*, 45 (5), 558-567.

Ma, L. (2008). *La coarticulation en français et en chinois: étude expérimentale et modélisation*. Thèse de Doctorat. Université de Provence - Aix-Marseille I.

Ma, L., Perrier, P. & Dang, J. (2015). Strength of syllabic influences on articulation in Mandarin Chinese and French: Insights from a motor control approach. *Journal of Phonetics*, 53, 101-124.

MacDonald, E. N., Johnson, E. K., Forsythe, J., Plante, P. & Munhall, K. G. (2012). Children's development of self-regulation in speech production. *Current biology*, 22 (2), 113-117.

Macé, M., Montaldo, G., Cohen, I., Baulac, M., Fink, M. & Tanter, M. (2011). Functional ultrasound imaging of the brain. *Nature methods*, 8, 662-664.

MacNeilage, P. F. (1998). The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences*, 21 (04), 499-511.

MacNeilage, P. F. & Davis, B. (1990). *Acquisition of speech production: Frames, then content*.

MacNeilage, P. F. & Davis, B. L. (2000). On the origin of internal structure of word forms. *Science*, 288 (5465), 527-531.

Mampe, B., Friederici, A. D., Christophe, A. & Wermke, K. (2009). Newborns' cry melody is shaped by their native language. *Current biology*, 19 (23), 1994-1997.

Mandel, D. R., Jusczyk, P. W. & Pisoni, D. B. (1995). Infants' Recognition of the Sound Patterns of Their Own Names. *Psychological Science*, 314-317.

Manuel, S. Y. & Krakow, R. A. (1984). Universal and language particular aspects of vowel-to-vowel coarticulation. *Haskins Laboratories Status Report on Speech Research*, 77 (78), 69-78.

Masapollo, M., Polka, L. & Ménard, L. (2015). When infants talk, infants listen: pre-babbling infants prefer listening to speech with infant vocal properties. *Developmental Science*.

Matthies, M. L., Svirsky, M., Perkell, J. & Lane, H. (1996). Acoustic and articulatory measures of sibilant production with and without auditory feedback from a cochlear implant. *Journal of Speech, Language, and Hearing Research*, 39 (5), 936-946.

Matyear, C. L., MacNeilage, P. F. & Davis, B. L. (1998). Nasalization of vowels in nasal environments in babbling: Evidence for frame dominance. *Phonetica*, 55 (1-2), 1-17.

Max, L., Wallace, M. E. & Vincent, I. (2003). Sensorimotor adaptation to auditory perturbations during speech: Acoustic and kinematic experiments. In *Proceedings of the 15th International Congress of Phonetic Sciences*, 1053-1056.

McGowan, R. W., McGowan, R. S., Denny, M. & Nittrouer, S. (2014). A longitudinal study of very young children's vowel production. *Journal of Speech, Language, and Hearing Research*, 57 (1), 1-15.

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.

Meltzoff, A. N. & Borton, R. W. (1979). Intermodal matching by human neonates. *Nature*, 282, 403-404.

Meltzoff, A. N. & Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. *Science*, 198 (4312), 75-78.

Ménard, L. (2002). *Production et perception des voyelles au cours de la croissance du conduit vocal: variabilité, invariance et normalisation*. Thèse de Doctorat. Université Stendhal - Grenoble III.

Ménard, L. & Noiray, A. (2011). The development of lingual gestures in speech: Comparing synthesized vocal tracts with natural vowels. *Faits de langue*, 37, 189-202.

Ménard, L., Aubin, J., Thibeault, M. & Richard. G (2011). Measuring Tongue Shapes and Positions with Ultrasound Imaging: A Validation Experiment Using an Articulatory Model. *Folia Phoniatrica et Logopaedica*, 64 (2), 64-72.

Ménard, L., Perrier, P. & Aubin, J. (2016). Compensation for a lip-tube perturbation in 4-year-olds: Articulatory, acoustic, and perceptual data analyzed in comparison with adults. *The Journal of the Acoustical Society of America*, 139 (5), 2514-2531.

Ménard, L., Perrier, P., Aubin, J., Savariaux, C. & Thibeault, M. (2008). Compensation strategies for a lip-tube perturbation of French [u]: An acoustic and perceptual study of 4-year-old children. *The Journal of the Acoustical Society of America*, 124 (2), 1192-1206.

