

INSTITUT NATIONAL POLYTECHNIQUE DE GRENOBLE

N° attribué par la bibliothèque
□□□□□□□□□□

T H E S E

pour obtenir le grade de

DOCTEUR DE L'INPG, Spécialité : « Sciences Cognitives »

préparée au laboratoire
Institut de la Communication Parlée, UMR CNRS 5009

dans le cadre de l'Ecole Doctorale
« Ingénierie pour la Santé, la Cognition et l'Environnement »

présentée et soutenue publiquement par

Jihène Emena Serkhane

le 25 novembre 2005

Un bébé androïde vocalisant: Etude et modélisation des
mécanismes d'exploration vocale et d'imitation
orofaciale dans le développement de la parole

Directeurs de thèse : Jean-Luc Schwartz, Pierre Bessière,
Louis-Jean Boë

JURY

J. Hérault
J.-F. P. Bonnot
Y. Laprie
B. Davis
J.-L. Schwartz
P. Bessière
L.J. Boë

Président
Rapporteur
Rapporteur
Examinatrice
Directeur de thèse
Co-directeur de thèse
Co-directeur de thèse

A photograph of a young child with dark skin and hair, sitting on a light-colored floor. The child is smiling and looking towards the camera. They are wearing a white diaper. The background is plain white.

Un bébé androïde vocalisant:
Etude et modélisation des
mécanismes d'exploration vocale et
d'imitation orofaciale dans le
développement de la parole

A celle qui s'est battue contre la prison misogyne de l'ignorance,

Chérifa Mohdhi,

[ʃɛ:fa mɔhði:]

à Jean-Luc et Marie, ...

...et à celles qui n'existent qu'à travers le regard des autres

Introduction générale: D'un moule à connaissances

L'étymologie (*infans*) des termes "enfant", en Français, et "infant", en Anglais, signifie "qui ne parle pas". Contrairement à "enfant", la motivation sémantique de "infant" reste d'actualité¹, en désignant, pour l'espèce humaine, l'individu de la naissance à deux ans, soit le nourrisson.

La question de savoir comment des êtres humains apprennent à parler est une énigme dont l'histoire ne date pas d'aujourd'hui. La première étude rapportée en la matière est probablement celle d'un roi de l'Égypte du septième siècle avant Jésus Christ, Psammetichus, qui ordonna d'élever deux nouveau-nés² sans exposition à la parole, pour savoir dans quelle langue serait leur premier mot (Barko Gleason, 2001 ; Cyrulnik, 1989)³. Depuis, ce genre de procédé est, fort heureusement, proscrit.

Traditionnellement, les recherches sur la parole et son ontogenèse⁴ utilisent l'observation et l'expérimentation, mais la fin du XX^{ième} siècle a vu l'émergence d'un troisième type de méthode d'investigation, celle de la modélisation à base de calcul. La modélisation, qu'elle soit conceptuelle ou computationnelle, se prête bien aux travaux relatifs aux *phénomènes biologiques complexes*, dont la compréhension nécessite l'intervention de plusieurs disciplines (Legay, 1997). Le *développement de la production de la parole* est un phénomène biologique complexe, et son étude peut faire appel à des domaines aussi divers que la phonétique, la psycho-acoustique, la psychologie du développement et les sciences du traitement de l'information. Les vertus de la modélisation résident en particulier dans sa capacité à structurer de manière synthétique et cohérente les connaissances et hypothèses que différentes disciplines ont sur l'objet. C'est un *outil de recherche* à part entière, en ce qu'elle est un moyen de comprendre comment fonctionne le système étudié et de construire de nouvelles connaissances à son sujet. Dans le cas de la modélisation à base de calcul, il est possible de simuler, de manière *quantitative*, le phénomène à l'étude et de confronter les résultats obtenus aux mesures prélevées sur le réel: l'écart des prédictions du modèle aux données permet de construire de nouvelles questions et hypothèses qui viendront à leur tour alimenter la forme du moule⁵ à connaissance.

¹ Le changement de sens du terme francophone est lié au fait que la notion d'enfance au sens où nous la connaissons est récente dans l'histoire occidentale (voir Firestone, 1970/2002).

² **Nouveau-né:** Pour l'espèce humaine, ce terme désigne l'individu de la naissance à 1 mois, et l'on parle de **période néonatale**.

³ Nous avons plus de détails sur l'expérience conçue par Frédéric II, au XII^{ième} siècle. Il "avait acheté quelques nouveau-nés pour les confier à des gardiens silencieux. " Il voulut connaître par l'expérience quelle sorte de langue et d'idiome avaient les enfants, quand ils grandissaient sans parler à personne. Et il commanda aux servantes et aux nourrices de donner du lait aux bébés, de les faire sucer à la mamelle, de les baigner et de les nettoyer mais de ne les cajoler d'aucune manière ni de leur parler ; car il voulait savoir s'ils parleraient l'hébreu, première langue qui fût, ou le grec ou le latin ou l'arabe, ou bien la première langue de leurs parents, de ceux dont ils étaient nés. Mais il s'efforçait en vain car les enfants ou les bébés mouraient tous" [Duby, 1979]" (Cyrulnik, 1989, p. 197).

⁴ **Ontogenèse:** Au sens large, l'ontogenèse désigne l'"ensemble des processus de développement individuel et les étapes depuis la fécondation de l'œuf jusqu'à la fin de la vie adulte" (Vauclair, p. 15). Dans ce mémoire, elle réfère au développement individuel jusqu'aux alentours de deux ans.

⁵ En vieux français, *modelle* signifie "moule" et comprend l'idée de mesure (Legay, 1997, p. 22).

Les travaux de thèse présentés dans ce manuscrit ont porté sur l'élaboration d'une modélisation computationnelle *réaliste* des premiers pas de la parole, dont les résultats de simulation ont été confrontés à des données quantitatives relevées auprès de *vrais* nourrissons, à l'âge de 4 et 7 mois. Cette démarche est en un certain sens originale.

Nos prédécesseurs ont proposé des modèles d'acquisition de la production de la parole, dont la conception et l'implémentation peuvent être qualifiées de techniquement très "puissantes", mais aucun modèle n'a été, jusque là, comparé à des *mesures* sorties de la bouche d'individus *préverbaux* (e.g. Laboissière, Schwartz et Bailly, 1991; Markey, 1994; Guenther, 1995; Bailly, 1997). De plus, l'intégration de connaissances, hypothèses et théories contemporaines sur le développement de la production de la parole, rapportées en phonétique et en psycho-phonétique, consiste en une référence assez *abstraite* à des phénomènes vraisemblablement en jeu, tels que le babillage et l'imitation vocale.

L'implémentation du babillage prend, dans ces travaux, la forme d'une exploration *exhaustive* et *aléatoire* de représentations sensori-motrices, qui permet l'apprentissage d'un "modèle direct" complet dictant *toutes* les associations possibles entre les domaines moteur et perceptif, *sans faire cas des répertoires réels de productions infantiles*. Il vient que l'hypothèse implicite sous-jacente correspond à un postulat du linguiste Jakobson (1968) qui stipule que l'inventaire des productions de babillage est *infini* et ne montre *pas de préférence*; mais, cette idée ne résiste pas à l'épreuve des faits.

L'imitation *vocale*, quant à elle, vient après l'établissement du modèle direct et consiste en la spécification et éventuellement l'*apprentissage* d'un "modèle inverse" qui sélectionne une réponse parmi *toutes* les commandes articulatoires possibles pour produire *un* son donné (la cible). Outre les faits que ce type d'implémentation demande une quantité considérable de calculs et que la résolution du problème que pose l'inversion fasse appel à diverses heuristiques pour sélectionner la ("bonne") réponse imitative, les choix de modélisation (i) ne sont jamais explicitement motivés par des connaissances effectives sur les phénomènes d'imitation orofaciale précoce et (ii) ne mènent pas à la comparaison des résultats de simulations aux performances de sujets réels.

Les fondements d'un modèle-outil à base de calcul résident, d'après nous, (i) dans le choix d'une implémentation qui se prête à la comparaison entre les résultats de simulation et des *données quantitatives* issues de l'étude directe du phénomène, et (ii) dans une intégration la plus effective possible de connaissances et hypothèses produites dans des domaines de *terrain*, tels que la phonétique et la psycho-phonétique du développement. Le second point implique la prise en compte d'un problème caractéristique de la transdisciplinarité: *des termes en apparence identiques cachent souvent des significations distinctes*.

Pour répondre à ces exigences, nous avons conçu un *agent sensori-moteur vocalisant*, en posant dans le même moule les ingrédients que nous avons jugés nécessaires à une modélisation computationnelle réaliste du développement précoce de la production de la parole et qui permettent de comparer les résultats de simulation à des données développementales rapportées dans la littérature. Notre composition comprend (i) des hypothèses et connaissances *explicites* sur l'ontogenèse de la parole, *définies* et *décrites* dans des termes appropriés à l'étude du développement de la parole (ii) un modèle statistique de production de la parole qui intègre la croissance non-linéaire et non-uniforme du conduit vocal, (iii) des modèles de perception de la parole, (iv) un modèle de mécanisme d'apprentissage de liens perceptuo-moteurs et d'imitation de cibles perceptives, (v) des données quantitatives de phonétique et de psycho-phonétique du développement, publiées par d'autres auteurs.

Le premier élément est un puzzle inachevé de l'ontogenèse précoce de la production verbale, dont la base est un alliage entre des apports de la théorie Frame/Content de MacNeilage et Davis (voir MacNeilage, 1998) et des réflexions menées par des membres de

l'ICP dans ce contexte (voir Schwartz, Abry, Boë et Cathiard, 2002). Deux pièces du jeu constituent, d'après nous, des clés de voûte du chemin qui mène de la non-parole à la parole: l'*exploration* (vocale) et l'*imitation* (orofaciale). Attention, ces deux notions sont, au moins au niveau conceptuel, à distinguer de leur acception dans les travaux de modélisation classiques. L'exploration des capacités articulatoires permet la formation d'une carte cognitive de liens entre espace perceptif et espace articulatoire. Au niveau computationnel, l'exploration mène à l'apprentissage d'un "modèle direct" du conduit vocal, à l'instar de nos prédécesseurs. Toutefois, elle *inclut* le babillage, parce que (i) elle commence *avant* le babillage et (ii) le babillage désigne, *spécifiquement*, des séquences d'ouverture-fermeture de la mâchoire qui émergent aux alentours de 7 mois, comme défini en phonétique développementale dans la théorie Frame/Content...et non les patrons de mouvements articulatoires aléatoires implémentés en général par les modélisateurs/trices.

Dans notre conception de l'ontogenèse de la parole, l'imitation orofaciale précoce de modèle verbaux permet d'orienter l'exploration du conduit vocal sur les objets de la/les langue-s de l'environnement. Bien que son implémentation fasse, ici aussi, appel à un processus d'inversion, nous en proposons une modélisation en adéquation avec le peu de données disponibles sur le phénomène en psycho-phonétique du développement. En effet, si le babillage fait l'objet de nombreuses études et a pu être clairement défini, dans le domaine de la phonétique, il existe très peu de travaux sur l'imitation précoce de modèles de parole au cours du développement infantile. Nous ancrons la plausibilité d'une imitation orofaciale au service du développement de la production de la parole dans une recherche bibliographique transdisciplinaire, qui raconte et discute ce que le terme d'imitation désigne et suscite comme problématiques, dans les traditions de la psychologie du développement, de l'éthologie et des neurosciences.

L'androïde résultant de notre démarche est une plate-forme expérimentale virtuelle qui nous a permis de simuler le processus d'exploration avant et au début du babillage, aux âges respectifs de 4 et 7 mois, et le mécanisme d'imitation à 4 mois.

Dans le cas de l'exploration, nous avons cherché à estimer, dans une analyse par modélisation, les capacités articulatoires des nourrissons (ou leur *espace d'exploration*) sur la base de données acoustiques décrivant leurs vocalisations, et testé des hypothèses, formulées dans le domaine de la parole, sur le développement précoce du contrôle du système articulatoire. L'imitation a été implémentée sous la forme d'un modèle probabiliste qui apprend des éléments issus de l'espace d'exploration estimé à 4 mois. La carte des liens perception-production qui en résulte a été utilisée pour simuler une expérience, rapportée dans la littérature, sur l'imitation précoce de voyelles [i a u]. Nous avons cherché à reproduire les performances accomplies par les nourrissons, pour interroger les caractéristiques de la carte cognitive et de l'apprentissage potentiellement en jeu dans ce comportement.

Ce mémoire se décline en trois parties. La première décrit les différents éléments que nous avons emprunté à d'autres auteurs: le chapitre 1 donnera les connaissances et hypothèses relatives au développement de la production de la parole et le chapitre 2 présentera les modèles computationnels de production, de perception, d'apprentissage et d'inversion de notre androïde. La seconde partie (chapitre 3) plaidera pour une imitation orofaciale au service du développement de la production de la parole, et veut être jugée comme un résultat à part entière des travaux de recherche qui font l'objet de cette thèse. La troisième partie, qui s'ouvrira sur une définition des notions d'exploration et d'imitation, présentera nos travaux de modélisation computationnelle de l'exploration (chapitre 4) et de l'imitation orofaciale précoce (Chapitre 5). En guise de clôture, nous ferons un bilan des questions et hypothèses résultant de notre démarche, et tenterons d'en proposer quelques perspectives pour qu'elles puissent alimenter la forme de notre moule.

INTRODUCTION GENERALE

1

PARTIE I: Des emprunts

Chapitre 1: Du développement de la production de la parole

7

Introduction

De la datation du début du développement de la parole: des premiers mots au premier souffle

I Du premier souffle aux premiers mots

1. Les apports de la phonétique du développement traditionnelle
 - a. Des idées de Jakobson
 - b. Des modèles classiques de développement par stade
 - c. Des inventaires de productions pré-verbales
2. De la croissance du conduit vocal

De "l'abaissement du larynx" de Lieberman: the show must stop!

3. Dans le courant "underground"
 - a. La classification sensori-motrice des productions précoces
 - Description de la phonation

Le cycle respiratoire.

La mécanique du larynx.

La coordination pulmo-laryngée.

Du codage de la phonation

- Description de l'articulation

Du codage de l'articulation

- b. Un modèle de développement par stade pour la coordination oro-laryngée
- c. Des coordinations intra-utérines

II Babillage canonique et Théorie Frame/Contant

1. A un niveau périphérique
 - a. Du BC comme précurseur de la parole

Des résultats à base de transcription en accord avec les prédictions

- b. Un test des cooccurrences par la modélisation
2. A un niveau central

Un chemin développemental du cadre au contenu

III Du développement de la coordination articulatoire

Conclusion

Chapitre 2: Des outils pour une modélisation réaliste du développement de la production de la parole

36

Introduction

La robotique cognitive de la parole

I Le modèle articulatoire

II Les modèles sensoriels

1. Le modèle auditif
2. Le modèle visuel
3. Le modèle tactile

Pas de parole sans somesthésique

Un modèle de perception somesthésique pour la parole

4. Des modèles perceptifs simplifiés

III Le modèle d'apprentissage sensori-moteur

Conclusion

PARTIE II: Une revue de questions sur l'imitation

Chapitre 3: De l'imitation dans le développement précoce humain et chez des non-humains

52

Introduction

De l'influence précoce de l'environnement...

...dans la prosodie

...dans les voyelles

...dans les consonnes

L'imitation: de l'ajustement de la production à la perception

I. Qu'est-ce que l'imitation?

1. La boîte noire vue du dehors
2. La boîte noire vue du dedans

II. Théories développementales et modèles de l'imitation

1. L'imitation dans la théorie du développement de l'intelligence de Piaget

Des bases pour comprendre Piaget

Un développement par stades de l'imitation

Qu'en retenir?

2. AIM: base d'une nouvelle théorie du développement

Le monde des nourrissons selon Meltzoff et Moore (1997)

Le modèle AIM ou comment les nouveau-nés pourraient imiter des modèles faciaux

Des changements développementaux dans l'imitation

III. Données ontogénétiques et phylogénétiques sur l'imitation

A. Imitation orofaciale chez les nourrissons humains

Pourquoi parler d'imitation oro-faciale de parole plutôt que d'imitation vocale?

1. L'imitation visuelle précoce de modèles oro-faciaux

Du doute

- a. Dès la naissance: imitation du but d'actes corporels variés
 - Procédure standardisée pour l'imitation immédiate

Des cibles croisées...

Architecture.

Motivations.

Interprétation.

...pour remédier à des doutes

- Données complémentaires sur la période néonatale
- b. De 1 à 6 mois: de l'imitation différée d'actes corporels nouveaux à l'imitation d'actions sur des objets
 - Procédure standardisée pour l'imitation différée

Architecture.

Motivations.

Interprétation.

Une alternative au doute

- Données complémentaires sur la période de 1 à 6 mois
- c. De 6 à 24 mois: de l'imitation différée d'actes sur des objets à l'inférence du but

2. Coordination intermodale

Au-delà de croyances: des données

- a. Période néonatale: équivalence visuo-haptique et équivalence audio-motrice

Équivalence visuo-haptique précoce

Équivalence audio-motrice précoce

- b. De 1 à 8 mois: coordination audio-visuelle pour la parole

Faisons le point

Proposition pour un développement intra-utérin d'une coordination audio-proprio-visuelle

3. L'imitation audio-visuelle précoce de modèles oro-faciaux

B. L'imitation chez les animaux non-humains

1. L'imitation chez les non-primates
2. L'imitation chez les primates non-humains: Hominés vs. non-Hominés
 - a. Monkeys

Les Anthropomorphes

- b. Orang-outan *Pongo pygmaeus*
- c. Gorille *Gorilla gorilla gorilla*
- d. Chimpanzé, *Pan troglodites*, et Bonobo, *Pan paniscus*

IV. Données de neurosciences sur des primates adultes

A. Système d'appariement perception-production d'actions de Macaques adultes

1. Vues d'ensemble
 - a. Du cortex perceptuo-moteur

Comme d'une mosaïque

- b. De neurones de l'aire F5

Des neurones perceptuo-moteurs

Notions de sélectivité et de congruence

2. Neurones perceptuo-moteur miroirs
 - a. Miroirs visuo-moteurs d'actions sur des objets
 - Etude de mise en évidence: équivalence perception-production
 - Paradigme expérimental de base

Une variété d'actions sur des objets...

- Propriétés caractéristiques

...pour une variété de miroirs VM d'actions sur des objets...

Sélectivités motrice et perceptive.

Congruence perceptuo-motrice.

...et une architecture de buts et de moyens

- Neurones miroirs VM d'actions manuelles sur des objets et but inféré

- b. Miroirs visuo-moteurs d'actions corporelles orofaciales

- Architecture expérimentale
- Propriétés caractéristiques

Congruence articulatoire des neurones miroirs communicatifs

- c. Miroirs audio-visuo-moteurs et audio-moteurs d'actions manuelles sur des objets
 - Equivalence audio-motrice
 - Equivalence audio-visuo-motrice et coordination intermodale
- d. Un circuit de neurones miroirs PFG/PF-F5c

PFG/PF

STS

3. Une cerise sur le gâteau?
 - a. Neurones miroirs et comportements perceptuo-moteurs précoces
 - b. Neurones miroirs visuo-moteurs orofaciaux et Théorie Frame/Content

De "l'évolution des gestes communicatifs"

B. Système d'appariement perception-production d'actions d'Humains adultes

1. Broca et F5 au niveau fonctionnel
2. Une architecture fonctionnelle minimale pour l'imitation adulte

Conclusion

PARTIE III: Modélisation des mécanismes d'exploration et d'imitation

Introduction à la troisième partie

Deux processus de base: l'exploration et l'imitation

Chapitre 4: Modélisation de l'exploration avant et au début du babillage 132

Introduction

I. Comment caractériser la production de la parole en devenir?

1. Par des méthodes traditionnelles: un manque d'information articulatoire
 - a. Mesures physiologiques
 - b. Mesures perceptives: les transcriptions
 - c. Mesures acoustiques
 - d. Mesures perceptives et acoustiques
2. Par la modélisation articulatoire: proposition d'une stratégie alternative

II. Avec des données acoustiques de référence

1. Le corpus de pré-babillage
2. Le corpus de début de babillage

III. Par un ajustement du modèle articulatoire-acoustique aux données acoustiques

1. Cadrage acoustique
 - a. Méthode
 - b. Résultats
2. Cadrage articulatoire
 - a. Méthode
 - b. Résultats
 - c. Quelques propositions pour traiter le cas de la nasalisation à 7 mois
3. Cadrage géométrique
 - a. Méthode
 - b. Résultats

Le [u] adulte est-il palatal pour des raisons développementales?

IV. Prédications du modèle à 7 mois et confrontation à la réalité

1. Le cas des cooccurrences dans le formalisme de la théorie Frame/Content
 - a. Méthode
 - b. Résultats
 - Un point de vue en surface
 - Un point de vue en profondeur
2. Le cas de l'émergence du contrôle du système articulateur au niveau acoustique dans le paradigme du locus

Un "point info" sur le paradigme du locus

- a. Des éléments pour une modélisation de l'émergence du contrôle du système articulatoire dans le formalisme du locus

Les tendances de cooccurrence de la théorie Frame/Content dans le formalisme du locus

De l'émergence du contrôle local des contacts dans le formalisme du locus

- b. Un mouvement exclusivement mandibulaire pour le babillage canonique
- c. Le contrôle de contacts consonantiques par un pur geste labial et un pur geste apical

V. Discussion

1. Le portrait cohérent d'une exploration articulatoire croissante
2. Les intérêts et les limites de l'approche par modélisation

Conclusion

Chapitre 5: Modélisation probabiliste de l'imitation orofaciale avant le babillage

177

Introduction

Au sujet de l'estimation des capacités articulatoires à 4 mois

I. Test de l'imitation visuelle pure

1. Méthode
2. Résultats

II. Implémentation de la carte des correspondances perception-production

1. Spécification des connaissances préalables à l'apprentissage
 - a. Choix des variables motrices et perceptives
 - b. Décomposition de la probabilité conjointe
 - c. Paramétrisation des distributions composant la structure de dépendance
2. Identification des paramètres libres de la structure de dépendance par apprentissage

Principe

III. Implémentation de l'imitation auditive et audiovisuelle

1. Des questions Bayésiennes
 - Une question pour l'imitation auditive
 - Une question pour l'imitation audio-visuelle
2. Evaluation des capacités d'imitation
 - a. Etude du compromis entre DDG et TBA pour l'inversion audio-articulatoire
 - b. Simulation de l'expérience d'imitation de Kuhl et Meltzoff

IV. Résultats

1. Inversion Bayésienne et compromis DDG-TBA
2. Résultats de la simulation de l'expérience de Kuhl et Meltzoff

Compromis DDG/TBA et sous-apprentissage

Cibles internes vs. externes : le biais du sur-apprentissage

L'imitation précoce de modèle de parole ne nécessite pas beaucoup d'apprentissage

Des matrices de confusion

V. Discussion

Les données vues à travers le modèle

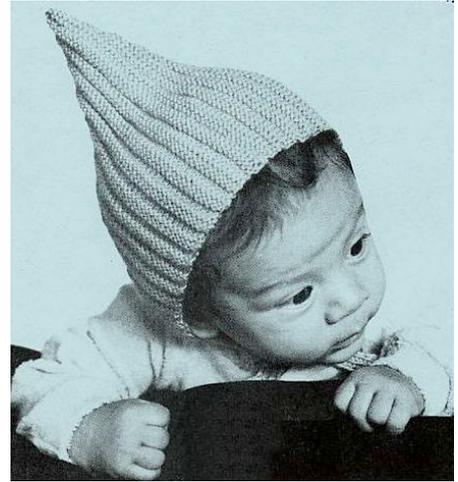
Le modèle vu à travers les données

DISCUSSION GENERALE 199

BIBLIOGRAPHIE 205

ANNEXES

ANNEXE A1: Des détails sur le calcul de P(Mi/D) (Chapitre 4)	225
ANNEXE A2 – Classement des modèles selon la procédure de cadrage articulatoire à 4 mois (Chapitre 4)	227
ANNEXE A3 – Figures du cadrage articulatoire à 3 paramètres pour 7 mois (Chapitre 4)	241
ANNEXE A4 – Classement des modèles selon la procédure de cadrage articulatoire à 7 mois (Chapitre 4)	243
ANNEXE A5 – Spécification de la frontière entre palatales et vélaires sur VLAM à 7 mois (Chapitre 4)	251



PARTIE I: Des emprunts

Chapitre 1: Du développement de la production de la parole dans la première année de vie

INTRODUCTION

De la datation du début du développement de la parole: des premiers mots au premier souffle

D'un point de vue historique, bien que les recherches relatives au développement de la parole soient de fait liées à la reconnaissance que la parole *adulte* est un objet d'étude à part entière, ce champ dérive, dans un premier temps, de celui de l'acquisition du *langage*. La littérature montre que, de Psammetichus dans l'Égypte antique (Barko Gleason, 2001) à Jakobson (1968/1972) et Piaget (1945/1976), la question du début de l'acquisition du langage s'est posée en termes relativement abstraits de *langue-s*. En effet, sans rentrer dans le problème de la définition de ce qu'est une langue¹, l'étude de l'acquisition de la production de langue-s a d'abord centré son attention sur la production des premiers **mots**², **proto-mots**³ ou *signes*, dont la reconnaissance de la forme et/ou de la valeur sémantique dépend-ent, en particulier, de l'expérience linguistique de l'observateur/trice.

En phonétique, la parole étant au moins un comportement moteur audible, les recherches subséquentes, et à ce jour, se sont majoritairement appliquées à étudier sa formation à partir d'un comportement qui apparaît aux alentours de 7 mois pour sa proximité d'apparence avec la parole adulte: le **babillage canonique** ou **babillage multisyllabique**⁴. D'un point de vue relativement naïf, le babillage est perçu par les adultes comme étant une succession de "syllabes", donnant une impression de production verbale dans une langue plus ou moins "étrangère": l'on parle de la qualité *speech-like* du babillage canonique, et la présence de "phonèmes" marque pour certains/aines le début du développement de la phonologie (eg. Kent, 1992c, dans Kent et Miolo, 1995; voir Menn et Stoel-Gammon, 2001, pp. 85-90, pour un résumé des grandes approches théoriques sur le développement de la phonologie). Par le formalisme phonétique qui permet de décrire tous les sons des langues du monde (tel que le système de l'Alphabet Phonétique International), la tendance a été, et est encore, de *transcrire* (ou *décrire*) en termes abstraits de "voyelles" et de "consonnes" (ou "phonèmes") les productions infantiles précoces. Bien que cette tradition méthodologique soulève au moins le problème de l'utilisation d'entités psycho-motrices adultocentrées⁵, les inventaires qui en découlent restent une source d'informations précieuse.

Depuis, la découverte des mécanismes physiologiques basiques de la production de la parole, qui sont indépendants de la langue, a influencé l'étude de l'acquisition de la production verbale. En effet, grâce à ces connaissances, la parole s'inscrit dans l'ensemble des comportements *psycho-moteurs* humains: son substrat fondamentalement biologique ne fait plus de doute (voir les travaux de Vallée et collègues qui développent l'idée de contraintes biologiques communes dans la structuration de la phonologie des langues du monde, e.g.

¹ Nous ne faisons pas, là, allusion à la différence faite entre langue et dialecte au niveau politique.

² **Mot** (traduit de Crystal, 1980/1995): "unité d'expression qui a une *reconnaissance* universelle par les locutrices natives, dans le langage parlé et écrit", ou encore "unité de vocabulaire d'une langue", ou **lexème** (pp.379-381), i.e. l'unité distinctive minimale dans le système *sémantique* d'une langue (p.199).

³ **Proto-mot**: mot inventé par l'individu, ou ressemblant à un mot de la langue ambiante mais ayant un sens ambigu, dans la transition du préverbal au verbal (Menn et Stoel-Gammon, 2001, p. 85).

⁴ Nous donnerons des définitions plus précises du babillage (canonique) aux sections I.3.b. et II..

⁵ Voir dans ce chapitre les sections I.1.c. et II.1.a., et au chapitre 4, section I..

Vallée, 1994, Schwartz, Boë, Vallée et Abry, 1997; Abry, 1997 ; Bonnot, 1989, 1990). De là, la production de la parole étant reconnue comme basée sur la coordination fine des mouvements de et entre différentes parties du corps (telles que le diaphragme, le larynx et la langue), le problème des débuts de l'*acquisition* de la production de la parole a pu être reformulé en termes de *développement* précoce du *contrôle* du *système de production* de la parole. C'est ainsi que la seconde moitié du 20^{ème} siècle a vu l'émergence lente mais sûre de la vision selon laquelle (i) le développement du contrôle du système de production de la parole peut commencer dès le premier souffle, voire *in utero*⁶, et donc bien *avant* le babillage, et (ii) la description des productions précoces en des termes linguistiques adulto-centrés ne suffit pas, à elle seule, pour comprendre les mécanismes en jeu dans l'émergence de la parole mature.

Notre petite histoire de la datation du début du développement de la parole montre que la question du commencement est fonction de la conception de la langue et/ou de la parole à laquelle les divers auteurs adhèrent. Cette idée se retrouvera en filigrane tout au long de ce chapitre qui se décline comme suit. Nous partirons d'un schéma global du développement de la parole du premier souffle aux premiers mots, pour arriver à un modèle intégratif de l'émergence du contrôle du système articulatoire, en passant par la théorie Frame/Content.

I. DU PREMIER SOUFFLE AUX PREMIERS MOTS

1. Les apports de la phonétique du développement traditionnelle

a. Des idées de Jakobson

L'une des premières théories linguistiques sur l'acquisition du langage est celle d'un des fondateurs de la phonologie structuraliste, Jakobson (1968/1972)⁷. L'on y trouve les idées suivantes.

(Ji) Dans la période du babillage, *tous les sons articulés concevables* dont ceux des langues du monde sont observés (Jakobson, 1968/1972, p. 21, où il reprend Grégoire, 1937): le répertoire des productions de babillage est *infini* et ne montre *pas de préférence* (voir aussi Jakobson, d'après MacNeilage et Davis, 2001).

(Jii) Le babillage est suivi de la *perte* de presque toute capacité de produire des sons articulés⁸, dont des sons communs au babillage et à la langue ambiante, bien que les sons de l'environnement servent globalement de *modèle à imiter*⁹. Cette perte marque, selon lui, la transition de la période pré-verbale à celle de l'acquisition des premiers mots, qui correspondrait au premier stade de la langue.

(Jiii) Dans sa vision, la sélection des sons de l'environnement au cours de la transition du babillage à la langue est due au fait que les réalisations acoustiques prennent la fonction de sons de parole (adulte), en acquérant une *valeur phonémique*¹⁰. Or, "l'apparition de phonèmes dans [le] système linguistique [de l'enfant] n'a rien de commun avec les productions sonores

⁶ Voir section I.3.c..

⁷ Pour une revue nord-américaine d'autres courants théoriques de l'acquisition du langage, voir Bohannon III et Bonvillian (2001), et pour une réponse contre l'idée de Chomsky d'un développement génératif de la phonologie, voir MacNeilage et Davis (2001).

⁸ D'après Menn et Stoel-Gammon (2001) cette proposition viendrait de l'existence, chez certains individus, d'une période dite "de silence", "entre" le babillage et la production des premiers mots.

⁹ Nous reviendrons sur ce point dans la seconde partie de ce mémoire.

¹⁰ **Phonème**: "unité minimale du système des sons d'une langue" (Crystal, 1995, p. 258).

éphémères de la période de babillage"¹¹ (Jakobson, 1968/1972, p. 28). En d'autres termes, la production des premiers mots ne se développerait pas à partir de celle du babillage: on parle de *discontinuité* entre babillage et premiers mots (Menn et Stoel-Gammon, 2001).

(Jiv) Toutefois, les réalisations acoustiques de la langue ambiante sont recouvertes sous la forme de sons de parole (adulte), après des efforts qui s'étendent parfois sur plusieurs années, parce que, pendant le babillage, l'individu s'est imprégné de leur image motrice et de leur image acoustique à force de répétition (Jakobson, 1968/1972, p. 28), mais ces "images" restent sans lien avec la langue puisque "la langue en tant que *système de signes arbitraires* [*sic*] [...] ne doit pas être confondue avec les gestes sonores"¹² (Jakobson, 1968/1972, p. 29).

Nous verrons que, malgré la grande influence de ses réflexions, parties des bases du domaine du développement de la parole, des éléments de sa vision seront remis en question par des études quantitatives. En particulier, les postulats de répertoire de babillage infini (*Ji*) et de discontinuité entre babillage et premiers mots (*Jiii*) ne résistent pas à l'épreuve des faits.

a. Des modèles classiques de développement par stade

Sous l'influence des théories du développement par **stades**¹³ de la tradition de la psychologie développementale (voir Harris et Butterworth, 2002, pp. 38-42), différentes propositions ont été faites pour le cas de l'acquisition du langage dans la première année de vie¹⁴ (Tableau 1-I).

Bien que, de nos jours, le babillage soit le plus souvent pris comme point de départ de l'acquisition de la parole, plusieurs auteurs ont aussi rendu compte du "développement des vocalisations" de la période antérieure (pour reprendre l'expression de Menn et Stoel-Gammon, 2001, e.g. p. 80).

D'après Kent et Miolo (1995), les différents modèles de développement par stades s'accordent sur la séquence d'événements suivante:

(i) Jusqu'à 1 mois, l'on observe surtout des **vocalisations réflexives**¹⁵.

¹¹ "[T]he appearance of phonemes in [the child's] linguistic system has nothing in common with the ephemeral sound productions of the babbling period".

¹² "[L]anguage as a *system of arbitrary signs* [*sic*][...] must not be confused with sound gestures".

¹³ L'idée-phare des théories par stades est que les changements, qu'implique la notion de développement, (i) ont lieu de manière discontinue plutôt que continue, et (ii) correspondent à des modifications *qualitatives* du comportement de l'individu. Un exemple de changement qualitatif s'illustre par le développement de la locomotion qui passe de "marcher à quatre pattes" à "marcher debout". En tête de proue des théories développementales par stades figure celle de Piaget (voir chapitre 3, section II.1.).

D'après Flavell (1963), les **stades** de développement, au sens strict, remplissent les critères suivants (Harris et Butterworth, 2002, p. 40-41). D'un stade à l'autre:

1. il y a des *changements qualitatifs*.
2. la transition est marquée par **des changements simultanés dans plusieurs autres aspects du comportement** de l'individu. L'exemple donné par Harris et Butterworth (2002) est que, dans le même temps que le sujet commence à comprendre l'association entre mot et sens, il se comporte comme si des objets avaient eux aussi des propriétés symboliques (jeux de faire-semblant qu'un objet est une personne): "il y a un effet généralisé de l'acquisition de la capacité de traiter le monde en termes de systèmes symboliques" ("there is a widespread effect of acquiring the capacity to treat the world in terms of symbol systems", p.41).
3. la *transition est typiquement rapide* comme dans le cas de la poussée de croissance et de poids à l'adolescence, ou du phénomène d'explosion de la taille du vocabulaire, au cours du développement de la parole, où aux environs des 30 premiers mots le nombre de nouveaux mots appris augmente de manière exponentielle.

¹⁴ Ici, le terme de "stade" est à prendre dans un sens plus large qu'en psychologie du développement, vu que les auteurs s'intéressent uniquement à l'acquisition du langage (Cf. note 14, critère (ii)).

¹⁵ **Vocalisations réflexives** : sons d'origine végétative, d'expression d'inconfort ou de stress, qui incluent les pleurs, les toux, grognements et rots (Menn et Stoel-Gammon, 2001).

(ii) Entre 1 et 2 mois, un premier changement a lieu: les **vocalisations non-réflexives**¹⁶ apparaissent. L'on parle notamment de phonation et de cooing/gooning. Pour information, les termes de **coo** et **goo** (prononcés [ku:] et [gu:], respectivement) désignent des sons qui semblent être articulés vers l'arrière du conduit vocal, et qui sont transcriposables en termes de "consonnes" *vélaires* (e.g. [g]) et de "voyelles" *arrières* (e.g. [u]) (Menn et Stoel-Gammon, 2001).

(iii) Entre 3 et 6 mois, il y a diversification des comportements du conduit vocal, avec des *constrictions de plus en plus fréquemment supra-glottales*. C'est là que l'on trouve le stade de **jeu vocal**, et, d'après Menn et Stoel-Gammon (2001, p. 81) qui s'appuient sur Oller (1980) et Stark (1980), "[d]ans cette période [qui va, ici, de 4 à 6 mois] c'est comme si les bébés testaient leur appareil vocal pour déterminer le domaine des qualités vocales qu'ils/elles peuvent produire. La période est caractérisée par l'apparition de sons très intenses et très doux (yells [*inTd*¹⁷] et chuchotements), et de sons très hauts et très bas (squeals [*inTd*] et grognements). Certains bébés produisent de longues séries de raspberries [*inTd*] (trills bilabiaux) et des voyelles soutenues, et occasionnellement quelques syllabes rudimentaires faites de consonnes et de voyelles ont lieu"¹⁸.

(iv) Entre 6 et 8 mois, le *babillage canonique* commence. Des auteurs, tels que Stark et Elbers proposent que le babillage canonique procède en deux phases successives (Kent et Miolo, 1995), en prenant d'abord la forme (a) du **babillage redoublé**, qui correspond à des séquences de type syllabique dans lesquelles le même motif est répété (eg. [ba ba ba]), puis celle (b) du **babillage varié** où les séquences sont constituées de motifs syllabiques différents (eg. [ba wa wə]) (Menn et Stoel-Gammon, 2001). Signalons de suite qu'en réalité ces deux formes de productions coexistent dès le début du babillage canonique (voir Kent et Miolo, 1995, pp. 325-326 pour des références).

(v) Entre 9 et 12 mois, il y a complexification des vocalisations en prenant la forme d'un **jargon** (ou **babillage conversationnel** ou **babillage modulé**), qui est caractérisé par la production de chaînes de sons et de "syllabes" dont les patrons accentuels et prosodiques sont très variés (Menn et Stoel-Gammon, 2001). De plus, c'est souvent vers la fin de la première année que les premiers mots, ou proto-mots, sont produits.

Ces grandes lignes du décours développemental se retrouvent de manière plus ou moins claire dans les études à base de transcription phonétique des productions infantiles.

¹⁶ **Vocalisations non-réflexives** : sons d'origine non végétative qui contiennent plusieurs caractéristiques phonétiques trouvées dans les langues (adulte) (Menn et Stoel-Gammon, 2001).

¹⁷ *inTd*: intraduisible, i.e. je n'en n'ai pas trouvé la traduction.

¹⁸ "In this period it seems as though babies are testing their vocal apparatus to determine the range of vocal qualities they can produce. The periode is characterized by the appearance of very loud and very soft sounds (yells and whispers), and very high and very low sounds (squeals and growls). Some babies produce long series of raspberries (bilabial trills) and sustained vowels, and occasionally some rudimentary syllables of consonants and vowels occur."

Tableau 1-I: Les modèles classiques de développement par stades en phonétique du développement (emprunté à et traduit de Kent et Miolo, 1995, p. 324).

Age (mois)	0-1	1-2	2-3	3-4	4-5	5-6	6-7	7-8	8-9	9-10	10-11	11-12
Stark (1979)	Pleurs réflexifs et sons végétatifs		Cooing et rires				Babillage redupliqué					
						Jeu vocal			Babillage non-redupliqué et jargon expressif			
Oller (1980)	Phonation		Gooing		Expansion		Babillage canonique			Babillage varié		
Holmgren <i>et al.</i> (1986)	Phonation continue sans articulation		Phonation interrompue sans articulation	Phonation continue ou interrompue avec une articulation		Variants phonatoires avec ou sans articulation		Phonation continue ou interrompue et articulation redupliquée				
Elbers (1982)	Vocalisation								Répétitif	Concaténation		Mélange
Kent (1990)	Phonation précoce		Phonation tardive	Articulation simple avec variants phonatoires			Babillage multisyllabique (les babillages redupliqué et varié se développent en parallèle)					
Nakazima (1980)	Pleurs réflexifs		Développement des mécanismes phonatori-articulatori-auditifs du babillage				Babillage répétitif		Réorganisation des mécanismes phonatori-articulatori-auditifs			

b. Des inventaires de productions pré-verbales

Bien que les productions préverbales ressemblent à des sons de parole, elles ne remplissent pas les caractéristiques données aux phonèmes adultes, qui supposent notamment l'utilisation de mots. Dans le but de distinguer les productions consonantiques et vocaliques matures des productions pré-verbales, l'on utilise depuis Martin (1981) les termes respectifs de **closants** (*C*) et **vocants** (*V*) (voir aussi ce chapitre, section III.1.a. et chapitre 4, section I). Dans le même ordre d'idée, les productions de type syllabique sont désignées par le terme de **proto-syllabes**.

De nos jours, le point de vue dominant en phonétique faisant du babillage canonique le début du développement de la parole, les inventaires à base de transcription phonétique décrivent le plus souvent les sons observés dans cette période développementale. Toutefois, quelques études rapportent des transcriptions faites sur l'ensemble de la première année de vie: le Tableau 1-II regroupe des informations fournies par Kent et Miolo (1995). Le bilan peut se résumer comme suit.

(i) Les vocants sont de composition relativement constante au cours de la première année, avec des sons majoritairement perçus comme *antérieurs* ([ɪ], [ε], [æ]), *centraux mi-ouverts* ([ə/ʌ]) et *postérieurs* ([ʁ/ɑ])¹⁹. En particulier, Kent et Bauer (1985) trouvent, dans les productions d'un individu de 12 mois, une dominance des vocants centraux [ə ʌ] et bas [ɑ] dans les proto-syllabes CV du babillage canonique (Kent et Miolo, 1995). La prépondérance de l'ensemble de ces types de productions vocaliques est vérifiée dans des suivis longitudinaux (a) par Davis et MacNeilage (1995), sur un groupe de 6 sujets dans la période allant du début du babillage aux premiers mots, et (b) par Liberman (1980) qui, avec 5 sujets de 16 semaines à 5 ans, ne trouve pas de discontinuité du répertoire vocalique entre le babillage et les premiers mots. En d'autres termes, au cours de la première année de vie, *les individus ont tendance à ne pas utiliser des positions de langue extrêmes* ([i], [u]) *sur l'axe avant-arrière*.

(ii) Les closants seraient plus souvent d'articulations *laryngales* (i.e. [ʔ] et [h]) que supra-glottiques, avec, au cours du développement, une augmentation de la fréquence des sons du second type²⁰. Plus précisément, les lieux d'articulation supra-laryngaux montrent à l'initiale une préférence vélaire ([k], [g]) qui tend à devenir bilabiale ([b m]) et alvéolaire ([d])²¹, en particulier à partir de 8 mois, i.e. avec les débuts du babillage canonique. En accord avec cette tendance, Davis et MacNeilage (1995) trouvent une prépondérance des lieux labial et alvéolaire sur l'ensemble de leur période d'étude. Au niveau des modes d'articulation, le *voisement* semble être la règle. De plus, si le son constrictif voisé [h] est très présent jusqu'aux alentours de 8 mois, les occlusifs voisés deviennent majoritaires sur l'ensemble du babillage canonique (Davis et MacNeilage, 1995). En effet, cette période est marquée par la production d'*occlusives voisées orales* [b d] *et nasales* [m n], bien que des *semi-consonnes* (glides) [w j]²²

¹⁹ Les auteurs précisent que la variabilité inter-individuelle du jeu des productions est relativement élevée.

Notamment, bien que les sons vocaliques extrêmes [i] et [u] soient globalement rares, certains individus en produisent fréquemment. Cela illustre le fait que les inventaires phonétiques donnent plutôt des indications *statistiques* que des informations déterministes.

²⁰ Le début de ce phénomène correspond à l'événement (iii) dans la section I.1.b. De plus, la diversité des vocants étant relativement constante aux cours de la première année de vie, l'on comprend au moins en partie la décroissance de l'**indice V/C**, qui correspond au rapport du nombre de réalisations vocaliques sur celui des réalisations consonantiques.

²¹ [k]: *fr.* comcombre, [g]: *fr.* gargarisme, [b]: *fr.* bagnole, [m]: *fr.* mousse, [d]: *fr.* dire.

²² [n]: *fr.* nom, [w]: *angl.* women, [j]: *angl.* year.

soient aussi observées. Stoel-Gammon (1985) trouve que ces closants sont aussi les plus fréquents dans les premiers mots: ils apparaissent avant les fricatives (e.g. [s ʃ]), les mi-occlusives (e.g. [tʃ]) et les liquides (e.g. [l])²³ dans la parole précoce (Menn et Stoel-Gammon, 2001).

Tableau 1-II: Les vocants et closants les plus fréquemment relevés dans les productions préverbaux de la première année de vie d'après Kent et Miolo (1995). Sources: (A) Mowrer (1980), (B) Irwin (1947b), étude menée sur 95 individus, (C) Kent et Bauer (1985), étude menée sur un individu de 12 mois.

Age (mois)	Indice V/C	Vocants	Closants
0-2	4.5 ^A	ε : 43% ^A ɪ : 27.4 % ^A ʌ : 25.2 % ^A	h ʔ: 87 % ^B k g : 11.6 % ^B suivent: m ø b t ʃ n p w ^B
2-4	3.6 ^A	ε ɪ ʌ : 81.5% ^A	h : 59.8 % ^B ʔ: 15.5 % ^B g: 11.7 % ^B suivent: k d b w l j m θ t ð n ŋ p v ø z ʃ s χ ^B
4-8	2.8 ^A	ε ɪ ʌ : □70% ^A	h : 59.9% ^B reste dominant ʔ: 10.5 % ^B g: 6.4 % ^B m b d augmentent ^B
8-12	2 ^A	ε ɪ ʌ : □70% ^A ə/ʌ ε ɑ æ γ ^C	d : 18.1 % ^B m b : 16 % ^B h : 36.5 % diminue ^B ʔ: 4.8 % diminue ^B

Globalement, les études à base de transcription phonétique mettent en évidence que les productions juvéniles les plus fréquemment collectées représentent un jeu relativement restreint des sons trouvés dans les langues du monde (voir Vallée, 1994; Rousset, 2004): le postulat de Jakobson selon lequel le répertoire de babillage est infini et ne montre pas de préférence (Cf. section I.1.a., point *Ji*), ainsi que celui de perte de toute capacité de production (*Jii*), ne sont pas vérifiés. De plus, les résultats obtenus par Liberman (1980) et Stoel-Gammon (1985) donnent une première remise en question de la discontinuité supposée entre babillage et premiers mots (*Jiii*). Dans le même ordre d'idée, nous verrons que les caractéristiques des closants et des vocants du babillage canonique, traditionnellement répertoriés de manière disjointe, ne sont vraisemblablement pas indépendants les uns des

²³ [s]: support [ʃ]: fr. chose, [tʃ]: churos, en Espagnol, [l]: fr. liquide.

autres et que cela permettra de renforcer la remise en question de l'idée de discontinuité (voir section II.1.a.).

Des tendances mises en évidence par le formalisme phonétique sont aussi compréhensibles en des termes proprement articulatoires, en particulier dans le cadre du *développement de la coordination orolaryngée*. Mais avant tout, nous désirons donner quelques éléments basiques autour de la croissance du conduit vocal dont la morphologie participe à la détermination de *l'espace acoustique des possibles articulatoires* sur lequel nous reviendrons au chapitre 4 (voir section III.1.a., la notion d'espace vocalique maximal).

2. De la croissance du conduit vocal

Une bibliographie détaillée sur la croissance du conduit vocal en relation avec ses conséquences acoustiques et auditives se trouve dans Ménard (2002). Toutefois, nous tenons à rappeler les points-clés suivants.

(i) A la *naissance*, la taille du conduit vocal (qui va du larynx aux lèvres) est plus petite qu'à l'âge adulte (Kent et Murray, 1982; voir Figures 1-1 et 1-2). Toujours en référence à l'adulte, la longueur de la cavité pharyngale rapportée à celle de l'ensemble du conduit est plus courte ("le larynx est haut"), permettant au velum et à l'épiglotte de se toucher, et l'angle entre les cavités buccale et pharyngale est plus élevé (la courbure du conduit vocal est moins prononcée). De plus, il n'y a pas encore de dentition, la masse linguale est plus antérieure qu'à maturité et remplit la cavité orale au repos²⁴ (Bosma, 1961, cité par Mackenzie Beck, 1999).

(ii) La morphologie globale du conduit vocal se rapproche de la norme adulte vers l'âge de *4 mois* (voir Kent et Miolo, 1995, pour des références). A environ *6 mois*, le velum et l'épiglotte ne se touchent plus sauf lors de la déglutition, et l'absence de contact entre ces deux structures est permanente à partir de *12-18 mois* (Mackenzie Beck, 1999). La datation de l'apparition de la dentition primaire ("dents de lait") est très variable: la première dent émerge à 6 mois en moyenne et toutes les dents sont généralement présentes autour de 2 ans et demi (Mackenzie Beck, 1999).

(iii) A un niveau plutôt synchronique, les différentes structures du conduit vocal, notamment les tissus osseux, la langue et les cartilages du larynx, se modifient chacune à sa manière et à son rythme au cours de la première décennie (Kent et Miolo, 1995): l'on parle de croissance non-uniforme du conduit vocal. Il en résulte, notamment, que *la taille de la cavité pharyngale augmente plus rapidement que celle de la cavité buccale*²⁵.

(iv) A un niveau plus diachronique, l'on observe que le décours temporel du processus de croissance d'une structure donnée n'est pas régulier vu qu'il est marqué par des périodes de croissance rapide ("poussées de croissance") et des périodes de croissance relativement lente: l'on parle de croissance non-linéaire. Comme pour d'autres parties du corps, le conduit vocal aurait deux poussées de croissance: l'une a lieu *entre 0 et 5 ans*, l'autre à la *puberté*, celle-ci étant fréquemment plus prononcée pour les individus masculins que pour ceux féminins (Mackenzie Beck, 1999). En particulier, les données anatomiques recueillies par Goldstein (1980), qui vont de quelques mois à 20 ans avec une analyse transversale du phénomène de croissance, ont permis à Boë (1999) de vérifier que *la longueur de la cavité pharyngale rapportée à celle de la cavité buccale suit une double-sigmoïde* (Figure 1-3) pour les

²⁴ Pour information, selon Mackenzie Beck (1999), "l'absence de dentition dans les premiers mois de vie influence le potentiel articulatoire et pourrait avoir un effet sur la *posture de la langue* [sic]" (p. 277, "the lack of teeth during the first months of life influences articulatory potential and may have an effect on tongue posture").

²⁵ Précisons que les modifications du conduit vocal ne sont pas qu'une question de "pure maturation". En particulier, les forces appliquées par les muscles semblent aussi avoir un rôle dans les modifications morphologiques du conduit vocal (Koopman-van Beinum et van der Stelt, 1998).

individus mâles²⁶. Ce décours développemental se retrouve aussi dans le cas de la longueur des "cordes" vocales (Goldstein, 1980, Figure 1-4).

(v) Les modifications anatomiques qui ont lieu entre le début et la fin de la croissance ont des répercussions au niveau acoustique. Notamment, la croissance de l'appareil vocal va de pair avec une *décroissance monotone de la fréquence fondamentale (f_0)*²⁷ (Mackenzie Beck, 1999; Figure 1-5), et une *contraction relativement homothétique de l'espace occupé par les fréquences de résonance (ou formants) du conduit*²⁸ (Boë, 1999; voir chapitre 2, section I, Figure 2-4).