Ménard, L., Schwartz, J. L., Boë, L. J. & Aubin, J. (2007). Articulatory–acoustic relationships during vocal tract growth for French vowels: Analysis of real data and simulations with an articulatory model. *Journal of Phonetics*, 35 (1), 1-19.

Ménard, L., Schwartz, J. L., Boë, L. J., Kandel, S. & Vallée, N. (2002). Auditory normalization of French vowels synthesized by an articulatory model simulating growth from birth to adulthood. *The Journal of the Acoustical Society of America*, 111 (4), 1892-1905.

Ménard, L., Schwartz, J.-L. & Boë, L.-J. (2004). The role of vocal tract morphology in speech development: Perceptual targets and sensori-motor maps for French synthesized vowels from birth to adulthood. *Journal of Speech, Language, and Hearing Research*, 47 (5), 1059-1080.

Mielke, J. (2015). An ultrasound study of Canadian French rhotic vowels with polar smoothing spline comparisons. *The Journal of the Acoustical Society of America*, 137 (5), 2858-2869.

Miller, C. L. & Morse, P. A. (1976). The "heart" of categorical speech discrimination in young infants. *Journal of Speech, Language, and Hearing Research*, 19 (3), 578-589.

Mitsuya, T., Samson, F., Ménard, L. & Munhall, K. G. (2013). Language dependent vowel representation in speech production. *The Journal of the Acoustical Society of America*, 133 (5), 2993-3003.

Miyawaki, K. (1974). A study of the musculature of the human tongue. *Annual bulletin of the research institute of logopedics and phoniatrics*, 8, 23-50.

Miyawaki, K., Hirose, H., Ushijima, T. & Sawashima, M. (1975). A preliminary report on the electromyographic study of the activity of lingual muscles. *Ann. Bull. RILP*, 9 (91), 406.

Mol, H. (1970). *Fundamentals of phonetics* (Vol. 2). Mouton, La Haye.

Mol, H. (1970). *Fundamentals of Phonetics II : Acoustical modes generating the formants of the vowel phomenes*, Mouton, La Haye.

Monnin, J. (2010). *Influence de la langue ambiante sur l'acquisition phonologique: une comparaison du français et du drehu*. Thèse de Doctorat, Université de Grenoble ; Université de Nouvelle Calédonie.

Munhall, K. G. & Jones, J. A. (1998). Articulatory evidence for syllabic structure. *Behavioral and Brain Sciences*, 21 (04), 524-525.

Murphy, K., Corfield, D. R., Guz, A., Fink, G. R., Wise, R. J. S., Harrison, J. & Adams, L. (1997). Cerebral areas associated with motor control of speech in humans. *Journal of Applied Physiology*, 83 (5), 1438-1447.

Naidoo, Y., Van der Merwe, A., Groenewald, E. & Naudé, E. (2005). Development of speech sounds and syllable structure of words in Zulu-speaking children. *Southern African Linguistics and Applied Language Studies*, 23(1), 59-79.

Nasir, S. M. & Ostry, D. J. (2006). Somatosensory precision in speech production. *Current Biology*, 16 (19), 1918-1923.

Nasir, S. M. & Ostry, D. J. (2008). Speech motor learning in profoundly deaf adults. *Nature neuroscience*, 11 (10), 1217-1222.

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.

Nijland, L., Maassen, B., Meulen, S. V. D., Gabreëls, F., Kraaimaat, F. W. & Schreuder, R. (2002). Coarticulation patterns in children with developmental apraxia of speech. *Clinical Linguistics & Phonetics*, 16 (6), 461-483.

Nittrouer, S. (1993). The Emergence of Mature Gestural Patterns Is Not Uniform: Evidence From an Acoustic Study. *Journal of Speech, Language, and Hearing Research*, 36 (5), 959-972.

Nittrouer, S., Studdert-Kennedy, M. & McGowan, R. S. (1989). The Emergence of Phonetic Segments: Evidence from the Spectral Structure of Fricative-Vowel Syllables Spoken by Children and Adults. *Journal of Speech, Language, and Hearing Research*, 32 (1), 120-132.