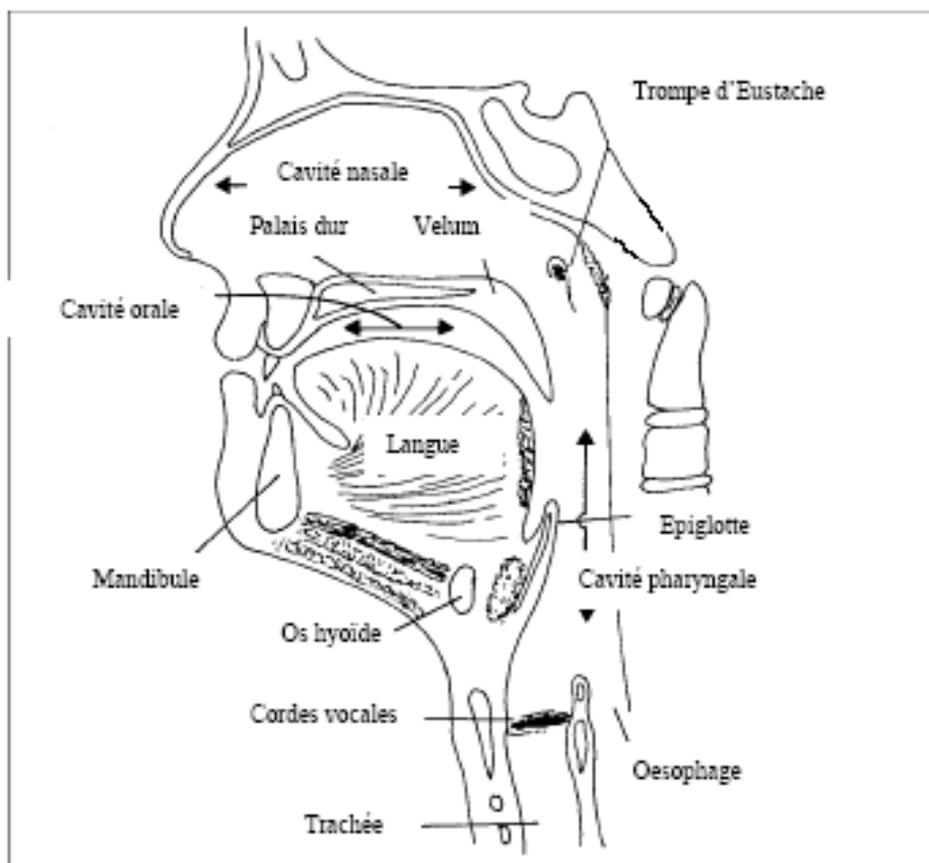


Figure 1-1: Schéma légendé d'un conduit vocal adulte (issu et traduit de Kent, 1997, p. 142).

²⁶ En accord avec Goldstein (1980), Boë (1999) considère que la croissance de la longueur du conduit vocal des individus de sexe féminin est identique à celle des individus de sexe masculin jusqu'aux alentours de 16 ans, âge auquel le processus viendrait à maturité chez les premières.

²⁷ La **fréquence fondamentale** (F_0) correspond à la fréquence de vibration des cordes vocales et détermine la hauteur de voix, mais aussi l'espacement entre les harmoniques dans les spectres des sons vocaliques. Plus la voix est aiguë, plus la fréquence est élevée, et plus les harmoniques sont espacées. De plus, si l'on écarte les facteurs culturels (Mattingly, 1966, cité par Ely, 2001), la hauteur moyenne de la voix d'une personne est en premier lieu fonction de la longueur et de l'épaisseur de ses cordes: plus elles sont fines et petites, plus la valeur de F_0 tend à être élevée. L'on comprend pourquoi au début du développement de la parole F_0 est élevée et le spectre des sons produits est "pauvre en harmoniques": les formants sont relativement difficiles à identifier avec certitude dans le cas des productions précoces.

²⁸ En effet, la taille du conduit vocal étant bien plus petite au début du développement de l'individu qu'à maturité, le résonateur de la parole tend à filtrer le son d'entrée issu du larynx, en amplifiant les fréquences élevées qui s'y trouvent.

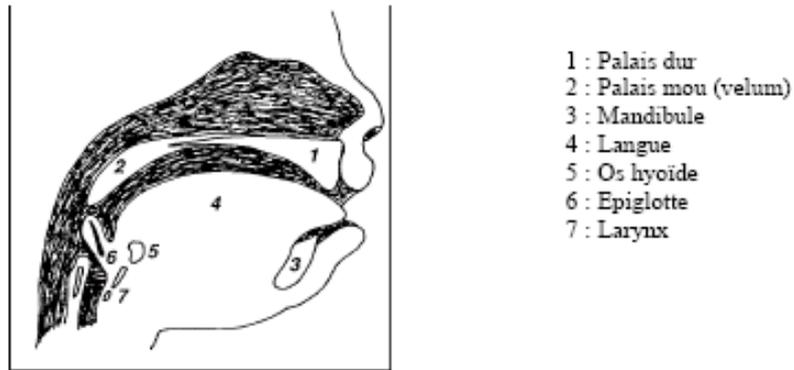


Figure 1-2: Schéma légendé d'un conduit vocal néonatal (issu et traduit de Kent, 1997, p. 145).

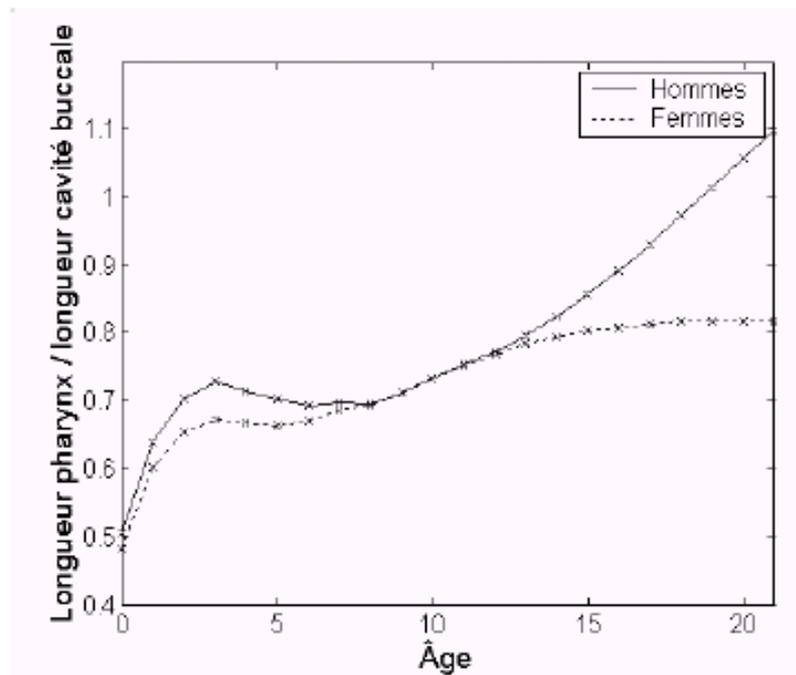


Figure 1-3: Rapport entre les longueurs des cavités pharyngale et buccale en fonction de l'âge et du sexe, selon Goldstein (1980; figure tirée de Ménard 2002).

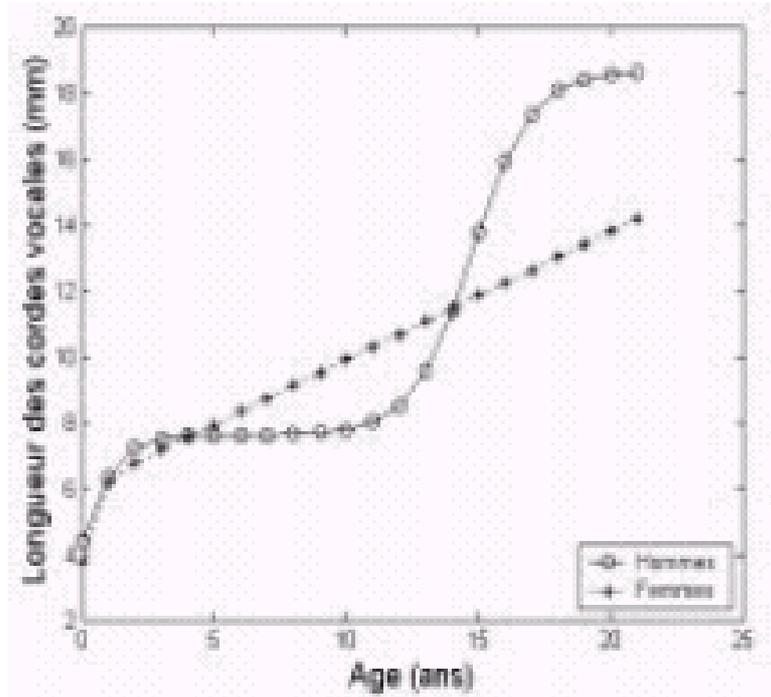


Figure 1-4: Evolution de la longueur des cordes vocales de la naissance à l'âge adulte, selon Goldstein (1980; figure tirée de Ménard 2002).

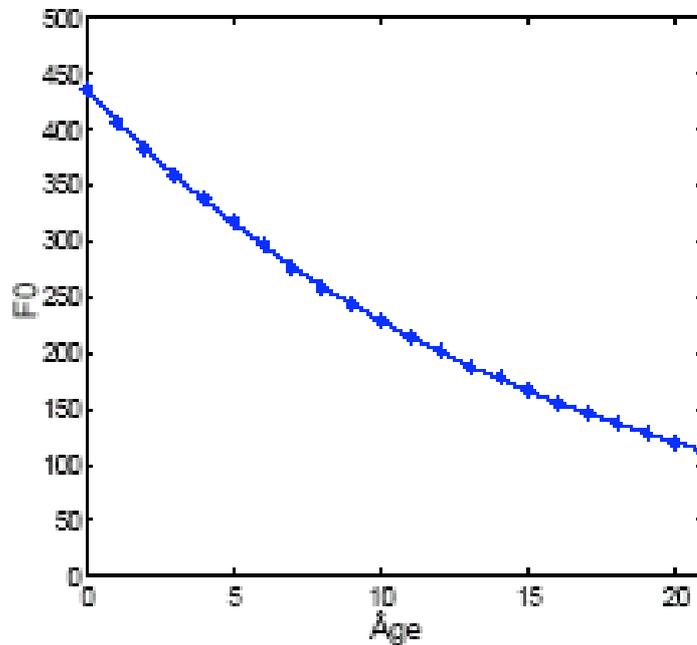


Figure 1-5: Evolution de la fréquence fondamentale (f_0) au cours du développement pour un locuteur masculin d'après Mackenzie Beck (1999; figure tirée de Ménard 2002)

De "l'abaissement du larynx" de Lieberman: *the show must stop!*

Nous avons vu qu'à la naissance, "le larynx est haut", que la cavité pharyngale est petite et qu'elle forme un angle ouvert avec la cavité buccale. Au cours de la croissance, la cavité pharyngale grandit plus vite que la buccale: il y a "descente du larynx". De plus, l'angle entre ces cavités tend à devenir droit (ou "coudé").

D'après Lieberman (1971), la "descente du larynx", et donc la présence d'une cavité pharyngale de grande taille, est un prérequis anatomique à la production de sons de parole, que cela soit au niveau ontogénétique ou phylogénétique (Boë, 1999; Boë, Heim, Honda et Maeda, 2002). En particulier, les nouveaux-nés ne seraient pas capables de produire les sons [i a u] parce que la hauteur de leur larynx les empêcherait de former les constrictions correspondantes au "bon endroit" dans leur conduit vocal.

Cette hypothèse onto-phylogénétique, que l'on rencontre presque systématiquement dans la littérature relative au développement de la parole et dans les manuels universitaires, a été réfutée par Boë et collègues sur la base de résultats de modélisation (Boë, 1999; Boë *et al.*, 2002). En effet, les auteurs montrent que si les nouveaux-nés avaient les mêmes capacités articulatoires que les adultes, même avec un larynx relativement court, ils seraient capables de produire un espace des voyelles au moins aussi riche que celui de leurs aînés, et que cet espace inclurait les voyelles [i a u]. Pour illustration, la production du son [i] nécessite simplement de former la constriction plus en avant à l'intérieur du conduit vocal dans le cas d'un "larynx haut" que dans celui d'un "larynx bas".

Il vient que, si les nouveaux-nés ne produisent pas les voyelles [i a u], ce n'est vraisemblablement pas dû à la position de leur larynx. Les auteurs proposent que des facteurs centraux (et non-périphériques), tels que le manque de coordination des différents articulateurs du conduit vocal pour la production de sons de parole, peuvent rendre compte de la non-observation de ces vocants dans les premiers mois de vie. Toutefois, l'idée selon laquelle avoir un conduit coudé pourrait mener à un contrôle plus simple de la configuration articulatoire par rapport au cas d'un conduit non-coudé reste un argument à démontrer... ou à démonter.

3. Dans le courant "underground"

a. La classification sensori-motrice des productions précoces

Pour sortir l'étude du développement de la production de la parole des sentiers de la phonétique orientée-adulte, Koopmans-van Beinum et van der Stelt (1998) proposent une *classification sensori-motrice* des productions précoces indépendante des systèmes phonémiques. D'après elles, il est possible de décrire les productions pré-verbales, avec plus de précision qu'avec des unités phonologiques, en termes de phonation²⁹ et d'articulation au cours du cycle respiratoire. Il s'agit ici de revenir aux mécanismes basiques de production de la parole, en s'appuyant sur le modèle source-filtre.

Pour rappel, la *source* est générée par le système respiratoire (poumons et trachée) et le système laryngé (larynx) où se trouvent les "cordes vocales". En phase d'expiration³⁰, l'air expulsé des poumons passe par l'espace qui sépare les cordes vocales (la **glotte**): si elles sont accolées, il y a production d'un son périodique (source de voisement), sinon, la source est de bruit et se situe plus en amont dans le conduit vocal. Ce son de base excite le volume délimité par les parois du reste du conduit qui constitue le *filtre*. La position relative des différents

²⁹ La **phonation** correspond à la production du son périodique qui sous-tend le voisement.

³⁰ La majorité des vocalisations précoces et des sons des langues du monde est produite en phase d'expiration.

articulateurs³¹, ou **configuration articuloire**, détermine l'effet de filtrage acoustique (sur la source) du conduit vocal dont la forme joue le rôle de résonateur. Globalement, le type de phonation informe sur le comportement du larynx, le type d'articulation sur celui du système articuloire.

La parole se développe vraisemblablement à partir de productions qui ont lieu *dans* et *par* le cycle respiratoire. De là, le cycle respiratoire est choisi pour unité de segmentation du système de codage des productions précoces: on parle de **groupe** ou d'**unité respiratoire (UR)**. L'on peut remarquer que cette métrique est *physiologique*, et non phonologique. De plus, la parole nécessite au moins une coordination entre respiration, configuration du larynx et configuration articuloire: l'on parlera de **coordination orolaryngée**. Cette capacité n'est pas présente dès la naissance, mais se développerait selon un emboîtement hiérarchique de la maîtrise des différents (sous-)mécanismes de la production verbale. Ces (sous-)systèmes se départagent en ceux relatifs à la phonation et ceux à l'articulation.

➤ Description de la phonation

La phonation nécessite le contrôle du cycle respiratoire, de la configuration du larynx, et de la coordination des deux.

Le cycle respiratoire. Chez l'adulte, le cycle respiratoire de base, qui est contrôlé de manière réflexive pour alimenter le corps en oxygène, présente un rapport des durées des phases d'inspiration et d'expiration de 1:1. Lors de la production (mature) de la phonation, le cycle est tel que le rapport des durées est de 1:6. Au cours du développement, l'individu doit (i) d'abord apprendre à stabiliser son cycle selon un patron de type 1:1 pour la respiration physiologique car celui-ci n'est pas stable dès la naissance (voir Koopman-van Beinum et van der Stelt, 1986, p. 38, pour des références), (ii) puis apprendre à réguler ce cycle par le contrôle volontaire de la contraction du diaphragme pour la production de sons (patron de type 1:6) (Koopman-van Beinum et van der Stelt, 1998).

La mécanique du larynx. Au repos, la configuration du larynx est telle que les cordes vocales sont séparées l'une de l'autre pour laisser l'air librement entrer dans les et sortir des poumons pour la respiration physiologique. Lors de la phonation, les cordes vocales s'accolent en phase d'expiration: il y a fermeture du larynx, et le débit d'air à la glotte se fait par impulsions régulières (production du voisement). Cela demande au moins une maîtrise du mécanisme de *fermeture-ouverture du larynx*. D'après les auteurs, la présence dans les premiers mois de vie d'occlusions, [ʔ], et de constrictions, [h], glottales selon le formalisme classique témoignent du manque de maturité de cette habileté (Koopman-van Beinum et van der Stelt, 1998).

De plus, au niveau individuel, les variations de hauteur de voix dans les productions verbales (prosodie) résultent de l'action de cartilages laryngaux qui permettent de faire varier l'état de tension des cordes vocales. L'indice de développement du contrôle de la *tension des cordes vocales* suit, d'après les auteurs, le degré de diversité des patrons prosodiques pré-verbals (Koopman-van Beinum et van der Stelt, 1998).

La coordination pulmo-laryngée. Le contrôle de la phonation nécessite l'apprentissage de la coordination de la mécanique du larynx avec le cycle respiratoire et le débit d'air issu des poumons, dont les corrélats dans la parole mature sont (i) l'alternance de sons voisés et non-voisés, au cours d'un même cycle respiratoire, et (ii) la présence de patrons de prosodie et d'intensité sonore ("volume") variés, dans le répertoire des productions.

³¹ L'ensemble des articulateurs, i.e. la langue, les lèvres, le vélum et la mandibule, forme le **articuloire** ou **système supra-laryngé**.

Du codage de la phonation

Il vient que la phonation, i.e. les caractéristiques de la source de base, peut être décrite selon 7 variables³²:

- l'occurrence de la phonation dans l'UR, qui s'exprime en terme de présence/absence;
- la direction du flux d'air lors de la phonation qui peut être sortant (égressif), dans le cas d'un développement "normal", ou entrant (ingressif) ;
- la *continuité de la phonation* dans l'UR, qui est soit continue soit interrompue, sachant que la capacité d'alterner initiations et arrêts de la phonation, dans une même UR, demande un degré élevé de coordination pulmo-laryngée, menant à la production de série de pseudo-syllabes de type [ʔV];
- la durée de la phonation dans l'UR, qui va de courte à longue, et qui sous-tend, d'après les auteurs, l'émergence de l'opposition phonologique entre voisement et non-voisement;
- le type du début du voisement, qui peut être aspiré, normal ou de type attaque/occlusion glottique et est lié au contrôle de la mécanique de fermeture du larynx;
- la qualité du voisement qui se décline en des termes relatifs à la clarté et à l'intensité sonore de la voix;
- le *patron prosodique*, qui est soit simple (stable, montant ou descendant), soit complexe (combinaisons des patrons simples).

➤ **Description de l'articulation**

Dans la période néonatale, le système articuloire sert surtout à se nourrir sur un mode réflexif (Koopman-van Beinum et van der Stelt, 1998): (i) l'activité pharyngale est péristaltique (lors de la déglutition) ou de type sphincter³³, (ii) le velum est le plus souvent au repos (il est baissé), permettant de respirer par le nez tout en s'alimentant, (iii) la langue a surtout des mouvements d'alternance abaissement-repos de son corps pour la tétée (Perrier, com. pers.), et ses muscles intrinsèques ne lui permettent pas d'adopter une grande variété de formes, et (iv) les lèvres sont globalement passives³⁴.

Le développement du contrôle de la configuration articuloire pour la production verbale procède en une *réorganisation des patrons moteurs* à fonction alimentaire en des mouvements et combinaisons de mouvements nécessaires à la parole.

Du codage de l'articulation

Les caractéristiques du filtre sont décrites en termes de **mouvement articuloire**, que les auteurs définissent comme "tout type de constriction ou friction qui peut être formée au moyen d'organes articuloires [supra-glottiques]"³⁵ (Koopman-van Beinum et van der Stelt, 1998, p. 119). Toutefois, elles précisent que, par défaut, le "mouvement articuloire" ne s'applique pas aux productions dont la configuration articuloire ne produit pas de turbulence ou de cloison, excluant ainsi les sons de type vocalique de leur description de base: ceux-ci peuvent néanmoins faire l'objet d'une description de type phonétique selon leurs degrés d'antériorité (et d'aperture).

Le codage de l'articulation peut se décliner en 7 variables:

- l'occurrence de l'articulation dans l'UR qui s'exprime en termes de présence/absence;
- le nombre de mouvements articuloires dans l'UR, qui peut être supérieur ou égal à 1;

³² Ces variables et leurs valeurs possibles sont données à titre d'exemple par les auteurs car le formalisme proposé reste adaptable aux nécessités de l'étude menée. Il en est de même pour la description de l'articulation.

³³ Ces réflexes disparaissent progressivement.

³⁴ A titre indicatif, le sourire apparaît autour de l'âge de 6 semaines (van der Stelt et Koopman-van Beinum, 1986).

³⁵ "[A]ll kinds of constrictions or frictions that can be made by means of articulatory organs."

- le moment du mouvement articuloire dans l'UR, i.e. initial, médian ou final;
- le *patron des mouvements articuloires* dans l'UR pour un nombre de mouvements supérieur à 1, qui est soit répétitif soit varié;
- le *lieu d'articulation* du mouvement, i.e. antérieur, médian ou postérieur;
- le mode d'articulation, i.e. nasale, occlusif, fricatif, glide ou trill ;
- la différenciation des sons vocaliques: articulation antérieure, neutre ou postérieure³⁶.

Pour faire un pont avec la tradition phonétique, notons que la production de pseudo-syllabes peut résulter soit de l'interruption de la phonation, soit de la modification de la source de voisement par des mouvements articuloires (Gillis, Koopmans-van Beinum et van der Stelt, 1988). De là, trois patrons de production donnent un effet de *syllabification* (Koopman-van Beinum et van der Stelt, 1986): (i) la phonation interrompue sans mouvement articuloire, (ii) une phonation continue ou interrompue combinée à un mouvement articuloire, et (iii) une phonation continue ou interrompue combinée à plusieurs mouvements articuloires.

a. Un modèle de développement par stade pour la coordination oro-laryngée

En utilisant le formalisme ci-dessus dans le suivi longitudinal de deux individus, Koopman-van Beinum et van der Stelt (1986) ont pu mettre en évidence un décours développemental de la coordination oro-laryngée³⁷ dans la première année de vie (i) *ordonné*, au niveau temporel, et (ii) suivant un jeu *d'emboîtements hiérarchisés* du contrôle des sous-systèmes de production, au niveau fonctionnel. Bien qu'elles considèrent toutes les productions précoces comme étant des précurseurs de la parole (Koopman-van Beinum et van der Stelt, 1998), dans l'étude de 1986, l'analyse se base uniquement sur les sons dits "*comfort*" qui sont non-végétatifs et ne correspondent pas à l'expression d'un stress (eg. pleurs).

Cinq changements qualitatifs dans les productions précoces (ou **repères** pour "*milestones*") ont permis de distinguer 6 stades, initiés chacun à l'apparition d'un nouveau comportement au niveau phonatoire et/ou articuloire (et/ou psycholinguistique)³⁸ (Figure 1-6 et Tableau 1-III):

- Stade I (0 sem.). La *phonation est continue* et de *patron prosodique simple*, et il n'y a *pas de mouvement articuloire*, à l'intérieur de l'UR. Au cours des premières semaines de vie les vocalisations sont donc presque exclusivement générées par l'activité du larynx, menant à la production de sons aspirés (type [h]) et d'occlusions glottales (type [ʔ]).
- Stade II (6 sem. □ 1,4 mois). La *phonation interrompue* apparaît à l'intérieur de l'UR. Autrement dit, l'individu est capable d'initier et d'arrêter la phonation, produisant ainsi des séries rythmiques d'occlusion glottale, initialement sans beaucoup de variation d'intensité et d'intonation.
- Stade III (10 sem. □ 2,3 mois) : L'individu commence à produire *un mouvement articuloire unitaire* au sein de l'UR, cette action étant *combinée à une phonation continue* (issue du stade I) *ou interrompue* (issue du stade II). Des aspects du *filtre* sont

³⁶ Cette variable n'est pas incluse dans l'analyse de base, mais peut être ajoutée au gré des besoins de l'étude.

³⁷ Le terme de coordination oro-laryngée inclut aussi ce que nous avons appelé plus haut coordination pulmo-laryngée.

³⁸ L'âge mis entre parenthèse donne une indication de l'âge de début du stade, exprimé en semaine (sem.). De plus, le modèle de développement par stade de ces auteurs décrit aussi l'émergence de l'association entre vocalisation et sens à partir de l'interaction parent-enfant. Toutefois, cette dimension est ici mise de côté (comme bien d'autres) par souci de clarté: cela ne signifie en rien qu'elle est moins importante à la parole que celle du développement du contrôle (sensori-)moteur.

explorés, les lieux d'articulation passant de postérieur à antérieur. Ce stade correspond dans la culture classique à la période des *coos*.

- Stade IV (20 sem. □ 4,6 mois) : C'est le début de la *diversification des patrons de phonation* en termes d'intonation, de durée et d'intensité, bien qu'il semble y avoir une régression en surface. En effet, l'on observe une diminution marquée de la fréquence des mouvements articulatoires (issus du stade III) et une augmentation de celle de la phonation continue, sans articulation (issue du stade II). D'après les auteurs, l'on retrouve ce qui est classiquement défini comme étant la période de *jeu vocal*, qui peut être vue comme la "pratique" (*training*) intense des possibilités de la *source*.
- Stade V (26 sem. □ 6 mois) : Il y a apparition de *mouvements articulatoires multiples, de type répétés et variés* dans l'UR, combinés à une phonation continue ou discontinue (combinaisons issues du stade III), avec une grande variété des patrons de phonation (diversité issue du stade IV). L'on retrouve le **babillage**, que van der Stelt et Koopman-van Beinum (1986) définissent comme étant une "répétition ou une combinaison de mouvements articulatoires au cours de la phase d'expiration d'un unique cycle respiratoire avec une phonation continue ou interrompue"³⁹ (p. 164). D'après les auteurs, une pratique d'aspects du filtre suivie de celle d'aspects de la source est nécessaire à l'émergence de la coordination de ces comportements: le babillage correspond à une *convergence* de capacités de variation sur la source et sur le filtre, donnant lieu à toutes les caractéristiques basiques de la parole adulte. L'âge d'apparition de ce patron comportemental montre un décours temporel de même forme que celui d'autres fonctions motrices (eg. s'asseoir, ramper, se mettre debout), au niveau populationnel (van der Stelt et Koopman-van Beinum, 1986)⁴⁰. De plus, l'initiation du babillage ne semble pas être indépendante du développement des autres activités motrices de l'individu: par exemple, un babillage tardif peut être associé à la précocité de l'émergence d'autres comportements (van der Stelt et Koopman-van Beinum, 1986). Nous désignerons ce phénomène de liaison potentielle entre le développement de différents comportements par "**compromis développemental**", par analogie avec la notion de **trade-off** en biologie évolutive⁴¹.
- Stade VI (40 sem. □ 9.3 mois) : Il y a émergence de l'*usage référentiel* de tous les types d'expression des stades antérieurs souvent pour désigner, dans des situations spécifiques, des événements. Il s'agit des **proto-mots**, dont la forme est reconnue et interprétée par des membres de l'entourage qui peuvent les utiliser à leur tour (voir Koopman-van Beinum et van der Stelt, 1998, pour des références).

Il est important de préciser que, ci-dessus, l'apparition d'un comportement n'implique pas la disparition de ceux déjà expérimentés (Koopman-van Beinum et van der Stelt, 1998): en général, la fréquence des comportements nouveaux est inférieure à celle des connus (voir Figure 1-6).

³⁹ "[R]epetition or a combination of articulatory movements during one single expiration cycle with interrupted or continued phonation."

⁴⁰ Pour information, van der Stelt et Koopman-van Beinum (1986) trouvent que l'âge de début du babillage varie entre 18 et 48 semaines, pour une moyenne de 31 ± 6.5 sem. (N=51).

⁴¹ La notion de **trade-off** en biologie évolutive réfère au principe selon lequel les ressources, i.e. temps et énergie, de l'individu sont finies et que, par conséquent, celles allouées à une fonction deviennent indisponibles aux autres fonctions. Pour illustration, ce principe peut rendre compte notamment de la liaison entre taille de la tige et réserve d'une plante: la sélection artificielle de blés à gros grains va de pair avec le raccourcissement de leur pied. Autre exemple, les ressources allouées à la progéniture et la taille de cette progéniture semblent liées: les espèces ayant une grande progéniture (beaucoup de petits) tendent à moins s'en occuper. Dans le même ordre d'idée, il est difficile pour les aviaires d'avoir beaucoup de gros œufs: il semblerait qu'il faille "choisir" entre leur grand nombre et leur grande taille.

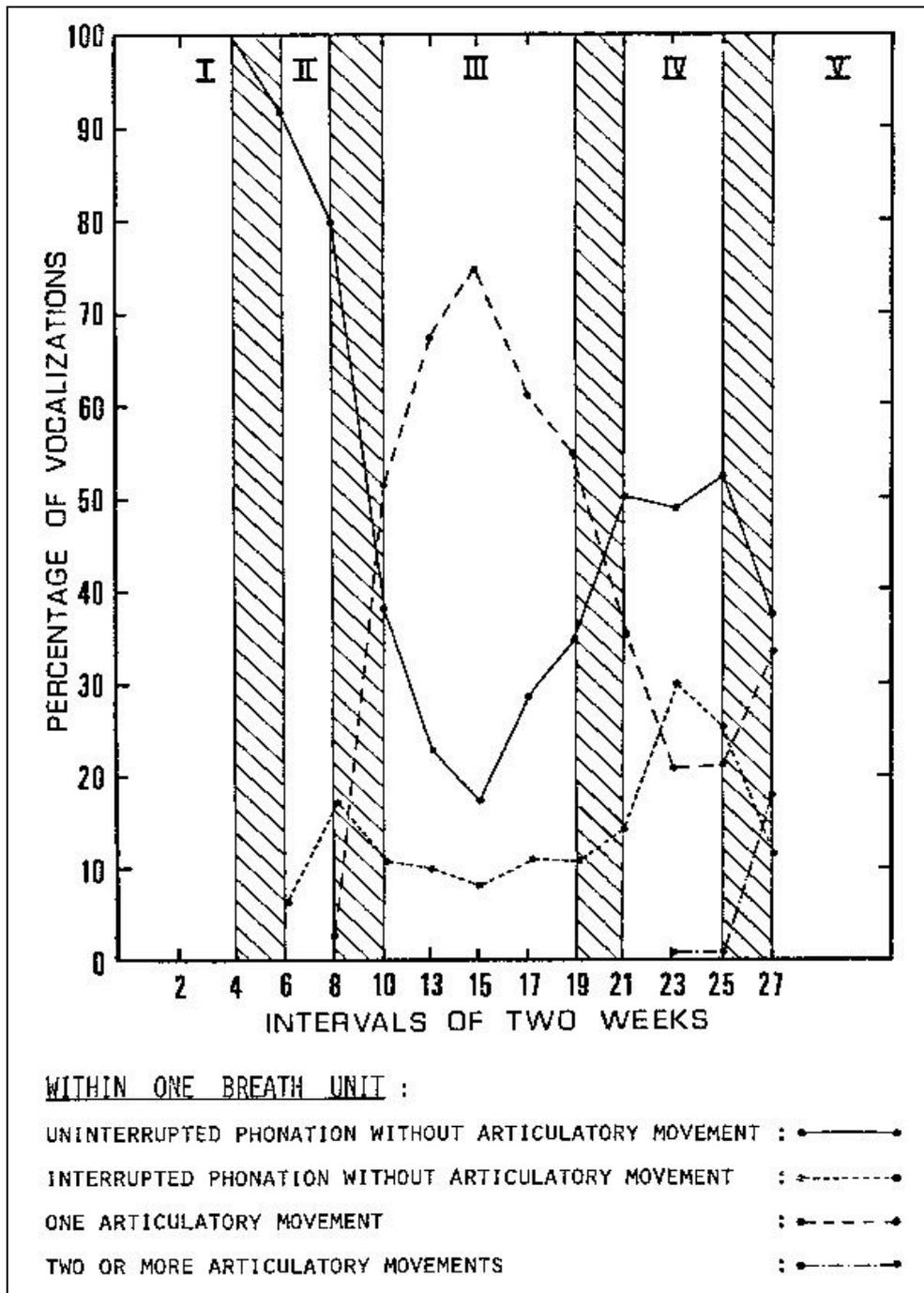


Figure 1-6: Résumé des données phonatoires et articulatoires relevées sur un sujet de l'étude de Koopman-van Beinum et van der Stelt (1986, p. 43), dans ses six premiers mois de vie. La douzième semaine est manquante. Les aires hachurées indiquent le passage à un nouveau stade.

Tableau 1-III: Description phonatoire et articulatoire des productions sonores de deux individus dans les six premiers mois de vie, selon Koopman-van Beinum et van der Stelt (1986, p. 40). Les chiffres 1, 2 et 3 correspondent aux patrons de productions qui donnent un effet de syllabification.

AGE IN WEEKS	01-05	06-09	10-14	15-19	20-26	27-
<u>PHONATION:</u>						
HIGH FUNDAMENTAL FREQ.	X					
CREAKY	X	X				
SHORT	X	X	X	X	X	X
FLAT	X	X	X	X	X	X
FALLING INTONATION	X	X	X	X	X	X
ASPIRATED		X	X	X	X	X
RISING INTONATION		X	X	X	X	X
GLOTTAL STOPS IN SERIES		⊗1	X	X	X	X
COMPLEX INTONATION			X	X	X	X
VARIATIONS IN LOUDNESS				X	X	X
VARIATIONS IN DURATION					X	X
RHYTHMIC SERIES						⊗3
<u>ARTIC. POSITION:</u>						
PHARYNX	X	X				
RAISED VELUM		X				
CLOSED MOUTH		X				
<u>ARTIC. MOVEMENT:</u>						
			⊗2	X	X	X
PHARYNX			X	X	X	X
VELUM			X	X	X	X
UVULA			X	X	X	X
BACK OF THE TONGUE			X	X	X	X
LOWER JAW			X	X	X	X
MIDDLE OF THE TONGUE				X	X	X
TONGUE TIP				X	X	X
LIPS				X	X	X
REPETITIVE ARTIC. MOV.						⊗3

La Figure 1-7 résume ce modèle de développement par stades: les flèches épaisses donnent l'ordre de leur succession temporelle, les flèches en pointillé celui de leur succession fonctionnelle. Ces dernières font ressortir le caractère *hiérarchique* et *emboîté* de l'émergence du contrôle des différents mécanismes physiologiques en jeu dans la coordination oro-laryngée. Il vient que la discontinuité entre babillage et premiers mots postulée par Jakobson (section I.1.a., point *Jiii*) cède la place à une *continuité des productions préverbaux du premier souffle aux premiers mots...*

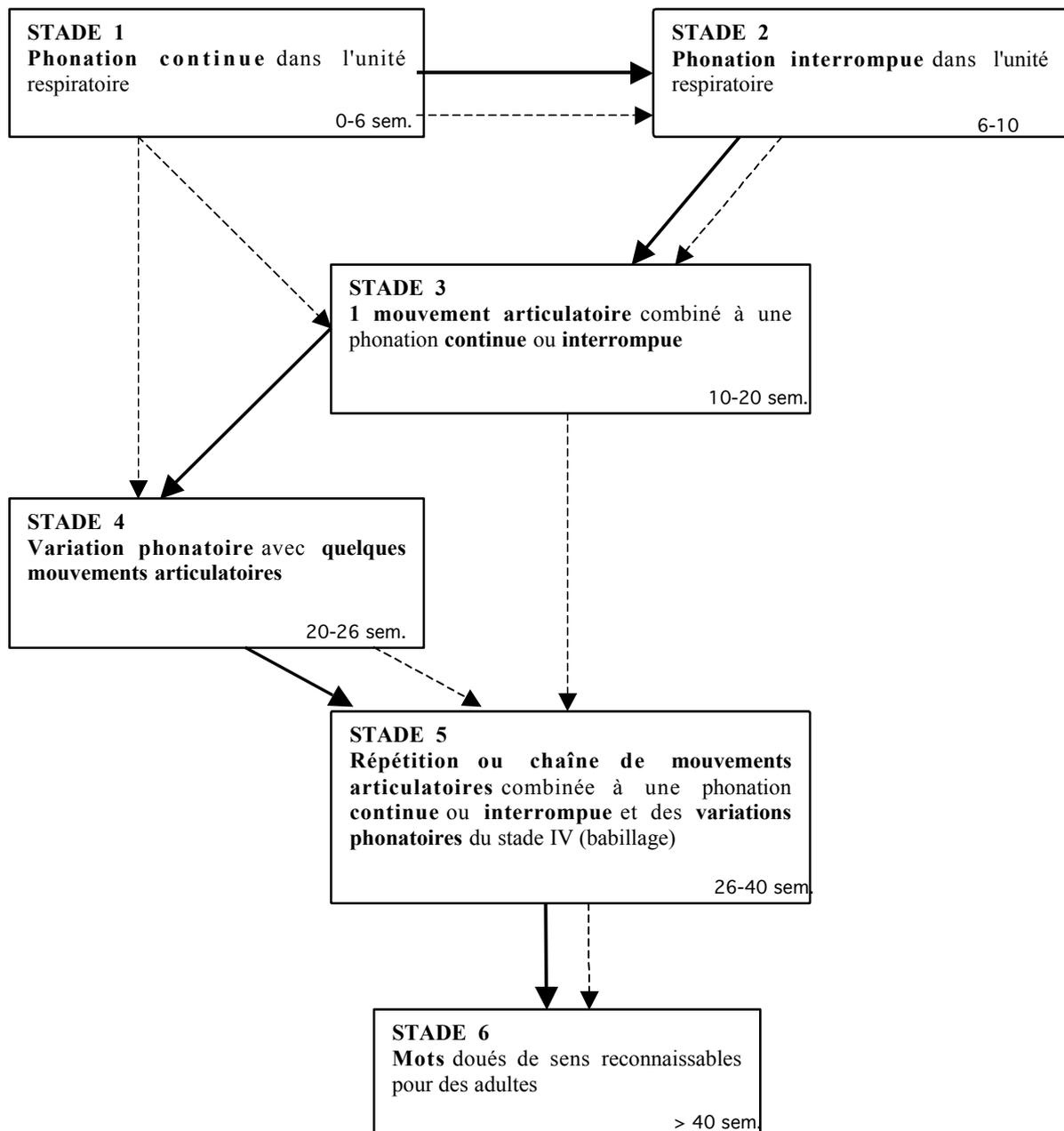


Figure 1-7: Diagramme résumant le modèle de développement par stades de la coordination oro-laryngée de Koopman-van Beinum et van der Stelt (1998, 1986). Les flèches épaisses donnent l'ordre de leur succession temporelle, les flèches en pointillé celui de leur succession fonctionnelle.

Le formalisme proposé par Koopman-van Beinum et van der Stelt est remarquable au moins (i) pour sa capacité à avoir mis en évidence un décours développemental de la coordination oro-laryngée cohérent avec les résultats de leurs prédécesseurs et avec ce que l'on sait des mécanismes de production de la parole, et (ii) pour avoir donné une description non-ambiguë du mode de production des différentes vocalisations précoces, permettant ainsi de faire ressortir une continuité dans le développement de la production verbale. Toutefois, à notre connaissance, ce formalisme n'a pas encore été exploité pour décrire dans le détail l'émergence du contrôle et de la coordination des différents articulateurs du système articulatoire. En particulier, il n'y a pas de description du lien spatio-temporel des mouvements articulatoires dans l'UR des productions du babillage. Or, le babillage semble être un tournant dans l'émergence du contrôle du système articulatoire, comme nous allons le voir par la suite.

b. Des coordinations motrices intra-utérines

D'après Koopman-van Beinum et van der Stelt (1986), "[l]a communication verbale humaine se développe dès le jour de la naissance, [et] probablement même avant"⁴² (p. 37). Les données relatives au développement intra-utérin humain sont rares, mais quelques études trouvent la présence de mouvements coordonnés, en particulier entre les systèmes articulatoire et respiratoire (pour des références voir Butterworth, 1999, et Harris et Butterworth 2002, pp. 69-70).

En effet, à 9-10 semaines de gestation⁴³, des fœtus produisent des séries de hoquets, i.e. des contractions abruptes du diaphragme. De plus, l'on peut observer l'*ouverture de la mandibule et des mouvements de langues* en accompagnement de mouvements de tête, mais aussi des *comportements de succion*.

Aux environs de 10-11 semaines, des *mouvements respiratoires* sont relevés⁴⁴: il s'agit de patrons de mouvements réguliers du diaphragme, du thorax, et de l'abdomen, *parfois en combinaison avec l'ouverture de la mandibule* et la déglutition de liquide amniotique. La coordination entre respiration et mouvements de la mâchoire peut aussi prendre la forme de *bâillements*.

A 15 semaines, la fréquence des mouvements est estimée à 20 000 par jour: ils comprennent, notamment, des *mouvements rapides et lents d'ouverture-fermeture de la mandibule*.

A 24-25 semaines, l'on observe des *comportements de succion du pouce*, avec des *mouvements mandibulaires répétitifs*, mais aussi la présence d'*expressions faciales* finement modulées.

Bien que parcellaires, ces informations montrent que la coordination au sein et entre des systèmes en jeu dans la production verbale a une histoire *antérieure* à la naissance. Même si les comportements sus-cités pourraient ne relever que de coordinations de type réflexe, ces synergies n'en sont pas moins présentes au cours du développement *intra-utérin*, et sont donc vraisemblablement le substrat à partir duquel des comportements moteurs, dont le développement est décrit par le modèle de Koopman-van Beinum et van der Stelt comme *menant à la parole*, pourraient émerger par des phénomènes de différenciation. De là, le continuum s'étend du premier souffle d'air à celui de liquide.

⁴² "[H]uman speech communication develops from the very day of birth onwards, possibly even before."

⁴³ Dans cette section, tous les âges correspondent à l'âge de gestation.

⁴⁴ Ici, le fluide en jeu est le liquide amniotique et non l'air.

II. BABILLAGE CANONIQUE ET THEORIE FRAME/CONTENT

La théorie Frame/Content (F/C) propose que la parole émerge, aux niveaux phylogénétique et ontogénétique, à partir de la cyclicité mandibulaire, qui prend la forme du babillage canonique chez les humains et de gestes de communication orofaciale chez des primates non-humains. Cette thèse s'appuie sur la convergence de données issues, notamment, de la phonétique du développement, de la psycholinguistique, de la neuropsychologie et de la primatologie (MacNeilage, 1998; MacNeilage et Davis, 1990a). Toutefois, dans cette section, notre attention se portera sur le versant *développemental* de cette théorie à portée fédératrice⁴⁵.

Un élément fondamental pour les auteurs: en des termes articulatoires, le **Babillage Canonique (BC)** est *une cyclicité endogène de la mandibule accompagnée de phonation*, qui fait partie intégrante du développement moteur de tout l'organisme⁴⁶. En particulier, le babillage canonique fait parties des différentes **rythmies**⁴⁷ observables dans la première année de vie (voir Thelen, 1981, et Thelen et Smith, 1995 pour plus d'informations sur les rythmies).

1. A un niveau périphérique

a. Du BC comme précurseur de la parole

Dans cette vision, le BC correspond au *précurseur* de la parole. En particulier, la cyclicité mandibulaire est vue comme un **cadre (Frame)**, qui contraint la production, parce qu'elle impose ses phases d'ouverture-fermeture aux sons des séquences de babillage et que cela se retrouve (plus tard) dans la syllabe: en phase d'ouverture, le son est vocalique, en phase de fermeture, il est consonantique. De plus, dans la parole mature les sous-unités syllabiques sont *contrôlées indépendamment du cadre*: l'on parle de **contenu segmental (Content)**⁴⁸. Plus précisément, chez l'adulte, le contenu, ou réalisations phonémiques, est généré par des mouvements de la mandibule, des lèvres, de la langue et du velum qui sont actifs indépendamment les uns des autres au cours de la séquence verbale. A l'opposé, au début du babillage canonique, le seul articulateur significativement actif est la mandibule parce que les mouvements des articulateurs qu'elle porte, i.e. la langue et la lèvre inférieure, ainsi que le velum et la lèvre supérieure, sont vraisemblablement passifs. Ainsi, la seule différence articulatoire significative entre les parties vocaliques et consonantiques contiguës dans les séquences de début de babillage seraient-elles la position haute *versus* basse de la mandibule.

Il vient qu'une séquence de babillage donnée peut être vue comme une forme statique, ou **presetting**, du conduit vocal à laquelle (seule) l'oscillation mandibulaire se superpose. Au niveau phonétique, ce phénomène se traduirait par un même lieu d'articulation pour les vocants et closants contigus, vu que, en particulier, la langue ne devrait pas changer de position suivant l'axe avant-arrière au cours du cycle. Il en résulterait une prépondérance des

⁴⁵ Le versant phylogénétique de la théorie concerne, entre autre, l'hypothèse selon laquelle l'aire homologue à celle de Broca, d'abord vouée au contrôle de processus d'ingestion, aurait été recrutée par les mécanismes de communication orofaciale de certains Primates non-humains (voir chapitre 3, section IV).

⁴⁶ Le caractère *cyclique* du babillage s'ajoute donc à la définition qu'en donnent Koopman-van Beinum et van der Stelt (section I.3.b).

⁴⁷ D'après Streri (1998), les **rythmies** "sont des conduites originales, stéréotypées, observables dès le deuxième jour après la naissance et exécutées de manière répétitive dans un délai temporel de l'ordre de la seconde. Plus de 40 formes sont distinguées qui occupent jusqu'à 40% du temps de la vie du BB durant la première année. L'ouverture-fermeture de la main, l'extension-flexion du bras ou de la jambe, les coups de pieds, les roulements de tête, les balancement du tronc sont les plus remarquables." (p. 31).

⁴⁸ Cela renforce l'idée qu'il faille utiliser des termes distincts pour les productions adultes (voyelle et consonne) et les productions pré-verbales (vocant et closant), comme proposé par Martin (1980, voir section I.1.c.).

patrons de **cooccurrence** (ou types de cadre) suivants dans les répertoires de début de babillage:

(i) Si la langue est en avant, le closant est coronal⁴⁹ et suivi d'un vocant antérieur. Un exemple de réalisation serait [nɛ̃]. Il s'agit du **cadre antériorisé** (*fronted frame*).

(ii) Quand la langue est vers l'arrière, le closant est palatal ou vélaire et couplé à un vocant postérieur. La réalisation est du type [gu]. Cette structure est le **cadre postériorisé** (*backed frame*).

3) La langue en position de repos associe un closant labial à un vocant central. Cette configuration articulaire peut produire [mã]. On parle de **cadre pur** (*pur frame*).

De plus, puisque la configuration de repos de la langue est le cas idéal de sa passivité, les auteurs prévoient que le type de cadre le plus fréquent doit être le "cadre pur" en début de babillage (MacNeilage et Davis, 1990b). A cela s'ajoute le fait que le velum étant baissé au repos, si le closant est nasal ([m n]), le vocant aura tendance à être nasalisé ([ã ẽ]), et l'on parle de **cadre nasalisé** (*nasalized frame*).

Des résultats à base de transcription en accord avec les prédictions

Ces hypothèses de cooccurrence sont globalement confirmées par Davis et MacNeilage (1995) en utilisant la méthode de transcription phonétique, et par Brosda, Laboissière et Abry (2001) sur la base de données CHILDES. De plus, Matyear, MacNeilage et Davis (1998) trouvent une tendance à la nasalisation des vocants associés aux vocants nasals. Il vient que:

- (i) les closants de lieu supra-glottique et les vocants, qui sont traditionnellement inventoriés indépendamment les uns des autres dans la littérature (section I.1.c), sont vraisemblablement *liés* aux niveaux articulaire et temporel via le mouvement mandibulaire au moins au début du babillage.
- (ii) les répertoires de BC montrent des *préférences* en termes d'associations possibles entre types de réalisations contiguës, en désaccord avec le postulat *Ji* de Jakobson (section I.1.a.; voir aussi MacNeilage et Davis, 2001).

De plus, en ligne avec Stoel-Gammon (1995) qui montre que les closants majoritaires du babillage le sont aussi dans les premiers mots (voir section I.1.c.), Davis et MacNeilage (2002a) trouvent que les cooccurrences prépondérantes dans les proto-mots sont les mêmes que dans le babillage. Le postulat de discontinuité de Jakobson (*Jiii*) n'est donc vérifié ni au niveau pseudo-segmental, ni au niveau pseudo-syllabique.

Toutefois, sans rien enlever à la valeur des propositions et résultats avancés par Davis et MacNeilage, nous tenons à rappeler que si la qualité speech-like des productions précoces trouve sans doute ses débuts saillants dans le BC, cela n'implique pas que le BC émerge d'une *tabula rasa*. En effet, comme en témoigne le développement de la coordination orolaryngée dans la première année de vie (section I.3.b.), le BC s'inscrit vraisemblablement dans un processus de développement qui commence *avant* lui, et qui est *nécessaire* à son émergence. De là, le babillage canonique ne peut être *le* début du développement de la production de la parole.

⁴⁹ L'articulation des closants coronaux a lieu dans les zones dentale, alvéolaire et/ou post-alvéolaire.

b. Un test des cooccurrences par la modélisation

Les hypothèses de cooccurrences de Davis et MacNeilage ont été testées par Abry, Badin, Vilain, Stéfano et Boë (1997) sur trois modèles articulatoires anthropomorphiques, Bergame, Gentiane⁵⁰ et SMIP qui sont basés sur les productions de trois adultes différents. A partir de la configuration articulatoire de 10 voyelles prototypiques, les auteurs ont déterminé le lieu d'occlusion du conduit vocal en ne montant que la mandibule de ces modèles.

Leurs résultats (Tableau 1-IV) montrent que les positions antérieure et centrale de la langue peuvent générer des labiales, des coronales et des palatales. Sur les trois modèles, la position arrière de la langue n'a donné lieu qu'à des occlusions labiales et à aucune articulation postérieure.

Tableau 1-IV : Récapitulatif des résultats du test des cooccurrences sur les modèles Bergame, Gentiane et SMIP. Légende: **coro**: C. coronale. **labi**: C. labiale. **palat**: C. palatale. #: absence d'occlusion. NT : Non testé par les auteurs.

VOYELLES											
	i	e	ɛ	y	ø	œ	ə	a	u	o	ɔ
BERGAME	coro	coro	palat	labi	labi	labi	NT	coro	labi	labi	labi
GENTIANE	coro	#	#	labi	NT	labi	labi	labi	labi	#	#
SMIP	coro	coro	coro	labi	NT	coro	palat	coro	labi	labi	#

V. Antérieures
V. Centrales
V. Postérieures

Par conséquent, la détermination du type de cadre doit au moins tenir compte de *la morphologie du système effecteur* (voir Vilain, 2002, pour plus de détails). Nous reviendrons sur ces simulations au chapitre 4.

2. A un niveau central

Dans la théorie F/C, la séparation entre cadre et contenu trouve son corrélat au niveau cérébral.

En effet des études en neuropsychologie suggèrent, notamment sur la base d'une double dissociation chez des adultes (voir MacNeilage, 1998, pour des références), que la cyclicité mandibulaire serait sous le contrôle du Système Prémoteur Médian (*SPM*), incluant le cortex cingulaire antérieur⁵¹ et l'Aire Motrice Supplémentaire⁵² (*AMS*). Ce système cortical est

⁵⁰ Bergame et Gentiane sont élaborés selon un mode de construction similaire à celui de SMIP (voir chapitre 2, section I).

⁵¹ Le cortex cingulaire antérieur fait partie du système limbique, impliqué dans les phénomènes motivationnels. Il se projette sur les ganglions de la base et le cervelet qui sont des structures recrutées pour le contrôle du mouvement volontaire (Purves, Augustine, Fitzpatrick, Katz, Lamantia et McNamara, 1997/1999, pp. 338-340 et pp. 518-519).

⁵² L'AMS joue un rôle dans la planification, la commande et le guidage des mouvements volontaires. En particulier, elle est active lors de l'exécution et la simulation de séquence d'actions (Purves *et al.*, 1999, pp. 291-340 et pp. 316-319). C'est une « aire du langage » au même titre que celles de Broca et de Wernicke (voir aussi chapitre 3, section IV).

impliqué dans la production de mouvements *autogènes* et fonctionnerait sur un mode *prédictif* (voir Grèzes, 2000).

De la même manière, le contenu phonologique de la syllabe (environnement-dépendant), serait géré par le Système Prémoteur Latéral (SPL), où l'on trouve en particulier les aires de Broca⁵³ et de Wernicke⁵⁴. Ce réseau expliquerait, au moins en partie, notre capacité d'*apprentissage par imitation*, et serait recruté pour la production de mouvements *exogènes* (ou stimulus-induits). Son mode de fonctionnement est de type *responsif* (voir Grèzes, 2000, et aussi chapitre 3, section IV).

Un chemin développemental du cadre au contenu

Nous avons vu que, pour la théorie F/C, en début de babillage canonique, seul le cadre est présent. Les auteurs proposent que le contenu segmental apparaît lors du passage d'*une phase endogène à une phase exogène*, où la première serait sous la dépendance du système prémoteur médian et la seconde sous celle du système prémoteur latéral.

D'après Schwartz *et al.* (2000), Cette transition du cadre au contenu trouve appui dans le fait que les connexions dendritiques de Broca à l'aire motrice primaire orofaciale ne se développeraient qu'entre 12 et 15 mois (Greenfield, 1991; voir aussi Abry, Boë, Laboissière et Schwartz, 1998), soit vers l'émergence des premiers mots (qui sous-tendent l'émergence de la distinction de sens qu'opèrent les phonèmes)⁵⁵.

Cependant, le rôle vraisemblable de l'aire de Broca dans l'acquisition de la phonologie ambiante n'exclut pas l'intervention d'autres structures cérébrales dans les capacités d'imitation orofaciale précoce (voir chapitre 3, section III.A.3.). En effet, des travaux suggèrent que, *bien avant* le babillage, l'on est capable de donner une équivalence motrice à des gestes de paroles audibles et/ou visibles perçus sur autrui. Ainsi l'imitation orofaciale précoce a-t-elle droit à une place dans l'ensemble des mécanismes potentiellement à la base de l'émergence d'une phonologie.

Autrement dit, le développement d'une phonologie, vraisemblablement sous-tendue par des capacités d'imitation et qui requiert un minimum de coordination des mouvements, notamment, mandibulaires, labiaux et linguaux, n'implique pas que les phénomènes d'imitation orofaciale, observés aux périodes antérieures au babillage canonique, ne contribuent pas à l'émergence de cette phonologie. En particulier, l'on peut proposer que les presettings auxquels se superpose la cyclicité mandibulaire du babillage sont influencés par l'apprentissage imitatif qui précède le babillage.

Ces remarques plaident en faveur de la *non-séparation des phases* endogène et exogène, pour la prise en considération de leur *chevauchement total ou partiel*. Toutefois, cette nuance n'enlève rien à l'idée fondamentale que deux mécanismes, fonctionnant potentiellement d'abord séparément puis de concert, soient requis au développement de la

⁵³ L'aire de Broca peut jouer plusieurs rôles à la fois. Selon les auteurs, on la trouve impliquée dans la programmation motrice en relation avec la phonologie, le traitement de la syntaxe, la prise en charge de séquences perceptives ou motrices complexes (Schwartz, Abry et Escudier, 2000).

Son homologue potentiel, chez le Macaque, est l'aire F5 où se trouvent une représentation des mouvements de la bouche et du larynx ainsi que des neurones actifs lors de la perception et de l'exécution d'actions manuelles et orofaciale sur des objets (voir chapitre 3, section IV pour plus de détails). Cette donnée est à la base de l'hypothèse selon laquelle F5 jouerait un rôle dans les phénomènes d'imitation, et donc dans l'émergence phylogénétique de l'apprentissage par imitation des gestes de parole.

⁵⁴ L'on considère communément que l'aire de Wernicke est en charge des mécanismes de compréhension auditive du langage.

⁵⁵ Précisons qu'il est difficile de savoir si des zones corticales ont les mêmes fonctions chez les nourrissons et chez les adultes. Toutefois, Dehaene-Lambertz, Dehaene et Hertz-Pannier (2002) trouvent des activations en réponse à des stimuli de parole dans des régions corticales pré-frontales congruentes avec celles de l'adulte, sur des individus de 3 mois.

parole. L'un de type prédictif serait recruté pour une mise en relation de capacités articulatoires du conduit vocal avec leur conséquences perceptives. L'autre, responsif, orienterait les productions de soi vers celles d'autrui. Parallèlement, le premier processus garderait à sa charge l'oscillation mandibulaire d'où émergerait un contenu segmental qui par essence a besoin du second processus. Dans la troisième partie de ce mémoire, nous proposerons une implémentation de ces deux mécanismes, sous les vocables d'"exploration" et d'"imitation".

La théorie F/C propose que le développement de la parole se fait du cadre au contenu. La section suivante clôt ce chapitre sur un scénario pour l'émergence du contenu par le développement de la coordination des différents articulateurs non-mandibulaires à partir de leur dépendance au cadre.

III. DU DEVELOPPEMENT DE LA COORDINATION ARTICULATOIRE

Abry, Cathiard, Vilain, Laboissière, Loevenbruck, Savariaux et Schwartz (2005) proposent que le développement du contrôle du conduit vocal pour la parole suit les mêmes principes que celui du développement du contrôle des segments du corps (e.g. de la tête au train) pour d'autres comportements moteurs (e.g. la locomotion). Les auteurs avancent un *modèle intégratif de l'émergence de la coordination des segments du système articulatoire* à partir de la cyclicité mandibulaire du babillage jusqu'aux mots, dont le but est de rendre compte des données trouvées dans la littérature. Le chemin avancé peut s'énoncer comme la succession des événements suivants.

(i) En accord avec la théorie F/C, pendant le babillage canonique, *seul le cycle mandibulaire est contrôlé* de manière *globale*, tandis que les articulateurs portés (i.e. la lèvre inférieure et la langue) ainsi que la lèvre supérieure et le velum tendent à être passifs (Munhall et Jones, 1998, pour les lèvres, Sussman, Duder, Dalston, et Cacciatore, 1999, pour la langue, et Matyear, *et al.*, 1998, pour le velum). Pour illustration, Munhall et Jones (1998) montrent, sur un individu de 8 mois, qu'au début du babillage canonique, dans la phase de fermeture d'une séquence [baba], la lèvre supérieure se déplace passivement sous l'effet du contact de la lèvre inférieure dont le mouvement semble suivre (passivement) celui de la mâchoire inférieure tandis que, chez l'adulte, la lèvre supérieure résiste au contact de l'inférieure et se lève en phase d'ouverture (Figure 1-8).

Le couplage spatio-temporel du mouvement des lèvres via le cycle mandibulaire du babillage est vérifié de manière statistique sur des sujets de 1 ans par Green, Moore, Higashikawa et Steeve (2000). De plus, lors de la production répétée de séquences CVCV, où C est une bilabiale, et V est [a], le patron cinétique de l'oscillation mandibulaire du babillage à 1 ans est (i) très stable au niveau individuel, et (ii) plus proche de celui de l'adulte que le patron de mouvement des lèvres inférieure et supérieure (Green, Moore et Reilly, 2002). Cela renforce l'idée selon laquelle le cycle mandibulaire est le mouvement dont le contrôle se développe en premier dans le système articulatoire, et que l'alternance de sons consonantique et vocalique du babillage est le simple sous-produit de cette oscillation qui constitue le cadre de la future syllabe.

(ii) Le mouvement indépendant des articulateurs portés vis-à-vis de celui de la mâchoire inférieure, ainsi que le mouvement actif du velum et de la lèvre supérieure, émergent: c'est le début d'un développement progressif du *contrôle local des contacts pendant la phase de closure*, et donc l'apparition d'un premier contenu segmental de type consonantique.

Toutefois, ce développement s'amorcerait de manière séquentielle, i.e. chaque articulateur à son tour. En particulier, l'activation du velum en phase de fermeture semble venir avant l'émancipation des lèvres et de la langue vis-à-vis de la mandibule.

En effet, une des caractéristiques de l'évolution des productions de babillage des individus normo-entendants est la raréfaction des closants nasals ([m n]) en faveur de leur homologue oral ([b d]), d'après leur transcription phonétique (Davis et MacNeilage, 2002b) : le contrôle de l'ouverture-fermeture du port nasal, grâce au mouvement actif de levée du velum, est vraisemblablement le premier à émerger dès le début du babillage (Lalevée et Vilain, 2003). De plus, l'étude de Green *et al.* (2002) montre qu'à 1 et 2 ans le mouvement de la lèvre inférieure est plus proche de celui de l'adulte que celui de la lèvre supérieure, et cela reste vrai à 6 ans, bien que les patrons cinétiques des deux lèvres se soient grandement rapprochés de la référence adulte.

(iii) Un *contrôle de la forme globale* du conduit vocal dans la *phase d'ouverture* apparaît progressivement, permettant l'émergence d'un premier contenu segmental de type vocalique, notamment, grâce à une émancipation des mouvements de la langue vis-à-vis du cycle mandibulaire. En effet, à la même période que les proto-mots, la posture vocalique du conduit vocal semble être de mieux en mieux anticipée dans la phase de fermeture (Sussman, *et al.*, 1999; voir chapitre 4, section IV.2.). Toutefois, ce phénomène reste sous le joug du contrôle local nécessaire à la production du closant. Plus précisément, les formes de conduit vocal semblent plus rapidement anticipées dans le cas des bilabiales, dont la production contraint relativement peu celle du vocant (puisque la configuration de la langue pour le vocant peut être formée en même temps que le pincement des lèvres pour le closant), que dans le cas des alvéolaires et des vélares. Abry *et al.* (2005) reprennent l'idée de Sussman *et al.* (1999) selon qui l'anticipation du vocant dans les labiales nécessite l'intégration de la forme vocalique de langue avec et indépendamment du contact des lèvres. L'hypothèse avancée pour les alvéolaires et les vélares est que (a) l'individu est d'abord incapable de prendre une configuration de langue dans le vocant qui soit indépendante du closant, (b) mais pourrait acquérir par la suite un contrôle moteur plus indépendant des différents degrés de liberté linguaux en utilisant probablement le retour auditif de ses productions et (c) surexploiterait ses capacités de coarticulation fraîchement acquises avant de les ajuster aux normes de la langue ambiante (pour une revue de questions sur le développement de la coarticulation, voir Ménard, 2002).

(iv) Un *contrôle de la forme locale* du conduit vocal *pour la production des voyelles* se développe. En particulier, il permet la production de sons vocaliques nécessitant la formation de petites constriction qui demandent un contrôle fin des articulateurs pour que le flux d'air reste de type laminaire, comme dans le cas de [u] et [y] au niveau des lèvres et de la langue.

(v) Les auteurs ajoutent une dernière étape de contrôle fin de la posture consonantique globale du conduit vocal, pour des contrastes acquis tardivement tels que /ʃ/ vs. /s/.

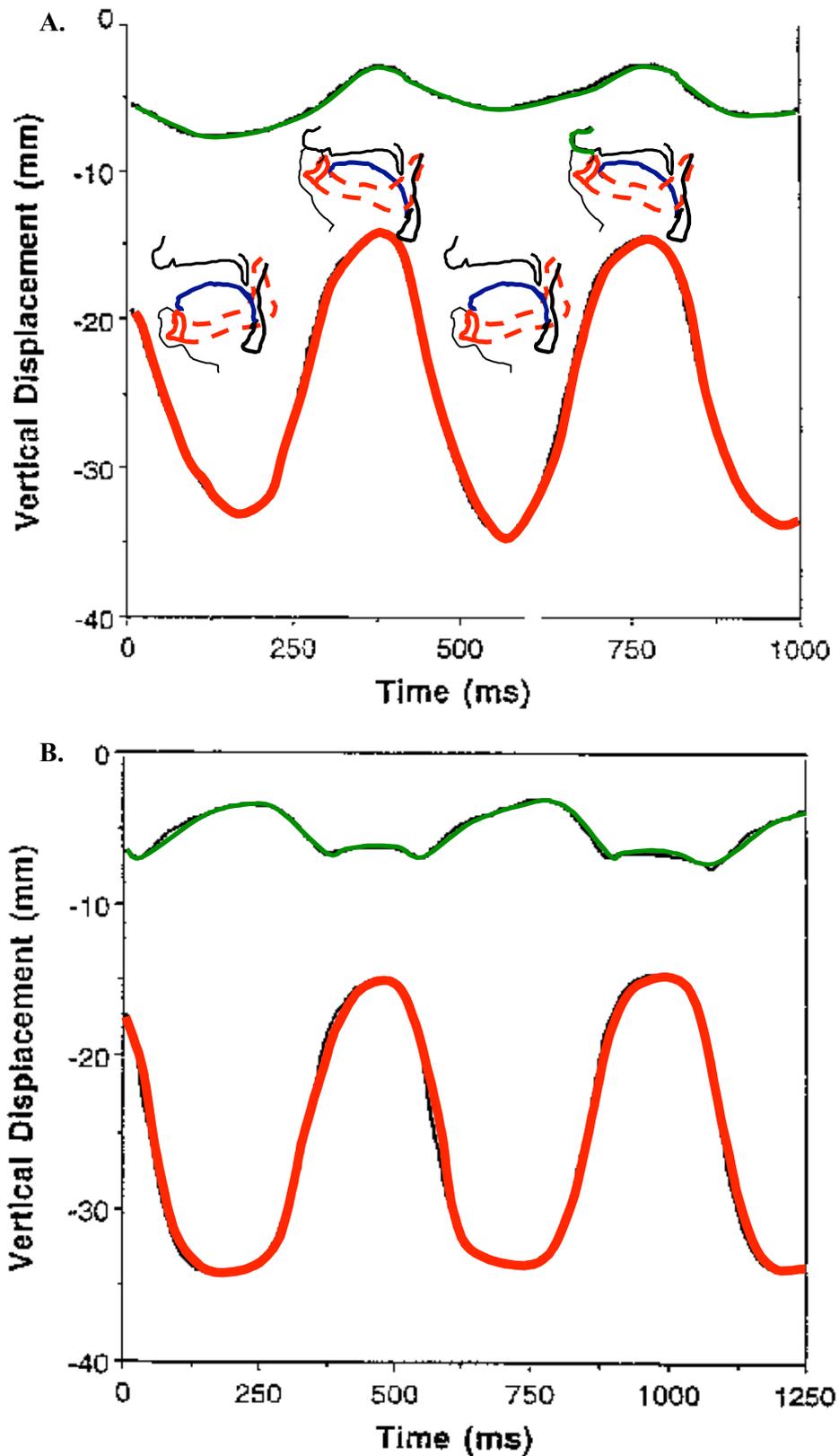


Figure 1-8 : Représentation des mouvements sagittaux des lèvres supérieure (en vert) et inférieure (en rouge) au cours du temps, lors d'une séquence de de type [baba] produite par un individu de 8 mois (A) et un adulte (B) (figures empruntées et modifiées, de Munhall et Jones, 1998).

Conclusion

En conclusion, nous garderons pour chemin de développement de la production de la parole la mise bout à bout des modèles (i) de l'émergence de la coordination orolaryngée de Koopman-van Beinum et van der Stelt, (ii) de la cyclicité mandibulaire comme cadre de la future syllabe de Davis et MacNeilage et (iii) du développement du contrôle du système articulatoire d'Abry, Cathiard, Vilain, Laboissière, Loevenbruck, Savariaux et Schwartz.

Toutefois, dans le cadre de ce mémoire, nous focaliserons nos efforts sur la modélisation de deux mécanismes généraux dont la fonction se retrouve dans les systèmes prédictif et responsif de la théorie Frame/Content: l'exploration et l'imitation. Le babillage canonique étant un tournant dans l'émergence du contrôle du système articulatoire et la période la plus décrite par nos prédécesseurs, il nous servira de repère à l'implémentation (i) du comportement d'imitation orofaciale antérieure au babillage qui contribue vraisemblablement à ses presettings (chapitre 5), et (ii) de la cyclicité mandibulaire, en tant que substrat d'où émerge la coordination des segments du système articulatoire (chapitre 4).

Chapitre 2: Des outils pour une modélisation réaliste du développement de la production de la parole

INTRODUCTION

La robotique cognitive de la parole

La perception et la production de la parole sont souvent étudiées indépendamment l'une de l'autre. Toutefois, la parole est, incontestablement, un *phénomène sensori-moteur*. Cela est le point de départ de la Théorie de la Perception pour le Contrôle de l'Action (ou *PACT*, pour *Perception for Action Control Theory*) (Schwartz, Abry, Boë et Cathiard, 2002), dans laquelle *la perception est vue comme un ensemble d'outils, de traitements et de représentations qui permet de contrôler l'action*. Sachant que les représentations perceptives et motrices émergent en parallèle au cours du développement, la PACT propose qu'elles se contraignent les unes les autres, au moins à l'âge adulte, bien qu'elles puissent relever de domaines cognitifs (et surtout d'investigation) différents. L'idée-phare est que, pour étudier les représentations perceptives et motrices qui sous-tendent la parole adulte et qui mettent en forme les langues du monde, une stratégie pertinente est de se centrer sur la manière dont elles se développent ensemble au cours de l'ontogenèse de la parole.

Dans cette approche, un cadre de modélisation computationnel est fourni par la robotique cognitive, dont l'un des promoteur est Brooks dans le projet Cog, qui se focalise sur les notions "[d']incarnation et [de] couplage physique, [d']intégration multimodale, [d']organisation développementale, et [d']interaction sociale."¹ (Brooks, Breazeal, Marjanovic, Scassellatti et Williamson, 1999).

L'incarnation, la multimodalité, le développement et l'interaction sociale sont aussi au cœur de la "Robotique de la Parole" (Abry et Badin, 1996; Laboissière, 1992), un programme de recherche dans lequel nous essayons:

1. d'élaborer un agent virtuel, vu comme un système sensori-moteur capable d'articuler et de percevoir des gestes de parole (*axe incarnation*, voir Boë, Gabioud, Perrier, 1995a; Schwartz et Boë, 2000), et d'apprendre des liens multisensori-moteurs (*axe multimodalité*, voir Schwartz, Robert-Ribes et Escudier, 1998) en parallèle avec la croissance de son appareil vocal;
2. de déterminer les stratégies potentielles par lesquelles cet agent pourrait passer de la production préverbale au contrôle de gestes de parole complexes (*axe développement*, voir Abry *et al.*, 2005);
3. d'explorer comment des principes de communication dans une société composée de tels agents pourraient modeler les structures acoustiques et articulatoires des langues (*axe interaction*, voir Berrah, Glottin, Laboissière, Bessière et Boë, 1996).

La présente étude est une étape préliminaire à ce programme de recherche. Le but est de donner les bases d'une modélisation du développement de la parole, par l'implémentation d'un agent, en tant que système sensori-moteur en croissance capable d'apprendre et d'interagir

¹ "[...] embodiment and physical coupling, multimodal integration, developmental organization, and social interaction."

(Schwartz *et al.*, 2002). Ce chapitre présente les différentes composantes de l'agent, dont tout ou partie a été utilisé dans nos travaux de thèse.

Dans un premier temps, nous verrons le versant production de l'androïde: un modèle articulatoire fournit à l'agent un conduit vocal virtuel qui intègre la croissance non-linéaire et non-uniforme du conduit vocal humain (voir chapitre 1, section I.2.). Pour ce qui est de la perception, les modalités auditive, visuelle et tactile sont disponibles sous la forme d'un modèle pour chacune. Un troisième volet concerne l'apprentissage et l'utilisation de liens entre les mouvements du conduit vocal et leurs conséquences perceptives: ces processus empruntent le formalisme de la programmation Bayésienne et seront la base computationnelle de la simulation du mécanisme d'imitation.

I. LE MODELE ARTICULATOIRE

Pour évaluer le jeu de configurations de conduit disponibles aux nourrissons pour vocaliser et babiller, nous avons utilisé le *Variable Linear Articulatory Model* (ou *VLAM*) (Boë, 1999), une version du modèle *Speech Maps Interactive Plant* (*SMIP*) (Boë *et al.*, 1995a) qui intègre la croissance non-linéaire du conduit vocal. Pour l'essentiel, le SMIP est issu d'une analyse en composantes principales (ACP) de la description de coupes cinéradiographiques d'un conduit oral adulte en action (Maeda, 1990, locutrice francophone PB). Cette analyse statistique a révélé que 5 facteurs expliquent 92% de la variance totale dans les données observées. Ils s'avèrent être liés à des axes de mouvements concrets, et décrivent les degrés de liberté articulatoires suivants: le mouvement vertical de la mandibule (noté *Jaw*, ou *J*), la protrusion-rétraction du corps de la langue (*Tongue Body*, ou *TB*), la courbure-applatissage de la langue (*Tongue Dorsum*, ou *TD*), le mouvement vertical de la pointe de la langue (*Tongue Tip*, ou *TT*) et l'élévation du larynx (*Larynx*, ou *Lx*). Notons que TB correspond aux axes de contraction des muscles génioglosse postérieur et hyoglosse, tandis que TD correspond à ceux des muscles génioglosse antérieur et styloglosse (Maeda et Honda, 1994; Payan et Perrier, 1997). Un modèle de forme des lèvres (Abry et Boë, 1986) a été adapté au modèle de Maeda, ajoutant ainsi deux degrés de liberté (Figure 2-1): la séparation des lèvres (*Lip Height*, ou *LH*) et la protrusion labiale (*Lip Protrusion*, ou *LP*). Les facteurs de pondération des 7 variables articulatoires ont été normalisés sur la base de données d'où est tiré le SMIP: leur domaine de variation s'exprime en nombre d'écart-types (*e-t*) centré sur 0. Ces paramètres servent d'entrée au modèle pour synthétiser une coupe sagittale et la fonction d'aire associée, à partir de laquelle il est possible d'obtenir la réponse harmonique (ou fonction de transfert), les fréquences formantiques (ou fréquences de résonance) et le signal de parole (Badin et Fant, 1984). Précisons que la coupe sagittale est décrite selon une grille semi-polaire fixe (Figure 2-2). La largeur de chaque section d'une forme de langue donnée est une combinaison linéaire des paramètres de commande linguaux, selon le principe de l'ACP. Puisque les facteurs de l'ACP sont orthogonaux et que les sections de la grille ne sont pas mobiles de par la construction du modèle, les variables articulatoires sont indépendantes les unes des autres et la contribution des différents paramètres articulatoires dans la détermination de la largeur de chaque section est fixe. Il en résulte qu'un recul de la langue (par l'action de TB ou TD) ne modifie pas la contribution relative du paramètre de commande de sa pointe (TT) dans les sections antérieures de la coupe sagittale. De plus, l'extraction de facteurs orthogonaux n'étant jamais "parfaite", le paramètre de pointe de la langue influence la largeur des sections antérieures de la coupe, mais aussi celle des sections proches de la glotte pour des valeurs extrêmes de cette commande. Un dernier point concerne l'adaptation du modèle de lèvres au modèle de langue: l'amplitude du déplacement de la mandibule entraîne un déplacement de moindre amplitude de la lèvre inférieure.

Dans VLAM, le processus de croissance est introduit par deux facteurs d'échelle qui redimensionnent la longueur et la largeur des parties buccale et pharyngale de la coupe sagittale adulte, en interpolant la zone qui les sépare. La variation des deux facteurs d'échelle de la naissance à 21 ans suit un modèle dérivé de mesures crânio-faciales recueillies par Goldstein (1980; voir chapitre 1, section I.2.). De manière similaire, la valeur de la fréquence fondamentale (f_0) varie selon une fonction de l'âge ajustée aux données de Mackenzie Beck (Mackenzie Beck, 1996). Ainsi l'âge du conduit vocal virtuel dont le comportement articulatoire-acoustique est modélisé par VLAM doit-il être spécifié avant usage: ce paramètre fixe la taille des cavités avant et arrière ainsi que la fréquence de la source de voisement qui excite le conduit. La Figure 2-3 récapitule l'architecture fonctionnelle de VLAM. La Figure 2-4 illustre l'effet de la croissance du conduit vocal sur la contraction de l'espace occupé par les fréquences de résonance du conduit vocal exprimées en Hertz à 0 et 21 ans (voir chapitre 1, section I.2.).

Une présentation plus détaillée du modèle ainsi que l'évaluation de sa pertinence vis-à-vis de mesures acoustiques et morphologiques prélevées à des âges variés peuvent être trouvées dans Ménard (2002), Ménard, Schwartz, Boë, Kandel, et Vallée (2002) et Ménard, Schwartz et Boë (2004). En particulier, il a été utilisé pour la conception d'expériences sur la normalisation perceptive des voyelles au cours du développement (Ménard *et al.* 2002).

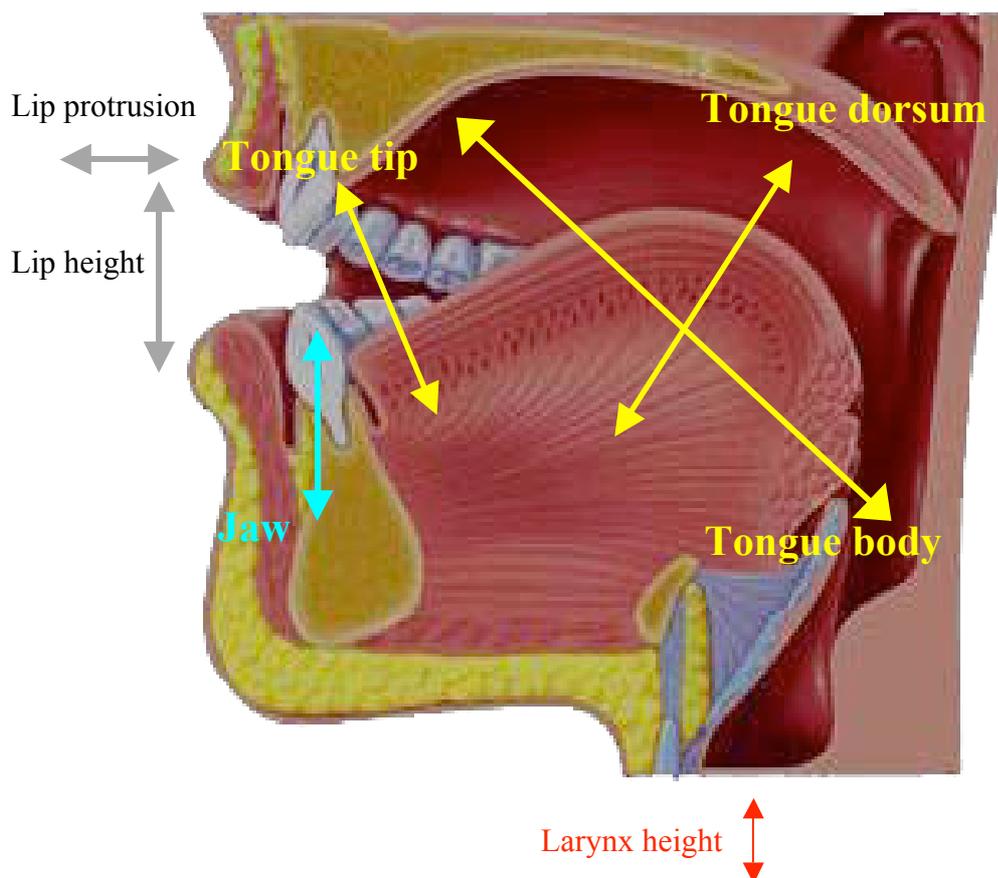


Figure 2-1: Les degrés de liberté du modèle de production SMIP (et VLAM).

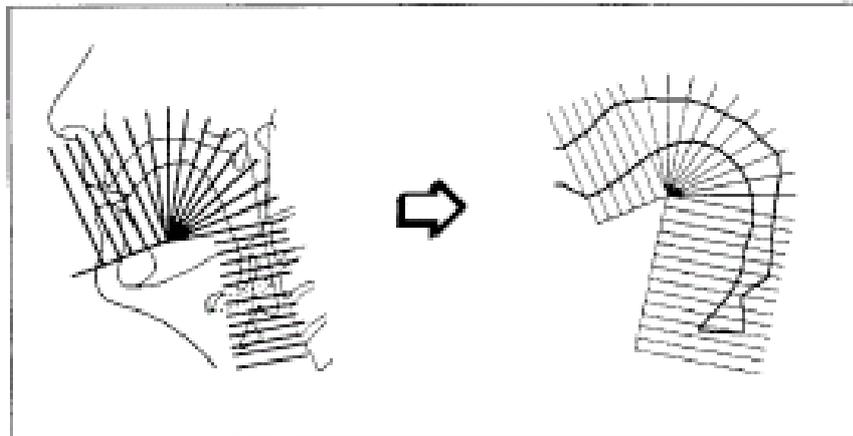


Figure 2-2: Représentation d'une coupe sagittale de conduit vocal à partir de mesure prise sur la grille semi-polaire (figure tirée de Ménard, 2002).

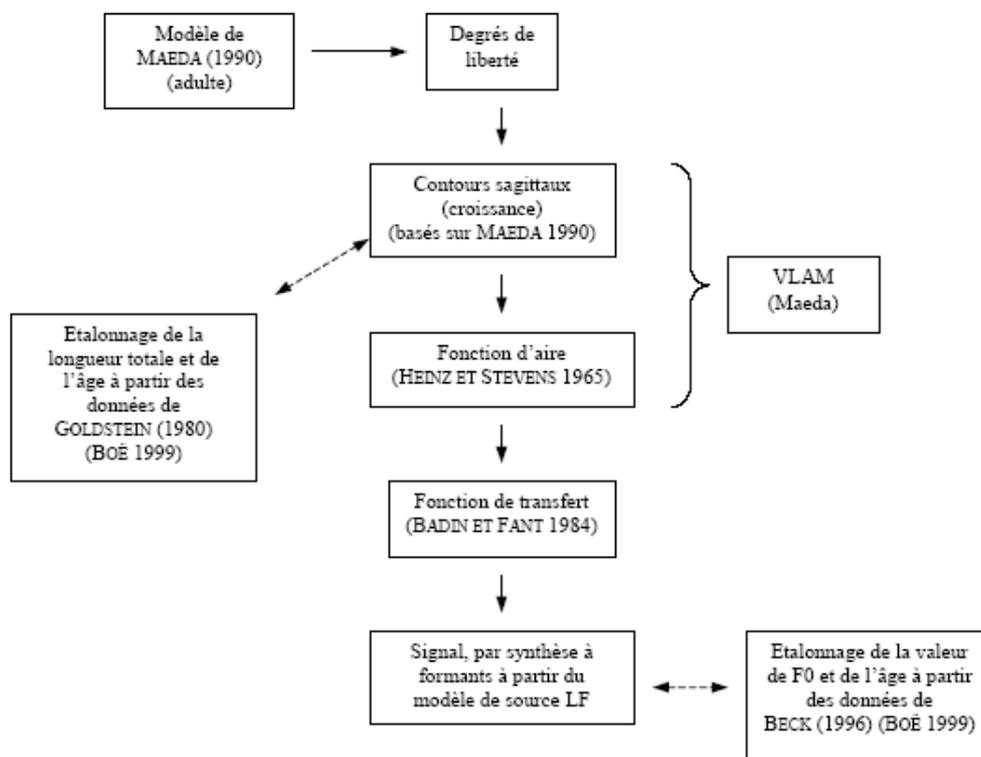


Figure 2-3: Architecture fonctionnelle de VLAM (diagramme issu de Ménard, 2002).

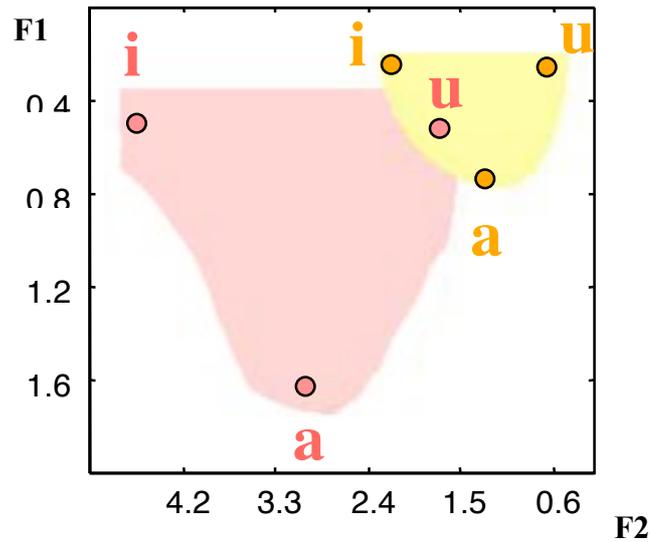


Figure 2-4: Effet de la croissance du conduit vocal sur la contraction de l'espace occupé par les fréquences de résonance du conduit vocal exprimées en kilo-Hertz à 0 (en rose) et 21 ans (en jaune), d'après VLAM.

II. LES MODELES SENSORIELS

1. Le modèle auditif

Le suivi des gestes de parole doit mettre en jeu un moyen de capturer et de caractériser les composantes fondamentales des actions vocales du/de la locuteur/trice, i.e. leur relation temporelle, ou *phasage*, et leur *cible* (Schwartz, Arrouas, Beautemps et Escudier, 1992). Une série de travaux influents, réalisés à l'Institut Pavlov de Leningrad dans les années 70s, ont mené Chistovich à proposer une architecture noyau pour le traitement auditif des sons de parole. Cette structure consiste en deux systèmes, l'un étant spécialisé dans le traitement et la détection d'événements temporels acoustiques dans l'information d'entrée, l'autre dans la fourniture continue d'analyses de son contenu spectral (Chistovich, 1976, 1980). Les bases neurophysiologiques de ces mécanismes sont présentes dès les neurones primaires du nerf auditif, ou dans les neurones secondaires du noyau cochléaire (qui est le premier centre de traitement auditif du système nerveux central) (pour une introduction au fonctionnement du système auditif, voir Romand, 2000). Cet ensemble fournit la base du système auditif de l'agent (Figure 2-5).

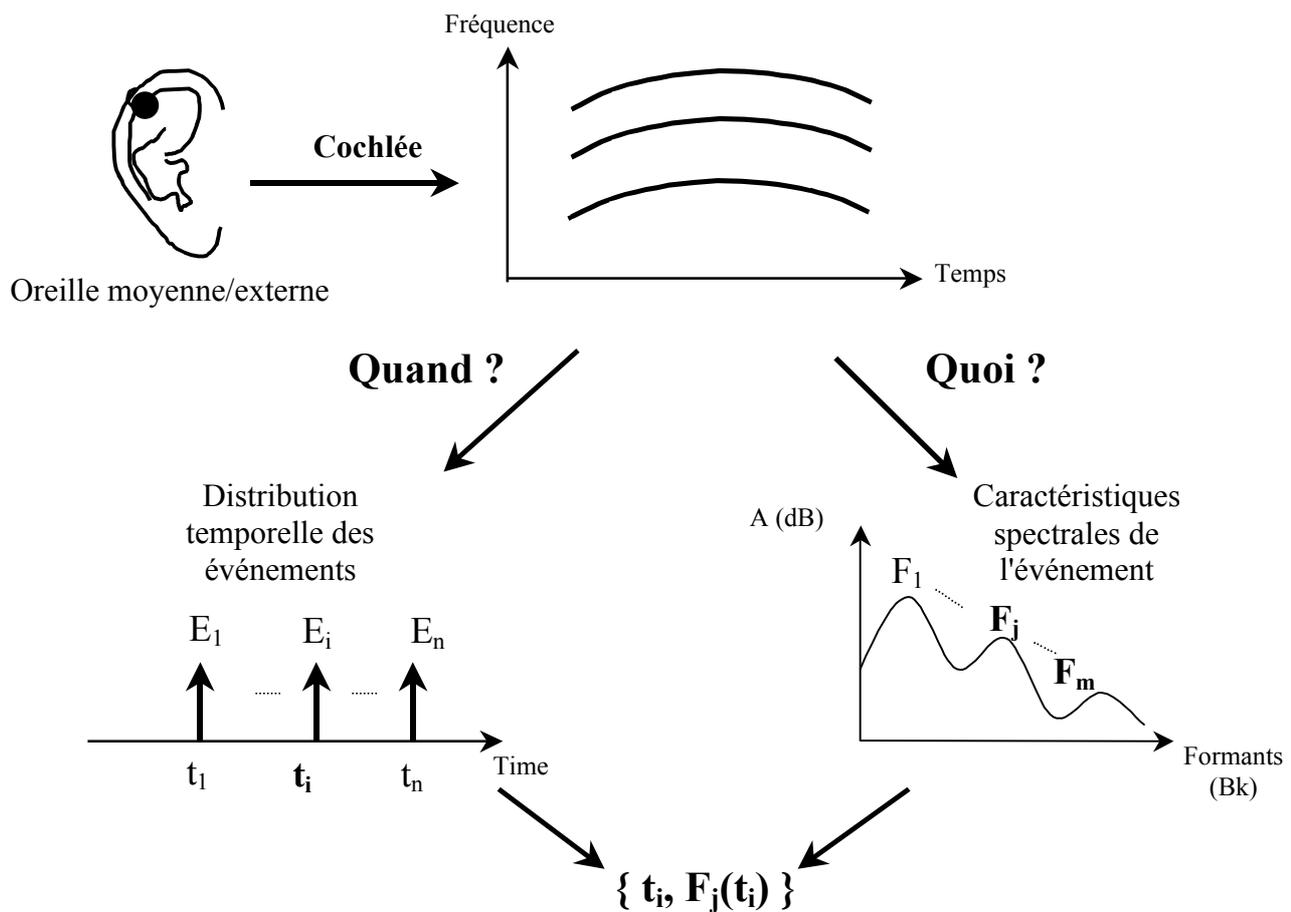


Figure 2-5: L'architecture du système auditif complet de l'agent.

Le système spécialisé dans la *détection d'événements* ("quand?") s'appuie sur des unités dites "phasiques" du système nerveux central, i.e. les cellules "on" et "off" qui répondent uniquement à des variations rapides de la stimulation sensorielle d'une zone donnée du spectre des fréquences. Des membres de l'ICP ont développé un modèle physiologiquement plausible pour la détection d'événements articulatoire-acoustiques dans le noyau cochléaire, tels que le début et la fin du voisement, les bruits de plosion, le début et la fin des productions vocaliques, (Piquemal, Schwartz, Berthommier, Lallouache et Escudier, 1996; Wu, Schwartz et Escudier, 1996). Ces événements, qui permettent l'étiquetage des discontinuités majeures dans le signal de parole, sont cruciaux pour le contrôle du phasage dans la production de la parole (Abry, Benoît, Boë et Sock, 1985; Abry, Orliquet et Sock, 1990).

Le système dédié aux *analyses spectrales* ("quoi?") se base sur des unités dites "toniques": elles répondent de manière continue à un stimulus donné, permettant ainsi d'établir des statistiques et traitements précis sur les variations de stimulation en fonction de leurs caractéristiques fréquentielles. Bien que le débat sur le rôle des formants dans le traitement auditif de la parole soit loin d'être clos (e.g. Bladon, 1982; Pols, 1975), il semble que des ingrédients neurophysiologiques élémentaires soient disponibles pour la détection de formants dans le nerf auditif, par des statistiques spatio-temporelles (Delgutte, 1984), mais aussi à un niveau plus central, dès le noyau cochléaire, par des mécanismes d'inhibition latérale pour le renforcement des contrastes. Ainsi les formants sont-ils les paramètres spectraux de base pour la caractérisation des sons de parole dans l'architecture de notre agent.

2. Le modèle visuel

Pour satisfaire le caractère multimodal de la perception de la parole, l'androïde a autant besoin de la vision que de l'audition. En effet, il est assez bien connu que la parole est non seulement entendue mais aussi vue (e.g. Dodd et Campbell, 1987; Campbell, Dodd et Burnham, 1998). La lecture labiale (i) permet de suivre en partie des gestes de parole en l'absence d'audition, (ii) augmente l'intelligibilité de la parole native en condition audio bruitée et des langues étrangères en condition écologique, (iii) intervient dans le recouvrement du geste, même quand l'entrée visuelle est en conflit avec l'entrée auditive, comme dans le cas de l'effet McGurk (McGurk et MacDonald, 1976), et (iv) serait impliquée dans le développement du contrôle de la parole, et l'acquisition d'une phonologie en conjonction avec le langage parlé complété chez les personnes mal-entendantes (voir Schwartz *et al.*, 2002, pour une revue sur la fusion audio-visuelle dans le contexte de la PACT). Le capteur visuel de l'agent doit être capable de prélever l'information visible sur le visage du/de la locuteur/trice, i.e. la forme des lèvres, la position de la mandibule, et probablement certaines informations sur la configuration de la langue. A ce jour, les entrées visuelles de l'agent sont les caractéristiques labiales de base, i.e. l'écartement, la séparation, l'aire et le degré de protrusion.

3. Le modèle tactile

Pas de parole sans somesthésique

Le **canal oro-sensoriel**, qui comprend les sensations tactiles aux niveaux des lèvres, de la langue et du palais, est souvent absent de la modélisation des représentations perceptives des gestes de parole. Cependant, l'être humain détient, au moins à l'âge adulte, une représentation très développée de l'espace oral. Pour illustration, dans des expériences relatives à la stéréognosie orale, des sujets sont capables d'intégrer des informations tactiles et motrices (donc haptiques) pour identifier les objets tridimensionnels placés dans leur bouche (Bosma, 1967). La pointe de la langue et les lèvres sont des parties parmi les plus sensibles de la surface du corps, comme en témoignent des données sur la discrimination à deux points (Landgren et Olsson, 1982).

La neurophysiologie du système somesthésique oral a été décrite, chez l'adulte, dans plusieurs revues (voir par exemple Hardcastle, 1976; Landgren et Olsson, 1982; Kent, Martinet et Sufit, 1990). La majorité de la muqueuse orale, et tout particulièrement celle de la langue, est douée de différents types de mécanorécepteurs capables de fournir une information détaillée sur la position de la mandibule, des lèvres et de la langue, et sur la vitesse et la direction de leur mouvement. Des données histologiques montrent que la densité des terminaisons sensorielles décroît progressivement de l'avant vers l'arrière de la bouche: la pointe de la langue semble être la mieux lotie en récepteurs dans le système oral, en accord avec son acuité tactile élevée.

Plusieurs études montrent l'importance de la somesthésie dans le contrôle de la parole. MacNeilage, Rootes et Chase (1967) citent le cas d'une patiente ayant un déficit congénital généralisé de la perception somesthésique: le sujet produit une parole totalement inintelligible bien qu'elle n'ait aucun trouble auditif ou moteur apparent. Hoole (1987) et Lindblom, Lubker, et McAllister (1977) ont mis en évidence l'influence de la sensibilité orale dans la production de parole perturbée (expériences de bite-block).

Un modèle de perception tactile pour la parole

Les faits ci-dessus ont motivé l'élaboration d'un modèle pour la prédiction des contacts palataux de la langue à partir des commandes articulatoires du SMIP (Schwartz et Boë, 2000).

Dans ce modèle, les patrons de contacts palataux sont décrits selon 5 variables (L_i , avec $i=1..5$), qui donnent le nombre de contacts par ligne selon cinq lignes allant de la périphérie à l'axe de symétrie du palais (Recasens, 1991; voir Figure 2-6). La prédiction des valeurs L_i se fait à partir des commandes articulatoires P_j par un assocateur linéaire à seuil:

$$L_i = f(\sum_j (w_{ij} * P_j + w_{i0}))$$

où w_{ij} et w_{i0} sont, respectivement, les poids et le biais d'apprentissage, et f , une fonction à seuil qui limite la valeur de L_i à son domaine de variation, qui va de 0 (pour aucun contact palatal sur la ligne i) à sa valeur maximale (qui est de 9, 8, 7, 5 et 4, pour les lignes 1, 2, 3, 4 et 5, respectivement). Les valeurs de w_{ij} et w_{i0} ont été ajustées en minimisant la somme des carrés des écarts entre les valeurs de L_i prédites et celles observées (Figure 2-7a).

Pour tester les performances de ce modèle, un jeu de contacts palataux prédits a été généré pour un grand nombre (1 000 environ) de configurations articulatoires, menant chacune à un patron de fréquences formantiques, dans les régions acoustiques des voyelles [i], [a] et [u]. Bien que ces configurations soient très variables en termes de valeurs des paramètres articulatoires, il s'avère que les contacts palataux prédits sont assez cohérents entre eux (Figure 2-7b) et avec la variabilité des contacts observée par Recasens (1991) pour des voyelles produites dans des contextes consonantiques variés. Il vient que le modèle semble capable de fournir des prédictions pertinentes en termes de structure acoustique et articulatoire du geste de parole.

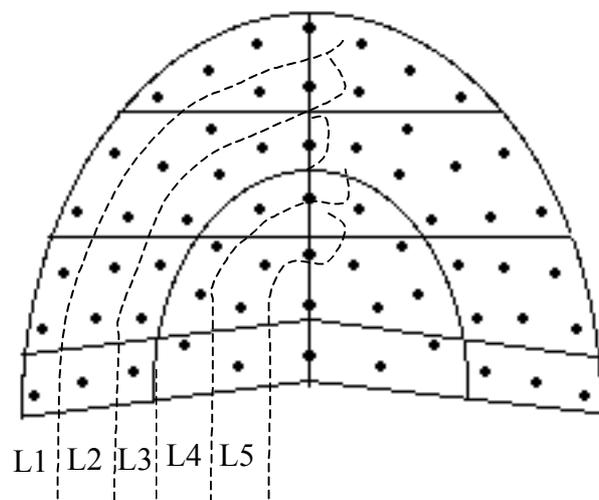


Figure 2-6: Les lignes de perception tactiles du palais dur de l'agent

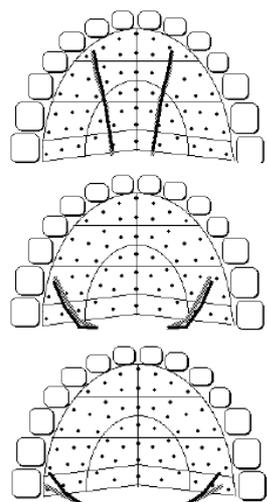


Figure 2-7a: Les zones de contacts palatal prédites (en noir) et observées (en gris) pour les configurations articulatoires prototypiques [i], [a], [o] (de haut en bas)

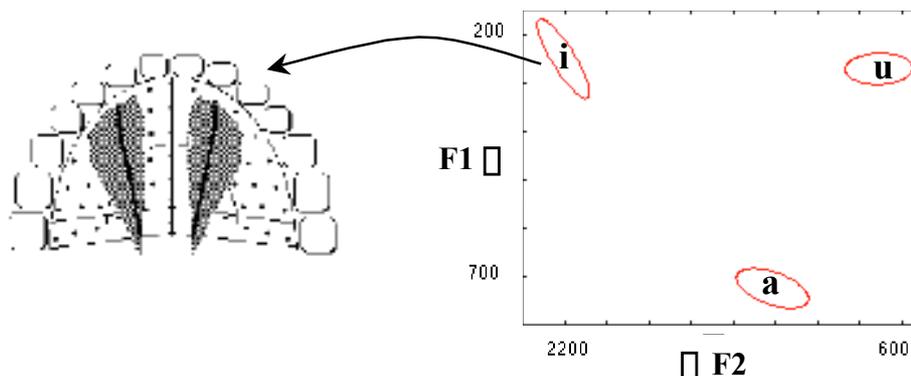


Figure 2-7b: Les zones de contact palatal prédites (gauche) pour une centaines de configuration articulatoire voisines de [i] (à droite, les formants; la même procedure a été suivie pour [a] et [u]).

4. Des modèles perceptifs simplifiés

Pour centrer notre étude sur des problèmes d'apprentissage, et non de perception, nous avons choisi de prendre en compte un jeu restreint et simplifié de variables sensorielles, d'interprétation aisée en des termes phonétiques (Figure 2-8).

- Les variables auditives comprennent les deux premiers formants (F1, F2) exprimés en **Bark**, i.e. une échelle perceptive des fréquences² (Schroeder, Atal et Hall, 1979) dont le calcul à partir des Hertz fait appel à la formule suivante:

$$z(\text{Bark}) = 7 \text{ Arg sh} \left[\frac{F(\text{Hz})}{650} \right]$$

² Pour rappel, le **Bark** (*Bk*) est une unité perceptive des fréquences. Plus précisément, l'échelle des Bark correspond à une échelle de fréquence semi-logarithmique. Pour illustration, un bark correspond à la différence minimale entre deux sons purs pour que leur perception respective n'interagissent pas entre elles : c'est la largeur de la **bande critique**. Cette échelle des fréquences permet donc de modéliser le processus par lequel l'oreille humaine distingue les stimuli de type voyelles.

- Le système tactile simplifié se base sur la géométrie du conduit vocal, qui peut être décrite par les systèmes suivants (Boë, Gabioud, Perrier, Schwartz et Vallée, 1995b): (i) l'aire (A_c) et la distance à la glotte (X_c) de la constriction principale le long du conduit vocal, ainsi que l'aire inter-labiale (A_l) produite par l'agent (Fant, 1960), et (ii) les coordonnées (X_h, Y_h) du plus haut point de la langue dans un référentiel fixe (Boë, Perrier et Bailly, 1992).
- Le système visuel estime uniquement l'aire aux lèvres (A_l) produite par autrui.

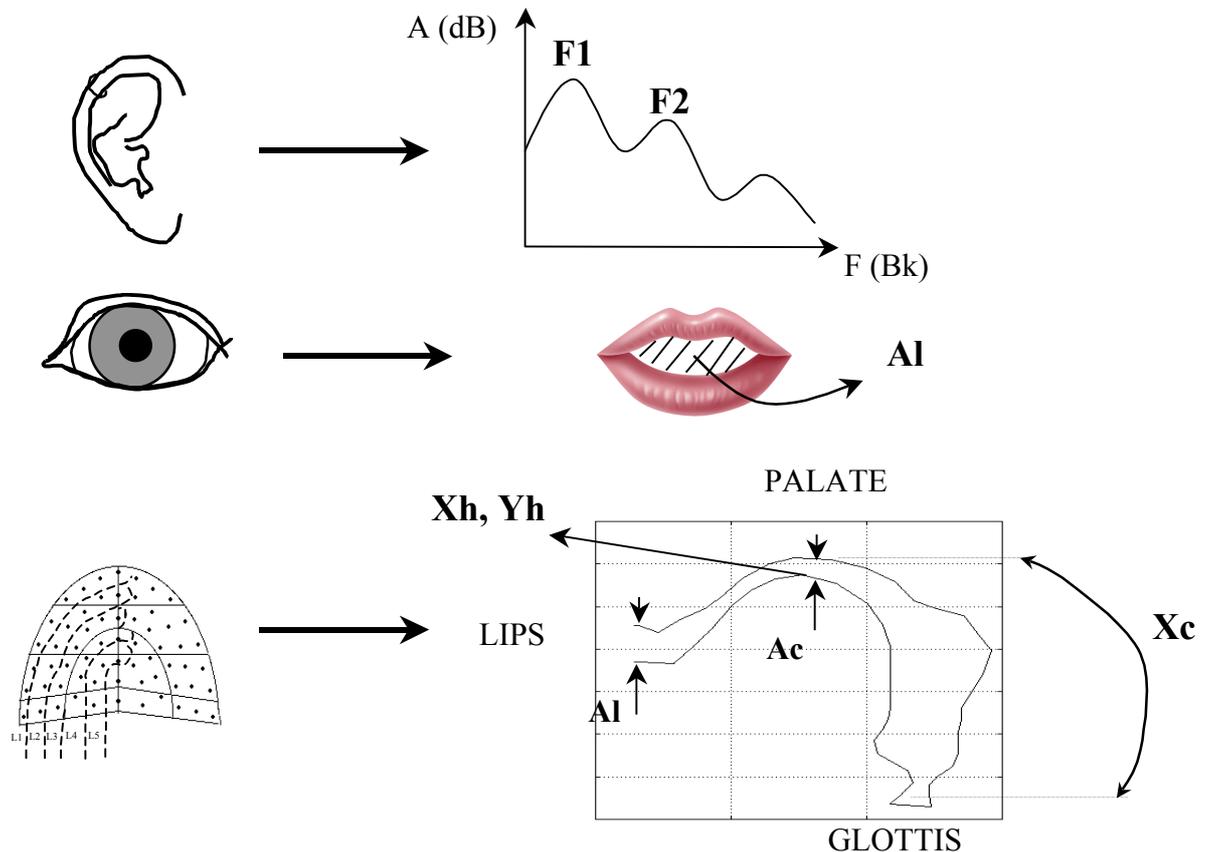


Figure 2-8: Les modèles perceptifs simplifiés de

III. LE MODELE D'APPRENTISSAGE SENSORI-MOTEUR ET D'INVERSION

La mise en œuvre de notre système sensori-moteur se fait en deux temps, dans le cadre de cette étude (voir aussi Jordan, 1990; Jordan et Rumelhart, 1991).

1. Au cours d'une *phase endogène*, le modèle apprend les relations basiques entre commandes motrices et entrées sensorielles, en explorant ses capacités articulatoires: il apprend un modèle direct de son conduit vocal.
2. Dans une *phase exogène*, il essaie d'imiter un son produit par le conduit vocal d'un congénère : l'on parlera de modèle inverse. C'est ce mode de fonctionnement qui intéressera la modélisation de l'imitation³.

³ Attention, les phases endogène et exogène ne sont pas synonymes de celles décrites par la théorie Frame/Content (chapitre 1, section II) : elles correspondent simplement à une implémentation en deux étapes des mécanismes prédictif et responsif, qui se chevauchent très probablement au cours du développement, sans prétendre à la modélisation des phénomènes neuronaux en jeu dans la réalité.

La programmation Bayésienne de robots proposée par Bessière et collègues (Bessière, Dedieu et Mazer, 1994 ; Bessière, 2000 ; Lebeltel, Bessière, Diard et Mazer, 2003) est un moyen de mettre en œuvre leur proposition pour une *théorie probabiliste des systèmes cognitifs sensori-moteurs* (Bessière, Dedieu, Lebeltel, Mazer et Mekhnacha 1998a).