Nittrouer, S., Studdert-Kennedy, M. & Neely, S. T. (1996). How Children Learn to Organize Their Speech Gestures: Further Evidence From Fricative-Vowel Syllables. *Journal of Speech, Language, and Hearing Research*, 39 (2), 379-389.

Noiray, A., Ménard, L. & Iskarous, K. (2013). The development of motor synergies in children: Ultrasound and acoustic measurements. *The Journal of the Acoustical Society of America*, 133 (1), 444-452.

Öhman, S. E. (1966). Coarticulation in VCV utterances: Spectrographic measurements. *The Journal of the Acoustical Society of America*, 39 (1), 151-168.

Öhman, S. E. (1967). Numerical model of coarticulation. *The Journal of the Acoustical Society of America*, 41 (2), 310-320.

Oller, D. K. & MacNeilage, P. F. (1983). Development of speech production: Perspectives from natural and perturbed speech. In *The production of speech*, Springer New York, 91-108.

Ostry, D. J., Darainy, M., Mattar, A. A., Wong, J. & Gribble, P. L. (2010). Somatosensory plasticity and motor learning. *The Journal of Neuroscience*, 30 (15), 5384-5393.

Pascalis, O., de Haan, M. & Nelson, C. A. (2002). Is face processing species-specific during the first year of life? *Science*, 296 (5571), 1321-1323.

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.

Payan, Y. & Perrier, P. (1997). Synthesis of VV sequences with a 2D biomechanical tongue model controlled by the Equilibrium Point Hypothesis. *Speech Communication*, 22 (2), 185-205.

Pentz Jr, A. L. & Gilbert, H. R. (1983). Relation of selected acoustical parameters and perceptual ratings to voice quality of Down syndrome children. *American journal of mental deficiency*, 88 (2), 203-210.

Perkell, J. S. (1969). *Physiology of speech production*. Cambridge, MA, MIT Press.

Perkell, J. S. (1974). *A physiologically-oriented model of tongue activity in speech production*. Doctoral dissertation, Massachusetts Institute of Technology.

Perkell, J. S. (2012). Movement goals and feedback and feedforward control mechanisms in speech production. *Journal of Neurolinguistics*, 25 (5), 382-407.

Perkell, J. S., Cohen, M. H., Svirsky, M. A., Matthies, M. L., Garabieta, I. & Jackson M.T.T. (1992). Electromagnetic midsagittal articulometer system for transducing speech articulatory movements. *The Journal of the Acoustical Society of America*, 92, 3078-3096.

Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Perrier, P., Vick, J., ... & Zandipour, M. (2000). A theory of speech motor control and supporting data from speakers with normal hearing and with profound hearing loss. *Journal of Phonetics*, 28 (3), 233-272.

Perkell, J. S., Matthies, M. L., Tiede, M., Lane, H., Zandipour, M., Marrone, N., ... & Guenther, F. H. (2004). The Distinctness of Speakers' /s/-/ʃ/ contrast is related to their auditory discrimination and use of an articulatory saturation effect. *Journal of Speech, Language, and Hearing Research*, 47(6), 1259-1269.

Perkell, J. S., Zandipour, M., Matthies, M. L. & Lane, H. (2002). Economy of effort in different speaking conditions. I. A preliminary study of intersubject differences and modeling issues. *The Journal of the Acoustical Society of America*, 112 (4), 1627-1641.

Perkell, J., Matthies, M., Lane, H., Guenther, F., Wilhelms-Tricarico, R., Wozniak, J. & Guiod, P. (1997). Speech motor control: Acoustic goals, saturation effects, auditory feedback and internal models. *Speech Communication*, 22 (2), 227-250.

Perrier, P. (2012). Gesture planning integrating knowledge of the motor plant's dynamics: A literature review from motor control and speech motor control. In *Speech planning and dynamics*, 191-238.

Perrier, P. (2014). "GEPPETO": A target-based model of speech production including optimal planning and physical modeling. In *Adventures in Speech Science*.

Perrier, P., Lœvenbruck, H. & Payan, Y. (1996). Control of tongue movements in speech: The equilibrium point hypothesis perspective. *Journal of Phonetics*, 24 (1), 53-75.