A la question « Comment utiliser un modèle incomplet de son environnement pour percevoir, inférer, décider et agir de manière efficace ? », ils répondent par l'utilisation du *principe de maximum d'entropie* pour prendre en considération l'incomplétude de nos connaissances et de l'*inférence Bayésienne* pour tenir compte, dans le raisonnement, de l'incertitude des données dont on dispose (Bessière, Dedieu, Lebeltel, Mazer et Mekhnacha, 1998b).

Tout modèle d'interaction robot-environnement est par définition incomplet par rapport au phénomène réel. Cette situation désarçonne le paradigme symbolique qui ne peut faire face au problème de l'ancrage, dans le réel, des symboles utilisés par le robot pour agir⁴.

Le paradigme Bayésien permet de transformer ce qui n'a pas été modélisé explicitement (l'incomplétude) en une incertitude qui s'intègre dans une description⁵ probabiliste du phénomène.

Cette description est la représentation qu'aura le système sensorimoteur de son interaction avec l'environnement. Il se construit cette connaissance statistique générale, dans une phase d'apprentissage, à partir de connaissances préalables et de données expérimentales fournies par le/la concepteur/trice. Le choix de cette matière première se fait selon le critère de maximum d'entropie. Etant données des connaissances préalables et des données expérimentales, la distribution qui maximise l'entropie est celle qui représente le mieux ce couple⁶. Ce principe permet d'ancrer la description dans le réel en donnant une signification précise, mathématique et quantifiable à son « observabilité » (pour des justifications du principe de maximum d'entropie, se référer à Bessière *et al.* 1998b).

Il en résulte que la description ne donne aucune certitude mais fournit un moyen de prendre la *meilleure décision possible en accord avec les informations disponibles*. Des connaissances préalables, même imparfaites et incomplètes, sont pertinentes et fournissent des indications intéressantes sur le phénomène observé.

Ainsi, le substrat du raisonnement du robot est-il la représentation probabiliste qu'il a de son interaction avec l'environnement. Il l'exploite par les règles de conjonction et de marginalisation⁷, propres à l'inférence Bayésienne.

⁴ Le lien entre les signaux de bas niveau qu'utilise le robot et les symboles qu'il doit manipuler pour agir suppose une interprétation des données sensorielles...mais le modèle utilisé pour interpréter ces données n'est jamais suffisamment précis et complet pour pouvoir rendre compte de toutes les variations possibles des mesures.

⁵ Dans ce contexte, « description » est synonyme de « distribution ».

⁶ Le **principe du maximum d'entropie** permet de proposer une modélisation statistique faisant intervenir le minimum d'informations non contenues dans les données. C'est par exemple ce qui conduit à proposer, pour une variable aléatoire dont on ne sait rien, une *distribution uniforme*. Ou pour une variable dont on ne connaît que la moyenne et la variance, une *distribution gaussienne*. Concrètement, le maximum d'entropie conduit à supposer que la variable aléatoire essaie d'occuper au maximum l'espace disponible, sous la contrainte des données existantes. Ainsi, s'il n'existe pas de données, la variable occupe uniformément tout son espace disponible (maximum d'entropie), d'où la loi uniforme. S'il y a des données résumées par une moyenne et une variance, l'occupation maximale du terrain (maximum d'entropie) qui tient compte de la moyenne et la variance est la **loi gaussienne**.

⁷ Soit deux variables discrètes et exhaustives dont les valeurs, dans leur ensemble de réalisations, sont mutuellement exclusives. La probabilité de la **conjonction** X et Y est : $P(X, Y) = P(X) * P(Y/X)$.

La règle de **marginalisation** postule que : $\sum_x P(X, Y) = P(Y)$.

La méthode de programmation proposée par l'équipe de Bessière comprend les deux grandes étapes suivantes :

1. *Phase de Description* : Elle fournit un moyen effectif pour calculer la distribution conjointe de l'ensemble des variables attribuées au système sensori-moteur, sachant des connaissances préalables et un jeu de données expérimentales.

1. a. *Spécification* : L'on fournit au système des connaissances préalables structurelles, de dépendance et *a priori*.

* *Choix des variables* pertinentes, de leur domaine de variation et de leur discrétisation. Ce sont les connaissances préalables de la structure du système.

* *Décomposition de la probabilité conjointe* de l'ensemble des variables en un produit de distributions plus simples. Des connaissances préalables d'indépendance conditionnelle entre certaines variables permettent de choisir la décomposition la plus représentative du problème au sein de l'univers des possibles. La distribution de probabilité conjointe prend le nom de **structure de dépendance**.

* *Définition des formes paramétriques* de chacune des distributions composant la structure de dépendance. Ce choix se base sur leur connaissance *a priori*.

1. b. *Identification* des paramètres libres des sous-distributions de la structure de dépendance par un apprentissage qui fournit au robot ses connaissances expérimentales. La structure de dépendance prend le nom de **description**. Elle rend compte des dépendances *observables*.

2. *Phase d'Utilisation* : Une question probabiliste est soumise à la description obtenue dans la première phase. La représentation dont dispose le robot lui permet d'inférer (i) le résultat sensoriel d'une action donnée (mode de fonctionnement direct) et (ii) les commandes motrices pour atteindre une situation sensorielle désirée (mode de fonctionnement inverse), et l'on parlera d'*inversion*. Dans tous les cas, la connaissance des valeurs de certaines variables permet de prédire les valeurs probables des autres variables.

En résumé, les compétences du système sensori-moteur sont modélisées par la distribution de probabilité conjointe des paramètres articulatoires et perceptifs qui le constituent. La décomposition de cette distribution fournit une structure statistique exprimant les relations d'interdépendance entre ses variables motrices et sensorielles. Ces relations perception-production se basent sur des considérations de dépendance conditionnelle entre variables et s'expriment, essentiellement, dans le choix de la décomposition de la distribution conjointe.

Le modèle s'ancre dans le « réel » via un ensemble de données d'apprentissages : il lui permet d'instancier les distributions de probabilité figurant dans sa loi de décomposition (ou *structure de dépendance*). Le robot résultant est interrogé par l'intermédiaire de sa *description* effective. Les *performances* de l'androïde sont intrinsèquement liées à son expérience et à la « finesse » de sa structure⁸.

⁸ Cette « finesse » est liée au **degré de discrétisation** des variables spécifiées. Pour illustration, dans le cas du système visuel périphérique, le degré de discrétisation d'une variable "coordonnée d'un point sur la rétine" correspondrait à l'acuité visuelle ou, plus schématiquement, à la capacité de distinguer deux points de l'espace se projetant sur la rétine.

Conclusion

Nous disposons désormais de tous les éléments pour mettre en œuvre notre modèle basique du développement précoce de la production de la parole. Notre objectif est de faire passer cet agent par le même chemin que les nourrissons réels. L'enjeu sera pour nous de comprendre la co-structuration des domaines perceptifs et moteurs comme nous le verrons dans la troisième partie de ce mémoire. En attendant, la partie suivante veut approfondir le thème de l'imitation orofaciale au service du développement de la production de la parole.

PARTIE II:
Une revue de questions sur
l'imitation



Chapitre 3: De l'imitation dans le développement précoce humain et chez des non-humains

INTRODUCTION

De l'influence précoce de l'environnement...

Nous avons vu que d'après la théorie Frame/Content, le contenu du cadre, ou production contrôlée de sons de la langue ambiante, ne se développe qu'après le début du babillage canonique (chapitre 1, section III. 2.). L'expérience triviale montre que quel que soit l'environnement social, et ce sauf empêchement¹, l'individu apprend à parler la ou les langue(s) utilisée(s) par sa communauté linguistique et plus précisément par les personnes avec qui il y a communication dans une ou plusieurs langues données². Apprendre à parler une langue fait partie des apprentissages sociaux dont la majorité des êtres humains font l'expérience.

Cependant, cet apprentissage soulève deux questions fondamentales: quand et comment se déroule-t-il? A quel moment et dans quels domaines du développement de la parole l'influence de la langue commence-t-elle à se faire sentir?

...dans la prosodie

A partir de 4 jours, des individus arrivent à distinguer des phrases de la langue ambiante par rapport à des phrases d'autres langues, sur la base de leur **prosodie**³, avec une préférence pour celle de la première (Mehler, Jusczyk, Lambertz, Halsted, Bertoncini et Amiel-Tison, 1988). Dès 6 mois, la prosodie des productions de babillage canonique ressemble à celle de la langue native d'après le jugement perceptif de locuteurs/trices adultes de cette même langue (de Boysson-Bardies, Sagart et Durant, 1984).

...dans les voyelles

A partir de 6 mois, la perception d'un contraste vocalique non-natif perd en discriminabilité contrairement à un contraste vocalique natif (Polka et Werker, 1994; Kuhl, William, Lacerda, Stevens et Lindblom, 1992). De plus, la perception vocalique tend à assimiler les voyelles non-natives à des catégories vocaliques natives, telles qu'elles semblent se structurer selon les catégories phonologiques de la langue ambiante: Kuhl et collègues parlent de "magnétisme de la langue native" (*Native Language Magnet*). A 10 mois, la forme de l'espace vocalique des productions de babillage tend à ressembler à celle des voyelles produites par des locuteurs/trices adultes de la langue de l'environnement (de Boysson-Bardies, Hallé, Sagart et Durant, 1989).

¹ Le terme "empêchement" regroupe les différences par rapport à la majorité des individus dans des capacités psycho-biologiques, le plus fréquemment observées dans l'espèce humaine, qui entrent en jeu dans le développement de la parole. L'on y trouve par exemple les sensibilités réduites par rapport à la moyenne à l'intensité acoustique ("surdité") et/ou visuelle ("cécité"); mais cela va jusqu'aux capacités d'établir un contact social satisfaisant des critères sociaux comme dans le cas de personnes dont le comportement a été classé sous le terme général d'autisme.

² Et cela va jusqu'à la variété de langue, le dit "accent", les gestes de l'ensemble du corps, où l'on retrouve les "mimiques familiales" transmises d'une génération à l'autre par la communauté de vie.

³ **Prosodie**: désigne essentiellement les variations de hauteur, d'intensité, de débit et de rythme de la parole (Crystal, 1980/1995, p. 283).

...dans les consonnes

Dès 4 jours, il y a capacité de distinguer des contrastes phonétiques qu'ils fassent ou non partie de la langue ambiante (Kuhl, 1991). A 8 mois, la capacité de discrimination de paires de plosives qui se distinguent selon un contraste phonétique absent dans la langue native et assimilé à la même catégorie phonémique dans cette langue commence à diminuer, pour arriver à un niveau adulte à 10-12 mois (Polka et Werker, 1994). Dans le même temps, la capacité de discrimination de contrastes plosifs de la langue ambiante se maintient, et ce *a priori* jusqu'à l'âge adulte. A partir de 10 mois, les productions consonantiques de babillage sont imprégnées de caractéristiques de la langue de l'environnement en termes de distribution des fréquences de lieux et de modes d'articulation (de Boysson-Bardies et Vihman, 1991)

L'imitation: de l'ajustement de la production à la perception

Des faits ci-dessus l'on peut tirer que la langue ambiante laisse vraisemblablement des traces au niveau de la perception infantile de la prosodie (à 4 jours), des voyelles (à 6 mois) puis des consonnes (à 10-12 mois). De plus, le paysage articulatoire semble commencer à se structurer autour de caractéristiques de la langue de l'environnement un pas derrière le paysage perceptif: au maximum à 6 mois pour la prosodie et à 10 mois pour les consonnes et voyelles en devenir. Or, des auteurs rapportent dans le cas de la prosodie et dans celui des voyelles la présence de comportements imitatifs dans la période qui précède l'apparition dans la production de caractéristiques environnement-dépendantes (pour les voyelles, Legerstee, 1990; Lebib et Baudonnière, 1999, 2000; Kugiumutzakis, 1999; Kuhl et Meltzoff, 1982, 1984, 1996; pour la prosodie, Kessen, Levine et Wendrich, 1979, cités par Locke, 1983; Kuhl et Meltzoff, 1982).

Nous proposons donc que, au cours du développement de la parole, l'imitation serve à orienter l'exploration articulatoire vers les percepts de la langue ambiante et à ajuster ainsi les productions de soi à celles d'autrui, en les faisant converger sur des caractéristiques phonétiques propres à la parole ambiante telle qu'elle est perçue. En d'autres termes l'imitation peut vraisemblablement être tout ou partie du mécanisme d'apprentissage qui permet l'ajustement graduel des vocalisations produites par soi à la parole de notre environnement social, tandis que la maîtrise de la coordination (ou *contrôle*) des gestes oro-faciaux pour la communication, en particulier ceux du conduit vocal, s'affine. De plus, nous défendrons dans ce chapitre l'ancrage du développement de la parole dans le développement sensori-moteur de l'organisme en mettant en exergue une continuité ontogénétique et phylogénétique des mécanismes et phénomènes ayant pu permettre l'émergence de l'imitation orofaciale pour la communication chez les Primates, dont *Homo sapiens*.

La plausibilité de cette proposition se décline en quatre parties. La première veut apprécier ce que désigne le terme "imitation", en particulier dans les traditions de l'éthologie et de la psychologie. La seconde donnera la vision de Piaget (1945/1976) et celle de Meltzoff et Moore (1997) des mécanismes sous-jacents aux comportements imitatifs oro-faciaux des nourrissons. Dans la troisième partie, vous trouverez des données sur l'imitation chez les Primates humain et non-humains mais aussi chez d'autres Vertébrés. Dans le cas des êtres humains, nous donnerons plus de précision sur l'imitation précoce de modèles liés à la parole et nous tenterons de faire un lien entre développement sensori-moteur et imitation orofaciale. La dernière partie fournit des données de neurosciences en lien avec l'imitation dans une perspective phylogénétique.

I. QU'EST-CE QUE L'IMITATION?

Une assertion fait l'unanimité: imiter, c'est *copier* ce que fait l'autre. Au-delà de ce fait *a priori* trivial, la littérature qui traite de l'imitation reflète le désaccord qui règne sur ce qu'est "exactement" l'imitation.

Pour notre part, nous tenterons d'illustrer comment le terme d'"imitation" désigne à la fois *un comportement (de surface)* et *des mécanismes (cachés)* pouvant rendre compte de ce comportement. Notre but est de proposer une conception de l'imitation qui unifie des apports de l'éthologie et de la psychologie développementale pour apprécier ce que peut être l'imitation dans une perspective onto-phylo-génétique.

En éthologie, l'accent est mis sur le fait que l'imitation correspond à une forme d'*apprentissage social* (ex. Shumaker et Swartz, 2002; Blackmore, 1998). En psychologie, imiter, c'est surtout *apparier* son propre comportement à celui d'un autre individu (ex. Butterworth, 1999). Ces deux points de vue sur l'imitation sont complémentaires: l'un insistant plus sur le comportement observable, l'autre sur les mécanismes sous-jacents. C'est ainsi que nous allons regarder cette boîte noire du dehors puis du dedans.

1. La boîte noire vue du dehors

En éthologie, la quête des espèces capables d'imiter et/ou pratiquant l'imitation a permis de situer l'imitation par contraste avec d'autres formes d'apprentissage. Cette vision systématique procède en une première distinction (Blackmore, 1998): l'imitation est une forme d'apprentissage *social*, par opposition aux apprentissages individuels qui comprennent les **conditionnements classique**⁴ et **opérant**⁵, et la **formation de cartes cognitives**⁶.

"Le problème dans la démonstration de l'imitation [...] est l'exclusion des hypothèses alternatives" (Visalberghi et Fragazsy, 1990)⁷.

L'**apprentissage social** (ou apprentissage par observation), implique un ou plusieurs autre(s) *congénères* (Blackmore, 1998). Galef (1988) propose une définition plus précise (Shumaker et Swartz, 2002). Pour lui, l'apprentissage social est le processus tel qu'un individu, *l'observateur/trice*, acquière un comportement en observant un autre individu, *le/la démonstrateur/trice*, effectuer ce même comportement. Toutefois, d'après Blackmore (1998), l'apprentissage social est très souvent basé sur les conditionnements classique et opérant, mais la manière dont est appris le comportement fait intervenir au moins un autre individu. Par

⁴ **Conditionnement classique:** Apprentissage de l'association de deux stimuli par leur cooccurrence répétée (voir aussi le principe de contiguïté temporelle, ce chapitre, section III.A.2.).

⁵ **Conditionnement opérant:** Il a lieu quand le comportement d'un animal est soit "récompensé" soit "puni" de manière à augmenter vs. diminuer sa fréquence. Par exemple, les humains apprennent à faire du vélo par conditionnement opérant. Ce type de conditionnement est un cas d'*apprentissage par essai-erreur*. Il a lieu à chaque fois que l'individu augmente d'un essai à l'autre la fréquence des actions qui mènent à des succès et diminue la fréquence des actions qui mènent à de la douleur ou à l'échec.

⁶ **Formation de cartes cognitives** : La formation de cartes cognitives correspond à l'apprentissage par conditionnement de la situation d'éléments d'intérêt dans un espace d'exploration, menant à la élaboration ou la mise à jour d'une représentation de cet espace. En éthologie, il s'agit en particulier de cartes du milieu de vie, qui permettent à l'individu d'avoir une idée précise de la structure de cet espace, et qui comprennent, par exemple, la localisation des sources de nourriture chez les abeilles ou celle du bureau de poste chez les humains. Dans le cadre de ce mémoire, les espaces qui pourraient faire l'objet d'un tel apprentissage sont de types perceptif et articulatoire, et l'on parlera d'apprentissage par exploration ou d'**exploration** du comportement du conduit vocal (voir l'introduction de la troisième partie pour plus de détails).

⁷ "The problem of demonstrating imitation is [...] exclusion of alternative hypotheses" (Visalberghi et Fragazsy, 1990, p. 294, cités par Bard et Russell, 1999, p. 106).

exemple, Bard et Russell (1999) précisent le cas du **conditionnement par observation** qu'elles définissent comme l'apprentissage d'une association entre un objet et une évaluation émotionnelle, et qu'il faudrait distinguer, d'après elles, de l'imitation. Elles prennent pour exemple l'apprentissage de la réponse de peur de certains serpents chez de jeunes Macaques qui observent l'expression de cette réaction chez des congénères, en ajoutant que l'aspect contagieux de ce conditionnement semble lié aux débuts des réponses empathiques, telles que les pleurs dits "contagieux" chez les nourrissons humains.

D'ailleurs, la **contagion**⁸ (ou *imitation instinctive* ou suggestion instinctive ou facilitation sociale ou facilitation de la réponse ou coaction) représente pour Bard et Russell (1999) une autre source de confusion⁹ entre imitation et non-imitation. Cette imitation "instinctive" correspond d'après les auteurs à l'augmentation temporaire de la probabilité qu'un individu s'engage dans un comportement qui existe déjà dans son répertoire (tel que les rires, bâillements et pleurs), comme résultat de l'observation d'un congénère engagé dans ce même comportement. Leur distinction entre contagion et imitation insiste sur le fait que l'imitation porte sur un comportement *nouveau* (voir plus bas).

Heyes (1993) distingue l'imitation des autres types d'apprentissages sociaux, dit "non-imitatifs", par le fait que la première corresponde à l'apprentissage *sur un comportement* en observant autrui, tandis que les seconds sont l'apprentissage d'informations *sur l'environnement* à travers l'observation de congénères (Blackmore, 1998; Bard et Russell, 1998). Si l'on se conforme à ce point de vue, l'imitation est aussi à distinguer du **rehaussement du stimulus** (ou rehaussement local)¹⁰ qui correspond à apprendre à orienter son attention dans le sens opposé de ou bien vers un stimulus (tel qu'un objet ou un lieu) de l'environnement par l'observation du comportement de congénères vis-à-vis de ce stimulus (Blackmore, 1998; Bard et Russell, 1999). Bard et Russell (1999) précisent que, par exemple, si lorsque l'attention est orientée sur un stimulus une action donnée a lieu de manière systématique, alors l'expression de cette action n'est pas à attribuer à l'imitation (même si un individu l'exprime à la suite d'un autre comme dans le cas de l'imitation).

La dernière confusion à éviter entre l'imitation et les autres apprentissages sociaux, d'après Bard et Russell (1999), prend appui sur la distinction faite par Köhler (1926) entre deux types d'imitation: celle qui porte sur la 'substance' et celle sur la 'forme'¹¹. Depuis, l'imitation de la substance est désignée par l'émulation du but, celle de la forme par l'imitation. L'**émulation du but** a lieu lorsqu'un individu copie le but (ou résultat) du comportement d'un autre individu, sans en copier la forme (Blackmore, 1998). Pour Bard et Russell, l'émulation du but se distingue de l'imitation par le fait que, dans la première, le résultat final est atteint plus rapidement avec l'observation que sans l'observation du comportement chez un/une congénère, et cela sans que les actions exactes utilisées par le/la démonstrateur/trice ne soient utilisées par l'observateur/trice.

Pour Whiten et Ham (1992), l'**imitation** (ou copiage imitatif) est une forme d'apprentissage social dans laquelle "B apprend quelque(s) aspect(s) de la *forme* intrinsèque d'un *acte*"¹² [sic]

⁸ Bien que Blackmore (1998) l'exclut des apprentissages sociaux (et même individuels), le sort contraire lui est fait par Shumaker et Swartz (2002).

⁹ Piaget parlait de la "confusion des pleurs" au sujet de la contagion...

¹⁰ Le rehaussement local (*local enhancement*, en Anglais) et le rehaussement du stimulus (*stimulus enhancement*) sont considérés ensemble par certains auteurs (ex. Bard et Russell, 1999) et séparément par d'autres (ex. Blackmore, 1998), bien que la différence ne semble porter que sur le fait que le premier s'applique spécifiquement à un lieu et le second à tous types de stimulus.

¹¹ Nous reviendrons plus bas sur le sens que prend cette distinction en linguistique.

¹² Un **acte** (ou action) est un mouvement qui a un but, et est donc intentionnel (voir plus loin, les concepts d'action et d'intention, section I.2.).

de A"¹³ (Bard et Russell, 1999). Thorpe (1963) restreint l'imitation à une répllication *exacte* de cette forme, en ajoutant que le comportement n'est imité que s'il est *nouveau ou improbable* (i.e. non-familier) dans le répertoire comportemental de l'observateur/trice, et que ce/cette dernier/ière n'a pas une tendance intrinsèque à exprimer ce comportement (Bard et Russell, 1999). Whiten et Ham (1992) ajoutent que les actes pré-existants sont nouveaux (i) si des comportements exprimés auparavant sont combinés ou ajustés d'une nouvelle manière, ou (ii) si des comportements du répertoire de l'observateur/trice sont combinés d'une nouvelle manière dans une action sur des objets de l'environnement (Bard et Russell, 1999). Shumaker et Swartz (2002) reprennent la définition de Thorpe (1963) en précisant qu'un acte clairement imitatif supposerait que l'imitateur/trice puisse comprendre la situation du point de vue du/de la démonstrateur/trice. En clair, l'imitation au sens strict correspondrait à une répllication exacte des *fins* et des *moyens*.

Cependant, la définition stricte de l'imitation semble inapplicable car l'appariement *exact* des moyens *et* des fins a rarement lieu dans les comportements imitatifs spontanés que ce soit chez les Primates humains ou non-humains (Bard et Russell, 1999).

Dans le même ordre d'idées, une question reste en suspens, en particulier au regard du développement sensori-moteur: comment imiter de manière parfaite un acte nouveau sans maîtriser le comportement du ou des effecteur-s à utiliser? L'on ne peut, en effet, ignorer les contraintes imposées par le degré de développement du contrôle moteur aux comportements imitatifs.

A un autre niveau, la distinction entre imitation du but et imitation de la forme, soulevée par Köhler (1926), n'est pas sans rappeler, aux personnes sensibilisées à l'histoire de la linguistique, les propos de son père fondateur officiel, Ferdinand de Saussure. En effet, de Saussure (1916) faisait une distinction du même ordre entre les phonèmes (*la forme*, l'invariance) qui peuvent être vus comme le *but* du/de la locuteur/trice et leurs différentes réalisations (*la substance*, la variabilité) qui peuvent être vues comme les différents *moyens* mis en œuvre pour atteindre un même but (voir Vallée, 1994)¹⁴. Dans l'acquisition de la parole, le développement va-t-il vers l'imitation du but, de la forme (au sens de Köhler) ou des deux? Que se passe-t-il s'il existe un nombre restreint de moyens pour atteindre le même but, i.e. si les capacités sensori-motrices en cours restreignent les moyens à disposition? Parlera-t-on d'émulation du but ou d'imitation? En ce qui concerne le développement de la parole, nous pensons qu'imitation et émulation du but ne sont pas à distinguer dès le départ de l'investigation car cela pourrait cacher une progression potentielle de l'une à l'autre ou même une transformation de l'une en l'autre.

Ainsi avons-nous choisi de garder des apports ci-dessus les idées-clés suivantes: l'imitation est un *apprentissage social du comportement de l'autre* (le/la démonstrateur/trice) qui se fait en *observant ce comportement et en répliquant tout ou parties*, i.e. *la fin et/ou les moyens*, de ce comportement.

¹³ "B learns some aspect(s) of the intrinsic *form* of an act from A" (Whiten et Ham, 1992, p. 250, cités par Bard et Russell, 1999, p. 90).

¹⁴ Ainsi, la notion de forme au sens de Köhler diffère-t-elle de celle au sens de de Saussure. La forme de l'acte du/de la démonstrateur/trice selon Köhler correspond plutôt à la substance (réalisation phonémique) selon de Saussure. De même, le but de l'acte chez Köhler pourrait être synonyme de la forme (phonème) chez de Saussure.

2. La boîte noire vue du dedans

En psychologie, la quête des *mécanismes* pouvant rendre compte des comportements imitatifs et surtout de leur développement a été la pomme de discorde. Nous avons choisi de regarder l'intérieur de la boîte dans une perspective historique.

Au cours du 19^{ième} siècle, l'imitation était définie en psychologie comme étant la *répétition* par un individu du comportement d'un autre individu (Butterworth, 1999).

A la fin du 19^{ième} siècle, Baldwin (1895), pour qui l'imitation est un phénomène à la fois biologique et social, en a proposé une définition selon trois axes (Butterworth, 1999): (i) au niveau interpersonnel, l'imitation est "la *performance* [sic] dans le mouvement, la pensée ou le mouvement et la pensée, de ce qui vient par les *sens* [sic] ou par suggestion, comme appartenant à un autre individu"¹⁵. En d'autres termes, quand on imite, une autre personne sert à donner la copie, i.e. le *modèle*, à imiter, et l'imitation est (déjà) *au croisement de la perception et de la production*¹⁶; (ii) au niveau intrapersonnel, l'imitation est "[t]oute répétition de pensée, action ou les deux qui rétablit une copie"¹⁷. D'après Butterworth (1999) un lien entre imitation et *mémoire* est suggéré; (iii) au troisième niveau se trouve le mécanisme de *développement de l'imitation*, et cela en interaction avec la mémoire selon Butterworth: l'imitation est "[t]oute réaction organique de répétition du stimulus ou de type auto-maintenance [qui est] synonyme de la conception physiologique d'une réaction circulaire"¹⁸. Précisons qu'une **réaction circulaire** est un cycle répétitif d'activité dans lequel l'atteinte du but de l'acte initie un nouveau cycle d'activité du même type, comme dans le cas d'une séquence de succion de la main chez les nourrissons humains. Selon Baldwin, ce mécanisme permet de conserver par répétition le résultat intéressant d'une conduite qui est découvert par hasard. Autrement dit, pour lui, la répétition d'une action est imitative de l'acte dont elle est une répétition. Or, le répertoire comportemental des nourrissons est caractérisé par des patrons répétitifs (les rythmiques, voir chapitre I, section II). Ainsi, l'on peut comprendre que d'après lui le développement de l'imitation commence par l'imitation de soi avant celle d'autrui (Harris et Butterworth, 2002). Les idées suivantes se retrouvent dans la vision que Piaget (1945/1976) a de l'imitation dans sa théorie du développement de l'intelligence sensori-motrice: (i) l'imitation d'autrui naît des réactions circulaires qui sont des imitations de soi, dans un premier temps, (ii) l'imitation a à voir avec la perception et la production, idée que l'on retrouvera aussi chez Guillaume (1926), et (iii) l'imitation entretient des liens avec la mémoire.

Au début du 20^{ième} siècle, Guillaume (1926) a proposé le mécanisme d'**appariement visuo-kinesthésique**, ou mise en correspondance entre vision de soi et **kinesthésie**¹⁹, pour

¹⁵ "[P]erformance in movement, thought or both movement and thought, of what comes through the senses or by suggestion, as belonging to another individual" (Baldwin, 1901, p. 519, cité par Butterworth, 1999, p. 64).

¹⁶ Il y a, là, une différence de vocabulaire significative entre le discours de l'éthologie traditionnelle et celui de la psychologie: le premier parle "d'observation", le second de "perception". Au niveau théorique, cela marque une distinction fondamentale entre les deux conceptions: la psychologie postule l'existence de représentations (mentales), contrairement à l'éthologie traditionnelle (voir section I.2. pour une définition du concept de représentation mentale). En d'autres termes, ici, l'observation cède le pas à une "construction mentale" qu'est la perception, avec pour conséquence la question du "Que perçoit-on?", i.e. "Quelle information extrait-on de l'environnement, lorsqu'on perçoit quoi et pourquoi faire?".

¹⁷ "Any repetition in thought, action or both which reinstates a copy" (Baldwin, 1901, p. 519, cité par Butterworth, 1999, p. 64).

¹⁸ "An organic reaction of the stimulus repeating or self sustained type [which is] synonymous with the physiological conception of a circular reaction" (Baldwin, 1901, p. 519, cité par Butterworth, 1999, p. 64).

¹⁹ **Kinesthésie** : Elle correspond à la proprioception dynamique, i.e. à la perception des mouvements propres, et s'inscrit dans le système somesthésique.

expliquer l'imitation basée sur la perception visuelle, ou **imitation visuo-motrice**, d'actes corporels²⁰ non-visibles sur soi ainsi que la reconnaissance de soi dans un miroir (Mitchell, 2002). Plus précisément, selon Guillaume, le développement de l'imitation est motivé par les et issu des tentatives répétées de reproduire les *effets* des actions d'autrui sur les objets, permettant ainsi l'apprentissage de la correspondance entre l'expérience interne de ses propres actions et ce à quoi elles ressemblent visuellement d'après leur perception sur l'autre: ce n'est que graduellement que cette connaissance mène à l'imitation des mouvements et actions corporelles d'autrui. Tout en étant d'accord sur le fait que l'imitation corporelle du non-visible sur soi (et la reconnaissance de soi) dérive de l'appariement visuo-kinesthésique, Piaget propose plutôt que l'imitation des *mouvements* doit venir avant celle de leur but, car l'intérêt premier du sujet est la reproduction des mouvements *en soi* et non celle de leurs effets²¹ (Piaget 1945/1976, pp. 85-87; Mitchell, 2002). De plus, "dès que l'imitation triomphe et que sa technique atteint la virtuosité, elle s'automatise et alors les résultats à atteindre priment les mouvements [ou *moyens*] qui y conduisent, ceux-ci s'ajustant aux *buts* [*sic*] par associations immédiates": pour lui, l'imitation des moyens vient *avant* celle du but²². D'après lui, l'imitation indiquerait une *conscience intelligente* de la similarité entre l'expérience kinesthésique de ses propres actions et l'expérience visuelle des actions d'autrui, un indice de cette intelligence étant l'*intention* qu'a le sujet de reproduire ce qui est perçu chez l'autre, comme en témoigne ses tentatives répétées et les "*erreurs*"²³ qui en résultent. Nous reviendrons par la suite sur les hypothèses et prédictions de Piaget.

Au 21^{ème} siècle, l'**imitation** est, au moins à l'âge adulte, "l'usage **intentionnel**²⁴ de l'**action**²⁵ d'autrui pour servir de guide à une activité propre, dirigée vers un *but*. Le sujet

Pour information, en neurosciences, le **système somesthésique** comprend deux sous-systèmes, l'un spécialisé dans la détection des stimuli mécaniques, l'autre dans celle de la douleur et de la température. Les informations de chacun de ces deux sous-systèmes sont collectées par des récepteurs qui leur sont propres et empruntent des voies spécifiques, mais toutes convergent sur le cortex somesthésique primaire (dans le gyrus post-central du lobe pariétal). *A priori*, seul le premier sous-système concerne notre problème: il informe sur les contacts de la surface du corps avec elle-même ou d'autres surfaces (modalité tactile), sur les mouvements (proprioception dynamique) et la position relative (proprioception statique) des différents segments du corps (voir Purves *et al.*, pp. 147-163, pour des connaissances générales).

²⁰ Tout au long de ce chapitre, nous distinguerons les actes (ou actions) corporels, qui ne s'appliquent à aucun objet matériel, des actes (ou actions) sur des objets.

²¹ Plus précisément, "c'est la *possibilité* [*sic*] de reproduction qui intéresse le sujet, c'est-à-dire que l'intérêt n'est pas extérieur [tel que l'effet des gestes d'autrui] mais immanent au fonctionnement [du développement de l'intelligence sensori-motrice]" (Piaget 1945/1976, p. 86).

²² L'on retrouve dans le texte de Piaget les notions de fins et de moyens, de but et de forme proposées par Köhler (section I.1.).

²³ L'on verra que l'étude de la structure de ces erreurs a suscité, plus récemment, l'intérêt d'autres psychologues du développement.

²⁴ **Intention**: D'après Grèzes (2000), l'intention est la "représentation d'un but, [le] contenu d'une pensée, d'une croyance, d'un désir etc." (p. 5). Pour Lycan (2001), "[l]e terme *intentionnelle* est utilisé par les philosophes non comme s'appliquant en premier lieu à des actions, mais pour signifier "dirigée sur un objet". De manière plus familière, pour qu'une chose soit intentionnelle il faut qu'elle soit *relative à quelque chose*. De manière paradigmatique, les états mentaux et les événements sont intentionnels dans ce sens technique [...]. Par exemple, les croyances, désirs et regrets sont relatifs à des choses, ou "ont des objets d'intention": j'ai des croyances sur Boris Yeltsin, je veux une bière et la paix dans le monde, et je regrette d'avoir été d'accord pour écrire autant d'articles d'encyclopédie." (p.413: "The term *intentional* is used by philosophers, not as applying primarily to actions, but to mean "directed upon an object." More colloquially, for a thing to be intentional is for it to be *about something*. Paradigmatically, mental states and events are intentional in this technical sense [...]. For instance, beliefs and desires and regrets are about things, or have "intentional objects": I have beliefs about Boris Yeltsin, I want a beer and world peace, and I regret agreeing to write so many encyclopedia articles.").

²⁵ **Action** (motrice): En psychologie cognitive, "[l]'action motrice est conçue comme le déroulement d'un geste dont il est possible d'identifier trois étapes: la planification, la programmation et l'exécution motrice. Seule la dernière étape est directement observable et introduit une modification dans l'environnement. Les deux

observe les *moyens* mis en œuvre et les *buts* atteints par le **modèle**²⁶, il en infère la relation utile, qu'il mettra en œuvre à son tour" (Grèzes, 2000, p. 28, citant Bruner, 1983, et Winnykamen, 1990). L'on retrouve, comme chez Piaget (et Köhler), la notion de but (avec son pendant psychologique qu'est l'intention) et celle de moyen.

Une autre manière de définir l'imitation est donnée dans la description d'un **acte imitatif** de Meltzoff (2002a): "[l']observateur/trice perçoit les actes du/de la démonstrateur/trice, utilise sa perception comme la base d'un plan d'action²⁷, exécute la sortie motrice. Cela implique la vision [dans le cas de l'imitation visuo-motrice], la coordination intermodale²⁸, et le contrôle moteur²⁹. Si l'imitation a lieu après un délai significatif, la mémoire et la **représentation**³⁰ de

premières, mentalement élaborées et antérieures au déclenchement de l'action, déterminent le but et la stratégie à adopter (planification) ainsi que le déroulement du geste (programmation). Le bon déroulement d'une action suppose que le sujet traite deux catégories d'informations sensorielles [...]: 1/ les informations prélevées sur l'espace externe (extéroceptives); elles peuvent être de nature auditive, visuelle, olfactive ou somesthésique, et agissent à la fois comme des déclencheurs et des guides de l'action motrice dans l'environnement ; 2/ les informations prélevées sur l'espace du corps (proprioceptives). Pour agir efficacement, le sujet doit connaître sa position par rapport à l'espace externe [...] et sentir la position de ses différents segments corporels avant le déroulement de l'action (statesthésie) et pendant l'exécution du geste (kinesthésie). L'action motrice se distingue du mouvement réflexe, réponse élémentaire et rapide aux sollicitations externes dont le niveau d'intégration se situe à la moelle épinière, et de la réponse automatique [...] s'effectuant essentiellement au niveau du tronc cérébral et des ganglions de la base. L'action motrice est donc un mouvement volontaire [...] dans la mesure où elle exprime les intentions de l'individu. Elle a pour origine une commande nerveuse générée au niveau du cortex [...] consécutive ou simultanée à l'intégration des informations sensorielles extéroceptives et proprioceptives." (Streri, 1998, pp. 29-30).

²⁶ **Modèle** (dans l'imitation): Il peut désigner, en fonction du contexte, l'action du/de la démonstrateur/trice, mais aussi le/la démonstrateur/trice (l'agent de l'action) lui/elle-même.

²⁷ L'on parle de planification. **Planification**: "capacité d'un organisme à se donner des représentations anticipatrices schématiques et hiérarchisées de séquences d'actions, ou d'opérations, pour atteindre un but. Ces représentations sont abstraites et ne comportent pas de détails concernant le déroulement de l'action. La planification est la démarche de construction de plans." (Grèzes, 2000, p. 7).

²⁸ Voir les sections II.2. et III.A.2..

²⁹ Voir par exemple, Bizzi (2001) et Purves *et al.* (1999, pp. 290-374).

³⁰ **Représentation** (mentale): En psychologie, une représentation mentale est une "entité de nature cognitive reflétant, dans le système mental d'un individu, une fraction de l'univers extérieur à ce système" (Grèzes, 2000, p.5). "Représenter ne consiste pas simplement à reproduire. Tout processus de représentation implique l'idée d'une transformation appliquée aux entités faisant l'objet de la représentation. Le degré de la préservation de la structure de l'information de départ détermine le degré d'analogie de la représentation à l'égard de l'objet représenté. Il importe de faire la distinction entre la représentation comme processus (ou ensemble de processus) et les représentations comme produits de ses processus. La psychologie s'intéresse aux processus générateurs de représentations mentales, c'est-à-dire d'entités cognitives produites par le fonctionnement d'un support biologique, le système nerveux. Pour la psychologie cognitive, les représentations sont des modèles intériorisés que le sujet construit de son environnement et de ses actions sur cet environnement. Ces modèles sont utilisables comme sources d'information et instruments de régulation et de planification de ses conduites." *Définitions du Dictionnaire fondamental de la Psychologie, Larousse 1997.*" (*idem*)

"L'une des hypothèses essentielles des sciences cognitives est donc la conviction selon laquelle il est légitime, et en fait nécessaire, d'établir un niveau d'analyse que l'on peut appeler "niveau de représentation". La notion de représentation suppose l'existence d'événements mentaux intermédiaires entre les stimuli et le comportement. Ces événements mentaux sont conçus comme des représentations objectives mais non nécessairement conscientes et non comme des représentations subjectives conscientes. Ces représentations sont sous-tendues par des processus, liés entre eux par des régularités causales. Ces processus sont objectivables en deux sens: d'une part parce qu'ils sont supposés être calculables, ou computationnels au même sens qu'un ordinateur calcule et produit à partir d'informations selon des lois mécaniques ; d'autre part parce qu'ils sont supposés être sous-tendus causalement par des événements neurophysiologiques et neuronaux dans le cerveau (Gardner 1993). D'une manière générale, l'explication en neuropsychologie cognitive prend la forme typique, énoncée par Marr (1982), d'analyse entre trois niveaux: le niveau computationnel où est déterminée la nature de la tâche de traitement de l'information à accomplir, son but et la logique de la stratégie qui serait appropriée ; le niveau intermédiaire définit l'algorithme qui permet la transformation de l'entrée en sortie ; le troisième niveau spécifie comment la

l'action entrent en jeu."³¹ (Meltzoff, 2002a, p. 20). En psychologie du développement, l'imitation reste une "technique" pour l'apprentissage de comportements nouveaux. Dans un désir d'en élaborer une définition implicitement donnée par Piaget (1945/1976, p. 63-64), Meltzoff (1988a) considère qu'un modèle est **nouveau** au moins selon six sens: (i) l'acte n'a pas été vu ou entendu par l'individu auparavant, (ii) il n'a pas été exécuté auparavant, (iii) bien qu'ayant été potentiellement découvert par hasard, il n'est pas familier ou fréquemment pratiqué par le sujet, (iv) il n'a pas été imité par l'individu auparavant, (v) ce n'est pas un comportement antérieurement mis en relation avec un objet ou une classe d'objet particulier/ière³², et/ou (vi) l'acte a lieu avec une probabilité nulle ou quasi nulle dans le comportement spontané du sujet. La vision que Meltzoff et ses collègues ont de l'imitation au cours du développement sera développée par la suite.

De Guillaume à Meltzoff, les efforts de recherche ont plus porté sur l'imitation de cibles visuelles que sur celle de cibles auditives: l'imitation de gestes faciaux (c'est-à-dire non-visibles sur soi)³³ impose une énigme. Une vision traditionnelle veut que différents *canaux* perceptifs, ou d'**afférences**³⁴, véhiculent des informations de natures, i.e. de "dimensions psycho-physiques", différentes, impliquant une incapacité de communication horizontale. Dans le cas de l'imitation vocale, les sons prononcés par soi et par autrui sont tous deux entendus, et sont donc *comparables* dans la même modalité perceptive (l'audition): la comparaison *peut* être intramodale. Comparer deux quantités exprimées dans les mêmes unités ne semble pas une opération difficile. Par contre, dans l'imitation faciale, l'observateur/trice ne peut voir son visage et celui du/de la démonstrateur/trice en même temps sans miroir: la comparaison ne *peut pas* être intramodale. Ainsi, ce phénomène a-t-il mené à la considération, *contre-intuitive* mais forcée, de l'existence d'un appariement entre modalités perceptives et donc entre deux informations, visuelle et proprioceptive, qui sont conçues à la base comme étant dans des canaux isolés, comme exprimées dans des unités différentes. Et cela va jusqu'à penser une communication possible entre canaux dont les flux d'information vont à contre-courant l'un de l'autre, c'est-à-dire entre la production (efférente) et la perception (afférente).

L'étude de l'imitation faciale a donc remis en question la vision traditionnelle d'une parcellisation de la perception en des canaux isolés les uns des autres, jusqu'à celle d'un cloisonnement à la base entre perception et production. Il en découle qu'une histoire propre à l'imitation vocale reste à écrire chez l'être humain. Toutefois, nous aurons le plaisir de tirer profit des réflexions, théories et données suscitées par l'imitation faciale pour reposer la question des mécanismes potentiellement sous-jacents à l'imitation vocale.

Pour l'instant, nous garderons à l'esprit que (i) l'imitation correspond à *apparier* son propre comportement à celui d'un modèle (ex. Butterworth, 1999), (ii) l'imitation a à voir avec un

base physique (neurologique) du système cognitif réalise ou implémente les représentations et l'algorithme." (*idem*).

³¹ "[T]he observer perceives the demonstrator's acts, uses visual perception as the basis for an action plan, executes the motor output. This involves vision, cross-modal coordination, and motor control. If imitation takes place after a significant delay, memory and the representation of action come into play" (Meltzoff, 2002a, p. 20)

³² Il donne pour illustration que faire semblant de boire d'une calculatrice comme d'un verre est un acte nouveau.

³³ L'on parle d'"imitation faciale".

³⁴ **Afférences**: les afférences sont les informations qui vont de la périphérie aux centres (de traitement). Par exemple, dans le cas de la perception tactile, l'information de la stimulation de la peau est périphérique. Elle est envoyée à des ensembles de neurones qui se trouvent dans la moelle épinière et dans le cerveau pour être traitée. La moelle épinière et le cerveau constituent le **système nerveux central** des Vertébrés. Par opposition, les **efférences** vont du centre vers la périphérie: l'on y trouve les réponses aux stimulations. Par exemple, si la stimulation cutanée nous "brûle", en recevant cette information, la moelle répondra par un réflexe de rétraction de la partie corporelle concernée.

lien sensori-moteur, ou *perception-production*, (iii) elle a au moins la possibilité de faire intervenir un appariement inter-perceptif, ou *inter-modal*, (iv) l'imitation et la *mémoire* peuvent "communiquer", (v) l'imitation étant celle d'actions, elle implique des *buts* et des *moyens* pour les atteindre, et (vi) l'imitation de comportements nouveaux peut mener à leur *apprentissage*.

La section suivante pénètre plus en profondeur dans les mécanismes potentiellement sous-jacents à l'imitation au cours du développement dans la jeune tradition de la psychologie cognitive³⁵.

II. THEORIES DEVELOPPEMENTALES ET MODELES DE L'IMITATION

1. L'imitation dans la théorie du développement de l'intelligence de Piaget

Pourquoi Piaget avançait-il donc que l'imitation (faciale) indique une *conscience intelligente* de la similarité entre l'expérience kinesthésique de ses propres actions et l'expérience visuelle des actions d'autrui (section I.2.)? Pour le comprendre, nous avons besoin de quelques éléments sur la place qu'il donne à l'imitation au sein du développement³⁶.

Des bases pour comprendre Piaget

Dans sa théorie du développement de l'intelligence, dans les deux premières années de vie qu'il nomme *période de l'intelligence sensori-motrice*, Piaget (1966) voit dans l'imitation la passerelle entre intelligence sensori-motrice et intelligence représentative³⁷ (Dolle, 1999, p. 151). En effet, d'après lui, dans le développement, l'imitation est le pré-requis de la fonction symbolique. La **fonction symbolique** (ou sémiotique) est l'application qui associe un sens aux choses (objets ou événements), en les *représentant* sous la forme de signes ou de symboles (Dolle, 1999, p. 148). Or, selon Piaget, bien que l'**imitation** au sens strict (ou *imitation en acte*) soit "une action par laquelle un modèle est reproduit, que ce processus dépende de la perception ou de la représentation"³⁸ (Nadel et Butterworth, 1999, p.2), l' **imitation sensori-motrice**³⁹ est le mécanisme qui, au cours du développement, mène à la dissociation entre signifiant et signifié en donnant la capacité de *représentation*. De plus, la fonction symbolique, chez Piaget, prend cinq formes: l'image mentale, le dessin, le jeu symbolique, le langage et *l'imitation différée*. De là, il distingue, à un niveau observable deux *comportements* imitatifs: l'**imitation immédiate**, pour laquelle le modèle est *présent*, et l'**imitation différée**, dans laquelle le modèle est *absent* des champs perceptifs. L'imitation différée est la forme intériorisée de l'imitation sensori-motrice et cette intériorisation permet la représentation et donc la fonction symbolique.

³⁵ La psychologie cognitive s'intéresse, comme son nom l'indique, à la **cognition**, dont une définition peut être, la "[c]apacité du système nerveux central de porter attention à des stimuli complexes, de les identifier et d'agir sur eux." (Purves *et al.*, 1999, p. 565).

³⁶ Pour une vision très générale des travaux de Piaget, voir Harris et Butterworth (2002), pp. 22-26.

³⁷ L' **intelligence sensori-motrice**, selon Piaget, est celle qui agit sur des *objets présents*, sans besoin de représentation, sans langage, sans concept, tandis que l'**intelligence représentative** opère sur des *représentations* que l'on peut voir comme des évocations de réalités absentes (Dolle, 1999, p. 147).

³⁸ "[A]n action by which a model is reproduced, whether the process depends on perception or representation" (Nadel et Butterworth, 1999, p.2).

³⁹ L'imitation sensori-motrice est celle qui existe dans la période sensori-motrice: elle n'utilise pas de représentations (voir plus bas).

Un développement par stades de l'imitation

Ainsi le développement de l'imitation fait-il partie intégrante de sa théorie du développement de l'intelligence, en particulier dans les stades dits "sensori-moteurs". Une synthèse de sa vision peut s'énoncer de la manière suivante⁴⁰:

- Stade I (0-1 mois). *L'imitation est absente*, parce que ce stade se limite aux activités réflexes: l'individu est sous l'emprise d'une réalité immédiate dans laquelle soi et l'autre sont confondus. Cette confusion est d'après Piaget la raison pour laquelle un nourrisson pleure quand un autre pleure: ces pleurs sont déclenchés et maintenus par un comportement réflexe. De plus, Piaget (1952, 1954) stipule que les différentes *modalités perceptives ne sont pas liées* les unes aux autres à la naissance (Meltzoff, 2002a) et que le lien entre perception et production se construit dans l'expérience sensori-motrice ultérieure.
- Stade II (1-4 mois). *Imitation sporadique de modèles vocaux et de modèles de coordination tête-main* (Kugiumutzakis, 1999, p. 36), *s'ils font partie du répertoire comportemental et si le comportement en question vient d'être produit par le sujet* pour pouvoir reconnaître l'analogie entre le modèle perçu et ce qu'il/elle produit. Cela se fait sans qu'il y ait une réelle différenciation entre, ni une réelle mise en correspondance de, soi et l'autre. Cette **pseudo-imitation** de mouvements perçus dans la même modalité perceptive que le modèle quand ils sont produits par soi (gestes visibles et audibles) a lieu grâce à l'assimilation dans les "**schèmes**⁴¹ réflexes" d'éléments de l'environnement par

⁴⁰ Dans la théorie de Piaget, chaque **stade** conserve les acquis en termes de connaissances et de capacités du/des stade-s qui le précède-nt. Plus précisément, selon Dolle (1999), Piaget retient cinq critères de définition et de délimitation d'un stade: "

1. *L'ordre de succession des acquisitions doit être constant [sic]*. Bien noter que "l'ordre de succession" ne signifie pas "la chronologie", car celle-ci est variable; elle dépend de l'expérience antérieure du sujet et pas seulement de sa maturation, du milieu social qui peut "accélérer ou retarder l'apparition d'un stade, ou même en empêcher la manifestation" [...]. Les âges qui sont donnés sont donc relatifs aux populations étudiées; mais ce qui est le plus important est l'ordre de succession, en ce sens qu'un caractère n'apparaîtra pas avant un autre chez un ensemble de sujets et après chez un autre ensemble.
2. *Les stades ont un caractère intégratif [sic]*, ce qui signifie que les structures construites à un niveau donné sont intégrées dans les structures du niveau suivant [...].
3. *Chaque stade doit se caractériser par une structure d'ensemble [sic]*. "Une structure, ce sera, par exemple, au niveau des opérations concrètes, un groupement, avec les caractères logiques du groupement qu'on trouve dans la classification ou la sériation" [...].
4. *Chaque stade comporte "à la fois un niveau de préparation d'une part, et d'achèvement, de l'autre" [sic]* [...]. Exemple, au niveau des opérations formelles, la période allant de 12-13 ans à 14-15 ans sera considérée comme le niveau de préparation, et le palier d'équilibre qui suit sera l'état d'achèvement.
5. "Mais comme la préparation d'acquisitions ultérieure peut porter sur plus d'un stade (avec des chevauchements divers entre certaines préparations plus courtes et d'autres plus longues), et comme, en second lieu, il existe des degrés divers de stabilité dans les achèvements, il est nécessaire de distinguer, en toute suite de stades, les processus de *formation* ou de genèse et *les formes d'équilibre finales* (au sens relatif): les dernières seules constituent les structures d'ensemble dont il a été question sous 3, tandis que les processus formateurs se présentent sous les aspects de différenciations successives de telles structures (différenciation de la structure antérieure et préparation de la suivante)" [...]" (Dolle, 1999, p. 61-62).

⁴¹ **Schème**: Globalement, un schème correspond à une structure qui (i) sert de base à l'activité du sujet, (ii) lui permet d'incorporer, i.e. *assimiler*, des éléments du milieu et (iii) est sujette à transformations (e.g. différenciation en schèmes nouveaux) sous l'effet des modifications du milieu, i.e. à *accommodation* au donné de l'expérience, pour s'adapter aux conditions nouvelles (voir Dolle, 1999, pp. 66-69). Les réflexes peuvent être vus comme des "pré-schèmes", au sens où les premiers schèmes sont issus des activités réflexives néonatales (Dolle, 1999, p. 104). Les **schèmes d'action** sont "ce qu'il y a de commun aux diverses répétitions ou applications de la même action [...]" par exemple, les schèmes de balancer un objet suspendu, de tirer un véhicule, de viser un objectif etc. " (Piaget, dans *Biologie et Connaissance*, p. 16, cité par Dolle, 1999, p.66-67). Les **schèmes sensori-moteurs** sont "les organisations sensori-motrices susceptibles d'application à un ensemble de situations

réaction circulaire primaire (Dolle, 1999, p.150). La **réaction circulaire primaire** est le premier type de réaction circulaire au cours du développement: elle est issue de la *coordination de réflexes* (Butterworth, 1999) et concerne donc son propre corps. Elle permet par sa variabilité intrinsèque de découvrir, par hasard, de nouveaux résultats (Dolle, 1999, p.106) et d'effectuer les premières modifications des schèmes en place (i.e. de les accommoder) de manière à prendre en compte les nouvelles informations ainsi découvertes et répétées. Cela nous éclaire sur le fait que pour Piaget "[...] au cours du second et du troisième mois deux conduites incontestablement nouvelles se font jour: la *protrusion de la langue* [sic], la succion du pouce" (Dolle, 1999, p. 105) parce que, d'après lui, aucun "réflexe héréditaire" ne correspond à ces comportements.

- Stade III (4 – 8/9 mois). *Imitation systématique et intentionnelle* (Dolle, 1999, p. 151) *de sons connus et de gestes visibles* (Kugiumutzakis, 1999, p. 36). Le modèle doit toujours faire partie du répertoire comportemental. Globalement, les actes de ce stade commencent à être intelligents car la différenciation entre moyens et but (nécessaire à l'intentionnalité) est émergente, grâce à l'apparition des réactions circulaires secondaires. En effet, une **réaction circulaire secondaire** est une réaction circulaire où les résultats que l'individu conserve par répétition sont obtenus lors d'interaction avec des *objets* autres que son propre corps. Elle est l'émergence de la différenciation entre fins et moyens parce qu'elle éloigne le comportement de répétition de l'activité réflexe (Dolle, 1999, p. 108). Cependant, cette réaction n'a lieu que si le résultat nouveau est compris comme étant la conséquence de son propre comportement.
- Stade IV (8/9 - 11/12 mois). *Imitation de gestes non-visibles connus, faciaux inclus* (Kugiumutzakis, 1999, p. 36). L'imitation des gestes non-visibles devient possible car "[a]u stade IV le sujet parvient, par l'intermédiaire d'"indices" manifestement intelligents, à repérer la correspondance entre le visage du modèle et les mouvements propres invisibles" (Piaget, 1976, p. 69) en faisant le lien entre la perception visuelle et l'exploration tactile de son visage et de celui d'autrui (Piaget, 1976, p. 41, obs. 25; Butterworth, 1999). Ainsi y a-t-il possibilité d'*apprendre* l'équivalence entre la configuration faciale de soi et de l'autre au niveau de la même modalité perceptive: cela permet non seulement l'émergence de la mise en correspondance visuo-kinesthésique mais aussi la "traduction" du modèle facial ou de ces propres mouvements dans la même modalité. De plus, à la même période, l'individu se donne des buts et sait les atteindre en *coordonnant* les schèmes à disposition (on parle de "schèmes secondaires"). Il y a dissociation entre les buts et les moyens pour les atteindre, donc "vraie" intention.
- Stade V (11/12 - 18 mois). *Imitation des modèles nouveaux, non-visibles inclus* (Dolle, 1999, p. 151). Ici, l'imitation se fait *pour* modifier les schèmes en place, s'inscrivant dans une période de quête active de moyens nouveaux par expérimentation. Cette expérimentation s'exprime par les **réactions circulaires tertiaires** qui apparaissent à ce stade et dont la caractéristique est la répétition de l'action *pour faire varier* le résultat, permettant ainsi d'établir "une table des correspondances" entre les actions et leurs conséquences, ou les moyens *variés* et des buts potentiels. Dans la théorie piagetienne,

analogues et témoignant ainsi d'assimilations reproductrices (répétition des mêmes activités), [d'assimilations] recognitives (reconnaître les objets en leur attribuant une signification en fonction du schème) et [d'assimilations] généralisatrices (avec différenciations) en fonction de situations nouvelles" (Piaget, dans *Etudes d'épistémologie Génétique, II*, p. 46, cité par Dolle, 1999, p. 67).

l'imitation de modèles nouveaux commence au même stade que les débuts de la **causalité**⁴² (objective) (Dolle, 1999, p. 133).

- Stade VI (18 - 24 mois). *Imitation différée*⁴³ (exemple, "faire-semblant" de parler au téléphone). Cette imitation est basée sur l'*intériorisation* de l'imitation (sensori-motrice) jusque là contrainte par la nécessité de la présence immédiate du modèle. Par la capacité de représentation qu'elle inaugure par celle du modèle en mémoire, l'imitation de ce stade permet l'intériorisation de l'expérimentation action-perception physique sous la forme d'une expérimentation action-perception "simulée", avec structuration et coordination des schèmes sous forme de combinaisons mentales (Dolle, 1999, p.116). Par exemple, Piaget décrit sa fille, Jacqueline, anticiper le fait qu'il faille poser ce qu'elle a dans la main et où il ne faut pas le poser pour pouvoir ouvrir une porte: d'après lui, ce comportement est fondé "sur la représentation ou la conscience des relations" (Dolle, 1999, p. 116). Il en découle que: (i) des *moyens nouveaux* sont inventés, (ii) l'individu commence à séparer le réel du mental et le *sens* (fonction sémiotique) émerge, (iii) l'imitation différée marque le passage de l'intelligence sensori-motrice à l'intelligence représentative, dans laquelle les objets deviennent permanents et (iv) ayant la possibilité d'expérimenter au niveau mental, le sujet peut prédire les conséquences de ses actions.

Qu'en retenir?

Au départ, l'imitation est absente et ne commence à se développer qu'à 1-4 mois, en prenant la forme de l'imitation sporadique d'uniquement ce qui est perceptible sur soi. En effet, tant que des liens entre modalités sensorielles ne sont pas appris, les comportements imitatifs se basent sur la comparaison intra-modale, comme dans le cas des modèles vocaux (1-4 mois). Ce n'est qu'à 8-12 mois que l'appariement inter-modal se développe par l'apprentissage visuo-**haptique**⁴⁴ de l'équivalence entre visage de soi et visage de l'autre, et permet l'imitation faciale de gestes non-visibles sur soi mais connus. A 12-18 mois, la quête de nouveauté pousse à imiter des modèles nouveaux, même les non-visibles, grâce à la connaissance de correspondances visuo-haptiques⁴⁵. La non-visibilité de modèles nouveaux induit leur représentation en mémoire, à 18-24 mois, ou mieux l'intériorisation de l'imitation jusque là sensori-motrice sous la forme de l'imitation différée.

Des déductions de sa théorie au regard du développement de l'imitation seront remises en question, comme nous le verrons par la suite. Il est toutefois important de noter que l'influence de sa vision a eu une portée telle que des observations effectuées de son temps (Lebib et Baudonnière, 1999; Nadel et Butterworth, 1999), parce qu'elles contredisaient en particulier l'absence d'imitation dans le premier mois de vie, ont été discréditées de manière systématique voire ignorées jusqu'à la fin du 20^{ième} siècle⁴⁶. Depuis, d'autres conceptions de

⁴² **Causalité**: Rapport de la cause à l'effet qu'elle produit. Nous ne savons pas si Piaget lui-même faisait un lien logique entre imitation et causalité, mais si tel n'était pas le cas, leur relation temporelle dans le modèle de développement par stade proposé par Piaget a été vraisemblablement notée et exploitée par ses successeurs.

⁴³ Au cours du développement ultérieur, l'imitation différée cédera le pas à une imitation dite "représentative": "l'imitation différée est intermédiaire entre l'imitation sensori-motrice et l'imitation représentative" (Piaget, 1976, p.232).

⁴⁴ **Haptique**: La modalité haptique est la combinaison des informations kinesthésique et tactile lors de mouvements exploratoires (Purves *et al.*, 1998, p.154).

⁴⁵ L'on retrouve que pour Piaget l'imitation visuo-motrice dérive de l'appariement visuo-kinesthésique (section I.2.).

⁴⁶ Tout cela n'enlève, évidemment, rien au fait que sa théorie a eu des impacts fructueux au moins en psychologie du développement, en pédagogie et en éthologie (par exemple dans l'étude de primates non-humains, voir Bard *et al.*, 1999).

l'imitation précoce ont été proposées dans le domaine de la psychologie, mais toutes continuent de se référer à celle de Piaget (*la* théorie dite "classique" par des psychologues anglo-saxons).

Il reste qu'une caractéristique indéniable de la théorie piagétienne est sa cohérence (i) extrinsèque, puisqu'elle rend compte de presque toutes les observations de Piaget et (ii) intrinsèque, au sens où le développement temporel d'une capacité cognitive donnée est cohérent avec celui des autres domaines de l'intelligence (au sens de Piaget). Ce dernier point permet de mettre en évidence des liens, peut-être différents de ceux fournis par son interprétation, entre différentes capacités cognitives, telles que l'imitation et l'intention, ou l'imitation et la causalité. Cette lecture renouvelée des données de Piaget semble sous-tendre la vision de ses successeurs comme nous allons le voir.

2. AIM: base d'une nouvelle théorie du développement

Dans le but de rendre compte de la capacité d'imitation faciale des nouveaux-nés, mise en évidence dans des conditions minutieusement contrôlées dans les années 80, comme nous le verrons plus loin, Meltzoff et Moore (1997) ont proposé l'hypothèse de la cartographie intermodale active (*Active Intermodal Mapping*, ou *AIM*). Elle s'inscrit dans une théorie développementale de l'imitation qui veut rendre compte du développement d'autres capacités basées d'après eux sur les mécanismes sous-jacents à l'imitation précoce⁴⁷: l'attribution d'états mentaux⁴⁸ à autrui d'après son comportement de surface (ou utilisation d'une théorie de l'esprit), le langage, la mémoire (Meltzoff, 1999) et l'empathie (Meltzoff, 2002b).

L'axiome de départ énonce que *l'imitation est un processus d'appariement à la cible* qui est tout ou partie de l'acte à imiter: Cette idée se fait l'écho de leur description de l'acte imitatif (section I.2.). Toutefois, cet "axiome" trouve sa crédibilité à l'issue d'une démonstration, qui déduit, de dix phénomènes-clés⁴⁹, trois caractéristiques relatives à l'imitation précoce (Meltzoff et Moore, 1997): (i) elle est dirigée vers un but, l'appariement à la cible, et cela, sans confusion de l'effecteur d'où sa spécificité (ii) elle génère des actes nouveaux sur des chemins non-aléatoires d'atteinte de la cible, d'où le terme *processus* et (iii) elle est basée sur des représentations, sachant qu'elle n'est pas prisonnière de la perception immédiate de la cible.

Le monde des nourrissons selon Meltzoff et Moore (1997)

Les nouveau-nés peuvent, d'après ces auteurs, identifier visuellement des parties du corps: la tête, les sourcils, la *mandibule*, les *lèvres*, la *langue*, les bras, les mains, les doigts, le tronc, les jambes et les pieds. Ces objets seront désignés par le terme d'"**organe**". Les organes sont des objets intermodaux, au sens où ils sont perceptibles par différentes modalités à la fois, notamment la vision et la proprioception. L'**identification de l'organe** par les nouveau-nés est, dans cette conception, une information perceptive qui pourrait, notamment, émerger des caractéristiques dynamiques des mouvements de l'organe (sa "signature cinématique")⁵⁰.

⁴⁷ On retrouve la proposition de Piaget selon laquelle l'imitation a une place fondamentale dans le développement des structures de la pensée, ou, dans un langage plus contemporain, dans le développement cognitif (Cf. section II.1.).

⁴⁸ Ces états mentaux comprennent notamment les croyances, intentions, motivations et émotions.

⁴⁹ Ces phénomènes vous seront donnés, en note, au cours de la section III.A.1..

⁵⁰ Ce point n'est pas tranché par les auteurs. Ils répertorient trois manières dont l'identification de l'organe pourrait avoir lieu, tout en étant compatibles avec leur conception de l'imitation au cours du développement. 1/ Les organes pourraient être identifiés à partir de leur forme dès la naissance. Goren *et al.* (1975) et Morton et Johnson (1991), proposent que les résultats d'expériences montrant que les nouveau-nés préfèrent les visages à d'autres patrons visuels de caractéristiques semblables s'expliquent par le fait que "à travers l'évolution la

Un corps ou un visage correspond à un ensemble cohérent d'organes liés les uns aux autres. Ainsi une configuration corporelle ou faciale donnée peut-elle être décrite selon la position relative des organes qui le composent les uns par rapport aux autres, ou mieux en termes de **relations-entre-organes**.

Un mouvement de soi au niveau de la proprioception et un mouvement d'autrui au niveau de la vision (dans le cas de l'imitation visuo-motrice) aboutissent chacun à un **état final** qui est dans les deux cas descriptible en termes de relations-entre-organes. Si le mouvement est un acte, cet état final correspond au but de l'acte. Puisque les organes sont des objets intermodaux, il en est de même pour les relations-entre-organes. Or, d'après Meltzoff et Moore, la relation-entre-organes est utilisée par tous les canaux perceptifs pour décrire les configurations corporelles. Il en découle que différentes modalités perceptives utilisent la même métrique. Cela leur permet donc de "communiquer" entre elles, dans la *lingua franca* relation-entre-organes, et de se mettre "d'accord": on parle de **coordination intermodale**⁵¹.

Vu que la proprioception est le pendant perceptif des mouvements de soi, la coordination intermodale représente une passerelle entre la production de soi et la perception de l'autre. Autrement dit, il y a un *lien perception-production indirect* entre les mouvements de soi et ceux d'autrui. Ce lien se retrouve dans une **représentation supra-modale**⁵² des mouvements de soi et de l'autre. Ainsi les mouvements de soi et d'autrui peuvent-ils être comparés, et potentiellement appariés, dans ce cadre commun de représentation.

De plus, le lien perception-production dans le cas des mouvements de soi est nourri par la proprioception: l'expérience de ses propres activités associée à la coordination intermodale permet une **cartographie intermodale** entre les mouvements, dans une première phase du développement, et les états finaux de relation-entre-organes auxquels ils mènent. Toutefois, cette carte primaire n'intègre pas un mouvement donné et son état final en une même unité. Meltzoff et Moore proposent que ce processus de cartographie production-perception est analogue à celui du babillage canonique dans le développement de la parole: ils parlent de

structure d'un visage serait devenue une unité perceptive distinctive à laquelle on répond à la naissance" (Meltzoff et Moore, 1997, p. 183: "through evolution the structure of a face has become a distinctive perceptual unit responded to at birth"). La découverte que des sites cérébraux spécifiques s'activent pour la présentation visuelle de régions, telles que des parties du visage et des mains, chez des Macaques adultes irait dans ce sens. Dans ce cas, "l'identification de l'organe est un donné perceptif, préadapté par l'évolution" (Meltzoff et Moore, 1997, p. 184: "organ identification is a perceptual given, preadapted by evolution"). 2/ Les organes sont identifiables à partir de leur mouvement parce que chaque organe a un unique patron spatio-temporel de mouvement. "Il est possible que les nourrissons parviennent à individuer leurs organes à travers le contrôle proprioceptif [sic] de leurs propres actions. La proposition est que ces patrons spatio-temporels de mouvements uniques, que nous appelons 'signatures cinétiques', sont reconnus comme équivalents au niveau inter-modal quand ils sont effectués par soi et par autrui" (Meltzoff et Moore, 1997, p. 184: "It may be that infants come to individuate their organs through proprioceptive monitoring of their own actions. The claim is that these unique spatiotemporal movement patterns, which we call 'kinetic signatures', are recognized as cross-modally equivalent when done by self and other."). Meltzoff et Moore (1997) donnent pour données compatibles avec la possibilité d'une telle cartographie précoce des organes le fait que les mouvements biologiques humains, tels que la marche, présentés sous la forme de points lumineux sont des patrons auxquels les nourrissons sont particulièrement sensibles (e.g. Bretenthal, 1996). Ici, l'identification de l'organe émerge des mouvements uniques que chaque organe peut effectuer (Meltzoff et Moore, 1997, p. 184: "organ identification emerges from the unique movements each organ can make.") 3/ "Les actes humains biologiquement pertinents sont codés de manière redondante, spécifiés par les relations de configuration entre les organes et par des signatures cinétiques" (Meltzoff et Moore, 1997, p. 184: "[B]iologically relevant human acts are redundantly coded, specified both by the configurational relations between organs and by kinetic signatures").

⁵¹ Cette idée est en continuité avec Guillaume (1926, Cf. section I.2.)

⁵² Le terme "[s]upramodal" est utilisé, suivant T. Bower [...], pour signifier que la représentation n'est pas spécifique d'une unique modalité sensorielle" (Meltzoff et Moore, 2000, p. 175: "[s]upramodal" is used, following T. Bower [...], to denote that the representation is not particular to one sensory modality alone").

babillage corporel, et précisent que celui-ci peut commencer *in utero*, vu la motricité observée dans la période pré-natale⁵³ (voir, par exemple, chapitre 1, section I.3.c.).

Dans cette vision, l'expérience de ses propres activités aide à la décomposition des mouvements de soi et de l'autre en termes de relations-entre-organes: cette décomposition permet d'aboutir à l'intégration des différents états par lesquels la relation-entre-organes passe au cours du mouvement. En d'autres termes, un mouvement donné, qu'il soit produit par soi ou par autrui, sera décrit par la trajectoire d'une **transformation-entre-organes**, qui est un objet intermodal comme les états qui la composent. Si le mouvement est un acte, cette transformation correspond au moyen de l'acte. Plusieurs transformations-entre-organes peuvent mener au même état final (but): ces mouvements (moyens) sont équivalents. Cette équivalence peut être cartographiée, en intégrant ces transformations-entre-organes et l'état final de relation-entre-organe associé en une même unité, grâce à l'expérience de son propre corps et de l'observation de l'activité d'autrui. L'on obtient, ainsi, un **espace des actes** qui fournit la correspondance entre moyens multiples et buts potentiels. D'après Meltzoff et Moore, assez tôt dans le développement, cette nouvelle carte intermodale prend la place fonctionnelle de la première dont elle dérive et qui donnait la correspondance mouvement-état final. De plus, les dimensions de l'espace des actes⁵⁴ pourraient provenir de l'observation des états finaux de relation-entre-organes des actes d'autrui qui seraient depuis différenciés de ceux de soi.

Le modèle AIM ou comment les nouveau-nés pourraient imiter des modèles faciaux

Dans l'imitation, il y a plusieurs actes: (i) l'acte du/de la démonstrateur/trice dont le but est le modèle, (ii) l'acte imitatif dont le but est l'appariement au modèle (ou cible de l'acte imitatif), et dont le moyen est le processus d'appariement à la cible et, entre les deux, (ii) l'acte courant de l'observateur/trice (nourrisson). Ce dernier acte est mis à jour au cours du processus d'appariement: son but à chaque itération est de corriger l'écart entre le résultat perceptif de l'acte précédant et la cible de l'acte imitatif.

Pour rendre compte de l'imitation faciale des nouveau-nés, Meltzoff et Moore (1997) posent trois systèmes qui interagissent au cours du processus d'appariement (Figure 3-1): le système perceptif, le système de représentation supra-modal et le système de l'action.

Le **système perceptif** fournit (i) la perception proprioceptive de l'état des relations-entre-organes de son propre corps, avant d'agir ou à l'issue du dernier acte produit, et (ii) la perception visuelle⁵⁵ du *but* de l'acte du/de la démonstrateur/trice, qui correspond à l'identification de l'organe en mouvement et à l'état final des relations-entre-organes issu du mouvement. Cet état est la cible de l'acte imitatif.

Le **système de représentation supramodal** reçoit les deux perceptions ci-dessus et les compare entre elles, au niveau de la représentation supramodale des mouvements de soi et de l'autre. S'il y a appariement entre l'état de relation-entre-organe de soi et l'état final de relation-entre-organe de l'acte du/de la démonstrateur/trice, le processus d'appariement n'est pas déclenché (comme quand le/la démonstrateur/trice imite l'observateur/trice) ou s'achève s'il a eu lieu. Sinon, le non-recouvrement des états définit une nouvelle cible pour l'acte

⁵³ L'idée que l'expérience prénatale de ses propres activités contribue à l'émergence de l'imitation néonatale se trouve aussi chez Butterworth (1999).

⁵⁴ Dans Meltzoff et Moore (1997), les dimensions de l'espace des actes correspondent globalement à des transformations élémentaires des composantes vectorielles du mouvement de l'organe, telles que l'avancement, la latéralité et le degré du mouvement.

⁵⁵ La perception visuelle se trouve sur la figure au niveau du champ extéroceptif, i. e. là où sont perçus les événements et objets de l'environnement.

suyant telle qu'elle pourrait compenser l'erreur, ou écart entre l'état de relation-entre-organe de soi et l'état final cible de l'imitation.

Le **système d'action** reçoit cette nouvelle cible et la cherche dans la carte de correspondance entre états finaux de relation-entre-organes et mouvements, qui se trouve dans ce système. Si l'état cible existe, le mouvement correspondant est exécuté: les commandes sont envoyées par le système d'action au corps. Si l'état final d'intérêt n'existe pas, le système d'action essaie d'isoler une composante manquante, qui, elle, existe dans la carte intermodale, à partir de l'erreur : l'extraction de cette composante résulte d'une "soustraction" entre état courant de soi et état final cible⁵⁶. Le mouvement associé à cet état est exécuté. Dans tous les cas, l'exécution du nouvel acte met à jour la perception proprioceptive de l'état des relations-entre-organe de son propre corps, ce qui a pour effet d'amorcer un nouveau cycle dans le processus d'appariement.

En d'autres termes, à chaque cycle du processus d'appariement le sujet essaie *activement* d'apparier son état de relation-entre-organes à celui de la cible en la *situant sur la carte* mouvement-états finaux de relation-entre-organes. Or, cette carte est *intermodale* de par la métrique (de relation-entre-organes) qu'elle utilise. D'où le nom de "cartographie intermodale active" donné à ce modèle. De plus, le processus d'appariement décrit un processus manifeste d'**apprentissage moteur**⁵⁷ (du Lac, 2001). Il en découle que le nourrisson détient les outils lui permettant d'apprendre des moyens d'atteindre des cibles, contribuant ainsi à la formation progressive d'un espace des actes.

Il est bon de noter que le modèle proposé par Meltzoff et Moore s'oppose à une explication du type comportement réflexe ou conditionné pour rendre compte de l'imitation précoce: le processus d'appariement n'a pas besoin de la présence immédiate du modèle pour se faire. En effet, en vertu de sa nature représentationnelle, la cible de l'imitation reste en mémoire, dans le système supramodal, pour son imitation ultérieure, ou imitation différée, potentielle (section III.A.1.b.). De plus, la description d'une cible en terme de relations-entre-organes par le système perceptif est nécessaire et suffisante pour qu'elle soit représentée: une cible même nouvelle n'a pas besoin de passer par le système de l'action, i.e. d'être exécutée, pour être mémorisée et donc pour son imitation différée, ou apprentissage ultérieur des moyens de l'atteindre.

⁵⁶ D'après Meltzoff et Moore (1997), cette composante est isolée grâce à sa saillance, attirant ainsi l'attention du sujet.

⁵⁷ A un niveau plus général que l'imitation, du Lac (2001) nous informe que "[l]a plupart des mouvements ne sont pas effectués parfaitement la première fois, mais au contraire demandent de longues périodes de pratique. Durant la pratique, nous détectons des erreurs de performance motrice puis modifions les mouvements subséquents pour réduire ou éliminer ces erreurs. Le processus itératif d'amélioration de la performance motrice en exécutant des mouvements, en identifiant des erreurs, et en corrigeant ces erreurs dans les mouvements suivants est appelé **apprentissage moteur** [sic]." (p. 571: "Most movements are not performed perfectly the first time around, but instead require extensive periods of practice. During practice, we detect errors in motor performance and then modify subsequent movements to reduce or eliminate those errors. The iterative process of improving motor performance by executing movements, identifying errors, and correcting those errors in subsequent movements is called motor learning").

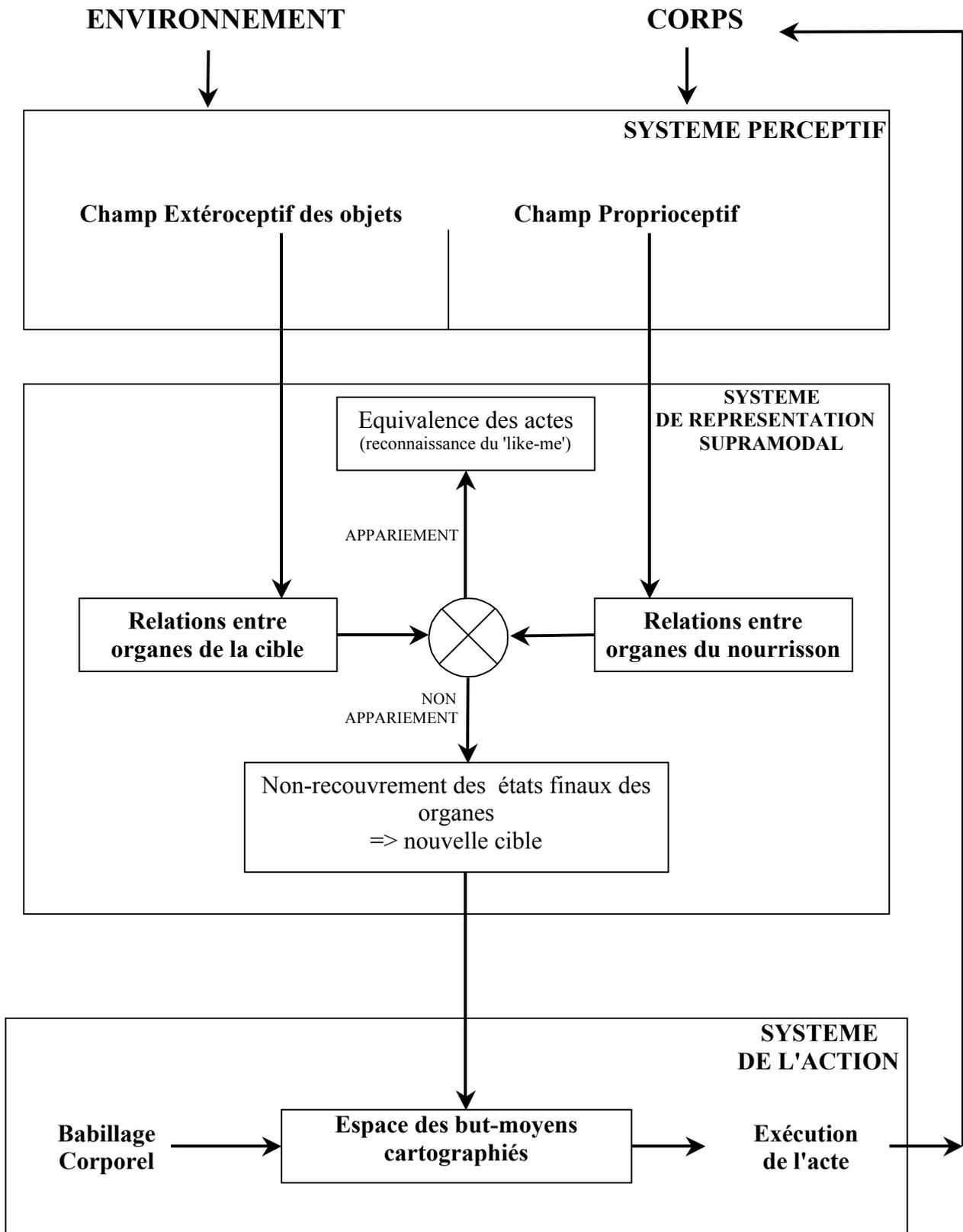


Figure 3-1: Le modèle de cartographie intermodale active (AIM), pour les mécanismes sous-jacents à l'imitation faciale précoce, de 0 à 2 mois (schéma tiré et modifié de Meltzoff et Moore, 1997).

Des changements développementaux dans l'imitation

Meltzoff et Moore (1997) proposent, données expérimentales à l'appui, qu'au cours du développement, l'être humain imite d'abord des actions purement corporelles, puis des actions sur des objets et seulement après l'utilisation d'objets comme outils pour agir sur d'autres objets⁵⁸.

En parallèle, ils donnent des mécanismes de l'imitation précoce un développement qui suit quatre changements⁵⁹ à partir de la naissance:

- Naissance: La *coordination intermodale* est présente⁶⁰ et les *mouvements* de soi et d'autrui sont décrits en terme de *relation-entre-organes*, ce qui fournit un lien perception-production indirect, via la proprioception, entre les deux protagonistes dont les mouvements sont différenciés⁶¹. De plus, (i) la cible de l'acte imitatif est *l'état final* de relations-entre-organes, ou but, de l'acte du/de la démonstrateur/trice et (ii) la cartographie intermodale porte sur la *correspondance entre les mouvements et les états finaux* de relation-entre-organes auxquels ils mènent, sans qu'états finaux et mouvements associés ne soient intégrés en une même unité. De là, les nouveau-nés peuvent apparier leur état de relation-entre-organes à la cible, en utilisant des moyens différents de ceux du/de la démonstrateur/trice.

- 1^{er} changement: Il y a *intégration* (i) des différents états du mouvement en la transformation-entre-organes correspondante, et (ii) entre un état final de la relation-entre-organes donné et les mouvements qui y aboutissent en une seule unité. Il en découle que (i) les mouvements de soi sont décrits en termes de *transformation-entre-organes* et (ii) la cible de l'acte imitatif est décrite en termes de transformation-entre-organes dirigée vers un but, ou d'*acte humain*⁶². De plus, la cartographie intermodale porte désormais sur la correspondance entre les transformations-entre-organes équivalentes et leur état final de relation-entre-organes, ou but, commun: un *espace des actes* émerge. Puisque le chemin des actes est spécifié, les nourrissons peuvent apparier leurs relations-entre-organes aux *moyens* utilisés par le/la démonstrateur/trice. Réciproquement, un acte imitatif de la part d'autrui à l'encontre de soi est désormais reconnu bien que les comportements de l'imitateur/trice soient pris pour une conséquence du comportement propre. De plus, l'espace des actes permet l'imitation de *cibles nouvelles*, à partir de moyens connus. En effet, vu que l'observation des états finaux de relation-entre-organes des actes des autres fournit les dimensions de l'espace des actes, les nourrissons peuvent trouver des moyens d'atteindre des états de relation-entre-organes non expérimentés par soi, mais inclus dans les limites de cet espace, en interpolant ces points inconnus à partir de liens connus entre états finaux et transformations-entre-organes. Autrement dit, grâce à la formation d'un espace des actes, il devient possible d'imiter des cibles nouvelles dans les limites de l'"expérience étendue": par une intégration des différents moyens connus, mis en œuvre au cours du processus de correction, l'individu découvre de nouveaux moyens d'atteindre ces cibles.

⁵⁸ On retrouve l'idée de Piaget selon laquelle le développement sensori-moteur va de soi vers le monde extérieur (Section II.1.).

⁵⁹ Ils ne font pas référence à des stades.

⁶⁰ Cette idée constitue le fondement d'où découle la majorité des distinctions entre la théorie en devenir de Meltzoff et Moore et celle de Piaget pour qui les modalités sensorielles ne sont pas coordonnées dès la naissance (Cf. section II.1.).

⁶¹ Ce point aussi se distingue de la vision de Piaget pour qui les nouveaux-nés vivent dans un "chaos perceptif" où les informations provenant du monde extérieur et du monde intérieur sont confondues (Cf. section II.1., ex. confusion des pleurs)

⁶² La cible est donc désormais tout l'acte du/de la démonstrateur/trice.

- 2^{ème} changement: L'imitation passe de l'appariement entre les actes de soi et ceux de l'autre à une *relation d'appariement* interpersonnel. Les actes d'autrui qui s'apparient à ceux de soi, notamment lorsque le sujet est imité, mènent à la compréhension que "l'autre agit *comme* j'agis": l'autre et soi sont compris comme étant deux *agents* distincts, dans un *référentiel égo-centré*. Leurs deux comportements n'entretiennent plus un lien de causalité, mais l'*appariement interpersonnel* aide au contrôle de l'interaction. Par exemple, un acte imitatif d'autrui à l'encontre de soi peut susciter des changements brusques de comportement, tout en observant l'imitateur/trice, pour tester si l'on est imité.
- 3^{ème} changement: Il y a un *regain d'intérêt pour les parties non-visibles* de son propre corps, avec une tendance à l'exploration haptique des organes homologues chez soi et l'autre. L'imitation d'autrui à l'encontre de soi devient l'occasion d'utiliser le corps de l'autre comme miroir biologique du sien. La *différenciation de chaque modalité* perceptive se développe sur la base des informations spécifiques à chacune d'elle, ainsi collectées⁶³. Cela permet de nourrir la perception supramodale du corps de soi de l'information sens-spécifique relative à: (i) ce à quoi les parties non-visibles ressemblent et (ii) l'*homologie entre soi et l'autre* au niveau de la perception que chaque protagoniste a de son corps. Il en découle une nouvelle compréhension modalité-spécifique (i) de soi dans un *référentiel autrui-centré*, telle que "ma protrusion de langue *ressemble* à cette protrusion de langue que je vois", mais aussi (ii) de ce qu'est être "*comme l'autre*". La relation d'appariement interpersonnel devient bidirectionnelle et contribue au développement de l'intersubjectivité⁶⁴.
- 4^{ème} changement: L'imitation d'un acte peut se passer de la perception de l'atteinte de son but. En d'autres termes, le but d'un acte peut être *inféré* sur la base de son propre espace des actes qui peut désormais être attribué à autrui. L'individu fait donc un pas en avant vers la compréhension des intentions d'autrui: un germe de théorie de l'esprit⁶⁵ émerge.

⁶³ Si pour Piaget, l'exploration visuo-haptique permet le développement d'une coordination intermodale (Section II.2.), pour Meltzoff et Moore, elle aide à une différenciation entre modalités.

⁶⁴ Pour une définition de l'intersubjectivité voir par exemple Trevarthen (2001).

⁶⁵ Pour des éléments sur la notion de théorie de l'esprit, voir par exemple Gordon (2001).

III. DONNEES ONTOGENETIQUES ET PHYLOGENETIQUES SUR L'IMITATION

En introduisant ce chapitre, à la question du quand et comment la parole ambiante peut être apprise, nous avons d'abord répondu par le constat d'une séquence développementale allant de l'influence de cette langue en perception⁶⁶ à son influence en production⁶⁷, en passant par la capacité d'imitation de stimuli audio-visuels de parole aux alentours de 4 mois. De là, nous avons proposé un lien de causalité schématique tel que l'imitation vocale précoce dirige la production *vers* la langue de l'environnement, d'après l'état en cours de la structure du paysage perceptif vis-à-vis de cette langue. Or, cette réponse ne saurait ignorer au moins trois questions qui en découlent et qui s'appellent les unes les autres:

Q1 Quand et comment l'imitation orofaciale humaine précoce a-t-elle lieu?

Q2 Quand et comment les mécanismes potentiellement sous-jacents à cette forme d'apprentissage auraient-ils pu se mettre en place à un "si jeune âge"?

Q3 Quand et comment avons-nous pu en arriver là, nous, animaux élus par le verbe qui descend du ciel quand nous descendions de l'arbre?

Ces questions constituent des graines de projets de recherche à part entière. Sans avoir la prétention d'y répondre de manière exhaustive, nous essaierons de mieux les cerner afin d'appuyer la plausibilité de l'imitation comme mécanisme d'apprentissage de la parole ambiante, mais aussi afin d'apporter une petite pierre à l'édifice d'une vision incarnée (ou *embodied*) du développement du langage et de son émergence évolutive.

Cette section a pour but explicite de situer le phénomène d'imitation orofaciale précoce en intégrant au mieux les apports de la psychologie développementale et de l'éthologie. Nous avons donc rassemblé des données sur l'imitation au cours du développement humain, et dans le règne animal. Le thème principal de l'imitation chez les nourrissons humains se déclinera en trois points: (a) l'imitation juvénile de modèles visuels et (b) la coordination intermodale que suppose l'imitation de modèles visuels orofaciaux alimenteront (c) notre appréhension du phénomène d'imitation orofaciale précoce de stimuli audio-visuels de parole. Le thème secondaire nous permettra de considérer la présence de la capacité d'imitation chez des espèces sociales autres que l'espèce humaine. Dans ce cadre, notre attention se penchera en particulier sur le cas des Hominés non-humains dont le génome est très proche du nôtre. L'idée sous-jacente à cette perspective phylogénétique, que nous recroiserons dans la section III.A.3., est que *le génome correspond à un ensemble de contraintes biologiques délimitant l'univers des possibles comportementaux*, et que ces contraintes ont pu être héritées de l'ancêtre putatif commun à tous les Hominés.

⁶⁶ A 4 jours pour la prosodie, 6 mois pour les voyelles, et 10-12 mois pour les consonnes.

⁶⁷ Au maximum à 6 mois pour la prosodie et à 10 mois pour les consonnes et voyelles en devenir.

A. Imitation orofaciale chez les nourrissons humains

*Pourquoi parler d'imitation oro-faciale de parole plutôt que d'imitation vocale?*⁶⁸

A cette question, une première réponse nous vient de l'histoire récente de l'imitation vocale précoce en psycho-phonétique. En effet, c'est au cours d'une étude expérimentale sur la capacité de détecter la **congruence audio-visuelle**, ou cohérence entre l'audition et la vision, de stimuli de parole que Kuhl et Meltzoff (1982) ont rapporté que des nourrissons de 3-4 mois imitaient ce qu'ils percevaient. Ce fait a peu de chance d'être le fruit du hasard, comme nous le verrons plus bas. Un second point concerne l'entrée dans la communication humaine. La première forme standard d'interaction sociale au cours du développement d'*Homo sapiens* s'ancre, sauf "empêchements"⁶⁹, dans le cadre de la *communication face-à-face*, avec ses corrélats du contact des regards et des échanges à tour de rôle (ou **turn-taking**) de vocalisations entre nourrissons et congénères (Abry, Vilain, Schwartz, 2004). De plus, une continuité avec l'âge adulte apparaît, si l'on s'éloigne des technologies de la communication parlée qui, ne l'oublions pas, n'existent pas dans tous les groupes sociaux humains ni même à toutes époques: la conversation face-à-face reste la situation d'occurrence la plus fréquente de la parole. Notre dernier argument fait appel aux études effectuées par plusieurs groupes de recherche, auxquels l'ICP participe: elles montrent que la perception de la parole est au moins audio-visuelle, si ce n'est audio-visuo-haptique (e.g. Schwartz, 2004).

Ainsi, avons-nous choisi d'aborder l'imitation orofaciale précoce de modèles visuels puis celle de parole audio-visuelle. Toutefois, ce choix n'est pas simplement motivé par une vision quelque peu réductionniste qui désire faire varier un puis deux paramètre-s perceptif-s. En effet, l'imitation orofaciale de stimuli visuels offre, de part ses apports théoriques et expérimentaux, et les remises en question qu'elle suscite (voir section I. 2.), un terreau riche pour appréhender l'imitation précoce de *modèles audio-visuels de parole* sous une lumière appelant à une continuité dans le développement cognitif (Cf. **Q1**).

1. L'imitation visuelle précoce de modèles oro-faciaux⁷⁰

Du doute

En 1957, Zazzo rapporte l'imitation de la protrusion de la langue, par un nouveau-né humain de 25 jours (Lebib et Baudonnière, 1999; Baudonnière, 1997). Toute observation du même ordre a été reçue avec beaucoup de scepticisme, pour ne pas rentrer dans le cadre piagétien (voir section II.1.). En particulier, de telles observations remettent en question l'état initial de non-coordination intermodale⁷¹ proposé par Piaget (1945/1976) ainsi que les apprentissages ultérieurs qu'il implique. Malgré de nombreuses études rapportant ce comportement, dont celle de Meltzoff et Moore (1977) qui utilisent d'autres modèles oro-faciaux et un modèle de la main, l'encre de la critique a coulé, et parfois coule encore (Lebib et Baudonnière, 1999).

⁶⁸ L'imitation faciale est basiquement *visuo*-motrice, et l'imitation vocale est par défaut *audio*-motrice.

⁶⁹ Le terme "empêchement" regroupe ici les différences par rapport à la majorité des individus dans des capacités psycho-biologiques, le plus fréquemment observées dans l'espèce humaine, qui entrent en jeu dans le développement d'une communication qui satisfait des critères sociaux. L'on observe, par exemple, que des individus dont le comportement a été classé sous le terme général d'autisme "évitent" le contact du regard.

⁷⁰ En réalité, cette section ne s'arrêtera pas aux modèles orofaciaux: elle présentera aussi des données développementales sur l'imitation d'actions sur des objets, et d'action sur des objets pour agir sur d'autres objets. En effet, nous pensons que des caractéristiques du décours développemental de l'imitation visuelle en général pourraient inspirer l'étude et la compréhension du décours développemental de l'imitation de modèles de parole.

⁷¹ Pour l'état initial se référer au stade I du développement de l'intelligence sensori-motrice (section II.1.).

Dans toute science, la mise en évidence d'un phénomène est conditionnée par la méthode utilisée qui est elle-même fonction de la vision que l'on a de ce phénomène et des outils à disposition pour l'étudier. Dans le cas de l'imitation précoce, le paradigme expérimental de base consiste en la présentation d'un geste suivi d'une période d'observation où l'on "attend de voir" si le sujet effectue le même geste. Ce protocole intuitif est le plus simple à mettre en oeuvre et a été utilisé dans les études pionnières de la psychologie du développement (eg. Piaget, 1945/1976). Cependant, cette procédure soulève quelques problèmes méthodologiques, car, comme en éthologie (section I.1.), la difficulté réside dans l'élimination des interprétations alternatives à l'imitation. En effet, dans le *paradigme intuitif*, l'on peut rendre compte de la similarité entre le comportement présenté et le comportement subséquent du sujet par l'imitation mais aussi par les phénomènes suivants:

(Di) La réponse observée fait partie de l'activité basale du sujet. Elle aurait donc eu lieu avec ou sans "modèle". En d'autres termes, l'effet est un artefact résultant d'une *pure coïncidence*. De là, l'on est en droit de s'interroger sur qui imite qui.

(Dii) La réponse fait partie d'un comportement d'*éveil global* due à la perception d'une partie de corps humain en mouvement.

(Diii) La réponse est le résultat d'un *conditionnement opérant* (Cf. section I.1.) que le sujet a subi de la part de l'entourage ou subit de la part de l'expérimentateur/trice. Ce/cette dernier/ère, après avoir présenté le comportement à l'individu, a renforcé ou renforce, par des sourires ou toute autre stimulation agréable, l'expression fortuite d'un comportement similaire. Cette hypothèse est défendue par les tenants/antes d'une vision *purement comportementaliste* du développement, ou du type "théorie de l'apprentissage", dont la tendance est **environnementaliste**⁷².

(Div) La réponse est un comportement *instantanément*, car pré-câblé, déclenché par un **stimulus** dit "**déclencheur**"⁷³, qui serait, ici, tout ou partie du modèle, et qui donnerait pour réaction tout ou partie du stimulus lui-même. Cette explication est en harmonie avec la présence de **réflexes**⁷⁴ dans la période néonatale mais aussi prénatale et cela dès 8 semaines de gestation (voir Harris et Butterworth, 2002, p. 67). Aussi, l'on s'attend à ce que l'imitation précoce soit sujette à *extinction* comme la majorité des réflexes néonataux. Cette hypothèse

⁷² Les théories du développement dites "**environnementalistes**" donnent plus de poids aux apports de l'environnement qu'aux mécanismes de développement "internes" au sujet.

⁷³ **Stimulus déclencheur** : Les stimuli déclencheurs sont des phénomènes physiques extérieurs à l'animal qui induisent des discontinuités *instantanées* dans la qualité des comportements exprimés (ou réponses). A un stimulus spécifique correspond une réponse spécifique. Les types de réponses à un stimulus déclencheur vont du réflexe à la séquence comportementale complexe. Cette réponse peut être un **comportement stéréotypé**, c'est-à-dire un comportement dont les éléments constitutifs s'enchaînent de manière invariable et qui se poursuit *même si le déclencheur est substitué* (contrairement aux comportements dits "non-stéréotypés"). L'on y trouve, par exemple, le comportement de roulement de l'œuf trouvé hors du nid pour l'y ramener chez les Oies. Toutefois, la substitution répétée du signal déclencheur (ici, l'œuf hors du nid) au cours de la réponse peut faire disparaître la réaction.

⁷⁴ **Réflexe**: "Réponse motrice *involontaire* [sic] et *stéréotypée* [sic] déclenchée par un stimulus particulier." (Purves *et al.*, 1997, p. 575). Par exemple, d'après Hooker (1939), la stimulation de la zone péri-orale d'un fœtus de 8 semaines (à l'aide d'un cheveu) déclenche l'ouverture de la bouche ainsi que des mouvements du tronc et des membres (Harris et Butterworth, 2002, p. 67). Le **réflexe tonique du coup** en est un autre exemple (Harris et Butterworth, p. 68). Ce réflexe est un des premiers réflexes posturaux humains, dans lequel le bras et la jambe s'étendent du côté vers lequel la tête se tourne, avec flexion du bras et de la jambe du côté opposé. Ce réflexe est facile à observer à partir de 28 semaines post-conception et est présent jusqu'à l'âge (post-natal) de 8 mois.

est défendue par les tenants/antes d'une vision *purement maturationaliste* du développement, dont la tendance est **nativiste**⁷⁵.

Vu les interprétations alternatives qu'il suscite, sans pouvoir trancher, nous comprenons que le protocole intuitif ait depuis laissé la place à d'autres paradigmes expérimentaux. Notamment, la première étude qui ait permis de contrôler la possibilité que l'effet d'imitation orofaciale précoce soit le résultat d'une pure coïncidence a été conçue et mise en œuvre par Meltzoff et Moore (1977; voir Butterworth, 1999). Une série d'expériences a depuis affiné la procédure de base et permis d'explorer ce phénomène sous différents aspects, en allant de l'imitation immédiate à l'imitation de l'intention du/de la démonstrateur/trice. Nous verrons ces protocoles de psychologie expérimentale au fur et à mesure afin que vous puissiez juger par vous-même de la validité de résultats dont nous ferons état. Pour des raisons pratiques, nous avons regroupé les données par période: cela ne signifie pas que nous défendons implicitement une conception discontinue du développement.

a. Dès la naissance: imitation du but d'actes corporels variés

Deux types de procédures de mise en évidence de l'imitation immédiate précoce peuvent être distinguées: les *standardisées* (Meltzoff et Moore, 1983) et l'*interactive* (Kugiumutzakis, 1999) mais nous ne décrivons ici que les premières car ce sont les plus contrôlées.

➤ **Procédure standardisée pour l'imitation immédiate**

Meltzoff et Moore (1983) mettent en évidence que des sujets âgés de 42 minutes à 71 heures (N=40) sont capables d'imiter la protrusion de la langue (*TP*) et l'ouverture-fermeture de la bouche (*MO*), présentées par le visage (réel) d'un adulte situé à 30 cm de celui du sujet⁷⁶. Le paradigme utilisé correspond à un *protocole de cibles croisées* où l'ordre de présentation répétée des modèles est contrebalancé: cette procédure est devenue un standard de la psychologie du développement.

Des cibles croisées...

Architecture. Au départ, le/la démonstrateur/trice découvre son visage⁷⁷ et, dans le même temps, le premier modèle qu'il/elle doit présenter (Figure 3-2). De là, chaque modèle (ici, $i=[1, 2]$) dispose d'un cycle, ou *période*, comprenant 6 fois de suite la séquence suivante:

- (a) 20 secondes de présentation qui consiste en quatre fois le doublet 4 secondes du geste (*Gi*) suivies de 1 seconde de visage passif⁷⁸,
- (b) puis 20 secondes de pause où le visage est passif.

Le cycle du modèle suivant (ici, $i=2$) s'enchaîne directement après le cycle du modèle précédent (ici, $i=1$). De plus, l'ordre de présentation des modèles est contrebalancé: la moitié des sujets est exposée à TP en premier et l'autre moitié à MO en premier. Tout au long de l'expérience, le visage de chacun/une des protagonistes est filmé sur une bande qui lui est exclusivement réservée.

⁷⁵ Les théories du développement dites "nativistes" donnent plus de poids aux mécanismes de développement "internes" au sujet qu'aux apports de l'environnement.

⁷⁶ Cette distance permet une perception visuelle optimale, car à cet âge le système visuel n'est pas encore mature. En particulier, la fovéa, l'accommodation visuelle et la convergence binoculaire ne sont pas établies avant l'âge d'environ 3 mois (voir Burnham, 1998, pour des références).

⁷⁷ Le visage du/de la démonstrateur/trice est au début caché derrière un drap blanc. Quand l'attention du sujet s'y fixe un certain temps, le visage est découvert et entre dans un flux de lumière qui le rend plus saillant.

⁷⁸ L'expression "visage passif" correspond à sa configuration de repos, avec notamment les lèvres closes.

L'enregistrement vidéo des sujets est séquencé en autant d'extraits qu'il y a de périodes sur l'ensemble des sujets. Ces extraits sont codés dans un ordre aléatoire par une personne naïve quant au geste présenté pendant la période traitée: l'on parle de **codage en aveugle** en ordre aléatoire. La tâche est de repérer les instants de début et de fin de l'occurrence des gestes cibles sur la bande, et cela strictement d'après leur définition opérationnelle qui est non-ambiguë et pré-établie. Dans cette étude, deux variables ont été estimées: la fréquence d'occurrence des gestes cibles et leur durée respective, au cours de chaque période et donc pour chaque sujet.

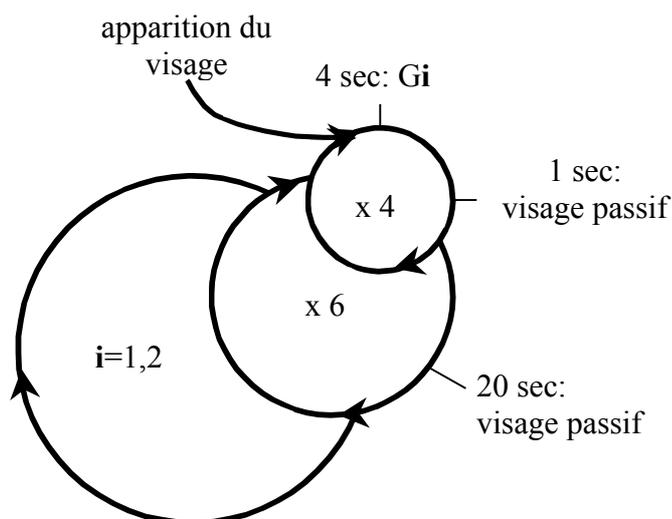


Figure 3-2: Les cycles emboîtés dans le paradigme expérimental de cibles croisées pour mettre en évidence l'imitation visuelle immédiate de modèle orofaciaux (Meltzoff et Moore, 1983)

Motivations. Un point crucial de ce paradigme réside, d'après les auteurs, dans l'*alternance* entre visage actif, sur les 20 premières secondes d'une séquence, et visage passif, sur les 20 secondes suivantes. En effet, cette alternance permet de montrer le geste assez longtemps pour que les sujets aient le temps de le percevoir et d'*organiser leur réponse motrice*, tout en minimisant leur perte d'*attention*⁷⁹. De plus, le contraste entre les deux phases-stimuli qui en résulte faciliterait l'*isolement de l'acte d'intérêt* au sein des variations de l'état du visage du/de la démonstrateur/trice. L'alternance peut aussi être vécue par le sujet comme une situation sociale qui, par sa structure de type *turn-taking*, appelle à la réaction plutôt qu'à la fixation en retour.

Interprétation. Globalement, vu que, pour un sujet donné, le seul paramètre qui varie est le modèle, si l'individu produit un nombre significativement plus élevé de TP (respectivement MO) dans la période où ce geste est le modèle que dans la période où il ne l'est pas, l'on peut estimer qu'il y a équivalence entre le stimulus (donné par le/la démonstrateur/trice) et la réponse. Ainsi, pour un/une participant/ante donné/ée la condition expérimentale définie par les périodes d'un modèle a-t-elle pour *contrôle-s* la/les condition-s expérimentale-s correspondant aux périodes du/des autre-s modèle-s; d'où l'utilisation de l'expression "cibles croisées" pour qualifier ce paradigme.