Perrier, P., Ostry, D. J. & Laboissière, R. (1996). The equilibrium point hypothesis and its application to speech motor control. *Journal of Speech, Language, and Hearing Research*, 39 (2), 365-378.

Perry, T. L., Ohde, R. N. & Ashmead, D. H. (2001). The acoustic bases for gender identification from children's voices. *The Journal of the Acoustical Society of America*, 109 (6), 2988-2998.

Peterson, G. E. & Barney, H. L. (1952). Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America*, 24, 585–594.

Petitto, L. A., & Marentette, P. F. (1991). Babbling in the manual mode: Evidence for the ontogeny of language. *Science*, 251 (5000), 1493-1496.

Petitto, L. A., Holowka, S., Sergio, L. E., & Ostry, D. (2001). Language rhythms in baby hand movements. *Nature*, 413 (6851), 35-36.

Pettinato, M., Tuomainen, O., Granlund, S., & Hazan, V. (2016). Vowel space area in later childhood and adolescence: Effects of age, sex and ease of communication. *Journal of Phonetics*, 54, 1-14.

Pierrehumbert, J. B. (2001). *Exemplar dynamics: Word frequency, lenition and contrast. Frequency and the Emergence of Linguistic Structure*, 45, 137.

Pike, K.L. (1946). *The Intonation of American English*. University of Michigan Press, Ann Arbor.

Pinker, S. (1995). *The language instinct: The new science of language and mind* (Vol. 7529). Penguin, UK.

Polit, A., & Bizzi, E. (1978). Processes controlling arm movements in monkeys. *Science*, 201 (4362), 1235-1237.

Purcell, D. W. & Munhall, K. G. (2006). Adaptive control of vowel formant frequency: Evidence from real-time formant manipulation. *The Journal of the Acoustical Society of America*, 120 (2), 966-977.

R Foundation for Statistical Computing. (2013). R: A Language and Environment for Statistical Computing. Vienna, Austria.

Recasens, D. (1984). Vowel-to-vowel coarticulation in Catalan VCV sequences. *The Journal of the Acoustical Society of America*, 76 (6), 1624-1635.

Recasens, D. (1989). Long range coarticulation effects for tongue dorsum contact in VCVCV sequences. *Speech Communication*, 8 (4), 293-307.

Recasens, D., Pallarès, M. D. & Fontdevila, J. (1997). A model of lingual coarticulation based on articulatory constraints. *The Journal of the Acoustical Society of America*, 102 (1), 544-561.

Repp, B. H. (1986). Some observations on the development of anticipatory coarticulation. *The Journal of the Acoustical Society of America*, 79 (5), 1616-1619.

Riecker, A., Ackermann, H., Wildgruber, D., Dogil, G. & Grodd, W. (2000). Opposite hemispheric lateralization effects during speaking and singing at motor cortex, insula and cerebellum. *Neuroreport*, 11 (9), 1997-2000.

Robb, M. P., Chen, Y. & Gilbert, H. R. (1997). Developmental aspects of formant frequency and bandwidth in infants and toddlers. *Folia Phoniatrica et Logopaedica*, 49 (2), 88-95.

Rochet-Capellan, A. & Schwartz, J. L. (2007). An articulatory basis for the labial-to-coronal effect: /pata/ seems a more stable articulatory pattern than /tapa/. *The Journal of the Acoustical Society of America*, 121 (6), 3740-3754.

Romeo, R., Hazan, V. & Pettinato, M. (2013). Developmental and gender-related trends of intra-talker variability in consonant production. *The Journal of the Acoustical Society of America*, 134 (5), 3781-3792.

Rossato, S., Badin, P. & Bouaouni, F. (2003). Velar movements in French: an articulatory and acoustical analysis of coarticulation. In *Proceedings of the 15th International Congress of Phonetic Sciences*. Barcelona, Spain, 3141-3144.

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

Savariaux, C., Perrier, P. & Orliaguet, J. P. (1995). Compensation strategies for the perturbation of the rounded vowel [u] using a lip tube: A study of the control space in speech production. *The Journal of the Acoustical Society of America*, 98 (5), 2428-2442.

Schwartz, J. L. (2009). Eléments d'une morphogenèse perceptuo-motrice des unités du Langage. In *Séminaire "Perception Sémiotique et Socialité du Sens"*, Collège de France, Paris.

Schwartz, J. L., Boë, L. J., Badin, P. & Sawallis, T. R. (2012). Grounding stop place systems in the perceptuo-motor substance of speech: On the universality of the labial–coronal–velar stop series. *Journal of Phonetics*, 40 (1), 20-36.

Schwartz, J. L., Boë, L. J., Vallée, N. & Abry, C. (1997). The dispersion-focalization theory of vowel systems. *Journal of Phonetics*, 25 (3), 255-286.

Sereno, J. A., & Lieberman, P. (1987). Developmental aspects of lingual coarticulation. *Journal of Phonetics*.

Sereno, J. A., Baum, S. R., Marean, G. C. & Lieberman, P. (1987). Acoustic analyses and perceptual data on anticipatory labial coarticulation in adults and children. *The Journal of the Acoustical Society of America*, 81 (2), 512-519.

Sharkey, S. G. & Folkins, J. W. (1985). Variability of Lip and Jaw Movements in Children and AdultsImplications for the Development of Speech Motor Control. *Journal of Speech, Language, and Hearing Research*, 28 (1), 8-15.

Shiller, D. M., Gracco, V. L. & Rvachew, S. (2010). Auditory-motor learning during speech production in 9-11-year-old children. *PLoS One*, 5 (9), e12975.

Shiller, D. M., Sato, M., Gracco, V. L. & Baum, S. R. (2009). Perceptual recalibration of speech sounds following speech motor learning. *The Journal of the Acoustical Society of America*, 125 (2), 1103-1113.

Siren, K. A. & Wilcox, K. A. (1995). Effects of lexical meaning and practiced productions on coarticulation in children's and adults' speech. *Journal of Speech, Language, and Hearing Research*, 38 (2), 351-359.

Smith, A. (2006). Speech motor development: Integrating muscles, movements, and linguistic units. *Journal of communication disorders*, 39 (5), 331-349.

Smith, A. (2010). *Development of neural control of orofacial movements for speech*. The Handbook of Phonetic Sciences, Second Edition, 251-296.

Smith, A. & Goffman, L. (1998). Stability and patterning of speech movement sequences in children and adults. *Journal of Speech, Language, and Hearing Research*, 41 (1), 18-30.

Smith, A. & Zelaznik, H. N. (2004). Development of functional synergies for speech motor coordination in childhood and adolescence. *Developmental Psychobiology*, 45 (1), 22-33.

Smith, B. L. & McLean-Muse, A. (1987). An investigation of motor equivalence in the speech of children and adults. *The Journal of the Acoustical Society of America*, 82 (3), 837-842.

Sonies, B. C., Shawker, T. H., Hall, T. E., Gerber, L. H. & Leighton, S. B. (1981). Ultrasonic visualization of tongue motion during speech. *The Journal of the Acoustical Society of America*, 70, 693–696.

Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics*, 17, 3-45.

Stevens, K. N. & Klatt, D. H. (1974). Role of formant transitions in the voiced-voiceless distinction for stops. *The Journal of the Acoustical Society of America*, 55 (3), 653-659.

Stone, M. (2005). A guide to analysing tongue motion from ultrasound images. *Clinical Linguistics & Phonetics*, 19:6, 455-501.

Stone, M., & Davis, E. (1995). A head and transducer support (HATS) system for use in ultrasound imaging of the tongue during speech. *The Journal of the Acoustical Society of America*, 98, 3107–3112.

Sussman, H. M., Duder, C., Dalston, E. & Cacciatore, A. (1999). An Acoustic Analysis of the Development of CV Coarticulation: A Case Study. *Journal of Speech, Language, and Hearing Research*, 42 (5), 1080-1096.

Sussman, H. M., Hoemeke, K. A. & McCaffrey, H. A. (1992). Locus equations as an index of coarticulation for place of articulation distinctions in children. *Journal of Speech, Language, and Hearing Research*, 35, 769-781.