⁷⁹ Notons que la tenue du même état du visage-modèle aurait tendance à obtenir pour réponse une simple fixation visuelle, comme par habituation sensorielle.

...pour remédier à des doutes

L'analyse des résultats de Meltzoff et Moore (1983) montre que la grande majorité des sujets produisent le geste TP (respectivement MO) *plus fréquemment* et *plus longtemps* dans la période du modèle correspondant que dans celle du modèle "opposé".

Or, l'imitation n'est l'interprétation la plus probable de l'effet obtenu que si les hypothèses alternatives sont écartées. Reprenons point par point les constituants du doute (Di à Div) cités plus haut, en parallèle avec l'architecture de cette procédure standardisée. Si la similarité entre le comportement du/de la démonstrateur/trice et celui du sujet était une pure *coïncidence* (Di) ou le résultat d'une réaction d'*éveil global* due à la perception d'un visage en mouvement (Dii), la fréquence d'un modèle donné aurait dû être constante d'une condition à l'autre. En effet, ni *a priori* l'activité basale des sujets, ni le nombre de gestes du/de la démonstrateur/trice ne varient d'une période à l'autre. Ces deux hypothèses sont donc très peu probables.

La réponse des sujets peut-elle être le résultat d'un *conditionnement opérant* (Diii)? Pour écarter la possibilité que l'expérimentateur/trice puissent renforcer un comportement fortuitement exprimé par le sujet, le protocole est *standardisé*: quoiqu'il se passe, les différentes phases de la procédure sont suivies à la lettre, rendant ainsi le comportement de l'expérimentateur/trice le plus hermétique possible à celui du sujet. De plus, le/la démonstrateur/trice ne sait qu'au dernier moment quel modèle sera présenté en premier et aucune expression faciale ne peut être communiquée au sujet avant l'apparition du visage de l'expérimentateur/trice⁸⁰. Meltzoff et Moore (1983) précisent que le fait que l'imitation précoce ne s'explique pas par un conditionnement n'implique pas que (i) l'individu ne peut pas être conditionné à imiter un geste donné ou que (ii) le panel des gestes imitables ainsi que le sens associé à ceux-ci ne peut pas s'étendre sous l'influence de renforcements vécus au cours d'interactions sociales⁸¹.

Un modèle peut-il correspondre à un stimulus déclencheur (Div)? Il est difficile de défendre une telle position au moins pour les raisons suivantes:

- (a) Meltzoff et Moore (1983) observent des *comportements de "correction"*⁸², au sens où les individus ne produisent *pas instantanément* leur réponse, mais y convergent (et cela fera l'objet d'une analyse plus poussée avec des sujets plus âgés en 1994). Or, par définition un stimulus déclencheur induit *instantanément* la réponse qu'il déclenche, et cela sans qu'il y ait "hésitation" de la part du sujet.
- (b) A notre connaissance, aucune étude dans le monde animal ne relate l'existence d'une réponse déclenchée qui soit similaire au signal qui la déclenche. En particulier, les **stimuli efficaces**⁸³ ne sont pas des gestes mais sont plutôt des objets perceptifs "basiques". Par exemple, le comportement de quémande alimentaire des goélands juvéniles vis-à-vis de leurs parents est déclenché par la vue de la tache rouge située sur le bec de l'adulte. Autre exemple, la ponte des œufs chez les épinoches est déclenchée par la perception d'un frottement au niveau du ventre.

⁸⁰ De plus, dans le cas de sujets plus âgés, les parents ne sont pas informés du but de l'expérience pour prévenir un conditionnement de leur part "pour que leur enfant ait de bons résultats" (Meltzoff, 2000).

⁸¹ Dans le même ordre d'idée, Zazzo (1957, cité par Baudonnière, 1997) pourrait avoir conditionné, sans le vouloir, son nourrisson à tirer la langue à la simple *apparition de son visage* dans le champ visuel de ce sujet, à force de vérifier s'il était bien capable d'imiter la protrusion de la langue. En d'autres termes, cet individu pourrait avoir associé le visage de son père à la production de la "grimace". Cependant, une autre explication potentielle réside dans le phénomène de l'imitation différée (section III.A.1.b.).

⁸² Ceci est le phénomène-clé n°5 à la base du modèle AIM (Cf. section II.2.).

⁸³ **Stimulus efficace**: partie minimale du stimulus-déclencheur nécessaire et suffisante pour que celui-ci soit efficace, c'est-à-dire pour qu'il déclenche la réponse qui lui est associé.

- (c) Il est difficile d'imaginer l'existence d'autant de stimuli déclencheurs que de gestes imitables dans la période néonatale. En effet, comme nous le verrons ci-dessous, le spectre de modèles imités par les nouveau-nés est large, en allant de la protrusion de la langue, à l'ouverture-fermeture de la main, en passant par des modèles d'expressions faciales émotionnelles.

Ainsi peut-on avancer que les *nouveau-nés sont capables d'imiter*⁸⁴ des modèles orofaciaux à partir de leur perception visuelle. Cela implique qu'ils/elles ne confondent ni l'organe à activer ni le geste qui le concerne: *l'imitation précoce est spécifique*⁸⁵. Par exemple, les données complémentaires ainsi que celles ci-dessus montrent qu'une protrusion de la langue n'induit pas celle des lèvres et que la protrusion des lèvres n'élicite pas leur ouverture par celle de la bouche (MO). Dans le même ordre d'idée, au cours de différentes études menées par Meltzoff et Moore (1983, 1994, 1997), les auteurs remarquent que les instances d'imitation commencent généralement par de petits mouvements de la partie corporelle d'intérêt avant que l'acte cible ne soit produit. De là, ils proposent un décours temporel de l'acte d'imitation constitué d'une phase d'identification de l'organe, qui sous-tendrait la *mobilisation de la partie appropriée du corps*⁸⁶, suivie de l'imitation *per se*.

➤ **Données complémentaires sur la période néonatale**

Kugiumutzakis (1985) vérifie l'imitation de TP et MO dès 40 minutes en utilisant un paradigme (interactif) différent de Meltzoff et Moore (1983) (Kugiumutzakis, 1999). Field, Woodson, Greenberg et Cohen (1982) mettent en évidence que des sujets de 2 jours imitent des configurations faciales interprétées par les adultes comme étant l'expression des émotions de joie, de tristesse et de surprise (Butterworth, 1999). Selon Vinter (1986), l'imitation néonatale a besoin de la dynamique du modèle d'après des données obtenues à 4 jours (Meltzoff et Moore, 1992). De plus, Meltzoff et Moore (1989) ont exploré le panel de gestes que des nouveau-nés de 13 à 67 heures sont capables d'imiter. En particulier, ils montrent que des sujets imitent la cible TP, mais aussi les mouvements non-oro-faciaux de translation et de rotation de la tête (dans les sens trigonométrique et opposé). Notons que l'imitation de mouvements céphaliques a déjà été observée par Piaget au stade II (1-4 mois) de la période sensori-motrice (Cf. Meltzoff et Moore, 1989) mais celui-ci l'interprétait comme la continuation du mouvement du/de la démonstrateur/trice: cela est peu vraisemblable vu que les sujets de Meltzoff et Moore n'avaient pas besoin d'imiter ces mouvements en même temps que de percevoir le modèle. Les mêmes auteurs rapportent la capacité d'imitation immédiate de TP, MO, la protrusion des lèvres (LP) et l'ouverture-fermeture de la main, à 12-17 jours (Meltzoff et Moore, 1977, Expérience 1). Fontaine (1984) ajoute qu'à 21 jours, le gonflement des joues et le clignement des yeux peuvent être imités.

Il découle des données sus-citées que l'imitation précoce s'étend à un *nombre appréciable de modèles*⁸⁷ arbitraires. Elle est donc une capacité vraisemblablement généralisée, et peut faire appel à un processus d'appariement à la cible *génératif*, pourvu que les modèles proposés fassent partie des capacités perceptives et motrices des sujets. Ainsi ces faits contribuent-ils à discréditer l'hypothèse selon laquelle l'imitation précoce correspondrait à un comportement pré-câblé induit par un stimulus-déclencheur (Div): l'existence d'un tel facteur déclencheur pour chaque mouvement imitable n'a que peu de crédibilité vu son manque de parcimonie.

Un dernier élément vient consolider le fait que les comportements imitatifs aient peu de chance d'être le résultat du déclenchement instantané d'une réponse pré-organisée innée (Div).

⁸⁴ Ceci est le phénomène-clé n°3 à la base du modèle AIM (Cf. section II.2.).

⁸⁵ Ceci est le phénomène-clé n°2 à la base du modèle AIM (Cf. section II.2.).

⁸⁶ Ceci est le phénomène-clé n°4 à la base du modèle AIM (Cf. section II.2.).

⁸⁷ Ceci est le phénomène-clé n°1 à la base du modèle AIM (Cf. section II.2.).

En effet, Meltzoff et Moore (1977, Expérience 2) mettent en évidence une *capacité d'imitation différée*⁸⁸ de 2,5 minutes de TP et MO à l'âge de 16-21 jours. Dans cette expérience, la réponse orale des sujets est inhibée au cours de la période de présentation des modèles grâce à l'insertion préalable d'une tétine dans leur bouche. Les individus devaient "attendre" le retrait de la tétine pour exprimer un comportement oral autre que celui de têter, ce retrait n'ayant lieu qu'après la disparition de la cible qui cédait la place à un visage passif. En particulier, les auteurs font remarquer que, contrairement au comportement prédit par l'hypothèse du stimulus-déclencheur, l'objet n'était ni expulsé de la cavité orale à l'aide de la langue lors de la présentation du modèle TP, ni lâché par un abaissement de la mandibule en réaction au modèle MO (Meltzoff et Moore, 2000). De plus, cette étude montre que la pratique de la cible au moment de sa présentation n'est pas nécessaire à son imitation ultérieure. Cela signifie que (i) les réponses imitatives obtenues ne peuvent pas correspondre à une imitation de soi du type réaction circulaire baldwino-pagetienne, (ii) il ne peut y avoir apprentissage (sensori-) moteur au moment de la présentation du modèle, et donc (iii) il semble raisonnable d'avancer un phénomène de rétention en mémoire d'une représentation de la cible uniquement basée sur sa perception extéroceptive (visuelle). Aussi, nous verrons en filigrane, par la suite, que le délai de rétention de cette représentation tend à croître au cours du développement en interaction avec la complexification de la tâche imitative.

b. De 1 à 6 mois: de l'imitation différée d'actes corporels nouveaux à l'imitation d'actions sur des objets

L'intérêt de l'imitation en tant que technique d'apprentissage réside au moins dans la possibilité d'apprendre de *nouveaux* comportements, en étant capable de les exécuter en *l'absence du modèle* dans les champs perceptifs. De plus, pour tirer profit de l'expérience accumulée par autrui, la capacité d'imiter les *moyens*, ou la forme⁸⁹, de l'acte a, elle aussi, son importance, vu la difficulté de savoir dès le départ quelle-s caractéristique-s d'un comportement nouveau est/sont nécessaire-s à son exécution correcte. De là, l'imitation semble efficace en particulier dans l'apprentissage d'acte *arbitraires*⁹⁰.

➤ **Procédure standardisée pour l'imitation différée**

Meltzoff et Moore (1994) mettent en évidence que des sujets âgés de 6 semaines (N=40) sont capables d'imiter, après un délai de 24 h, trois gestes présentées par le visage (réel) d'un/une adulte situé à 30 cm de celui du sujet : l'ouverture-fermeture de la bouche (*MO*) et deux types de protrusions de langue (*TP*), l'une médiane (*TPmid*), l'autre latérale (*TPside*). *TPside* est un *nouveau*⁹¹ geste au sens où vraisemblablement il n'apparaît pas dans le répertoire spontané des participants/antes⁹², tout en étant composé d'éléments *a priori* dans les capacités sensori-motrices de cet âge. Plus précisément, les imitations de *TPside* et de *MO* font l'objet d'une micro-analyse qui a permis de décrire le processus d'appariement à la cible, à la base du

⁸⁸ Ceci est le phénomène-clé n°7 à la base du modèle AIM (Cf. section II.2.).

⁸⁹ Nous retrouvons ainsi l'importance de la forme, à prendre au sens de Köhler (Cf. Section I.1.).

⁹⁰ Par exemple, la manière de prononcer un phonème "spécifique" d'une communauté linguistique (ou "l'accent") peut être vue comme une manière arbitraire de le produire. Or, c'est cette prononciation de "chez nous" qui est apprise au cours du développement et non celle de "chez eux".

⁹¹ Cf. section I.2. pour des définitions.

⁹² Les auteurs ont pris le soin d'observer, pendant 20 minutes juste avant l'expérience, les interactions entre les sujets potentiels et la ou les personnes qui les accompagne-nt: les individus ayant des jeux utilisant des comportements semblables à *TPside* n'ont pas été invités à participer à la suite de l'étude.

modèle AIM⁹³, en termes de convergence itérative de la réponse sur des caractéristiques de plus en plus semblables à celles du stimulus.

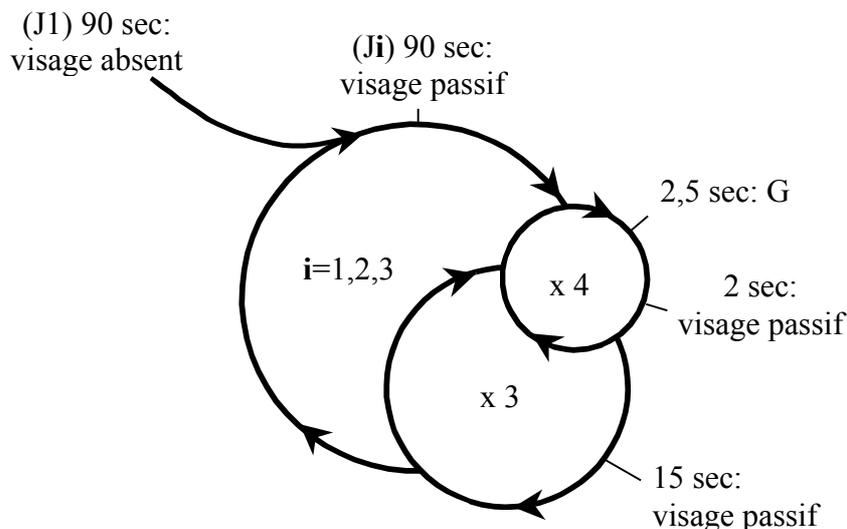


Figure 3-3: Les cycles emboîtés dans le paradigme expérimental de mise en évidence de l'imitation visuelle différée de modèle orofaciaux pour le groupe associé au geste G (Meltzoff et Moore, 1994)

Architecture. Le protocole expérimental s'étend sur trois jours (J_i ; $i=1..3$; Figure 3-3). 4 groupes de 10 individus chacun sont constitués. Au préalable, chaque sujet est aléatoirement affecté à un groupe de traitement, ou *condition expérimentale*, auquel ne sera présenté qu'un modèle (G) parmi les quatre configurations faciales suivantes: MO, TPside, TPmid et le visage passif (*NOM*)⁹⁴.

Le premier jour ($i=1$) commence par un temps d'acclimatation aux lieux, pendant lequel le visage de l'expérimentateur/trice est absent, suivi d'une période d'acclimatation au visage passif de l'expérimentateur/trice durant 90 secondes. De là, 5 temps expérimentaux (ou périodes) de 90 secondes chacun, répartis sur les trois jours, se succèdent en alternant test de l'imitation immédiate du geste G, où il est présenté aux sujets, et test de l'imitation de G différée au lendemain. Les 3 périodes relatives à l'imitation immédiate correspondent chacune à un cycle comprenant 3 fois de suite la séquence suivante:

- (a) 15 secondes de présentation qui consiste en quatre fois le doublet 2,5 secondes du geste (G) suivies de 2 secondes de visage passif,
- (b) puis 15 secondes de pause où le visage est passif.

Les 2 périodes relatives à l'imitation différée consistent en la simple exposition au visage passif de l'expérimentateur/trice et inaugurent le 2^{ème} et le 3^{ème} jour du plan expérimental ($i=2$ et $i=3$, respectivement).

L'enregistrement vidéo des sujets résulte en 120 extraits de 180 secondes chacun (40 sujets x 3 jours pour le nombre d'extraits, et 2×90 sec pour leur durée): ils sont codés en aveugle dans un ordre aléatoire. La tâche est de repérer les instants de début et de fin de l'occurrence des gestes cibles sur la bande, et cela strictement d'après leur définition opérationnelle qui est non-ambiguë et pré-établie. Dans cette étude, deux variables ont été estimées: la fréquence

⁹³ Voir section II.2.

⁹⁴ Les 3 premières sont les gestes cibles.

d'occurrence des 3 gestes cibles et la durée de MO, au cours de chaque période et donc pour chaque sujet⁹⁵.

De plus, les auteurs ont mené une micro-analyse du processus d'appariement potentiel à TPside et à MO. Au départ, les mouvements de type TP sont répartis selon 5 profils mutuellement exclusifs suivant l'étendue du geste et sa direction (médiante vs. latérale): l'on distingue la petite TP (*ST*) qui ne dépasse pas clairement le bord le plus externe de la lèvre inférieure et n'a pas de composante directionnelle classifiable sans ambiguïté, la petite TP non-médiane (*NMT*) qui émerge après avoir poussé le coin des lèvres ou la joue, la petite TP sur le côté (*STS*) dont la protrusion est suivie d'un mouvement latéral clair et définitif avant l'achèvement du recul de la langue, la grande TP (*LTP*) dont la protrusion dépasse clairement le bord le plus externe de la lèvre inférieure, et la grande TP sur le côté (*LTPS* ou *TPside*) qui émerge et se rétracte au coin de la bouche ou montre un mouvement latéral visible et définitif à l'extérieur de la bouche en glissant entre les lèvres. Toutes les TP produites par les sujets appartiennent à l'une de ces catégories. D'après le même principe, la durée des MO est classifiée selon l'écart à la durée de la cible (ici, 2,5 sec). Pour cela, 5 catégories ($j=1..5$) de durée sont définies d'après les 5 intervalles de temps déduits de la formule $[2,5 - j*0,5; 2,5 + j*0,5]$ sec. Par exemple, les occurrences de la 5^{ième} catégorie ont une durée distante de plus de 4 fois 0,5 sec de la durée cible, alors que celles de la 1^{ère} catégorie est à +/- 0,5 sec de la durée cible et correspondent donc au "bon" appariement au stimulus.

Motivations. D'après les auteurs, un aspect fondamental pour éliciter un comportement précoce d'imitation différée est de faire en sorte que le visage de l'expérimentateur/trice rappelle son comportement de la veille, et donc, avant tout, que les sujets ne confondent pas le/la démonstrateur/trice et la personne qui les accompagne. Pour cela le/la premier/ère se déguise en contraste avec la seconde, en jouant sur le port de lunettes et/ou d'un fichu. De plus, ces personnes, ou *objets animés*, entrent et sortent, sans chevauchement de leur transit dans le temps, par des points spatiaux opposés dans l'environnement expérimental et en prenant soin que le sujet en cours suive (sans interruption) du regard le déplacement de l'entité en mouvement. Aussi, le/la démonstrateur/trice n'est jamais en contact avec les participants/antes en dehors des périodes de test et ne montre sur les 3 jours qu'un seul modèle G à chaque sujet. Ainsi, chaque groupe de sujet a-t-il la possibilité d'associer le visage de l'expérimentateur/trice à son unique "grimace"⁹⁶ et de se rappeler son geste pour le produire plus tard.

Un geste, en particulier s'il est nouveau, n'est pas tenu d'être exécuté correctement dès la première tentative d'imitation: il doit être possible de mettre en évidence un chemin obéissant globalement à une suite d'étapes ordonnées pour l'atteinte de son appariement en termes d'actes, avec ses aspects de but et de moyen. La prédiction de base est que ce processus de correction se fait d'une exécution allant d'approximative à exacte; d'où la définition d'une hiérarchie de référence dans le degré de précision des réponses congruentes avec la cible. Dans le cas de TPside, l'ordre standard proposé correspond à la séquence temporelle NMT _ STS _ LTP _ LTPS, et dans celui des productions de MO, à la série des catégories de durée

⁹⁵ Notons toutefois qu'ici la durée des TP n'est pas analysée, contrairement à l'expérience de 1983 qui portait sur des sujets plus jeunes. Les auteurs signalent que ce geste est trop rapide pour que sa durée soit quantifiable de manière précise et reprennent les termes de Sackett (1978) et Backeman et Gottman (1986) pour qui la TP est "un mouvement momentané" où l'on s'intéresse plus à combien de fois il a lieu qu'à combien de temps il dure.

⁹⁶ Nous attirons votre attention sur le fait qu'une association visage de l'autre-geste de l'autre n'est pas synonyme d'une association geste de l'autre-geste de soi. En effet, il n'y a pas, a priori, de conditionnement classique à donner une réponse similaire au stimulus. En revanche, l'idée d'une association visage de l'autre-geste de soi sera considérée plus bas.

5 _ 4 _ 3 _ 2 _ 1, au cours des essais successifs⁹⁷. De plus, un élément clé de cette vision réside dans le fait que seule importe la tendance: si une étape vient à "manquer", tant que l'étape suivante s'apparie plus précisément à la cible, la tendance reste valide.

Interprétation. La répartition sur trois jours des 5 périodes expérimentales permet de tester la capacité de donner une réponse similaire à la cible même si elle est absente des champs perceptifs depuis 24h. En effet, durant les 2 périodes du test de l'imitation différée, la perception en cours est celle de NOM pour tous les groupes et le seul paramètre qui varie d'un groupe à l'autre est le geste G présenté 24h auparavant. Cela implique que si, pour une période de test de l'imitation différée donnée, le groupe TPmid ou TPside (respectivement MO) produit un nombre significativement plus élevé de TP (respectivement MO) que les groupes MO et NOM (respectivement TPmid, TPside et NOM) à cette même période⁹⁸, l'on peut estimer qu'il y a *équivalence* entre le stimulus (donné par le/la démonstrateur/trice) la veille et la réponse du jour. Ainsi, le groupe NOM sert-il de contrôle absolu de la nature imitative des réponses données par les trois autres groupes en présence et en l'absence du stimulus effectif.

Une alternative au doute

Toute période confondue, les résultats par groupe⁹⁹ révèlent que (i) les TP sont plus souvent produites par les sujets exposés à TPside et TPmid, avec des fréquences de même ordre quelque soit le type de stimulus TP et (ii) les gestes MO sont plus fréquents et durent plus longtemps dans la condition expérimentale correspondant à ce geste bien que la variable de fréquence de MO ne passe pas le seuil de signification statistique. De plus, sur l'ensemble des périodes de test de l'imitation immédiate, (i) les sujets exécutent MO plus fréquemment et plus longtemps en réponse à MO qu'en réponse à TPside, TPmid ou NOM, tandis que (ii) les sujets produisent plus de TP en réponse à TPside et TPmid qu'en réponse à MO ou NOM. De la même manière, sur l'ensemble des périodes de test de l'imitation différée, (i) les TP sont plus souvent produites par les sujets exposés la veille à TPside et TPmid, et (ii) les MO durent plus longtemps si la cible présentée 24h auparavant est MO, sans que la différence de fréquence moyenne de MO en fonction de la cible ne soit, là aussi, statistiquement significative¹⁰⁰.

L'interprétation selon laquelle la similarité entre les gestes courants du/de la démonstrateur/trice et ceux des sujets s'explique par un mécanisme d'imitation immédiate est la plus vraisemblable d'après les arguments fournis section III.A.1.a.. Par contre ces "remèdes aux doutes" ne suffisent pas à écarter la possibilité que la similarité entre les gestes du jour des sujets et le geste de la veille du/de la démonstrateur/trice puisse être due à un conditionnement classique¹⁰¹ qui relève d'une association entre le visage de l'expérimentateur/trice et le geste de la veille des sujets. Si tel est le cas, l'on s'attend à ce que les productions du jour des participants/antes soient similaires à leur production de la veille. Or, la micro-analyse de l'organisation motrice de leurs réponses ne vérifie pas cette prédiction: les sujets ne reproduisent pas "simplement" leurs propres gestes de la veille, ils/elles les *corrigent*. Il en découle que l'hypothèse la plus probable pour rendre compte de

⁹⁷ L'analyse du processus d'appariement à TPmid n'a pas été menée parce qu'il est difficile de distinguer de manière non-ambiguë de la qualité *exclusivement* médiane d'une protrusion de langue. Toutefois, les auteurs signalent qu'un tel processus de correction pour TP a déjà été rapporté dans l'étude de 1983 avec des sujets fraîchement nés (section III.A.1.a.).

⁹⁸ Dans l'analyse des résultats des imitations immédiate et différée, la distinction entre TPmid et TPside est ignorée, en regroupant toutes les productions du type TP sous le même label.

⁹⁹ Toutes les tendances décrites dans ce paragraphe portent sur des valeurs moyennes sachant le groupe.

¹⁰⁰ Meltzoff et Moore (1983, 1994) précisent que la durée est une variable plus apte à révéler l'effet d'imitation de MO que la fréquence.

¹⁰¹ Cf. section I.1.

l'équivalence entre le comportement des sujets et celui du/de la démonstrateur/trice est celle de l'imitation en modes immédiat et différé, et les résultats de la micro-analyse viennent en renfort à cette alternative.

Dans la condition MO, 7 individus sur 10 convergent sur la durée cible d'une manière similaire à celle de la séquence standard, contre 14/30 dans les autres groupes. De plus, dans la condition TPside, 8 sujets sur 10 vérifient la tendance prédite par le modèle progressif d'appariement à la cible, avec atteinte du modèle par 7 d'entre eux¹⁰², alors que cette tendance n'apparaît que chez 7 participants sur les 30 issus des groupes contrôles (TPmid inclus). Ce fait plaide pour une émergence de TPside ordonnée (et non du type essai-erreur aléatoire) et spécifique aux individus exposés à ce nouveau mouvement.

Aux résultats publiés en 1994, Meltzoff et Moore (1995, 1997) ajoutent les observations suivantes:

- (i) Dans le groupe TPside, les premières tentatives d'imitation sont centrées sur la composante latérale du geste, où la langue bute contre la joue ou encore sort légèrement de la bouche et glisse sur le côté au cours de sa rétraction: ces comportements ne figurent ni dans le stimulus ni dans l'activité basale, et font l'objet de corrections de la part des participants/antes.
- (ii) Des sujets exposés à des TP pleurent après avoir vainement essayé de tirer la langue à plusieurs reprises. Les auteurs découvrent que la morphologie de leur langue les empêche de produire cet acte : leur frein s'étend jusqu'à la pointe de la langue.
- (iii) Des "erreurs" créatives ont été observées dans la condition TPside, sans être incluses dans la micro-analyse de l'organisation motrice de la réponse imitative. Elles consistent en la protrusion médiane de la langue (TPmid) associée à une rotation de la tête. Ce comportement intrigant, qui n'apparaît ni dans le stimulus ni dans l'activité basale, trouve sa cohérence dans le fait que, comme l'acte TPside, il aboutit à une configuration latérale de la langue par rapport au tronc: il y a équivalence en termes de *relations entre des "organes"* entre le résultat de l'erreur créative et celui du geste modèle.

Cette étude met en évidence qu'à 6 semaines les êtres humains sont capables d'imiter des *gestes nouveaux*¹⁰³, sur des chemins d'appariements non-aléatoires, et que la réponse imitative peut être *différée*, sans besoin que le modèle fasse partie du répertoire comportemental spontané. Or, l'imitation d'un stimulus absent des champs perceptifs implique qu'il est représenté en mémoire. De là, une aptitude à la *représentation* semble exister de manière précoce: cela remet en question non seulement la notion même de période purement sensori-motrice proposée par Piaget, mais aussi l'idée que l'"apparition" de la représentation au cours du développement cognitif puisse être un achèvement, sachant que sa présence est rapportée dès la période néonatale (section II.A.1.a., Données complémentaires). De plus, les comportements de correction ordonnée convergeant sur le modèle et la frustration exprimée par les individus ne pouvant pas produire ce geste révèlent que les nourrissons détectent la différence entre l'acte propre en cours et celui d'autrui, et que leur satisfaction se trouve dans un appariement ou, mieux, une mise en *équivalence* des deux. Ainsi peut-on en déduire que (i) *l'imitation même précoce est un acte intentionnel* qui a pour but l'appariement au modèle, (ii) le processus d'appariement se base sur la comparaison de la représentation de l'acte de soi à celle de l'acte de l'autre, pour *construire* la réponse imitative, et (iii) la représentation de la configuration faciale de soi n'est pas confondue avec celle d'autrui¹⁰⁴. Plus précisément, vu

¹⁰² Ici, l'évaluation de l'achèvement de TPside est uniquement basée sur les données recueillies le dernier jour du test, afin de laisser aux sujets un maximum de temps pour organiser leur réponse à ce modèle.

¹⁰³ Ceci est le phénomène-clé n°6 à la base du modèle AIM (Cf. section II.2.).

¹⁰⁴ Vu que des comportements de correction ont été observés dans la période néonatale (section III.A.1.a.), l'idée selon laquelle les nourrissons confondraient les "mondes interne et externe", comme Piaget l'a proposé notamment pour expliquer la "contagion des pleurs" (section I.2.), est sérieusement mise en doute.

que l'imitation oro-faciale de modèles visuels porte sur la correspondance entre une information issue de la vision et une information issue de la proprioception, l'on est en droit d'avancer que les représentations en jeu sont *intermodales*, voire a- ou *supra-modales*, comme proposé par Meltzoff et Moore.

Les "erreurs" créatives suggèrent que l'espace de représentation de ces actes est susceptible de se faire en termes de relation-entre-organes, sachant que cette grille d'interprétation permet d'expliquer de manière cohérente ces "excentricités": en ces termes, l'état final, ou but, de l'acte à imiter est atteint, bien que les moyens utilisés soient différents de ceux du/de la démonstrateur/trice. De plus, les données sur l'imitation de la durée de MO et sur l'émergence de la réponse TPside indiquent qu'à 6 semaines (i) il est possible d'imiter des caractéristiques du geste cible et des données complémentaires présentées plus bas étayent cette idée, et (ii) le contenu des comparaisons successives, qui mènent à la mise en correspondance effective de moyens de soi avec des moyens utilisés par le/la démonstrateur/trice (moyens-cibles), peut être déduit du et exploité au cours du processus d'appariement par le sujet. Ainsi, aux alentours de 1 mois, l'imitation peut-t-elle vraisemblablement avoir pour cible l'*acte* du/de la démonstrateur/trice, compris dans ses dimensions de but mais aussi de moyens, et les mouvements de soi et d'autrui peuvent-ils être situés, ou cartographiés, dans un *espace des actes* potentiellement en devenir¹⁰⁵.

➤ Données complémentaires sur la période de 1 à 6 mois

Deux points particuliers animent la littérature relative à l'imitation précoce: l'information nécessaire à l'imitation et si la capacité d'imitation décline et/ou disparaît de la même manière que les **réflexes néonataux**¹⁰⁶. Dans les deux cas, la dispute tourne autour de résultats jugés contradictoires sans que la communauté ne se permette de trancher définitivement. En effet, la divergence des observations peut s'expliquer par des différences méthodologiques, aussi infimes puissent-elles paraître, entre la/les expérience-s de mise en évidence (positive) et celle-s de non mise en évidence. Toutefois, ne pas mettre en évidence un phénomène n'implique pas son inexistence: nous partirons du principe que ce qui a été observé est *observable sachant* au moins la méthode utilisée et sa mise en œuvre¹⁰⁷.

Vinter (1986) n'a pas pu éliciter l'imitation de postures faciales statiques à 4 jours (Harris et Butterworth, 1999, pp. 128-129). De là, il a été proposé que la dynamique du modèle est une information accessible à la perception du modèle et est nécessaire à l'imitation dans *les premier mois* de vie. Ainsi l'univers des imitables infantiles correspondrait-il à des configurations en mouvement. Cependant, Meltzoff et Moore (1992) montrent que des individus de 6 semaines imitent les postures faciales *statiques* correspondant à l'état final des gestes TP et MO¹⁰⁸, avec dans le cas de MO l'imitation de la durée de la pose¹⁰⁹. D'après eux, la différence entre les données à 4 jours et les leurs (à 6 semaines) pourrait résider dans le fait qu'à 1 mois, les individus seraient depuis devenus aptes à déduire d'un modèle statique le-s

¹⁰⁵ Le fait que l'imitation ait pour cible l' *acte* du/de la démonstrateur/trice, et non plus son but seul, est un des changements du décours développemental de l'imitation, qui constituent le phénomène-clé n°10 à la base du modèle AIM (Cf. section II.2.).

¹⁰⁶ Les **réflexes néonataux** sont les réflexes jugés spécifiques de cette période développementale (voir le point (Div) du doute en introduction de la section III.A.1.).

¹⁰⁷ Une prise de position prudente veut que seuls les phénomènes avérés, i.e. répliqués et publiés un "assez grand" nombre de fois (de préférence dans des journaux de renom) pour considérer leur existence, méritent crédit. Nous avons choisi une attitude différente, en racontant ce qu'il est *possible d'observer*, et non le contraire, d'après des études rapportées et jugées par des membres de la communauté scientifique, qui sont *a priori* de bonne foi et indépendants les uns des autres.

¹⁰⁸ Les modèles statiques correspondent ici au maintien de la posture de la langue ou de la bouche pendant 15 secondes de présentation, en alternance avec 15 secondes de visage passif.

¹⁰⁹ Ceci est le phénomène-clé n°8 à la base du modèle AIM (Cf. section II.2.).

mouvement-s qui y mène-nt contrairement aux plus jeunes qui auraient besoin de percevoir ce chemin, faute d'avoir associé l'état final de relation-entre-organes à au moins une transformation-de-relation-entre-organes correspondante (ou moyen). Donc le changement aurait vraisemblablement lieu au niveau de la "compréhension" des actes humains¹¹⁰. Cette explication permettrait aussi de rendre compte du fait qu'à 1 mois, il y a capacité de détecter le fait d'être imité¹¹¹ (pour des références, voir Meltzoff et Moore, 1999). En effet, face à deux personnes en mouvement toutes deux au même rythme que le sujet mais dont seulement une l'imité, l'individu est plus attentif/tive et sourit plus à cette dernière qu'à l'autre. De plus, le geste imité est reproduit à plusieurs reprises, comme si l'imitation d'autrui était prise pour une conséquence de soi, et si l'autre cesse d'imiter, le sujet se vexe¹¹². D'après Meltzoff et Moore, la possibilité de percevoir que l'autre nous imite est le germe de la croyance que l'autre est comme soi (ou *likemness*) et serait donc à la base du développement de l'empathie et d'une première théorie de l'esprit (voir Meltzoff, 2002b).

Des auteurs rapportent une impossibilité d'évoquer une réponse imitative à partir de 2 mois (voir Lebib et Baudonnière, 1999), d'autres un déclin de l'imitation de modèle faciaux à partir de 2 mois (eg. Fontaine, 1984) suivie d'une réapparition vers 5-6 mois ou 8-9 mois (voir Kugiumutzakis, 1999). Certains proposent que la disparition de l'imitation s'expliquerait par sa nature réflexive: la capacité "d'imitation" serait inhibée au même titre que les autres réflexes néonataux sous l'effets de centres nerveux en développement. D'autres avancent qu'une différenciation des modalités perceptives (initialement fusionnées) impliquerait la raréfaction de l'imitation selon le même décours que d'autres habiletés néonatales telles que la localisation audio-visuelle: l'imitation réapparaîtrait sous la forme d'une reconstruction basée sur la coordination des modalités désormais différenciées¹¹³. En désaccord avec ces hypothèses plutôt maturationnistes, Meltzoff et Moore (1992) montrent dans un protocole de mesure répétée que des individus de 2-3 mois imitent aussi bien, si ce n'est mieux, que ceux de 6 semaines les modèles MO et TP statiques et dynamiques. Ils observent, toutefois, qu'au premier essai les sujets de 2-3 mois, contrairement à ceux de 6 semaines, tentent de mettre en place des routines de jeux sociaux avec l'autre. Passée cette période, les participants/antes imitent. Ainsi ne semble-t-il pas y avoir de diminution de l'aptitude à imiter des modèles faciaux. Les auteurs proposent qu'à la différence des plus jeunes les individus de 2-3 mois ont déjà appris qu'autrui joue avec eux, et ont donc des attentes sociales qu'il faut leur laisser le temps de dépasser (sous l'effet de l'échec de leurs tentatives de jeu) pour les inciter à imiter. Dans la même étude, il est montré que, dès 6 semaines, le facteur de la familiarité du/de la démonstrateur/trice (parent versus étranger/ère) n'a pas d'effet sur les performances d'imitation, et que l'association modèle-démonstrateur/trice pourrait être utilisée par les sujets pour tester l'identité d'autrui via l'imitation différée de sa grimace "caractéristique"¹¹⁴.

Barr, Dowden et Hayne (1996) mettent en évidence qu'à 6 mois l'on peut imiter une à deux action-s sur un objet, telle qu'enlever le gant d'une marionnette¹¹⁵, immédiatement après leur

¹¹⁰ Meltzoff et Moore proposent aussi qu'un visage statique puisse être une situation que les fraîchement-nés ne peuvent pas comprendre comme étant sociale à la différence des individus plus âgés.

¹¹¹ Ceci est le phénomène-clé n°9 à la base du modèle AIM (Cf. section II.2.).

¹¹² Ces auteurs mettent en évidence ce patron de réaction jusqu'aux environs de 9 mois.

¹¹³ Kugiumutzakis (1999), quant à lui, propose que l'imitation faciale pourrait se raréfier parce qu'elle cède la place à l'imitation de modèles vocaux dans une "quête de stimuli frais" (en accord avec Guillaume 1925/1971), en rejetant dans le même temps des explications du déclin observé par (i) une perte momentanée de la capacité de traitement de l'information (Maratos, 1973), (ii) une perte de la capacité d'appariement (Jacobson, 1979), (iii) un non-renforcement des "grimaces" par l'entourage (Dunkeld, 1978) et un déclin de l'attention (Field, Goldstein, Vega-Lahr et Porter, 1986; Fontaine, 1984; Heimann, 1989). Les curieux/euses pourront lire Kugiumutzakis (1999) pour plus de détails et les références bibliographiques.

¹¹⁴ Les auteurs parlent de la fonction d'identification des personnes que pourrait servir l'imitation différée.

¹¹⁵ D'après Bushnell (1985), la capacité de préhension apparaît aux alentours de 4-5 mois (Barr, Dowden et Hayne, 1996).

présentation. En particulier, la probabilité d'imiter l'acte présenté à l'initiale d'une séquence de 3 actions sur le même objet est plus élevée que celle d'imiter les actes suivants. Meltzoff et Moore (1994) voient dans ces débuts de l'imitation d'action sur des objets le passage de l'appariement des actes de soi à ceux d'autrui à l'appariement de la relation soi-objet à la relation autrui-objet.

c. De 6 à 24 mois: de l'imitation différée d'actes sur des objets à l'inférence du but

A 6 mois, la capacité d'imiter une à deux action-s successives sur un objet peut se faire immédiatement après leur démonstration, mais aussi 24 h plus tard sans pratique immédiate avec les mêmes performances, pourvu que le nombre de démonstrations durant la session de présentation soit suffisant pour faciliter la mémorisation des actes-cibles (Barr *et al.*, 1996). Dans la même étude, des individus de 12 mois montrent les mêmes performances avec un nombre de présentations inférieur à celui nécessaire à 6 mois. De plus, la probabilité d'imiter une séquence de 2 actes et de 3 actes sur un même objet après 24 h de rétention tend à croître de 6 à 24 mois.

A 9 mois, Meltzoff (1988b) obtient des résultats en accord avec ceux de Barr *et al.* (1996) et ne trouve pas de différence significative en termes de performances entre imitation immédiate et différée de 24 h, pour une série de 3 actions impliquant chacune un objet spécifique et nouveau, présenté et mis à disposition séparément des deux autres¹¹⁶. De là, il y a aptitude à retenir et imiter en acte plusieurs buts d'actions sur des objets différents. Meltzoff et Moore rapportent qu'à partir du même âge, le fait d'être imité induit des "comportements de test" de la part des sujets: les individus tendent à changer brusquement d'activité, tant que l'autre suit, tout en l'observant le sourire aux lèvres¹¹⁷ (pour des références, voir Meltzoff et Moore, 1999). De là, les auteurs proposent que la notion de "likemness" initiale, où un acte donné imité par l'autre serait vécu comme une conséquence de l'acte propre¹¹⁸, prend désormais la forme d'une *relation d'appariement soi-autrui généralisée*, du type "tu feras ce que je fais (quels que soient mes actes)". Cette relation d'appariement est vraisemblablement à la base des jeux d'imitation en séquence d'actes arbitraires substituables à l'infini, que l'on observe entre nourrissons (voir Baudonnière, 1997).

A partir de 12 mois, l'on met en évidence une capacité d'imitation différée décontextualisée sans besoin de copiage immédiat. En effet, il y a imitation différée de 1 et 4 semaine-s d'actions sur des objets nouveaux même avec changement d'environnement spatial entre la période d'observation et celle de test¹¹⁹ (Klein et Meltzoff, 1999). Plus précisément, les performances sont les mêmes entre 1 et 4 semaines de délai, entre changement et non-changement d'environnement, et entre les 2 x 2 conditions expérimentales combinées, mais les performances sont meilleures si la rétention n'est que de 3 minutes.

A 14 mois, l'on peut imiter à la maison plusieurs actes réalisés par des pairs sur des objets nouveaux observés 2 jours avant, à la garderie ou dans une salle de laboratoire, sans besoin de toucher les objets le jour de l'observation¹²⁰ (Hanna et Meltzoff, 1993). De plus, Barnat, Klein et Meltzoff (1996) montrent au même âge une aptitude à imiter des actions sur des objets nouveaux avec changement de l'environnement spatial et des caractéristiques de taille et de

¹¹⁶ Dans l'expérience portant sur l'imitation différée, les sujets n'ont accès aux objets qu'après les 24 h de rétention.

¹¹⁷ Ce patron de comportement a été observé jusqu'à 14 mois.

¹¹⁸ Cf. section III.A.1.b. Données complémentaires.

¹¹⁹ Dans cette étude, le lieu de présentation correspond soit à une salle standard de laboratoire, soit à une salle de laboratoire tapissée d'orange à pois blancs ("non-maison"), pour les délais de 1 et 4 semaine-s, soit à la maison, dans le cas de 1 semaine de rétention, alors que le lieu de test est toujours la pièce standard de laboratoire.

¹²⁰ Dans cette étude, ce phénomène a été mis en évidence jusqu'à l'âge de 18 mois.

couleur des objets entre la période de présentation et celle de test qui a lieu 10 minutes plus tard. En particulier, les performances décroissent de la condition sans changements, à celle de modification du contexte et de la taille des objets, à celle de changement d'environnement spatial, de taille et de couleur des objets. A la même période de développement, sont rapportés des jeux dits "symboliques", qui sont par définition décontextualisés, tels que faire semblant, par exemple, de parler dans un téléphone miniature (voir Barnat *et al.*, 1996, pour des références). Les auteurs proposent que ces activités se construisent à partir du couplage entre les capacités d'imitation différée et de généralisation. Dans le même ordre d'idée, Meltzoff (1988c) met en évidence qu'à cet âge, il y a aptitude à imiter une action simple sur un objet nouveau vue à la télévision 24 h auparavant. Dans ce cas, la capacité de représentation infantile se rapproche de la définition saussurienne de la notion de forme vu qu'une image 2D, de petite taille et en noir-et-blanc, réfère à une situation 3D réelle, grandeur nature et en couleur. En d'autres termes, dans cette expérience, le sujet construit une représentation de l'acte-cible à partir d'une représentation culturelle de cette action perçue visuellement.

Les études sus-citées ne veulent pas minimiser la capacité de rétention dans le temps de modèles d'actions sur des objets des individus de 14 mois, ni même oublier que l'imitation est un outil potentiel pour l'apprentissage d'actions *nouvelles* et *arbitraires* sur des objets. En effet, Meltzoff (1988a, 1995a) montre qu'il est possible d'imiter une action qu'aucun contrôle ne produit spontanément avec des délais de 1 semaine, 2 mois et 4 mois, sans besoin de produire l'acte-cible le jour de la présentation. En particulier, dans les instances d'imitation de ce *nouvel acte arbitraire*, qui correspond à se pencher pour toucher une plate-forme avec le haut de la tête pour l'allumer, ce sont les *moyens spécifiques* du/de la démonstrateur/trice qui ont été utilisés, malgré la possibilité d'user, par exemple, de la main pour atteindre le même but d'allumage. De plus, l'étude de 1995a retrouve que, comme à 12 mois (Klein et Meltzoff, 1999), les performances sont meilleures en imitation immédiate qu'après 2 mois et 4 mois, et sont similaires entre ces 2 derniers délais. Hormis le test d'imitation différée d'1 semaine, l'ensemble de ces résultats ont été vérifiés avec des sujets de 16 mois (Meltzoff, 1995a).

A 18 mois, l'on peut imiter une *action sur un objet pour agir sur un autre objet* (Meltzoff, 1995b). Le modèle correspond à appuyer avec un bâton sur un bouton qui déclenche un son. Une fois de plus, c'est le *moyen arbitraire* spécifiquement employé par le/la démonstrateur/trice que les sujets utilisent, alors que l'usage d'un doigt aurait suffi à atteindre le même but. De plus, cette étude met en évidence qu'à cet âge, la capacité de représentation de l'action d'autrui s'étend de celle de l'acte effectivement observé à celle de l'acte dont le but est *inféré* d'après la perception de cette action amputée de son achèvement. En effet, les participants/antes imitent des actions sur des objets nouveaux¹²¹ que le/la démonstrateur/trice a l'*intention* de faire, sachant ses tentatives infructueuses, i.e. son comportement de surface¹²². Plus précisément, les performances sont les mêmes, que les sujets aient vu ou non l'atteinte du but, en termes de nombre d'actes imités et de temps de latence, et le fait d'occulter le but ne demande pas de phase du type essai-erreur: les sujets vont droit au but. De plus, 5 objets-composés étaient mis en jeu, avec pour chacun 2 actions différentes¹²³, dont l'une n'avait pas de "vrai" but. Par exemple, dans le cas de l'objet à outil, cette action était de déplacer le bâton horizontalement contre une face de la boîte où se trouvait le bouton. Les résultats montrent que même ces *actes totalement arbitraires* sont imités tels quels.

¹²¹ Ces actions ont été choisies de manière à faire intervenir différentes relations objet-objet, telles que "mettre dedans", "mettre autour" et l'utilisation d'un outil intermédiaire.

¹²² Par contre, si un bras de robot effectue les mêmes mouvements que l'acteur/trice, il n'y a imitation que si la totalité de l'action est montrée (Meltzoff, 1995b; pour plus d'information, voir aussi Meltzoff 1996, cité dans Meltzoff, 2002a). Il en découle qu'à 18 mois, les individus attribuent des états mentaux (en l'occurrence des intentions) aux personnes et ont donc une grille d'interprétation *psychologique*, plutôt que physique, du comportement des personnes, ou une "théorie de l'esprit".

¹²³ Une action de test, qu'elle soit montrée entièrement ou non, et une action contrôle.

A 2 ans, sont observés des "jeux de rôles", où des individus imitent des *rôles sociaux*, avec prise de la perspective du personnage¹²⁴ (voir Meltzoff et Moore, 1994). De plus, des individus intègrent spontanément à leur répertoire verbal des slogans publicitaires diffusés par la télévision (Lemish et Rice, 1986, dans Meltzoff, 1988c).

Un résumé chronologique des tendances développementales de l'imitation visuo-motrice dans les deux premières années de vie se trouve Figure 3-4..

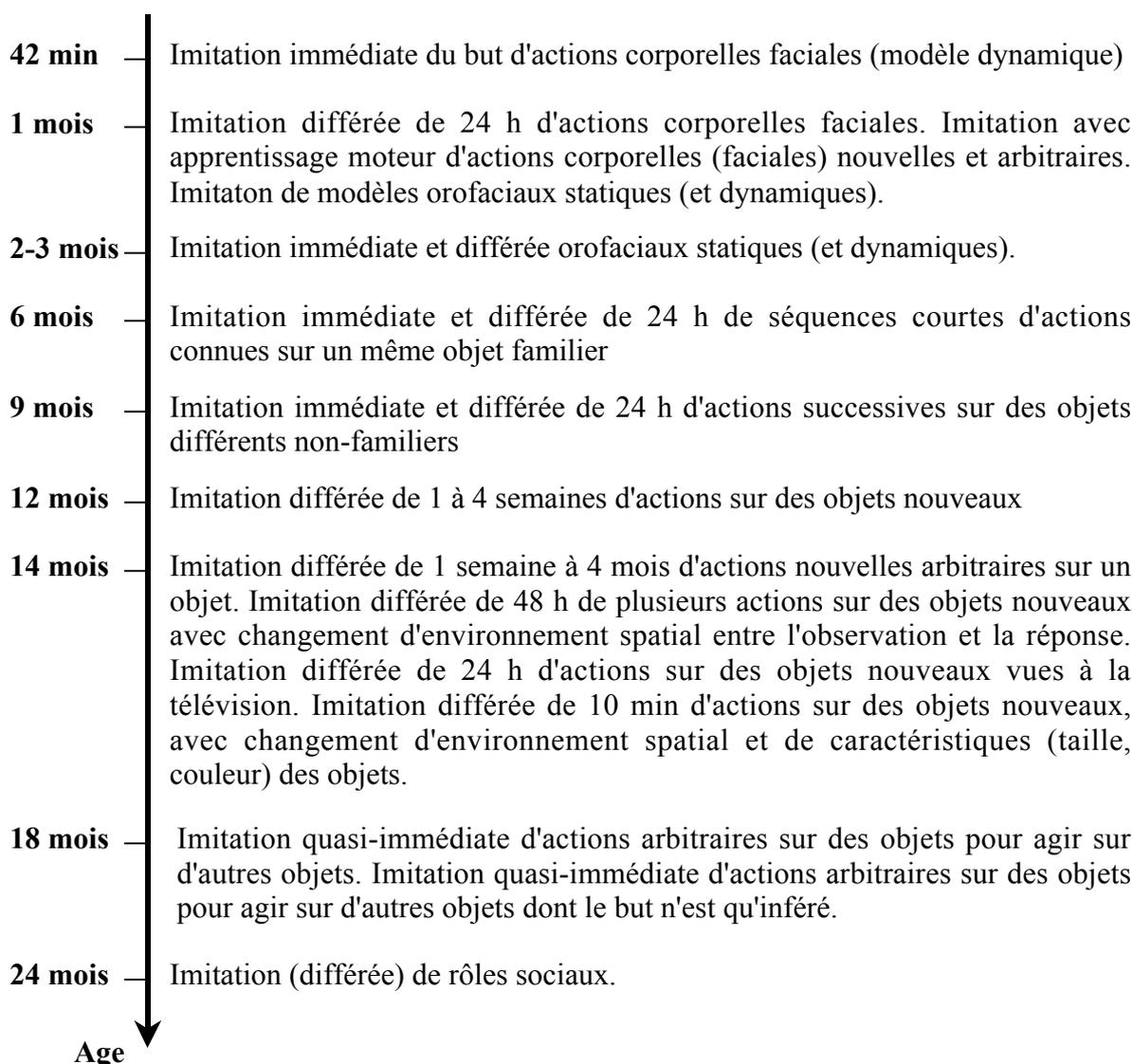


Figure 3-4: Un résumé chronologique des tendances développementales de l'imitation visuo-motrice dans les deux premières années de vie.

Globallement, l'ensemble des études mettent en évidence un décours développemental où l'individu est susceptible d'effectuer des tâches de plus en plus complexes. L'on peut dégager 8 axes¹²⁵ de complexification de la capacité d'imitation qui semble s'étendre, notamment, (i)

¹²⁴ L'imitation peut donc être tout ou partie du mécanisme d'apprentissage des comportements dits "générés", ou stéréotypes comportementaux liés au genre (en accord notamment avec Hanna et Meltzoff, 1993).

¹²⁵ Ces 8 "axes" ne sont vraisemblablement pas indépendants les uns des autres, tout comme l'imitation est une aptitude dont le développement interagit avec celui d'autres domaines (e.g. la/les mémoire-s, le contrôle moteur des différents segments du corps etc.).

de l'action corporelle (faciale) à l'action sur des objets pour agir sur d'autres objets, (ii) du connu au nouveau, (iii) de l'immédiateté de la réponse à des délais de rétention de la cible en mémoire de plus en plus élevés, (iv) du but de l'acte à l'acte dont le but n'est qu'inféré, (v) de l'action isolée à la séquence d'actions, (vi) de conditions de présentation allant d'identiques à différentes de celles de test (changements de contexte et de caractéristiques de l'objet), (vii) de modèles présentés dans la réalité à leur représentation culturelle (enregistrement vidéo noir et blanc), et (viii) d'actions concrètes à des représentations collectives de rôles sociaux.