Sussman, H. M., Minifie, F. D., Buder, E. H., Stoel-Gammon, C. & Smith, J. (1996). Consonant-Vowel Interdependencies in Babbling and Early Words: Preliminary Examination of a Locus Equation Approach. *Journal of Speech, Language, and Hearing Research*, 39(2), 424-433.

Tabain, M. (2000). Coarticulation in CV syllables: a comparison of locus equation and EPG data. *Journal of Phonetics*, 28 (2), 137-159.

Taine, H. (1870). *De l'intelligence*, 2 vols. Paris: Hachette, 408.

Takahashi, C. D., Nemet, D., Rose-Gottron, C. M., Larson, J. K., Cooper, D. M. & Reinkensmeyer, D. J. (2003). Neuromotor noise limits motor performance, but not motor adaptation, in children. *Journal of Neurophysiology*, 90 (2), 703-711.

Takemoto, H. (2001). Morphological analyses of the human tongue musculature for three-dimensional modeling. *Journal of Speech, Language, and Hearing Research*, 44 (1), 95-107.

Tang, L., Bressmann, T. & Hamarneh, G. (2012). Tongue contour tracking in dynamic ultrasound via higher-order MRFs and efficient fusion moves. *Medical Image Analysis*, 16 (8), 1503– 1520.

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

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

Trehub, S. E. & Rabinovitch, M. S. (1972). Auditory-linguistic sensitivity in early infancy. *Developmental Psychology*, 6 (1), 74.

Tremblay, A., University, D. & Ransijn, J. (2013). LMERConvenienceFunctions: A suite of functions to back-fit fixed effects & forward-fit random effects, as well as other miscellaneous functions, R package version 2.5.

Tremblay, S., Shiller, D. M. & Ostry, D. J. (2003). Somatosensory basis of speech production. *Nature*, 423 (6942), 866-869.

Turnbaugh, K. R., Hoffman, P. R., Daniloff, R. G. & Absher, R. (1985). Stop-vowel coarticulation in 3-year-old, 5-year-old, and adult speakers. *The Journal of the Acoustical Society of America*, 77 (3), 1256-1257.

Uno, Y., Kawato, M. & Suzuki, R. (1989). Formation and control of optimal trajectory in human multijoint arm movement. *Biological Cybernetics*, 61 (2), 89-101.

Vasudevan, E. V., Torres-Oviedo, G., Morton, S. M., Yang, J. F. & Bastian, A. J. (2011). Younger is not always better: development of locomotor adaptation from childhood to adulthood. *The Journal of Neuroscience*, 31 (8), 3055-3065.

Velleman, S. L. & Vihman, M. M. (2007). Phonology in infancy and early childhood: implications for theories of language learning. In *Phonology in context*. Palgrave Macmillan UK, 25-50.

Vihman, M. M. (1981). Phonology and the development of the lexicon: Evidence from children's errors. *Journal of Child Language*, 8 (02), 239-264.

Vihman, M. M. (1993). Variable paths to early word production. *Journal of Phonetics*, 21 (1), 61-82.

Vihman, M. M. (1996). *Phonological development: The origins of language in the child*. Blackwell Publishing.

Vihman, M. M. (1998). *Later phonological development. Articulation and phonological disorders*, 113-147.

Vihman, M. M., Ferguson, C. A. & Elbert, M. (1986). Phonological development from babbling to speech: Common tendencies and individual differences. *Applied Psycholinguistics*, 7 (01), 3-40.

Villacorta, V. M., Perkell, J. S. & Guenther, F. H. (2007). Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception. *The Journal of the Acoustical Society of America*, 122 (4), 2306-2319.

Vorperian, H. K. & Kent, R. D. (2007). Vowel Acoustic Space Development in Children: A Synthesis of Acoustic and Anatomic Data. *Journal of Speech, Language, and Hearing Research*, 50 (6), 1510 –1545.

Vorperian, H. K., Kent, R. D., Lindstrom, M. J., Kalina, C. M., Gentry, L. R. & Yandell, B. S. (2005). Development of vocal tract length during early childhood: A magnetic resonance imaging study. *The Journal of the Acoustical Society of America*, 117 (1), 338-350.