2. Coordination intermodale

Revenons au problème que pose l'imitation faciale juvénile (section I.2.). Cette aptitude rend évident et implique un lien entre vision et production, canaux dont les flux d'information sont de sens opposés par rapport au Système Nerveux Central (SNC)¹²⁶. Dans l'univers des scientifiquement (re)connus, une architecture neuronale peut être avancée comme "charpente" potentielle de ce lien: l'**arc réflexe**¹²⁷. Or, au regard des données d'imitation précoce désormais à notre disposition, un mécanisme de type réflexe n'a que peu de crédit (Cf. section III.1.) sachant, en particulier, qu'à notre connaissance, quand la réponse de l'organisme n'est pas perturbée en cours de réalisation, une réponse réflexe ne fait pas plus l'objet de *correction ordonnée* qu'une réponse stimulus-déclenchée. De là, le lien vision-production de l'imitation faciale a de forte chance d'être clairement indirect, et se doit donc d'être posé en des termes plus subtils que réflexifs. Toujours dans l'univers des scientifiquement (re)connus, il existe un lien, au moins temporel, indéniable entre le monde de la perception et celui de la production: le fait qu'à chaque exécution motrice (ou mouvement), une information proprioceptive soit (r)envoyée au système nerveux central. Par défaut, nous considérons que cela est vrai quelque soit l'âge. Il est très plausible que la répétition de cette concomitance fasse l'objet de l'apprentissage d'une correspondance production-proprioception, au cours du *développement intra-utérin*, par **contiguïté temporelle**¹²⁸. Si une telle relation existe au niveau architectural au moins dès la naissance, et qu'elle entre en jeu dans le mécanisme d'imitation visuo-motrice, le problème du lien vision-*production* se reformule en problème de lien vision-*proprioception*, c'est-à-dire en termes de coordination intermodale entre *perception* de stimuli internes et *perception* de stimuli externes à l'organisme, ne faisant donc "communiquer" que des flux d'informations allant dans le même sens. Or, si l'on conçoit que des informations portant sur les milieux interne (proprioception) et externe (vision) à l'organisme puissent partager une relation, une coordination intermodale entre toutes les modalités perceptives, comme proposée notamment par Meltzoff et Moore (1997), devient crédible.

Au-delà de croyances: des données

L'idée d'une coordination intermodale plus ou moins généralisée a ainsi été le point de mire de nombreuses études développementales. Vous ne trouverez pas ici une revue exhaustive de cette littérature: le problème est complexe, au moins parce que plusieurs facteurs entrent en jeu, tels que le nombre de modalités sollicitées et l'intensité et la complexité des stimuli (voir par exemple Maurer, 1993). Sans rentrer dans le détail, nous vous présentons un échantillon

¹²⁶ Cela se comprend si l'on garde à l'esprit que le perceptif va vers le SNC, que le "productif" part du SNC et que les trains de potentiels d'action, ou " message" véhiculé par un neurone, sont *toujours unidirectionnels*, et que dans un nerf, ou faisceau de neurones, ce message va toujours dans le même sens pour toutes les cellules.

¹²⁷ **Arc réflexe**: Voie de conduction des influx dans le système nerveux, reliant un récepteur et un effecteur. Elle comprend un récepteur, i.e. un neurone sensitif, un centre d'intégration situé dans le système nerveux central, un neurone moteur et un effecteur.

¹²⁸ **Contiguïté temporelle**: "[P]rincipe selon lequel deux événements se déroulant de façon rapprochée dans le temps auront tendance à être associés ou appris" (Vauclair, 2004, p.45).

de résultats montrant une capacité de *détection d'équivalence intra-perception*, et donc de coordination intermodale précoce. Leur intérêt est aussi d'être liés à la parole: ils concernent tous la partie orofaciale du corps et les modalités reconnues comme étant impliquées dans la communication orofaciale.

a. Période néonatale: équivalence visuo-haptique et équivalence audio-motrice

Pourquoi nous intéresser à l'"haptique" plutôt qu'au "proprioceptif"? La modalité haptique, qui comme la proprioceptive fait partie du système somesthésique au moins chez l'adulte, a l'intérêt d'être une information *tactile* collectée au cours de *mouvements*, dits "d'exploration", de l'objet avec lequel il y a contact cutané. L'haptique est vu comme du tactile associé, au moins temporellement, à du proprioceptif, et la répétition de cette concomitance peut mener à l'apprentissage d'un lien tactilo-proprioceptif, au cours du *développement intra-utérin*, par contiguïté temporelle. A cela s'ajoute le fait que la modalité haptique intéresse particulièrement le développement de la *parole*: les *mouvements* du conduit vocal induisent justement des *contacts* entre les surfaces cutanées des articulateurs (eg. lèvre-lèvre, lèvre-langue) et entre articulateurs et parois du conduit (eg. langue-palais, langue-membrane périodontale). De plus, le conduit vocal est connu pour être riche en **mécanorécepteurs tactiles** et **proprioceptifs**¹²⁹, au moins chez l'adulte (Gentil, 1990).

Equivalence visuo-haptique précoce

Suite à la mise en évidence de l'imitation visuo-motrice de Meltzoff et Moore (1977), ce premier s'associe à Borton (Meltzoff et Borton, 1979) pour tester l'existence d'un lien visuo-tactile néonatal, en adaptant le paradigme expérimental de la préférence visuelle précoce. Classiquement, ce protocole a pour principe d'exposer les sujets à deux stimuli visuels présentés simultanément, chacun d'un côté, en position contrebalancée d'une personne à l'autre. Le temps total de fixation de chaque stimulus est l'indice de mesure de la préférence pour l'un ou pour l'autre des stimuli visuels (voir Harris et Butterworth, 2002, p. 107). A la différence de la procédure originale, les auteurs ont, au préalable, exposé les individus à un stimulus *haptique* afin de tester si l'information d'origine somesthésique entraîne une préférence pour l'objet correspondant présenté dans la modalité *visuelle*.

Dans cette étude, les sujets ont 29 jours en moyenne (N=64). Le stimulus haptique est, pour la moitié des participants/antes, une tétine de surface lisse, et, pour l'autre, une tétine garnie de "nodules". L'objet est inséré dans la bouche du sujet et induit un comportement de tétée, et donc des mouvements oraux. 90 secondes plus tard, la tétine est retirée de manière à ne pas être vue par le/la participant/ante, puis les stimuli visuels sont présentés pendant 20 s. Chaque stimulus visuel correspond à une version polystyrène, agrandie et orange, d'un des deux types de tétines, et se trouve à la gauche ou à la droite du sujet. La majorité des individus fixent plus longtemps le stimulus visuel correspondant à la tétine perçue par l'haptique mais désormais *absente*.

Il en découle que des nouveau-nés ont une capacité (i) de *discrimination haptique buccale*, (ii) de discrimination visuelle, (iii) de garder en mémoire une *représentation* d'informations d'origine *oro-haptique*, et (iv) de faire le *lien* entre cette représentation et l'information visuelle correspondante et donc de détecter une *équivalence visuo-haptique*. Le premier point n'est pas sans rappeler la finesse de l'acuité tactile de la langue chez l'adulte (Gentil, 1990), ainsi que l'importance de l'information somesthésique dans le contrôle de la parole et son rôle

¹²⁹ **Mécanorécepteurs**: cellules sensorielles sensibles au toucher (**mécanorécepteurs tactiles**), et à l'état de contraction et de tension des muscles et à la position spatiale des parties du corps (**mécanorécepteurs proprioceptifs**) (Purves *et al.*, 1999, pp.147, 149 et 152).

probablement primordial pour le développement de cette maîtrise (Cf. Schwartz et Boë, 2000; MacNeilage, Rootes, et Chase, 1967) De plus, la congruence visuo-haptique de mémoire semble attirer l'attention des nouveau-nés, vu qu'elle oriente leur regard et de fait leur perception visuelle. La détection d'équivalence intermodale peut donc peser sur *l'orientation des sources perceptives* et donc sur la qualité de l'information ainsi mise à disposition de mécanismes d'apprentissage potentiels, au moins, au cours du développement précoce¹³⁰.

Equivalence audio-motrice précoce

L'imitation orofaciale de modèles visuels ainsi que des travaux entrepris par Kugiumutzakis (1985, voir 1999) sur le développement de l'imitation de modèle vocaux amènent Chen, Striano et Rakoczy (2004) à s'interroger sur la présence dès la naissance de quelque aptitude en lien avec l'imitation de sons de parole. Le "quelque aptitude" plutôt que "l'aptitude" en soi se pose en raison de la difficulté de définir, et donc de coder, par rapport à une cible adulte et de manière non-ambiguë, les productions vocaliques néonatales, sachant la difficulté que peut représenter une utilisation du conduit vocal d'un niveau de maîtrise comparable à celui de la parole mature à cette période du développement (voir chapitre 1, section I.3.b.). Toutefois, s'il existe une capacité d'appariement visuo-moteur, à la naissance, pourrait-on mettre en évidence une capacité néonatale de mise en équivalence audio-motrice?

Pour tester l'hypothèse en jeu, les auteurs empruntent le protocole de Meltzoff et Moore (1983) (voir Figure 3-2, section III.A.1.a.), appliqué à des cibles utilisées par Kugiumutzakis (1985)¹³¹: le geste est soit le son [a], soit le son [m], et est présenté 8 fois par le visage de l'expérimentateur/trice. Les sujets ont de 1 à 7 jours et en moyenne 3 jours (N=25). Les réponses comptées comme équivalentes à la cible sont les mouvements buccaux correspondant à la production du modèle. En d'autres termes, l'appariement à [m] est codé comme étant le pincement des lèvres, et celui à [a], comme étant l'ouverture de la bouche. En résultat, la majorité des participants/antes donnent plus de réponses équivalentes à la cible qu'à l'opposée, que leurs yeux soient ouverts ou fermés au cours des périodes de test. De plus, l'analyse de l'effet de la position des paupières, yeux ouverts *vs.* fermés, ne révèle pas de différence significative entre les deux groupes d'individus¹³².

Il vient qu'à la naissance, bien que la capacité d'imitation oro-faciale de stimuli de parole soit difficile à mettre en évidence dans le cas où elle existerait, les nouveau-nés n'en sont pas moins capables d'exprimer un comportement précurseur de cette aptitude, qui ne s'explique pas par leur capacité à imiter des modèles visuels. De plus, cette étude met en évidence que *l'audition est porteuse d'une information articulatoire* pour les nouveau-nés. Enfin, la croyance selon laquelle l'imitation de modèle de parole utiliserait une comparaison entre cible et production propre qui soit inexorablement audio-auditive est remise en question, sachant que les sujets de cette études ne produisaient *pas de son* qui puisse être comparé à celui du stimulus, au moins dans le cas de l'appariement à [m] (voir aussi Kugiumutzakis, 1999) et, notamment, même si leurs paupières étaient closes. Ainsi la question d'une coordination *inter-*

¹³⁰ Des capacités de coordination visuo-haptique ont été mises en évidence avant 1 mois. En effet, Streri et Gentaz (2003) rapportent, en utilisant une procédure différente de celle de Meltzoff et Borton (1979), qu'une capacité de mise en équivalence visuo-haptique au niveau de la *main* est présente dès la naissance: l'âge moyen de leur sujet va de 16 à 110 h, pour une moyenne de 62 h. Les auteurs trouvent une capacité de détecter visuellement la nouveauté de la forme d'un objet par rapport à celle d'un autre objet auquel les sujets ont été habitués au niveau tactile: il y a une capacité de *transfert* d'information de la modalité haptique à la visuelle (pour plus de détails, et une comparaison des résultats obtenus par Meltzoff et Borton avec ceux de Streri et Gentaz, voir Vauclair, 2004, pp. 96-98).

¹³¹ Voir plus bas, section III.A.3..

¹³² La plupart avaient les paupières closes.

modale, vraisemblablement de type audio-proprioceptive, pour l'imitation précoce de modèles de parole, peut-elle être avancée.

b. De 1 à 8 mois: coordination audio-visuelle pour la parole

En quête d'une capacité de coordination intermodale précoce généralisée, Kuhl et Meltzoff (1982, 1984) reprennent le paradigme de la préférence visuelle modifiée (Meltzoff et Borton, 1979) pour tester la présence de cette aptitude dans le cas de stimuli audio-visuels de parole. L'idée avancée est que la vision est porteuse d'information *articulatoire*, comme le montre le phénomène d'imitation faciale précoce, et que la perception de la parole au moins adulte sait utiliser cette information (Schwartz *et al.*, 1998).

Les sujets ont de 18 à 20 semaines et en moyenne 4.5 mois (N=64). Les stimuli auditifs sont les sons [a] et [i]. Pour la moitié des participants/antes, le stimulus auditif est le son [a], et, pour l'autre, [i], chacun passé en boucle pendant 2 minutes. Les stimuli visuels correspondent chacun à l'enregistrement vidéo muet d'une personne prononçant l'un des deux sons en synchronie avec l'enregistrement sonore. Au départ, les vidéos muettes sont présentées durant 10 secondes chacune en l'absence de l'autre puis toutes deux cachées brièvement. Le stimulus auditif est émis en découvrant les stimuli visuels. La majorité des individus fixent plus longtemps le stimulus visuel correspondant à la voyelle perçue auditivement. De plus, la possibilité que l'information d'origine auditive utilisée par les sujets pour orienter leur regard soit de type temporelle ou énergétique a été testée et rejetée (ici et dans Kuhl, William et Meltzoff, 1991).

Il en découle qu'à 4.5 mois il y a une capacité (i) de discrimination auditive de sons de parole, comme de nombreuses études en perception l'ont montré (voir, par exemple, Polka et Werker, 1994), (ii) d'extraire une information vraisemblablement spectrale liée à l'articulation, (iii) d'extraire une information visuelle d'un mouvement de parole, et (iv) de faire le *lien* entre le son et la vision de voyelles, et donc de détecter une *équivalence audio-visuelle*. De plus, comme dans le cas de la coordination visuo-haptique, la détection d'une congruence audio-visuelle peut peser sur *l'orientation des sources perceptives* et donc sur la qualité de l'information ainsi mise à disposition de mécanismes *d'apprentissage de parole*.

Ces résultats ont été répliqués dans le cas des paires [i]-[u] et [a]-[u] (Kuhl et Meltzoff, 1988, cité dans Kuhl, Williams et Meltzoff, 1991, pour [i]-[u]; Legerstee, 1990, en utilisant des visages réels, et Walton et Bower, 1993, avec un paradigme utilisant la succion non-nutritive, pour [a]-[u]).

La capacité de détecter une équivalence entre informations visuelle et auditive peut être le fruit d'un apprentissage basé sur la perception concomitante des deux en condition écologique. En effet, les sujets sus-cités ont eu l'occasion en quelques mois d'associer ces patrons perceptifs au cours d'interactions face-à-face avec des adultes (Cf. Introduction du III.A.), sachant que les voyelles [i], [a] et [u] sont présentes dans leur langue ambiante (une variété de l'Anglais), comme dans bien d'autres (Vallée, 1994). La question de l'origine de la coordination audio-visuelle a été posée par Walton et Bower (1993), en testant s'il y a capacité de détecter une équivalence visuo-auditive dans le cas d'un son de parole inconnu. Les auteurs montrent que des individus de 6-8 mois préfèrent l'association d'une vidéo de [u] avec les sons [u] et [y] (réalisations possibles sachant la vidéo) qu'avec [i] (réalisation impossible). Or, dans cette étude, [y] ne figure pas dans la langue ambiante des sujets. De là, l'on est en droit de s'interroger sur la possibilité que la coordination audio-visuelle de la perception de parole ne soit pas un apprentissage post-natal du type association son-image simple (conditionnement classique, voir section I.1.).

Une autre manière d'approcher l'origine de la coordination audio-visuelle a été testée par Patterson et Werker (2003), en répliquant l'étude de Kuhl et Meltzoff (1982, 1984) à l'âge le

plus précoce possible, sachant la méthode: 2 mois. Les sujets ont plus exactement de 7.8 à 11.1 semaines (N=32), et font preuve des mêmes capacités de *détection d'équivalence audio-visuelle* que ceux de Kuhl et Meltzoff (avec des performances parfois meilleures)¹³³.

Il en découle que, si la coordination audio-visuelle est le résultat d'un apprentissage associatif *post-natal* des informations auditive et visuelle concomitantes dans la parole, cet apprentissage ne semble pas nécessiter plus de 2 mois. Cependant, une autre hypothèse peut être avancée. Sans écarter pour autant la possibilité de mise en jeu de mécanismes d'apprentissage, l'aptitude à détecter des équivalences intermodales peut commencer son développement durant la gestation, en étant vraisemblablement le résultat d'une interaction organisme-environnement¹³⁴ (voir plus bas).

L'ensemble des études sus-citées sur la coordination audio-visuelle dans la perception de parole ont, en particulier, un point commun, un détail, qui n'enlève rien à leurs apports. En effet, à la différence de la procédure de Meltzoff et Borton (1979), les stimuli visuels sont toujours présentés en même temps que le stimulus auditif. Il vient que la question de la mémorisation d'une *représentation auditive* comme base de la détection d'une équivalence audio-visuelle ne peut être avancée. Toutefois, la présentation d'un son sans image pourrait ne pas être "comprise" comme étant de la parole à ces âges précoces, sachant qu'en condition standard, elle est intimement liée à la présence d'un visage. Il serait intéressant, d'après nous, de répliquer l'expérience de Kuhl et Meltzoff (1984) en passant le son en présence de deux visages passifs, puis ces visages mimant les voyelles choisies en l'absence du son.

Faisons le point

En résumé, dans les premiers mois de vie, l'on disposerait d'au moins 4 *relations* qui s'expriment soit en termes de capacité de détection d'une équivalence (ou congruence) intermodale, soit en termes de capacité de mise en équivalence (ou appariement) perceptuo-motrice: audio-oro-motrice et visuo-oro-motrice, dès la naissance, visuo-oro-haptique, à 1 mois, et audio-visuelle, à 2 mois. De plus, à l'idée que la vision précoce porte une information articulatoire, nous pouvons désormais ajouter que l'audition en fait de même. Que penser de l'idée d'une coordination intermodale généralisée?

Sans rentrer dans le débat du type "pour ou contre la synesthésie juvénile", ou confusion des sens à la naissance, ni même nous submerger de données suscitées par ce thème, nous voulons faire le point sur ce dont nous disposons et qui peut intéresser l'imitation orofaciale précoce pour le développement de la parole. Il nous semble clair que plusieurs modalités "communiquent" entre elles, mais que coordination ne signifie pas confusion. En ce qui concerne celle-ci, rappelons juste que d'après la capacité de correction ordonnée mise en évidence dans l'étude de l'imitation faciale néonatale, l'information visuelle n'est pas confondue avec l'information motrice et donc, *a priori*, proprioceptive.

Si nous acceptons toujours que le proprioceptif est le pendant perceptif du moteur, les 4 termes de coordination ci-dessus peuvent se reformuler de la manière suivante: visuo-oro-proprioceptive (visuo-oro-motrice), visuo-oro-tactilo-proprioceptive (visuo-oro-haptique), audio-oro-proprioceptive (audio-motrice) et audio-visuelle. Les 3 premières relations impliquent le partage d'une information avec la proprioception. De là, il apparaît que la proprioception "communique" avec la vision, le tactile et l'audition. Nous proposons qu'elle peut constituer le lien non seulement entre le domaine perceptif et le domaine moteur, mais

¹³³ Patterson et Werker (2003) ajoutent que le genre du visage et de la voix des personnages-stimuli n'est pas différencié à cet âge, en accord avec d'autres études qui montrent que cette caractéristique n'est pas détectée avant environ 8 mois.

¹³⁴ Nous ne proposons pas l'hypothèse d'un déterminisme génétique: comme pour tout autre phénomène comportemental, il nous semble difficile d'invoquer l'existence d'un gène de la coordination intermodale.

aussi entre modalités extéro-perceptives, en accord avec l'idée d'une métrique commune à/aux différentes modalités en termes de "relation-entre-organes" de Meltzoff et Moore (1997). Il en découle que la proprioception entre probablement en jeu dans "l'extraction" d'informations *articulatoires* notamment d'origine visuelle et/ou auditive. Peut-être sert-elle de base développementale à la coordination entre l'audition et la vision, sachant que la somesthésie semble le premier système perceptif à être fonctionnel au cours du développement et cela dès la 11^{ème} semaine post-conception (Vauclair, 2004, p. 53; Butterworth 1999).

Proposition pour un développement intra-utérin d'une coordination audio-proprio-visuelle

Nous disposons de très peu d'éléments pour rendre compte d'un développement d'une coordination audio-proprio-visuelle. Cependant, si l'on reprend les idées rencontrées section II.2., l'identification visuelle de l'organe pourrait se baser sur la connaissance de sa "signature cinétique" au niveau proprioceptif. Nous proposons que l'émergence de cette connaissance pourrait avoir lieu au cours d'un apprentissage pré-natal faisant appel, notamment, au cervelet¹³⁵. En effet, le cervelet est connu pour son intervention dans des mécanismes d'apprentissage et de contrôle sensori-moteur au moins chez l'adulte (voir Darlot, 2001): il est possible que des propriétés biomécaniques (sous la forme de modèles internes) d'organes de soi soient apprises au cours du développement intra-utérin, sur la base de l'expérience de l'activité propre en relation avec ses conséquences, notamment, tactiles et proprioceptives. Après la naissance, ces propriétés pourraient être détectées dans le flux d'information visuelle, permettant ainsi l'identification de l'organe, mais peut-être aussi une description de son mouvement, chez autrui¹³⁶.

Selon des principes similaires, une cartographie de correspondances audio-proprioceptives pourrait se faire entre la proprioception de configurations articulatoires basiques du conduit vocal de l'embryon et des conséquences auditives et tactiles auxquelles elles mènent, sachant que l'audition est fonctionnelle au moins à partir de 32 semaines post-conception (Vauclair, 2004, p. 53). Dans ce cas, la source d'excitation du conduit ne serait pas produite par le larynx (comme pour la parole), mais serait plutôt l'ensemble des ondes acoustiques dont le liquide amniotique est le support matériel¹³⁷. Dans le même ordre d'idée, le fluide amniotique étant plus visqueux que l'air, les valeurs des fréquences de résonance du conduit seraient vraisemblablement différentes de celles obtenues à l'extérieur. Toutefois, pour une forme de conduit vocal donnée, toute ou partie de la *relation des formants entre eux* (e.g. leur patron de focalisation) devrait être similaire dans les deux milieux. En d'autres termes, il semble plausible qu'une cartographie audio-proprioceptive pré-natale du comportement du conduit

¹³⁵ Au moins une partie de cette structure (le vermi) est fonctionnelle à la naissance, tout comme le tronc cérébral (où se trouvent les corps des motoneurones de la face), les noyaux sous-thalamiques et les cortex sensoriel et moteur primaires (Vinter, 1998, p. 6). Il est très probable que l'ensemble de ces structures soient déjà en activité avant la naissance.

¹³⁶ Cette idée trouve appui dans le fait que des sujets (adultes) aient une plus grande probabilité de reconnaître leur propre démarche (présentée sous la forme de point lumineux) que celle des personnes avec qui ils/elles habitent et dont la démarche est plus souvent vue que la leur (Beardworth et Bukner, 1981). Cela suggère que l'on peut reconnaître les règles de production de ses propres mouvements dans la modalité visuelle, vraisemblablement sur la base de leur connaissance proprioceptive.

¹³⁷ Ces sons comprennent la voix de la mère (si elle n'est pas aphone), des bruits intérieurs à son corps, tels que les battements du cœur, mais aussi extérieurs, tels que de la musique (voir Vauclair, 2004, p. 55). En ce qui concerne la source d'excitation du conduit vocal, nous avons vu que, dès 10-11 semaines de gestation, il y a des mouvements respiratoires (voir chapitre 1, section I.3.c.). Toutefois, nous pensons que si les sujets de Chen *et al.* n'ont pas activé leur larynx par un souffle d'air, ils/elles auraient probablement eu plus de mal à l'activer avec un souffle de fluide plus visqueux. D'une certaine manière, le fait qu'ils/elles ne l'aient pas activé en donnant leur réponse peut être vu comme un élément en accord avec la non-utilisation intra-utérine du larynx comme source d'excitation du conduit vocal.

vocal existe et serve de base à la capacité de mise en équivalence audio-motrice rapportée par Chen *et al.* (2004). De là, la coordination audio-visuelle pourrait se faire via la connaissance d'équivalences audio-proprioceptive et visuo-proprioceptive acquises *in utero*.

3. L'imitation audio-visuelle précoce de modèles oro-faciaux

Au cours de leur étude sur la coordination audio-visuelle de la perception précoce de parole, Kuhl et Meltzoff (1982, Expérience 1) rapportent la présence d'imitation des stimuli sonores proposés aux participants/antes : des vocalisations de sujets (4.5 mois) avaient des caractéristiques similaires à celles du stimulus auditif, au niveau de l'identité ([i], [a]) et de la durée (1 s) de la voyelle, mais aussi de son contour intonatif (prosodie) dont la forme montante-descendante est considérée peu fréquente dans le répertoire des productions de cet âge. Depuis, ces observations ont cédé la place à des expériences plus contrôlées.

Legerstee (1990) étudie l'influence de la capacité de détection d'équivalence audio-visuelle sur celle d'imitation de stimuli de parole. Dans ce but, elle reprend le protocole de Meltzoff et Moore (1983) (voir Figure 3-2, section III.A.1.a.) avec pour cibles les voyelles [u] et [a], et des sujets de 3 mois en moyenne (N=30)¹³⁸. Dans cette expérience, la cible est le résultat du couplage synchrone des composantes auditive et visuelle, variant indépendamment l'une de l'autre: la composante auditive correspond au son [a] ou [u] diffusé par un casque à écouteurs au sujet et au/à la démonstrateur/trice dont les mouvements faciaux forment la composante visuelle, avec une *protrusion des lèvres* (LP) pour [u] et une *ouverture-fermeture de la mandibule* (MO) pour [a]. Pour la moitié des participants/antes, le visage mime la même voyelle que celle perçue auditivement (*Condition d'appariement audio-visuelle*), pour l'autre moitié, le visage mime la voyelle opposée (*Condition de non-appariement*). Les sujets sont filmés de près et audio-enregistrés. La première bande sert au calcul de la fréquence de leurs gestes faciaux selon 3 classes: LP, MO et "autres". La seconde bande permet d'estimer la fréquence de leurs vocalisations en fonction de 3 catégories de transcription phonétique: pseudo-[a], pseudo-[u] et "autres". En moyenne, les sujets (i) produisent le geste MO (resp. LP) plus fréquemment quand le stimulus auditif est [a] (resp. [u]) que [u] (resp. [a]), (ii) produisent des vocalisations pseudo-[a] (resp. pseudo-[u]) plus souvent quand le stimulus auditif est [a] (resp. [u]) que [u] (resp. [a]), et (iii) les effets (i) et (ii) viennent principalement de la condition expérimentale d'appariement, celle de non-appariement audio-visuel donnant lieu globalement à moins de mouvements¹³⁹.

Il en découle qu'à 3 mois des êtres humains sont capables d'imiter sélectivement les sons vocaliques [a] et [u] en présence du mouvement facial correspondant, sans délai de rétention de la cible. Nous avons vu que, peu après la naissance, l'information auditive véhicule les informations articulatoires d'ouverture et de pincement des lèvres (Chen *et al.*, 2004): pourquoi des individus plus âgés auraient-ils besoin d'une information visuelle en accord avec l'auditive pour les inciter à imiter? Tout d'abord, si le son [a] suffit à induire une ouverture de la bouche de nouveaux-nés (Chen *et al.*, 2004), le son [u] pourrait nécessiter la vision de la protrusion des lèvres pour être imité. En effet, le [u] figure peu dans le répertoire spontané de cette période développementale et peut donc être plus difficile à produire que [a] en général. Toutefois, un regard sur la publication nous apprend que le son [a] est faiblement exprimé quand la voyelle est diffusée en condition de non-appariement: une difficulté de production pour [u] ne suffit pas à expliquer le peu de réponses imitatives face au non-appariement. Un

¹³⁸ Ici, l'expérience commence par une phase de familiarisation où les voyelles [a] et [u] sont mimées deux fois chacune.

¹³⁹ Aucun effet du stimulus sonore, de la condition expérimentale et de leur interaction n'a été trouvé pour les 2 classes de productions "autres". De plus, les effectifs d'individus en fonction du patron de réponse ne sont pas donnés (seules les *fréquences* correspondantes sont fournies).

autre point de vue s'impose. Si l'on ajoute au "sens" articulatoire que porte le son, celui porté par la vision (Cf. section III.A.1.), dans la condition de non-appariement, les sujets de Legerstee sont, somme toute, en présence d'une information auditive en *conflit* avec la visuelle: la difficulté de définir *une* cible "à imiter" pourrait contribuer à expliquer une fréquence des réponses imitatives, mais aussi de mouvements visibles et audibles, inférieure face à l'incongruence que face à la congruence. Cela n'enlève rien au fait que l'imitation orofaciale de modèles de parole serait apte à utiliser l'équivalence intermodale, que les sujets semblent détecter dans le flux d'information audio-visuelle mis à leur disposition, cette faculté de détection étant probablement présente depuis l'âge de 2 mois au moins (Patterson et Werker, 2003). De plus, l'apprentissage potentiellement sous-tendu par l'imitation orofaciale de modèle de parole a la possibilité de se servir de l'aspect multi-modal de la perception de cet objet pour étendre le répertoire gestuel du sujet. De là, la découverte de comportements imitatifs au cours d'expériences portant sur la coordination intermodale de la perception précoce de parole ne paraît pas être le fruit du hasard. Cependant, Legerstee donne toujours le stimulus sonore en présence d'un visage en mouvement, que ce mouvement visible soit congruent ou non avec le son: la vision de la face d'un/une congénère est-elle nécessaire à l'imitation précoce de parole?

Lebib et Baudonnière (2000) ont adressé (sans le vouloir) cette question à des individus de 3 mois (N=23), ayant pour cible le geste [aə], issu d'une bande son, pour le versant audible, et/ou d'un visage réel, pour le versant visible¹⁴⁰. L'expérience comprend une séquence de 4 conditions : chacune dure 1 minute, est séparée de la suivante d'une *minute de pause* et correspond à la présentation en boucle du stimulus dispensé toutes les 4 secondes. La première phase expérimentale présente le son sans le visage (*Condition contrôle*), la seconde le geste visible en synchronie avec le son (*Condition écologique*), puis les deux autres donnent le son associé à un visage passif (*Condition ventriloque*) et le geste visible sans son (*Condition de mime*). Les deux premières conditions sont en ordre invariant et les deux dernières en ordre contrebalancé¹⁴¹.

Les résultats montrent que la condition écologique, où l'auditif et le visuel s'accordent, est celle qui suscite le plus de réponses imitatives. De plus, le son en l'absence de visage de la condition contrôle n'incite pas les sujets à imiter. Ainsi, à 3 mois, aurait-on besoin de la *présence d'un visage*, de préférence dont le *mouvement s'apparie* au son ou, du moins, ne soit pas en "conflit actif" avec, pour imiter un modèle de parole. Cette nécessité indique que la parole est vraisemblablement perçue comme étant un *acte imitable* si le son est accompagné d'un visage. Et qu'est la parole si ce n'est, au moins, un signal sonore sorti d'une face?

Un élément subsidiaire que l'on peut tirer des résultats publiés de cette étude est que des sujets, après avoir perçu le modèle audio-visuel complet (condition écologique), l'imitent, bien que sa partie visible (condition ventriloque) ou sa partie audible (condition de mime) soit actuellement *absente* du champ perceptif associé *depuis plus d'une minute*. Cela pourrait s'expliquer, d'après nous, au moins de deux manières. La première est que les sujets imiteraient, sur le mode immédiat, le modèle auditif dans la condition ventriloque et le modèle visuel, en activant leur larynx, dans la condition de mime. La seconde interprétation possible serait qu'à 3 mois une *représentation visuelle* d'un modèle de parole, dans la situation

¹⁴⁰ A la base, les auteurs voulaient tester si la congruence entre audition et vision est un "contexte déclencheur" de l'imitation précoce de modèles de parole.

¹⁴¹ Pour la moitié des participants/antes la séquence se termine sur le mime puis le ventriloque, pour l'autre moitié sur le ventriloque puis le mime. Les deux premières conditions ne sont pas contrebalancées avec les deux dernières pour éviter que le désagrément d'une condition (non-congruente), pouvant susciter des réactions d'alarme (stress), ne soit associée au visage de l'expérimentateur/trice. En particulier, de nombreuses espèces animales, dont l'humaine, peuvent vivre comme une préparation à l'attaque, et donc comme une situation dangereuse, la fixation du regard par autrui.

ventriloque, et, surtout, une *représentation auditive* d'un modèle de parole, dans la situation de mime, peuvent être gardées en mémoire et servir à l'imitation "*semi-différée*" d'une minute du modèle complet. De plus, les auteurs remarquent que, dans ces deux dernières conditions, les instances d'imitations avaient lieu à l'initiale de ces phases expérimentales. Dans le cadre de la seconde interprétation, il semblerait que les individus aient associé le visage de l'expérimentateur/trice à son *geste de parole caractéristique* et qu'ils/elles essaient de l'éliciter chez cette personne muette ou immobile voilà plus d'une minute. De là, l'imitation orofaciale de modèle audio-visuel de parole pourrait, comme celle de modèle visuel, servir, en plus d'un apprentissage d'actions nouvelles, une fonction d'identification des personnes (Cf. III.A.1.b. Données complémentaires).

Quatorze ans après avoir observé des instances d'imitation orofaciale précoce de modèles audio-visuels de parole, Kuhl et Meltzoff (1996) entreprennent d'étudier l'évolution de cette aptitude en termes phonétiques et acoustiques entre 12 et 20 semaines. Leurs sujets sont de 3 classes d'âge: en moyenne 3 mois (*12 semaines*), 3-4 mois (*16 semaines*) et 4-5 mois (*20 semaines*) (N=24 par classe). Pour chaque classe d'âge, chaque individu est affecté aléatoirement à un groupe auquel ne sera présenté que la voyelle-cible [i], [a] ou [u], présentée par l'enregistrement audio-visuel d'un visage adulte, via un moniteur de télévision. La période de présentation correspond à la cible alternée avec une pause de 2-3 s, en boucle, durant entre 5 et 10 minutes. Chaque sujet est enregistré pendant 5 minutes au cours de la présentation. Ce traitement est dispensé une fois par jour pendant 3 jours consécutifs. Les vocalisations infantiles ayant lieu entre 2 stimuli sont transcrites et classées selon 3 catégories: pseudo-[i], pseudo-[a] et pseudo-[u]. Il en résulte que sur l'ensemble des 224 vocalisations ayant pu être transcrites, tous âges confondus, les sujets (N=45) produisent plus de vocalisations labellisées [i] (resp. [a], [u]) lorsque le stimulus est [i] (resp. [a], [u]) que [a] et [u] (resp. [i] et [u], [i] et [a]). De plus, les individus dont la plupart des productions ressemblent à la voyelle [i] (resp [a], [u]), d'après le jugement perceptif de deux adultes, sont majoritairement ceux étant exposés à la cible [i] (resp [a], [u]). Cette tendance se retrouve au niveau de la classe d'âge 20 semaines mais n'est significative pour les 2 plus jeunes qu'en fusionnant leurs données en une seule tranche d'âge (12-16 semaines), à cause du nombre relativement restreint de données. Au niveau acoustique, les productions sont globalement en faveur d'un comportement imitatif pour le groupe le plus âgé (4-5 mois).

Il vient que les nourrissons sont globalement capables d'imiter sélectivement les voyelles audio-visuelles [i], [a] et [u], vraisemblablement dans la mesure de leurs capacités articulatoires qui sont *a priori* fonction du "temps" d'expérience avec le conduit vocal propre. De plus, d'après les données exploitables, cette aptitude est visible au niveau acoustique à 4-5 mois. Sachant que des nouveau-nés peuvent donner l'équivalence articulatoire d'un son de parole et que des êtres humains vocalisent dès 4 semaines, pourquoi ces sujets plus âgés n'imiteraient-ils pas plus que cela?

L'on peut remarquer dans un premier temps qu'à la différence des sujets de Chen *et al.* (2004), ceux de 4-5 mois ont à *coordonner* les mouvements de leur larynx avec ceux du reste du conduit vocal et ceux de leur respiration¹⁴², ce qui peut représenter une difficulté de plus pour produire des sons qui ressemblent à la cible. De plus, au niveau méthodologique, une démarche faisant appel à l'analyse transcriptionnelle et acoustique de vocalisations infantiles a pour revers inévitable, sachant le temps d'expérimentation, un nombre relativement faible de vocalisations par classe d'âges, ne donnant droit qu'à une signification statistique restreinte. A cette difficulté s'ajoute peut-être la concision des pauses entre les stimuli: il est possible que l'utilisation d'un casque à écouteurs aurait permis l'analyse d'un plus grand nombre de

¹⁴² Voir chapitre 1, section III.3..

vocalisations¹⁴³. Ces éléments expliquent probablement pourquoi le rapport de cette étude ne fournit pas d'évolution des performances au cours des périodes de test et des 3 jours d'expérimentation. Il nous est donc impossible de savoir (i) si l'imitation a lieu de manière prépondérante en début ou en fin des périodes de test, et (ii) si ses performances s'améliorent aux deuxième et troisième jours. En d'autres termes, dans le premier cas, nous ne sommes pas en mesure d'apprécier le poids de l'habituation au stimulus: sa présentation en cycle sans pause de plus de 2-3 secondes pourrait avoir perdu l'attention des sujets. Dans le second cas, l'effet de l'apprentissage d'un jour sur l'autre ne peut pas être estimé¹⁴⁴. Tout cela n'enlève rien aux apports de ce travail. En particulier, à notre connaissance, il est le seul à fournir une analyse acoustico-phonétique de l'imitation oro-faciale précoce de modèles audio-visuels de parole. Nous retrouverons cette étude dans la troisième partie de ce mémoire, à travers les résultats de nos travaux de modélisation.

Kugiumutzakis (1985, cité dans 1999) fournit une étude longitudinale de l'imitation des modèles [a], [m] et "ang" présentés en condition écologique par un visage réel, de 45 minutes à 6 mois. Il donne de cette capacité un décours développemental en forme de cloche culminant à environ 3,5 mois. Toutefois, comme nous l'apprend l'imitation orofaciale de modèles visuels et les résultats obtenus par Chen *et al.* (2003), cela ne signifie pas que cette aptitude apparaîtrait pour disparaître. De plus, l'imitation orofaciale immédiate de modèle de parole pourrait céder la place à sa forme différée au cours du développement, expliquant ainsi la raison pour laquelle aucune étude, à notre connaissance, ne signale l'expression de cette aptitude au-delà de 6 mois, âge auquel un ajustement de la production est rapporté au niveau de la prosodie (de Boysson-Bardies *et al.*, 1984). Dans une étude d'observation longitudinale et naturaliste de l'imitation vocale immédiate spontanée, Kugiumutzakis (1993, cité dans 1999) nous apprend que de 15 jours à 6 mois, ce comportement a lieu en moyenne une fois toutes les 3 minutes.

B. L'imitation chez les animaux non-humains

Le problème du quand et comment au cours de l'évolution la capacité d'imitation est apparue chez les êtres humains en appelle à des apports de la psychologie comparée et de l'éthologie pour savoir quelles espèces *sociales* imitent. Cette approche phylogénétique n'a évidemment pas pour but de faire un lien direct entre génome et comportement, mais veut (juste) apprécier la répartition de cette faculté dans le règne animal, sachant que le génome correspond à un *ensemble de contraintes biologiques délimitant l'univers des possibles organiques et donc comportementaux*¹⁴⁵.

Une incursion dans ces domaines suffira à montrer que l'imitation n'est pas plus le privilège de l'Humain (*Homo sapiens*) qu'il n'est celui de ses adultes. En effet, la capacité d'imitation de modèles oro-faciaux est présente dès la naissance, dans le genre *Homo* mais aussi dans le genre *Pan*, en particulier dans l'espèce Chimpanzé. L'imitation n'est pas non plus propre à l'embranchement des primates: au moins une espèce de cétacé imite des modèles corporels et vocaux. Les mammifères ne sont pas non plus les seuls à pouvoir imiter:

¹⁴³ En effet, l'utilisation d'un casque permet de "gagner" en bande analysable car le stimulus auditif n'apparaît plus sur l'enregistrement des productions du sujet.

¹⁴⁴ Toutefois, Meltzoff (1999) déclare que dans cette étude les performances s'amélioreraient.

¹⁴⁵ Nous garderons à l'esprit que *seul* le génome est clairement héritable, et que les gènes codent, i.e. sont des instructions pour produire, *a priori*, uniquement des *molécules biologiques* (eg. protéines, acides ribonucléiques ou ARNs). Or, il est difficile de trouver un lien direct entre molécules biologiques et comportements bien que l'on sache l'effet déplorable que certaines formes de gènes peuvent avoir sur le comportement des individus (eg. la forme du gène IT15, codant la glutamine, qui contient une répétition du motif CAG, et que l'on soupçonne d'être à la base de la "maladie" de Huntington).

l'imitation vocale précoce est présente dans le répertoire comportemental de juvéniles aviaires¹⁴⁶. L'on constatera dans ce qui suit la présence d'une capacité et/ou de l'usage de l'imitation dans des espèces aux *génomes très différents les uns des autres* et, en particulier, proches autant que distants de celui de l'être humain.

1. L'imitation chez les non-primates

L'imitation visuo-motrice d'actions simples est une capacité mise en évidence chez des rats, des pigeons et des perroquets (Whiten, 1998 dans Butterworth, 1999).

Dans le phylum des Oiseaux, l'imitation "vocale" précoce des oiseaux chanteurs¹⁴⁷ est étudiée depuis plus d'un siècle (Blackmore, 1998). Chez les chanteurs, la transmission de notes/thèmes relève de la tradition au point que l'on parle de "*dialectes*": les juvéniles apprennent par imitation des notes/thèmes de leurs parents et voisins en les incorporant à leur chant. S'ils viennent à s'installer ailleurs, leur chant peut servir de *marqueur* de leur zone géographique natale, et semble leur donner moins de facilité pour leur "emménagement". Chez les pinsons, les juvéniles sont exposés au chant de leur père bien avant d'être capables de chanter. Quelques mois plus tard, ils commencent à produire une variété de sons (période du **préchant**) pour se restreindre graduellement¹⁴⁸ à ceux entendus quand ils étaient plus jeunes. L'apprentissage présente une **période critique**¹⁴⁸ et nécessite la perception (auditive) de leur propre chant pour l'apparier à celui qu'ils se rappellent (Blakemore, 1998): il semble donc y avoir capacité d'imitation vocale *différée*. Chez certaines espèces, la fonction du chant dicte si et quand la forme se fixe: si le chant sert à défendre son territoire, l'apprentissage continue à l'âge l'adulte, mais s'il entre en jeu dans l'appariement sexuel (*mating*), il prend sa forme définitive à l'âge de reproduction (Butterworth, 1999).

Toutefois, l'imitation vocale dans les espèces aviaires ne s'arrête pas à celle de congénères. Les perroquets et les mainates sont connus pour leur imitation de la voix humaine. De plus, selon Davis (1981), les productions d'une espèce d'oiseau-lyre comprennent en majorité des sons non-spécifiques tels que ceux issus d'aboiements canins et d'instruments de musique (Butterworth, 1999).

Du côté des mammifères, le grand dauphin, ou *Tursiops truncatus*, a une capacité d'imitation sur commande, par une instruction sonore ou visuelle symbolique, et sait généraliser cette opération à tous sons ou comportements *nouveaux* (Herman, 2002). En effet, les membres de cette espèce peuvent imiter un grand nombre de modèles sonores et corporels, et cela que les démonstrateurs/trices soient des congénères ou des humains. En particulier, dans le cas de l'imitation visuelle de modèles humains, ils/elles montrent une capacité de mise en correspondance du plan d'organisation de leur propre corps avec celui du corps humain. Par exemple, si l'expérimentateur/trice remue le bras, ils/elles remuent la nageoire pectorale; si il/elle lève ses membres inférieurs, ils/elles lèvent leur nageoire caudale, et si la configuration combine levée des bras et des jambes, ils/elles lèvent les nageoires pectorales et caudale. Il semble donc y avoir une capacité d'imitation *interspécifique*. Cette capacité de mise en correspondance entre visuel et moteur se retrouve dans le fait que des dauphins sachent utiliser leur reflet dans un miroir pour regarder une tache qui a été posée, sous anesthésie, sur une partie de leur corps à laquelle leur regard ne peut accéder directement: on parle de *reconnaissance dans le miroir* (pour des références, voir Gallup, Anderson et Shillito, 2002). De plus, Herman (2002) défend que les dauphins gardent en mémoire une

¹⁴⁶ Aviaire: relatif aux oiseaux.

¹⁴⁷ Les oiseaux chanteurs font partie de l'ordre des Passereaux.

¹⁴⁸ **Période critique**: "[P]ériode du développement, de durée limitée, pendant laquelle les organismes présentent une sensibilité particulière aux effets de l'expérience" (Purves *et al.*, 1999, p. 573).

représentation du comportement qu'ils viennent d'effectuer et que cette représentation est mise à jour au fur et à mesure de leurs faits et gestes. En effet, ils obéissent aux instructions abstraites signifiant "répète" *versus* "fais autre chose que ce que tu viens de faire", ce qui revient dans les termes classiques à s'imiter *versus* "s'anti-imiter". D'après le même auteur, l'imitation des dauphins, espèce qui vit en **groupe**¹⁴⁹, pourrait avoir un lien (i) avec la coordination de l'ensemble des individus par la synchronisation de leurs mouvements, qui a été observée en non-captivité, lors de tâches de recherche de nourriture (*foraging*) et de capture de proie, mais aussi (ii) avec l'expression de l'affiliation sociale.

Cette proposition se fait l'écho du rôle de l'imitation "vocale" dans le système de communication de cette espèce (Nguyen, 2002). En effet, pour que la reconnaissance interindividuelle se fasse, les individus développent une **signature sifflée** qui leur est propre, ainsi qu'un **répertoire de sifflements** partagés par les membres du groupe. Les juvéniles sont élevés à plusieurs au sein de groupes de mères, et dépendent de la leur jusqu'à l'âge de 3 à 6 ans. Au cours de la première année de vie, la signature qui est une répétition de signaux continus, ou "syllabes", se construit à partir d'une grande variété de productions sonores dont une partie est "perdue" par la suite d'après Blackmore (1998). La forme de ce sifflement est fixe sur plusieurs années (Nguyen, 2002). Dans des interactions sociales, les dauphins imitent la signature de congénères et peuvent l'incorporer dans leur répertoire de sifflements. Toutefois, le plagiat se distingue de l'original, par exemple, via un décalage en fréquence. Les éléments communs entre répertoires individuels servent, en particulier, à la reconnaissance des membres du groupe. De plus, au cours de jeux, des dauphins *imitent en différé*, de manière spontanée, des sons produits par des objets (Blackmore, 1998).

De ces "merveilles du monde", nous retiendrons que l'imitation se trouve dans des espèces d'embranchements phylogénétiques très différents. En particulier, dans le cas de modèles vocaux, les oiseaux, les cétacés et les humains ont des moyens de production bien distincts: les premiers ont une syrinx, les seconds n'ont pas de larynx et les derniers ont un conduit vocal. Des génomes très différents permettent l'expression d'un même type d'apprentissage: l'imitation semble pouvoir s'appuyer sur une architecture cérébrale très diverse et a peu de chance d'être génétiquement déterminée.

2. L'imitation chez les primates non-humains: Hominés vs. non-Hominés

Dans le phylum des Primates, les anglophones regroupent les Prosimiens, les Platyrrhiniens et les Cynomorphes (qui incluent les Macaques) sous le label de "Monkeys", tandis que les Anthropomorphes, dont fait partie l'espèce humaine, sont nommés "Apes" (Figure 3-5).

En général, les tests de l'aptitude à imiter des primates non-humains ont pour démonstrateur/trice un être humain, c'est-à-dire, un membre d'une autre espèce, et font appel à l'apprentissage d'actions arbitraires et/ou, le plus souvent, de tâches faisant intervenir l'utilisation d'*outil*, i.e. d'action sur un objet pour agir sur un autre objet. Sans rentrer dans le débat de l'imitation chez les primates non-humains, nous garderons à l'esprit que (i) la non mise en évidence n'est pas synonyme d'inexistence, et (ii) qu'il est difficile de savoir si les preuves négatives sont le résultat d'une incapacité à imiter des individus testés, de la difficulté

¹⁴⁹ **Groupe**: le groupe est une structure de socialité en biologie des populations. Legay et Debouzie (1987, p16-17) nous apprennent que "[l]e groupe est une communauté unie par un *lien d'attachement mutuel*; quelles que soient les circonstances, il y a identification personnelle des partenaires [...]. Il n'y a pas perte de l'agressivité, mais réorientation rituelle de celle-ci [...]. Non seulement il y a transformation de l'agressivité intra-spécifique en actes d'apaisement et d'amour, mais déviation de l'agressivité vers l'extérieur du groupe (un voisin territorial par exemple). Le lien qui s'établit entre deux individus se renforce progressivement et devient de plus en plus indépendant des conditions de milieu, et en particulier du territoire. *Un groupe se déplace sans perdre sa structure de groupe*". De plus, d'après ces auteurs, la structure de groupe apparaît souvent chez les Vertébrés dits "supérieurs".

de la tâche à imiter, de protocoles expérimentaux ne prenant pas en compte le point de vue particulier à l'espèce ou à l'individu¹⁵⁰ ou encore d'autres facteurs, tels que la période de développement interrogée de manière intentionnelle ou non (Bard et Russell, 1999; Shumaker et Swartz, 2002). Nous partirons donc du principe que ce qui a été observé est *observable sachant* au moins la méthode utilisée et sa mise en œuvre.

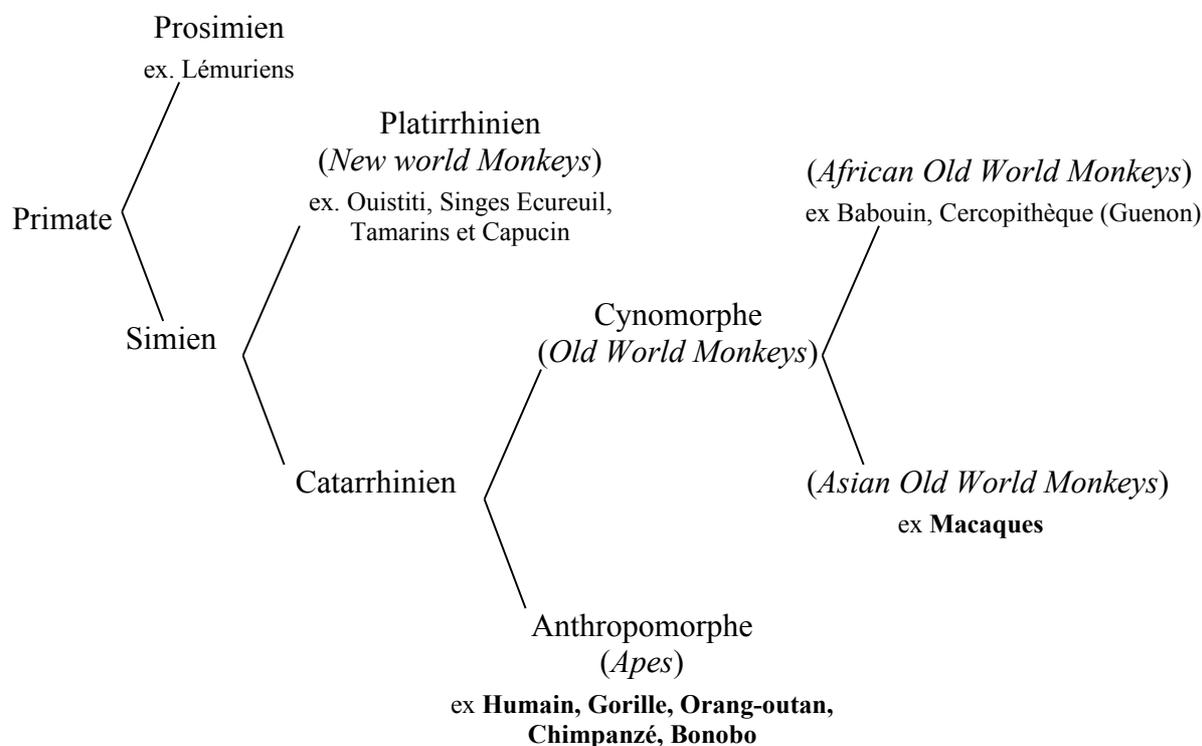


Figure 3-5: Une classification taxonomique de l'embranchement des Primates.

a. Monkeys

Bard et Russell (1999) soulignent le manque d'études en faveur d'une capacité d'imitation chez les monkeys dont les données ne peuvent s'expliquer par des alternatives à l'imitation (pour des exemples d'alternatives, Cf. section I.1.). Toutefois, Whiten et Custance (1996) nous apprennent qu'ils peuvent imiter des *actions sur des objets* rencontrés en liberté, telles que peler des fruits (Butterworth, 1999), bien qu'ils ne semblent pas apprendre par imitation l'utilisation d'outils (Bard et Russell, 1999; Shumaker et Swartz, 2002). De plus, des ouistitis seraient capables d'imiter une action bi-manuelle simple présentée par un congénère (Bugnyar et Huber, 1997, cité par Butterworth, 1999)

Notons que le cerveau des macaques (genre *Macaca*) est le modèle animal standard dans l'étude des fonctions cérébrales humaines, dont celles (re)connue pour faire intervenir un lien perception-production, en neurosciences. Nous verrons ultérieurement que le cortex des macaques montre, au moins, des prédispositions à imiter (voir plus loin, section IV).