Vorperian, H. K., Wang, S., Chung, M. K., Schimek, E. M., Durtschi, R. B., Kent, R. D., ... & Gentry, L. R. (2009). Anatomic development of the oral and pharyngeal portions of the vocal tract: An imaging study. *The Journal of the Acoustical Society of America*, 125 (3), 1666-1678.

Walley, A. C. & Flege, J. E. (1999). Effect of lexical status on children's and adults' perception of native and non-native vowels. *Journal of Phonetics*, 27 (3), 307-332.

Walsh, B. & Smith, A. (2002). Articulatory Movements in Adolescents: Evidence for Protracted Development of Speech Motor Control Processes. *Journal of Speech, Language, and Hearing Research*, 45 (6), 1119-1133.

Walsh, B., Smith, A. & Weber-Fox, C. (2006). Short-term plasticity in children's speech motor systems. *Developmental Psychobiology*, 48 (8), 660-674.

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.

Whalen, D. H. (1990). Coarticulation is largely planned. *Journal of Phonetics*, 18, 3-35.

Whalen, D. H. & Liberman, A. M. (1987). Speech perception takes precedence over nonspeech perception. *Science*, 237 (4811), 169-171.

Whalen, D. H., Giulivi, S., Goldstein, L. M., Nam, H. & Levitt, A. G. (2011). Response to MacNeilage and Davis and to Oller. *Language Learning and Development*, 7 (3), 243-249.

Whalen, D. H., Giulivi, S., Nam, H., Levitt, A. G., Hallé, P. & Goldstein, L. M. (2012). Biomechanically Preferred Consonant-Vowel Combinations Fail to Appear in Adult Spoken Corpora. *Language and Speech*, 55 (4), 503-515.

Whalen, D. H., Iskarous, K., Tiede, M. K., Ostry, D. J., Lehnert-Lehouillier, H., Vatikiotis-Bateson, E. & Hailey, D. (2005) The Haskins optically corrected ultrasound system (HOCUS). *Journal of Speech, Language, and Hearing Research*, 48(3), 543-553.

Whiteside, S. P. & Hodgson, C. (2000). Some acoustic characteristics in the voices of 6-to 10-year-old children and adults: A comparative sex and developmental perspective. *Logopedics Phoniatrics Vocology*, 25 (3), 122-132.

Wilhelms-Tricarico, R. (1996). A biomechanical and physiologically–based vocal tract model and its control. *Journal of Phonetics*, 24 (1), 23-38.

Wilhelms-Tricarico, R. (2000). Development of a tongue and mouth floor model for normalization and biomechanical modeling. In *Proceedings of the 5th speech production seminar and CREST Workshop on models of speech production*, 141-148.

Wilhelms-Tricarico, R. (1995). Physiological modeling of speech production: Methods for modeling soft-tissue articulators. *The Journal of the Acoustical Society of America*, 97 (5), 3085-3098.

Winkler, R., Ma, L. & Perrier, P. (2010). A model of optimal speech production planning integrating dynamical constraints to achieve appropriate articulatory timing. In *Cognitive*

and Physical Models of Speech Production, Speech Perception and Production-Perception Interaction - Part III Planning and Dynamics, 44-48.

Wolpert, D. M. & Kawato, M. (1998). Multiple paired forward and inverse models for motor control. *Neural networks*, 11 (7), 1317-1329.

Wolpert, D. M., Miall, R. C. & Kawato, M. (1998). Internal models in the cerebellum. *Trends in Cognitive Sciences*, 2 (9), 338-347.

Xu, Y. (2010). In defense of lab speech. *Journal of Phonetics*, 38 (3), 329-336.

Zahorian, S. A. & Jagharghi, A. J. (1993). Spectral-shape features versus formants as acoustic correlates for vowels. *The Journal of the Acoustical Society of America*, 94 (4), 1966-1982.

Zharkova, N., Hewlett, N. & Hardcastle, W. J. (2011). Coarticulation as an indicator of speech motor control development in children: An ultrasound study. *Motor Control*, 15 (1), 118-40.

Zharkova, N., Hewlett, N. & Hardcastle, W. J. (2012). An ultrasound study of lingual coarticulation in /sV/ syllables produced by adults and typically developing children. *Journal of the International Phonetic Association*, 42 (2), 193-208.

Zharkova, N., Hewlett, N., Hardcastle, W. J. & Lickley, R. J. (2014). Spatial and temporal lingual coarticulation and motor control in preadolescents. *Journal of Speech, Language, and Hearing Research*, 57 (2), 374-388.

Zmarich, C. & Bonifacio, S. (2005). Phonetic inventories in italian children aged 18-27 months: A longitudinal study. In *Ninth European Conference on Speech Communication and Technology*, 757-760.

Résumé

Ce travail de thèse étudie la production de la parole à l'âge de 4 ans, en comparaison avec l'adulte, sous l'angle du contrôle moteur. Un intérêt particulier a été porté à deux indices : la variabilité liée à la répétition de voyelles isolées et la coarticulation anticipatoire intra et extra-syllabique dans le cadre de la production de séquences de type V1-C-V2. Des données acoustiques et articulatoires ont été enregistrées via l'échographie linguale et le dispositif HOCUS. Les données acoustiques ont été analysées pour 20 enfants et 10 adultes et les données articulatoires ont été analysées pour 6 de ces enfants et 2 de ces adultes. En accord avec de nombreuses études de la littérature, la variabilité liée à la répétition est plus importante chez les enfants que chez les adultes. Une forte anticipation de V2 dans la réalisation de V1 a été observée pour tous les adultes. Chez les enfants, cette anticipation n'est pas systématique, et lorsqu'elle est observée, elle est d'amplitude plus faible que chez les adultes. Ainsi, seulement 5 enfants parmi les 20 étudiés montrent des signes d'anticipation, majoritairement dans la dimension antéropostérieure, manifestée dans l'espace acoustique par la dimension F2. La coarticulation anticipatoire intra-syllabique semble également être d'amplitude plus faible chez les enfants. Enfin, les gestes de parole des enfants sont plus lents que ceux des adultes. Dans leur ensemble, ces résultats sont interprétés comme la preuve de l'immaturité du système moteur pour la production de la parole chez l'enfant, selon deux perspectives : des patrons moteurs insuffisamment stables pour la production de gestes vocaliques et un manque d'efficacité dans l'anticipation les gestes futurs lors de la production d'une séquence de sons de parole. Selon une perspective empruntée aux théories du contrôle moteur optimal, nous proposons que la coarticulation anticipatoire repose sur l'usage de modèles internes, représentations sensori-motrices du système moteur de la parole dans le système nerveux central, et que son amplitude reflète la maturation progressive de ces représentations sensori-motrices à mesure que la production de la parole se développe.

Mots-clés : Développement, production de parole, contrôle moteur, coarticulation anticipatoire, planification, échographie linguale.

Abstract

This thesis work investigates speech production in 4-year-old children, in comparison with adults, from a speech motor control perspective. It focuses on two indices: measures of token-to-token variability in the production of isolated vowels and on anticipatory intra and extra-syllabic coarticulation within V1-C-V2 sequences. Acoustic and articulatory data were recorded thanks to ultrasound tongue imaging within the HOCUS system. Acoustic data from 20 children and 10 adults have been analyzed. Ultrasound data have been analyzed from a subset of these participants: 6 children and 2 adults. In agreement with former studies, token-to-token variability was greater in children than in adults. Strong anticipation of V2 in V1 realization was found in all adults. In children, anticipation was not systematic, and when observed, it was of smaller amplitude than in adults. In more details, only 5 children among the 20 studied showed a small amount of anticipation, mainly along the antero-posterior dimension, manifested in the acoustic F2 dimension. Anticipatory intra-syllabic coarticulation also seems to be of smaller amplitude in children than in adults. Last, children's speech gestures are slower than those of adults. These results are interpreted as evidence for the immaturity of children's speech motor control from two perspectives: insufficiently stable motor control patterns for vowel production, and a lack of effectiveness in anticipating forthcoming gestures. In line with theories of optimal motor control, we assume that anticipatory coarticulation is based on the use of internal models, i.e. sensori-motor representations of the speech production apparatus in the central nervous system, and that the amplitude of anticipatory coarticulation reflects the increasing maturation of these sensori-motor representations as speech develops.

Keywords: Development, speech production, speech motor control, anticipatory coarticulation, speech planning, ultrasound tongue imaging.