¹⁵⁰ Le "point de vue" désigne, notamment, l'incompréhension de la tâche ciblée par l'expérimentateur/trice, le fait que le sujet puisse être enclin à imiter certains individus et pas d'autres etc.

Les Anthropomorphes

Les contraintes génétiques d'*Homo sapiens* sont plus proches de celles des anthropomorphes non-humains que de celles des monkeys (Figure 3-6). La capacité à imiter des anthropomorphes non-humains a majoritairement été testée et/ou étudiée au sein des Homidés où l'on trouve les orangs-outans, les gorilles, les chimpanzés et les bonobos.

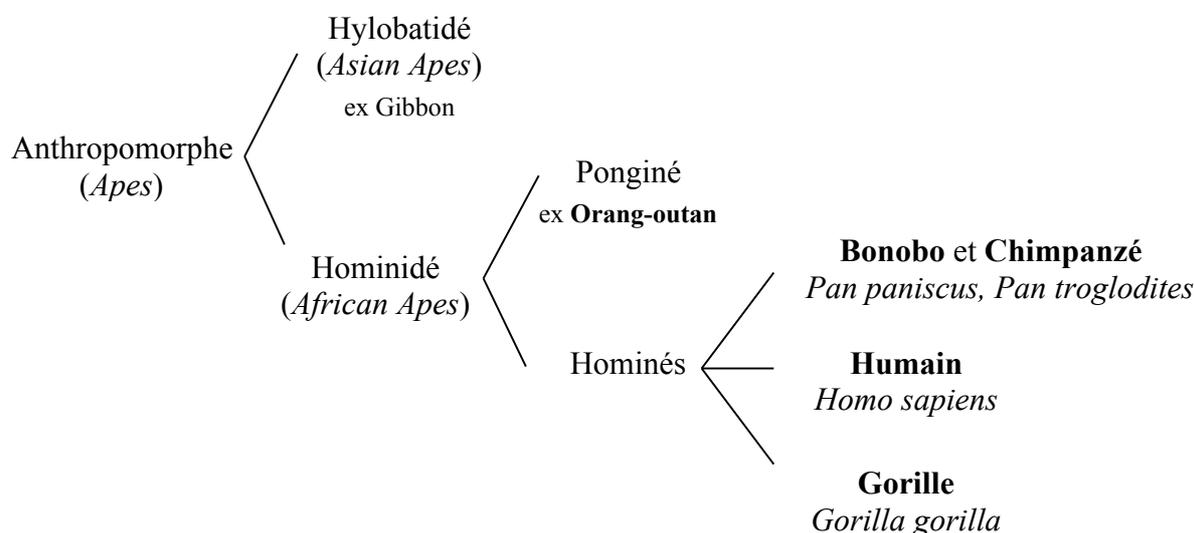


Figure 3-6: Arbre phylogénétique des Hominidés au sein des Anthropomorphes, basé sur une méta-analyse de plusieurs études en phylogénie moléculaire (Graur et Li, 2000; à l'origine le Gibbon ne sert qu'à orienter l'arbre).

b. Orang-outan *Pongo pygmaeus*

Russon et Galdikas (1995) rapportent l'apprentissage d'actions humaines par une orang-outan adulte que l'imitation *différée* est le phénomène le plus apte à expliquer (Bard et Russell, 1999). Ces actions comprennent se brosser les dents avec une brosse à dents et du dentifrice et allumer un camping-gaz, en produisant des détails de l'action des humains de son entourage, tels que cracher entre les barreaux de la véranda, dans le premier cas, et "vider le fuel contenu dans un récipient à l'aide d'une tasse, placer la pointe de la tige d'allumage dans le fuel et l'aplatir grâce à un couvercle en métal en le tenant verticalement et en le bougeant vivement d'avant en arrière"¹⁵¹ (Bard et Russell, 1999, p.105), dans le second cas.

De plus, Miles, Mitchell et Harper (1996) rapportent l'imitation différée d'*actions faciales* et d'*actions sur des objets* par un orang-outan. De plus, les auteurs lui ont appris à 4 ans à imiter quelques actions sous la directive "fais la même chose" et a pu étendre l'incitation à imiter à des actions sur des objets (Bard et Russell, 1999).

D'après Russon et Galdikas, (1995) la capacité d'imitation des orangs-outans a plus de chance de s'exprimer quand les démonstrateurs/trices sont des congénères (Bard et Russell, 1999).

En effet, Shumaker, Beck, Brown et Taub (1998) ont mené une expérience sur 3 paires dominant-subordonné au début du test (Shumaker et Swartz, 2002). L'individu dominant sert de démonstrateur/trice tandis que le/la seconde a l'opportunité d'observer toutes les

¹⁵¹ "[S]cooping fuel from a container with a cup, placing the tip of a smouldering stick in the fuel, and fanning the stick with a metal lid by holding it vertically and waving it briskly back and forth" (Bard et Russell, p.105).

interactions entre l'humain et le/la démonstrateur/trice qui apprend l'association arbitraire d'une nourriture spécifique à un symbole abstrait, au départ d'une étude sur l'acquisition du langage. L'expérimentateur/trice n'a donné aucune instruction directe à l'observateur/trice. Celui/celle-ci peut voir tous les aspects de la tâche, ainsi que les réponses "erronées" autant que les réponses "justes" (qui sont récompensées). A 10 occasions, deux des dominants/antes quittent brièvement l'aire de test: à chaque fois, l'observateur/trice approche immédiatement le dispositif expérimental, effectue la tâche selon les règles qui ne lui ont pas été directement enseignées, avec succès dans 9 cas, en répliquant exactement le *nouveau* comportement vu chez l'autre, pour tendre la main ou les lèvres en direction de l'humain dans l'attente de la récompense. En particulier, la *manière* dont les subordonnés désignent le symbole est le pointage à un doigt, comme le fait le/la dominant/ante. Or, ce comportement n'est pas typique des orangs-outans et il n'a été inculqué qu'au/à la dominant/ante. Il en découle que les subordonnés imitaient vraisemblablement leur dominant/ante¹⁵².

c. Gorille *Gorilla gorilla*

D'après Bard et Russell (1999) peu d'études portent sur les capacités d'imitation des gorilles. Toutefois, l'on rapporte que des gorilles de montagne en liberté adoptent des techniques particulières pour traiter une nourriture à épine. Bien qu'à un niveau global la méthode soit la même, une similarité, vraisemblablement intra-groupe, au niveau des actions indique un apprentissage par imitation (Bard et Russell, 1999).

De plus, des expériences montrent qu'un individu de 4 ans est capable d'imitation différée d'*actions faciales* et d'*actions sur des objets* (Chevalier-Sholnikoff, 1982, citée dans Bard et Russell, 1999).

d. Chimpanzé, *Pan troglodites*, et Bonobo, *Pan paniscus*

L'espèce chimpanzé est probablement la plus étudiée des anthropomorphes et du genre *Pan*. Les bonobos ont de fortes chances de montrer les mêmes habiletés que les chimpanzés qui restent notre source d'information principale sur le genre.

En "colonie", de Waal (1982) rapporte que de jeunes chimpanzés imitent la démarche caractéristique d'un individu (p. 80) et celle de blessés juste après une échauffourée (p. 144). Hayes et Hayes (1952) décrivent un chimpanzé se mettre du rouge à lèvres à l'aide d'un miroir, en incluant des détails vus chez le modèle, tels que sa manière d'étaler le produit.

En conditions contrôlées, des chimpanzés de 4 ans¹⁵³ sont capables d'imiter des actions arbitraires nouvelles après avoir appris à obéir à la commande "Fais ça!" (Custance, Whiten et Bard, 1995)¹⁵⁴. Des individus de 4 à 6 ans ont la faculté d'imiter en différé des actions faciales et des actions sur des objets (Mathieu et Bergeron, 1981; Mignault, 1985; Whiten, Custance, Gomez, Teixidor et Bard, 1996)¹⁵⁵. Comme dans le cas des orangs-outans, l'aptitude à imiter s'exprime mieux quand le modèle est un/une congénère (Bard et Russell, 1999).

Bard, Frigaszy et Visalberghi (1995) étudient l'apprentissage de l'utilisation d'outil par imitation entre un individu qui apprend la tâche et un autre qui observe son apprentissage (Bard et Russell, 1999). Elles analysent l'imitation en termes de transmission d'acquis au cours du développement. Plus précisément, jusqu'à 2 ans l'observation du modèle n'augmente pas les performances propres probablement parce que l'action cible est relativement difficile.

¹⁵² Les auteurs ne précisent pas si le statut social du/de la démonstrateur/trice par rapport à l'observateur/trice influence l'expression de la capacité d'imitation chez les orang-outans.

¹⁵³ Pour information, l'espérance de vie des chimpanzés est de 53 ans, et leur âge de sevrage de 4-5 ans.

¹⁵⁴ Voir Bard et (1999), pour la référence.

¹⁵⁵ Voir Bard et (1999), pour les références.

En revanche, à partir de 3 ans l'observateur/trice trouve la solution plus rapidement que le/la démonstrateur/trice, et à 4 ans, l'effectue dès le premier essai.

Les chimpanzés très jeunes peuvent ouvrir-fermer la bouche (*MO*), tirer la langue (*TP*) et produire des clicks linguaux (*TC*). Bard et Russell (1999) reprennent le paradigme de Meltzoff et Moore (1977) ainsi que celui de Kugiumutzakis (1985, section III.A.3.) pour l'appliquer à des nouveau-nés chimpanzés. Les sujets ont de 1 à 11 jours (N=10). Quatre des cinq individus de 1-3 jour ont imité au moins une des actions TP et/ou MO, tandis que tous les sujets de 7-11 jours ont imité ces deux actes faciaux. Une participante âgée de 7 jours a imité tous les modèles, dont *TC*.

Il en découle que, vraisemblablement, les chimpanzés partagent la même *capacité d'imitation oro-faciale néonatale de modèles visuels* que les nouveau-nés humains, avec une disposition à imiter des modèles audio-visuels. De plus, une exposition antérieure au visage humain n'est pas nécessaire à l'imitation précoce de modèles orofaciaux avec un/une démonstrateur/trice humaine¹⁵⁶.

¹⁵⁶ Notons qu'à notre connaissance jamais sujet humain juvénile n'a été testé sur ses capacités d'imitation avec pour démonstrateur/trice un membre d'une autre espèce. Toutefois, l'histoire des "enfants-loups", Kamala et Amala, (Harris et Butterworth, 2002, pp. 14-15) indique que cela est plus que concevable.

IV. DONNEES DE NEUROSCIENCES SUR DES PRIMATES ADULTES

Bien que les données que nous avons présentées jusque là traitent de l'intérieur de la boîte noire qu'est l'imitation, elles restent issues de sa surface. La présente section tire ses informations de l'intérieur de la boîte crânienne d'adultes macaques et humains, grâce à des techniques qui ne peuvent être appliquées à des juvéniles humains. Ces techniques sont celle de l'**enregistrement de cellules unitaires**¹⁵⁷, pour les macaques, et celles de la neuro-imagerie¹⁵⁸, pour les humains.

D'après Rizzolatti et Luppino (2001), l'imitation peut être sous-tendue par deux mécanismes alternatifs. Dans le premier cas, le sujet reconnaît l'action observée, puis la reproduit: le système moteur n'est impliqué que dans la phase de reproduction¹⁵⁹. Dans le second, les structures motrices responsables de la génération de l'action sont aussi celles responsables de sa reconnaissance, et il y a appariement *direct* entre la perception de l'action et des éléments équivalents dans le répertoire moteur du sujet.

La seconde vision trouve appui dans la découverte de neurones, au niveau du **cortex cérébral**¹⁶⁰ de macaques, dont l'activité est détectée quand le sujet perçoit une action spécifique sur un objet, action exécutée par autrui, mais aussi lorsque le sujet lui-même produit une action identique ou similaire. Ces cellules, qui ont pris le nom de "neurones miroirs", semblent impliquées dans un système de conversion de l'action observée en action exécutable par le sujet. Autrement dit, le premier intérêt de ces cellules nerveuses est qu'elles proposent une implémentation physique d'un lien *non-réflexif* entre perception et production dans l'embranchement des primates, lien vraisemblablement nécessaire à l'expression de comportements imitatifs, et cela à l'extérieur de la moelle épinière et du tronc cérébral¹⁶¹. Toutefois, le deuxième intérêt des neurones miroirs est loin d'être des moindres: ils se trouvent dans des régions cérébrales dont les contraintes génético-environnementales de formation embryonnaire pourraient avoir été "héritées" par l'espèce humaine de l'ancêtre putative qu'elle partage avec les macaques, comme base de l'émergence d'aires corticales impliquées dans la *parole* et l'*imitation* d'actes corporels arbitraires.

Cette section a pour but de décrire les systèmes miroirs de primates. Dans un premier temps, nous verrons celui des macaques à queue de cochon, avec une attention particulière portée aux propriétés des neurones miroirs. Un petit retour sur l'évolution nous permettra de rebondir sur l'architecture cérébrale qui sous-tend potentiellement l'imitation d'actes arbitraires chez des adultes humains.

A. Système d'appariement perception-production d'actions de Macaques adultes

Pour comprendre le système d'appariement perception-production mis en évidence sur des macaques à queue de cochon (*Macaca nemestrina*), nous apprécierons, par une vue d'ensemble, ce qu'est le cortex perceptuo-moteur dans lequel il s'inscrit, ainsi que deux

¹⁵⁷ La technique d'**enregistrement de cellules unitaires** consiste en l'insertion, dans le cerveau, d'une électrode à proximité de la membrane cytoplasmique d'un neurone situé dans la zone à l'étude du système nerveux, pour suivre de manière précise le décours temporel de son activité, ou fréquence à laquelle il *émet* des potentiels d'action (décharges), au cours d'une tâche comportementale dans laquelle on soupçonne lui ou la région dans laquelle il se trouve d'intervenir. C'est donc une *mesure directe de l'activité neuronale*. Son côté invasif la rend difficilement utilisable dans le cas de l'espèce humaine.

¹⁵⁸ Voir Purves *et al.* (1999), pp. 22-23.

¹⁵⁹ L'on retrouve la définition de l'imitation donnée par Meltzoff (2002a) et Grèzes (2000) (voir section I.2.).

¹⁶⁰ Le **cortex cérébral** est la couche superficielle, ou étymologiquement "l'écorce", du cerveau (*adj.* Cortical).

¹⁶¹ Le fait que ce lien se trouve au niveau cortical, et non au niveau de la moelle épinière ou du tronc cérébral, s'accorde avec l'idée selon laquelle les comportements imitatifs ne relèvent pas du domaine des réflexes.

notions par lesquelles seront caractérisés les neurones qui le composent. De là, une description des propriétés fonctionnelles des différents types de cellules miroirs sera donnée, pour en déduire des liens avec l'émergence évolutive de la parole et son développement.

1. Vues d'ensemble

a. Du cortex perceptuo-moteur

Comme d'une mosaïque

Sur la base de données cyto-architectoniques, immunochimiques et fonctionnelles, issues de macaques, Matelli, Luppino et Rizzolatti (1985, 1991) défendent une redéfinition de la parcellarisation des Aires de Brodman (BA) 4 et 6, qui se trouvent dans la partie du cortex frontal dépourvue de cellules granulaires (ou "cortex moteur"), en 7 aires, notées F1 à F7¹⁶² (Luppino et Rizzolatti, 2000). De plus, les auteurs proposent une vision intégrée du *système moteur cortical* des primates, où l'unité de base n'est plus l'aire "motrice isolée" dans le lobe frontal, mais le circuit formé par les connections privilégiées qu'elle partage (i) avec une aire (à dominante) "sensorielle" pariétale postérieure, et l'on parle d'aire motrice *pariéto-dépendante* ou (ii) avec une aire pré-frontale, et l'aire (à dominante) motrice est dite *préfronto-dépendante*¹⁶³ (Rizzolatti, Luppino et Matelli, 1998).

En d'autres termes, une aire frontale pariéto-dépendante reçoit la majorité de ses informations d'entrée, ou afférences, de l'aire pariétale (perceptive) dont elle dépend, et qui lui envoie le principal de ses projections, ou efférences : si les connexions sont réciproques¹⁶⁴ les auteurs parlent de *circuit pariéto-frontal*. Dans ce cas, les deux aires tendent à avoir des neurones caractérisés par des propriétés sensorielles, et parfois à la fois sensorielles et motrices, similaires. Chaque circuit cortical (i) comprend une représentation somatotopique, i.e. relative à des parties du corps, et (ii) est spécialisé dans une *transformation sensori-motrice spécifique* participant à l'élaboration et/ou la mise en œuvre de *l'action appropriée en direction d'une cible définie*¹⁶⁵.

¹⁶² Traditionnellement, il existe 2 types de **cartes** représentant l'organisation du cortex cérébral: (i) les **architectoniques**, telles que celle de Brodman, qui sont essentiellement basées sur la structure cellulaire des tissus cérébraux, et (ii) les **fonctionnelles** qui se basent sur l'implication des différentes zones cérébrales dans des tâches comportementales, en utilisant des méthodes de recueil de données indirectes (eg. effets des lésions, imagerie cérébrale) et directes (eg. enregistrement de cellule unitaire). Dans la nomenclature proposée, F1 est l'aire de Brodman 4 (BA4) plus une partie de BA6 et, au niveau fonctionnel, l'aire motrice primaire (M1). De plus, l'aire BA6 correspond dans sa partie: (i) *mésiale* à F3, ou Aire Motrice Supplémentaire (AMS) en termes fonctionnels, plus F6, ou Aire Prémotrice Supplémentaire (pré-SMA), (ii) *dorsale* à F2 plus F7 dont l'ensemble forme le cortex Prémoteur dorsal (PMd), et (ii) *ventrale* au cortex Prémoteur ventral (PMv) constitué de F4 plus F5 dont deux parties sont distinguées l'une se situant dans le Bord postérieur du sillon Arqué inférieur (F5ab) et l'autre sur la Convexité corticale immédiatement adjacente au sillon arqué (F5c) (voir aussi figure 1, p. 220, dans Luppino et Rizzolatti, 2000, disponible à l'URL:

<http://physiologyonline.physiology.org/cgi/content/full/15/5/219>).

¹⁶³ Les aires motrices *pariéto-dépendantes* sont les aires F1 à F5, tandis que les *préfronto-dépendantes* sont F6 et F7. Aussi, les secondes projettent sur les premières (voir plus bas).

¹⁶⁴ Au niveau cérébral, quand il y a **connexions réciproques** entre 2 aires A et B, cela signifie que des neurones de A projetant sur B, i.e. ayant des axones qui contactent B, transfèrent de l'information de A à B, et que des neurones de B projetant sur A transfèrent de l'information de B à A. Le transfert d'information entre A et B est donc bidirectionnel.

¹⁶⁵ Prenons pour exemple le circuit VIP-F4, où VIP est l'aire Ventrale Intra-Pariétale. VIP semble envoyer à F4 l'information de la localisation de l'objet à prendre en coordonnées égocentrées dans l'espace péripersonnel. De là, F4 transformerait ce donné en des mouvements (potentiels) de bras, cou, visage et/ou bouche appropriés pour contribuer à l'atteinte de cette cible sachant sa localisation par rapport au corps (Luppino et Rizzolatti, 2000; Rizzolatti, Luppino et Matelli, 1998).

Autrement dit, un circuit pariéto-frontal traduit une information perceptive déterminée en des termes moteurs, au moins *potentiels*, *appariés* au perçu, contribuant ainsi à la sélection, la planification et/ou le contrôle du mouvement adéquat en direction d'un but, sachant, par exemple dans le cas d'un objet pour cible, des caractéristiques de forme et de localisation spatiale de cet objet, le contexte et les motivations, notamment physiologiques, du sujet.

De là, le système moteur cortical comprend une mosaïque de circuits perceptuo-moteurs travaillant en parallèle mais aussi de concert, via les connections qui les relient, dans la construction de la réponse motrice volontaire¹⁶⁶. Une pièce de cette mosaïque est le circuit PFG/PF-F5c¹⁶⁷ qui semble implémenter un *système d'appariement perception-production d'actions*. Ce qui suit veut apprécier, en particulier, les caractéristiques de neurones de F5c qui de ces deux aires est la mieux connue dans le cadre de cette problématique.

b. De neurones de l'aire F5

L'aire F5 comprend, notamment, des neurones dont la micro-stimulation induit des mouvements coordonnés d'extrémités du corps, telles que les mains et la bouche¹⁶⁸, et dont l'enregistrement met en évidence une activation lors de la production de ces mêmes mouvements s'ils sont *dirigés vers un but*. Ces neurones s'intéressent donc aux *actes* ou actions¹⁶⁹, et les actions pour lesquelles ils déchargent lors de leur exécution sont qualifiées d'*effectives en production*. Plus précisément, les cellules nerveuses se trouvant dans les zones

¹⁶⁶ L'exécution effective des commandes motrices se fait par la convergence des informations motrices sur les motoneurones qui induisent la contraction des muscles d'intérêt. (Les motoneurones se trouvent, au moins chez l'humain, (i) dans les noyaux moteurs des nerfs crâniens au niveau du tronc cérébral, pour la contraction des muscles orofaciaux, et (ii) dans la moelle épinière, pour stimuler les muscles du reste du corps (Purves *et al.*, 1999, pp.319-325). C'est surtout le second type d'effecteur qui a été étudié chez les macaques). Toutes les aires "motrices" pariéto-dépendantes projettent sur des motoneurones. D'après les couches de la moelle épinière où elles aboutissent (Porter et Lomon, 1993, cité par Luppino et Rizzolatti, 2000), les projections de F1, F2, F3, F5ab et de parties de F4 semblent contacter (i) les motoneurones (couche VIII) des muscles proches de et sur l'axe de symétrie du corps (muscles proximaux et axiaux, resp.), (ii) les (inter)neurones (couche VII) dont la fonction est de faire l'association basique entre sensitif et moteur comme dans le cas des réflexes posturaux, et (iii) des neurones proprioceptifs (couche VI). En plus de cela, F1, qui contient une représentation complète du corps et qui décharge lors de mouvements passifs (information proprioceptive), est connectée aux motoneurones (couche IX) des muscles distaux qui induisent le mouvement des extrémités des membres. Aussi, l'on apprend que les aires F2 à F5 projettent sur F1. Pour ce qui est des aires "motrices" préfronto-dépendantes, F6 et F7, elles ne sont connectées ni à la moelle épinière, ni à F1, mais contactent de manière diffuse les aires F2 à F5. De là, Rizzolatti et collègues proposent que l'activation de F1 déterminerait la morphologie fine du mouvement, et les aires F2 à F5 seraient impliquées dans sa morphologie globale. De plus, F7 jouerait dans la sélection de mouvements conditionnés qui se trouvent dans la mémoire des actions passées, alors que F6 aurait pour rôle de faciliter/inhiber les actions potentielles issues des circuits pariéto-frontaux, en fonction de la situation, des motivations et plan à long terme du sujet, ayant ainsi une importance dans l'organisation séquentielle de la réponse motrice (Rizzolatti, Fogassi et Gallese, 2000; Luppino et Rizzolatti, 2000; Rizzolatti, Luppino et Matelli, 1998; voir aussi le cours d'anatomie en médecine de Gérard Outrequin pour la structure laminaire de la moelle épinière à l'URL: <http://www.anatomie-humaine.com/neuroa/moelle.html>).

¹⁶⁷ Rappelons tout d'abord que F5c est la partie de F5 se trouvant sur la Convexité corticale immédiatement adjacente au sillon arqué. Aussi, les initié-e-s sont ici informé-e-s que le circuit PF-F5c va prochainement céder la place au circuit PFG-F5c, suite à la dissection du lobule pariétal inférieur obtenue par le groupe de Parme dont les résultats qui ne sont pas encore publiés dans le détail défendent une nouvelle parcellarisation du cortex pariétal (Fogassi, com. pers., 21/01/2005; un avant goût se trouve dans Luppino, Belmalih, Calzavara, Matelli et Rozzi, 2004, disponible sur le WEB)

¹⁶⁸ Notons que F5 ne semble pas concerner les mouvements oculaires.

¹⁶⁹ Ici, le terme d'"action" sera, en fonction du contexte, soit synonyme d'actes, soit un mouvement composé de plusieurs actes ou actions plus simples, comme dans le cas d'une séquence motrice telle que prendre-déplacer-manger.

dorsale et ventrale de F5 tendent à être spécifiques d'actes, respectivement, manuels et orofaciaux¹⁷⁰ (Gentillucci *et al.*, 1988, citées par Ferrari, Gallese, Rizzolatti et Fogassi, 2003).

Des neurones perceptuo-moteurs

Un sous-ensemble des neurones spécifiques d'une ou plusieurs actions en production ont, en plus de cela, des propriétés perceptives: ils déchargent quand le sujet perçoit des stimuli en lien avec leur-s action-s effective-s en production. A ce jour, les modalités perceptives testées de manière systématique sont la vision et/ou l'audition¹⁷¹, bien que la première soit la mieux connue. De la même manière, la main est l'effecteur le plus étudié dans les recherches relatives à l'aire F5 des macaques.

Plus précisément, parmi les neurones perceptuo-moteurs de cette aire, il y a (i) ceux qui déchargent lors de la *vision d'un objet* auquel pourraient s'appliquer les actions effectives en production, et (ii) ceux qui déchargent lors de la *perception de l'exécution par un autre individu* des actions effectives en production ou d'actions similaires ou encore d'actions qui lui sont liées au niveau fonctionnel. Les premiers sont les neurones visuo-moteurs manuels *canoniques* et se situent surtout dans F5ab, les seconds sont les neurones audio- et/ou visuo-moteurs manuels et/ou orofaciaux *miroirs*¹⁷² et se trouvent en particulier dans F5c (Rizzolatti, Luppino et Matelli, 1998; Rizzolatti et Luppino, 2001). Cette section est dédiée à la description du second type de neurones perceptuo-moteurs, que deux notions servent à décrire: la sélectivité et la congruence.

Notions de sélectivité et de congruence

En fonction du nombre d'actes effectifs en production (resp. perception), l'on distinguera les miroirs *strictement* et *globalement sélectifs* en production (resp. perception): les premiers sont spécifiques de (ou actifs pour) un seul acte, les seconds de plus d'un. Ce degré de sélectivité pourrait non seulement simplifier l'exécution et le contrôle de l'action¹⁷³, mais aussi implémenter la représentation *des buts et des moyens* des actes de soi et d'autrui qu'ils soient donc produits ou perçus (Rizzolatti et Luppino, 2001).

La notion de *congruence stricte vs. globale* décrit le degré avec lequel la spécificité de la réponse en perception est congruente avec celle de la réponse en production. Plus précisément, un neurone miroir qui décharge pour l'exécution d'un seul acte et qui ne décharge en perception que pour l'audition et/ou la vision de ce même acte (identique au premier en termes de but et de moyens) sera dit "*strictement congruent*", tandis que s'il y a un lien entre la/les action-s effective-s en production et celle-s en perception, il sera qualifié de *globalement congruent*, sinon, on le dira non-congruent. La congruence perceptuo-motrice est à la base de l'implication potentielle des neurones miroirs dans un *système d'appariement* entre la perception et la production des actions qui intéresse l'imitation. La sélectivité et la congruence seront illustrées et précisées plus bas.

¹⁷⁰ La partie ventrale de F5 contient aussi une représentation du larynx (Gentillucci *et al.*, 1988, cité par Rizzolatti et Arbib, 1998).

¹⁷¹ Dans ce contexte, il est possible que F5 concerne aussi d'autres modalités perceptives en lien avec des actions. En particulier, la proprioception n'a pas été étudiée de manière systématique (Fogassi com. pers., 21/01/2005).

¹⁷² L'expression de "neurones miroirs" au sens strict s'applique aux neurones actifs pour la production et la perception *visuelle* d'une ou plusieurs actions spécifiques.

¹⁷³ En effet, une représentation motrice en termes d'actes et/ou d'actions potentiel-le-s peut être vue comme un stock structuré de "paquets de commandes prêts à l'emploi".

2. Neurones perceptuo-moteur miroirs

Sachant que les neurones miroirs concernent la vision et/ou l'audition, ainsi que la main et/ou la face, nous nous permettrons de préciser dans ce qui suit s'ils sont visuo-moteurs (*VM*), audio-moteurs (*AM*) ou audio-visuo-moteurs (*AVM*), et manuels et/ou oro-faciaux, bien que seuls ceux ayant des propriétés *visuelles* soient dits "neurones miroirs" dans la terminologie du groupe de recherche qui les a mis en évidence¹⁷⁴. La description de ces cellules nerveuses se décline de la manière suivante. Dans un premier temps, les *miroirs VM d'action sur des objets*, i.e. les plus connus, seront l'occasion de présenter le paradigme expérimental de base pour la mise en évidence et l'étude des neurones de type miroir, d'illustrer par des exemples les caractéristiques de ces cellules et de mieux comprendre leurs fonctions potentielles en termes de représentations cognitives des actions exécutables. De là, nous donnerons des informations sur les *miroirs VM d'actions (corporelles) orofaciales* et sur les *miroirs AM et AVM d'actions sur des objets*.

a. Miroirs visuo-moteurs d'actions sur des objets

➤ Etude de mise en évidence: équivalence perception-production

En 1996, le groupe de Parme rapporte l'existence de neurones de F5 ayant des propriétés visuelles qui *reflètent* leurs propriétés motrices, d'où leur nom de "neurones miroirs" (Gallese, Fadiga, Fogassi et Rizzolatti 1996; Rizzolatti, Fadiga, Gallese, Fogassi, 1996).

• Paradigme expérimental de base

Dans cette étude, deux sujets macaques adultes vivants en laboratoire ont pour tâche (i) de fixer du regard des objets de tailles et de formes diverses qui leur sont présentés un à un à des distances variables, et (ii) d'atteindre et prendre l'objet quand son accès leur est permis à la lumière et dans l'obscurité. En parallèle, l'activité de cellules nerveuses de leur aire F5 est suivie pour trouver celles qui s'activent à l'exécution et/ou l'observation d'actions¹⁷⁵. Dans un second temps, les propriétés visuelles des neurones sont étudiées plus en profondeur pendant que le sujet observe un autre individu, congénère ou humain, effectuer un panel d'actions.

Une variété d'actions sur des objets...

La définition de l'acte le plus effectif en production pour chaque neurone se fait par l'analyse fine du décours temporel du taux de décharge de la cellule pendant l'exécution (par le sujet) d'une variété de combinaisons de mouvements proximaux et distaux. Dans le cas de la *préhension manuelle*, la taille et la forme de l'objet permettent d'évoquer des actions de même but (prendre) mais dont les moyens varient, avec en particulier: (i) la **pince de précision**, où il y a opposition entre le pouce et l'index, pour les objets de petite taille, (ii) la **préhension digitale**, où l'opposition se fait entre le pouce et les autres doigts, pour récupérer des objets de taille moyenne d'un récipient profond et étroit, et (iii) la **prise de pleine main**, dans laquelle tous les doigts sont en flexion autour de l'objet qui est de taille importante.

¹⁷⁴ Ce groupe de recherche prendra parfois le pseudonyme de "groupe de Parme".

¹⁷⁵ L'on peut ajouter que ces comportements ont fait l'objet d'un apprentissage et que les deux hémisphères cérébraux ont été explorés. Plus précisément, pendant la période de fixation, l'enregistrement de cellule unitaire permet de différencier les neurones canoniques, qui déchargent à la simple vue de l'objet qui les intéresse, des autres neurones, dont les miroirs, qui ne déchargent pas dans ce cas, tandis que les propriétés motrices, communes aux neurones moteurs de F5 spécifiques d'actions, sont estimées dans la phase d'atteinte et de préhension par le participant. Notamment, s'il y a décharge lors de la préhension dans l'obscurité, la réponse est considérée comme ne pouvant être que motrice.

Quand l'objet est une nourriture, il est possible (i) de distinguer les moyens de la préhension en fonction de l'effecteur en jeu, i.e. s'il s'agit de la main (préhension manuelle) ou de la bouche (**préhension buccale**), mais aussi (ii) d'induire notamment l'**interaction entre les mains**, où, en particulier, la main porteuse de nourriture se rapproche de l'autre, et les actions de *porter à la bouche*, de tenir avec la main (*port manuel*) et de **manipuler**, qui comprend celles de *casser*, *déchirer* et *plier*¹⁷⁶.

Pour étudier dans le détail leur réponse visuelle, le taux de décharge des neurones est suivi tandis que le sujet observe l'exécution par autrui¹⁷⁷ d'*actions sur des objets*: à celles ci-dessus s'ajoutent *mettre* de la nourriture sur un support, la *donner* au ou l'*éloigner* du sujet ou un autre individu (macaque ou humain). De plus, des *actions corporelles manuelles* sont présentées: lever les bras, remuer les mains, menacer¹⁷⁸, et les mouvements correspondant à des actions sur des objets en l'absence d'interaction entre l'effecteur et l'objet qui peut être absent (**pantomimes**). Des *actions manuelles sur des objets pour agir sur d'autres objets* ont aussi été testées, notamment la préhension via des pinces¹⁷⁹.

- **Propriétés caractéristiques**

Sur 532 neurones suivis, 92 (17%) sont miroirs VM d'*actions sur des objets*, i.e. actifs lors de la production, par soi, et de la vision, chez autrui (humain ou macaque), d'actions manuelles et/ou buccales spécifiques¹⁸⁰.

...pour une variété de miroirs VM d'*actions sur des objets*...

Sélectivités motrice et perceptive. En production¹⁸¹, 60 des miroirs VM déchargent spécifiquement lors de la préhension manuelle, 9 pour la préhension buccale, 11 pour les préhensions manuelle et buccale¹⁸². Plus précisément, les 60 cellules spécifiques de l'exécution d'une préhension manuelle en comprennent 38 non-sélectives sur le type de pince, tandis que 11 sont sélectives de la pince de précision, 7 de la préhension digitale, 2 de la prise de pleine main et 2 à la fois de la pince de précision et de la préhension digitale.

¹⁷⁶ De même, la manière dont l'objet est présenté, eg. sur un support *versus* tenu par les doigts de l'expérimentateur/trice, permet de contrôler le sens de rotation ou de déplacement du poignet lors de la préhension (manuelle) et s'il existe une préférence pour la main gauche ou la main droite en perception et en production.

¹⁷⁷ L'agent des actions est humain et parfois macaque.

¹⁷⁸ Nous n'avons pas de précision sur ce à quoi correspond l'action de menacer.

¹⁷⁹ Pour passer sur la polémique soulevée par la mise en évidence de propriétés perceptives spécifiques d'actions dans des neurones d'aires connues pour être (pré)motrices, précisons que les contrôles de l'activité des miroirs en perception comprennent les vérifications suivantes: s'il y a décharge en perception, (i) elle ne relève pas d'une préparation motrice puisque le sujet n'a pas à agir suite à la perception des stimuli, (ii) elle a lieu alors que les muscles recrutés dans les actions perçues sont inactifs d'après leur électromyogramme (EMG). De plus, (iii) l'attention, estimée par la direction du regard des sujets, est la même quelle que soit l'action présentée (ce contrôle ne semble pas figurer dans la présente étude) et (iv) à chaque étude de mise en évidence, la localisation des neurones enregistrés est vérifiée, suite à l'euthanasie d'un des sujets, par la dissection de son cerveau.

¹⁸⁰ 25 neurones ont les mêmes propriétés visuelles que les miroirs VM mais ne semblent pas avoir de propriété motrice. Parmi eux (i) 17 sont spécifiques d'une action, dont 7 pour la préhension, 6 pour l'interaction manuelle, 3 pour le placement, 1 pour le port manuel, et (ii) 8 sont spécifiques de 2 actions, dont 4 pour prendre et placer, 2 pour la préhension et la manipulation manuelles, 1 pour prendre avec la main et la bouche, et 1 pour la manipulation et l'interaction manuelles.

¹⁸¹ Ici, les auteurs reprennent la classification proposée par Rizzolatti, Camarda, Fogassi, Gentilucci, Luppino et Matelli (1988) pour les neurones de F5 actifs lors de l'exécution d'actions manuelles sur des objets (Gallese *et al.*, 1996). Dans cette catégorisation, l'action de déchirer n'est pas comprise dans celle de manipuler.

¹⁸² La spécificité motrice des miroirs VM restants se décline en 8 cellules pour la manipulation, 2 pour déchirer et 2 pour porter à la bouche.

Au niveau de leurs propriétés visuelles, 51 neurones miroirs VM sont *strictement sélectifs*, avec dans les 30 cas sensibles à la préhension manuelle, 18 cellules spécifiques d'un seul type de pince¹⁸³. De plus, parmi les 41 miroirs *globalement sélectifs* en perception, l'on en trouve 36 actifs lors de l'observation de 2 actions dont 20 pour la préhension manuelle et le placement, et 3 pour la préhension que l'effecteur soit la main ou la bouche¹⁸⁴. La précision perceptive des miroirs VM peut aller plus loin: des cellules montrent une préférence quant au sens spatial de l'action observée, voire, vis-à-vis de la main gauche vs. droite utilisée par le/la démonstrateur/trice¹⁸⁵. De plus, tous les neurones miroirs semblent décharger selon un patron temporel précis par rapport au déroulement de leur-s action-s effective-s en perception¹⁸⁶.

La vision des pantomimes de la ou des action-s effective-s en perception ne suffit pas à stimuler les neurones miroirs VM d'actions sur des objets. Il en est de même pour les actions manuelles sur des objets pour agir sur d'autres objets qui ont induit des taux de décharge faibles à nuls.

Congruence perceptuo-motrice. Sur l'ensemble des 92 miroirs VM, 29 sont strictement congruents: il ne déchargent que pour un seul et même acte, en termes de but et de moyens, que cela soit en perception ou en production. Par exemple, un neurone sélectif en production pour l'extraction de nourriture avec l'index, ne s'activera en perception que pour la même action, et sera insensible, notamment, à une extension similaire de l'index figurant dans une pince de précision. Ce degré de congruence peut aller jusqu'à s'exprimer dans des neurones sélectifs sur le sens, trigonométrique vs. horaire, de l'acte de rotation du poignet lors d'une pince de précision.

Toutefois, la majorité (56) des miroirs VM sont globalement congruents. Trois catégories sont distinguées, par les auteurs, en fonction de leur degré de sélectivité en production. Les premiers (7) sont actifs pour l'exécution d'un acte précis en termes de but et de moyen (eg. la pince de précision), mais sont visuellement sensibles à plusieurs types d'actions utilisant le même effecteur et ayant un même but que l'acte effectif en production (ici, les préhensions manuelles). Les seconds (46) sont majoritaires et moins sélectifs en production que les précédents, en déchargeant lors de l'exécution d'actions ayant un même but (eg. préhension manuelle), et sont globalement sélectifs en perception mais uniquement sensibles à des actions manuelles (eg. prendre et manipuler). Les 3 derniers neurones miroirs globalement

¹⁸³ Nous n'avons pas plus de précision quant à la distribution des miroirs strictement sélectifs de pinces. Par contre, les 21 miroirs strictement sélectifs en perception restant se répartissent comme suit: 7 pour le placement, 7 pour la manipulation, 5 pour l'interaction et 2 pour le port manuel.

¹⁸⁴ Les 18 miroirs globalement sélectifs en perception restant en comprennent 3 pour la préhension et la manipulation, 3 pour la préhension et l'interaction manuelle, 5 pour la préhension et le port manuel, 1 pour le placement et le port, 1 pour le port et l'interaction manuelle, tandis que 5 déchargent lors de l'observation de 3 actions manuelles: la préhension, le placement et le port (4) ou la manipulation (1).

¹⁸⁵ En effet, sur 47 cellules testées sur le sens de l'action, 30 s'avèrent plus actives quand l'action va de gauche à droite ou de droite à gauche que l'inverse, et la plupart si l'action s'achève du même côté que l'hémisphère où elles se trouvent, ie. du côté gauche pour l'hémisphère gauche. Dans le même ordre d'idée, les auteurs ont pu mettre en évidence que, parmi 32 neurones testés, 12 montrent une préférence manuelle: 9 (resp. 3) cellules présentent une décharge plus marquée quand l'action est effectuée par la main du même côté que (resp. du côté opposé à) l'hémisphère où elle se trouve, ie. la main gauche (resp. droite) quand il s'agit de hémisphère gauche. Globalement, 7 (resp. 5) neurones préfèrent la main gauche (resp. droite).

¹⁸⁶ Par exemple, parmi des miroirs sensibles à la préhension manuelle qui commencent à décharger pendant la mise en forme (ou *shaping*) de la main, certains cessent de décharger quand la main a attrapé l'objet tandis que d'autres continuent de décharger un certain temps après la fin de l'action. Dans une autre étude, la même catégorie de miroirs comprend des cellules qui s'activent soit en début soit en fin d'action (Umiltà, Kohler, Gallese, Fogassi, Fadiga, Keyssers et Rizzolatti, 2001). De plus, les miroirs VM globalement sélectifs en perception qui répondent aux actions de prendre et tenir, ou prendre et manipuler commencent à décharger à la fin de la phase de transport de la main (Umiltà *et al.*, 2001).

congruents sont sélectifs pour la production de préhensions manuelles, quelle qu'en soit le type, et répondent indifféremment à la perception des préhensions manuelle et buccale¹⁸⁷.

...et une architecture de buts et de moyens

En perception, l'activation concomitante de plusieurs neurones miroirs (et perceptifs) spécifiques d'actions indique la possibilité de représenter, dans F5, le but d'actes d'autrui par recouplement des actions effectives associées à ces cellules.

De plus, que cela soit en perception ou en production, le niveau de sélectivité des cellules miroirs semble refléter une architecture de l'action en termes de buts et de moyens selon plusieurs dimensions¹⁸⁸. Dans le cas d'une sélectivité stricte, les neurones miroirs sont susceptibles de représenter les moyens précis d'actes précis, eg. rotation du poignet de sens défini dans une pince de précision.

Pour ce qui est de la sélectivité globale, le lien entre les actions effectives d'une même cellule réside dans le but commun de ces actions. Ce but pourrait se décliner au moins selon 2 dimensions croisées, ou *niveaux de généralisation du but*, dont l'une est plutôt "spatiale" dans la relation sujet-objet, et l'autre plutôt "temporelle", dans une séquence d'actions. Du point de vue spatial, l'on trouvera des neurones associées à des actions utilisant (i) soit le même effecteur (eg. la main) pour une même action (eg. la préhension manuelle) qui s'exprime sous différentes formes ou moyens (eg. la pince de précision et la prise de pleine main), (ii) soit des effecteurs différents, ici la main vs. la bouche, pour une même action, eg. *prendre* avec la main ou avec la bouche. Ces neurones pourraient avoir associé différents moyens menant à l'atteinte d'un même but. Au niveau temporel de la généralisation du but, des cellules miroirs paraissent lier différentes actions pouvant figurer dans une même séquence comportementale et faisant intervenir soit un seul effecteur, eg. *prendre puis manipuler*, soit deux effecteurs, eg. *prise manuelle puis buccale*.

Les neurones miroirs s'activent quand l'action perçue est en lien avec leur-s action-s effective-s en production. Ce phénomène suggère que la perception d'actions d'autrui évoque, dans F5c, des représentations motrices qui mettent l'observateur/trice dans une "configuration cognitive" proche de celle qui précède l'exécution d'actes identiques, similaires ou liés à ceux observés. La congruence perceptuo-motrice peut donc être vue comme le reflet d'un "rappel d'actions" ou un "appel à agir" en donnant une *équivalence entre perçu et exécutable*. De ce point de vue, une congruence stricte pourrait convertir les moyens précis observés en commandes motrices potentielles, tandis qu'une congruence globale donnerait différents moyens exécutables pour atteindre le même but que celui perçu. Pour les auteurs, le mécanisme d'appariement (miroir) perception-production joue un rôle dans la reconnaissance (et la compréhension) des actions d'autrui, et serait implémenté par le circuit formé de F5c et PFG où des neurones miroirs ont aussi été trouvés (Fogassi, Ferrari, Gesierich, Rozzi, Chersi et Rizzolatti, 2005). Pour nous, ce mécanisme d'appariement sensori-moteur *pourrait participer à l'imitation des actions d'autrui en donnant des équivalents moteurs des actions perçues*¹⁸⁹.

¹⁸⁷ Pour les neurones miroirs non-congruents (7), aucune relation fonctionnelle ou conceptuelle n'a été trouvée entre leurs actions effectives en production et celles en perception. Les auteurs précisent qu'ils n'en sont pas moins miroirs, sachant qu'ils ne sont pas canoniques et déchargent pour la production et la perception d'actions.

¹⁸⁸ La dimension temporelle au niveau d'une même action révélée par la stabilité du décours temporel de la décharge de chaque neurone miroir ne sera pas abordée plus en détail bien qu'elle soit vraisemblablement essentielle à l'organisation de l'action.

¹⁸⁹ D'après le groupe de Parme, le fait que les macaques n'imitaient pas, pourrait s'expliquer par une incapacité à utiliser les commandes "exécutables" générées par la perception des actions d'autrui (eg. Rizzolatti et Luppino, 2001). Toutefois la définition de l'imitation à laquelle ils se réfèrent est la version stricte de certains éthologues pour qui l'utilisation d'outil est le modèle le plus fréquent dans leur expérience sur la capacité d'imitation des

Selon nous, le fait que les neurones miroirs VM d'actions manuelles sur des objets soient insensibles à la perception d'actions manuelles corporelles (les pantomimes de leur-s action-s effective-s incluse-s) ainsi qu'à celle d'actions s'appliquant à des objets pour agir sur d'autres objets indique que ces cellules s'activent pour des éléments figurant dans le *répertoire comportemental* des sujets. D'après Rizzolatti et Luppino (2001), les neurones miroirs *se développeraient* par un apprentissage individuel de l'association entre l'exécution de l'action et la perception de ses conséquences, sans donner plus de précision sur la manière dont les conséquences perceptives expérimentées par soi pourraient être reconnues sur autrui¹⁹⁰.

Les pantomimes d'actions sur des objets suggèrent que la perception de l'atteinte du but de leur-s action-s effective-s serait essentielle à l'activation des neurones miroirs. Ce point a été approfondi dans l'étude suivante.

➤ **Neurones miroirs VM d'actions manuelles sur des objets et but inféré**

En 2001, le même groupe de recherche réplique la découverte des neurones miroirs VM de F5c et montre qu'une partie de ceux spécifiques d'actions manuelles sur des objets s'activent pour la perception de leur-s action-s effective-s même si l'atteinte du but n'est qu'inférée par le sujet (Umiltà, Kohler, Gallese, Fogassi, Fadiga, Keyssers et Rizzolatti, 2001).

Deux macaques adultes vivants en laboratoire ont pour tâche d'exécuter et d'observer des actions manuelles sur des objets selon le même protocole que Gallese *et al.* (1996). Des neurones miroirs visuellement spécifiques des actions de prendre, tenir, placer, manipuler et approcher sont isolés dans l'aire F5c¹⁹¹. De là, l'activité de ces cellules est suivie pendant que le sujet observe la production, par un/une expérimentateur/trice, de leur-s action-s effective-s en perception dans 2 conditions: vision complète *vs.* partielle de l'action et de sa pantomime. En condition de vision complète, le sujet voit de bout en bout l'exécution de l'action (sur un objet) ou de sa pantomime (sans objet). En condition de vision partielle, un écran opaque est glissé à mi-chemin du déroulement du geste pour en cacher la seconde partie, après que le sujet a vu la présence ou l'absence d'un objet-cible. L'information manquante, au niveau d'une analyse purement visuelle, est donc l'atteinte du but de l'action¹⁹² ou de la fin de la pantomime.

Il en résulte que sur 220 cellules de F5c, 103 sont des neurones miroirs VM d'actions manuelles sur des objets: 37 sont testés sur leur activité dans les conditions de vision complète et partielle. En perception, 12 sont strictement sélectifs, 25 le sont globalement¹⁹³. Comme tous les neurones miroirs d'actions sur des objets, ceux-ci ne s'activent pas lorsque le sujet perçoit la pantomime de leur-s action-s effective-s, que cela soit en condition de vision complète ou partielle. Par contre, la moitié de ces neurones déchargent lors de la perception

primates nn-humains (voir Rizzolatti, Fogassi et Gallese, 2001). Or, nous avons vu que les monkeys sont capables d'imiter des actions manuelles simples sur des objets (Section III.B.2.a.), mais pas d'actions sur des objets pour agir sur d'autres objets.

¹⁹⁰ L'on retrouve, ici, la problématique de la manière dont se fait le lien entre ce que le sujet perçoit lors de l'exécution de ses propres actes et ce qu'il/elle perçoit lors de l'exécution de ces mêmes actes chez autrui. La psychologie du développement propose globalement l'intervention de la proprioception dans la "table de conversion" (Sections I.2. et II.2.).

¹⁹¹ Seul l'hémisphère droit a été exploré, sans que la raison de ce choix ne soit précisée par les auteurs.

¹⁹² Cela est valable pour les actions qui s'achèvent par un contact main-objet. Les auteurs ne précisent pas le protocole utilisé dans les autres cas (eg. placer).

¹⁹³ Les miroirs strictement sélectifs se répartissent comme suit: 7 pour la préhension manuelle, 3 pour tenir et 2 pour placer. Les miroirs globalement sélectifs en comprennent 19 pour prendre et tenir, 2 pour prendre et manipuler, et 4 pour approcher et manipuler.

de leur-s action-s spécifique en vision partielle¹⁹⁴, pourvu que le sujet sache que l'objet-cible se trouve derrière le cache. En d'autres termes ces cellules miroirs d'actions sur des objets s'activent quand l'atteinte du but de leur-s action-s effective-s perçue-s est inférée par le sujet mais ne répondent pas à la pantomime partiellement perçue de cette/ces même-s action-s bien que l'analyse visuelle du mouvement visible soit identique qu'un objet soit présent ou non en condition cachée¹⁹⁵.

Ainsi des neurones miroirs sont-ils capables de représenter des actions dont le but est inféré par le sujet d'après la mémoire de l'objet-cible. Bien que les auteurs proposent que ce sont ces neurones miroirs eux-mêmes qui infèrent le but de l'action, nous proposons que les éléments de leur étude n'écartent pas la possibilité que l'information du but puisse être issue d'autres neurones/aires et transmise à ceux-ci. Toutefois, cela n'enlève rien au fait que *l'absence d'atteinte du but n'empêche pas des neurones miroirs de donner une équivalence motrice potentielle des actions perçues sur autrui*, pourvu que l'intention de l'acteur/trice soit explicite pour l'observateur/trice.

b. Miroirs visuo-moteurs d'actions corporelles orofaciales

Si les actions manuelles des macaques s'appliquent uniquement à des objets, il n'en est pas de même pour leurs actions orofaciales. En effet, dans ce cas, l'on peut distinguer les *actions ingestives*, qui ont des objets pour cible, des *actions communicatives* dont le but n'est pas une interaction effecteur-objet. Récemment, le groupe de Parme a cherché et trouvé des neurones miroirs spécifiques d'actions orofaciales communicatives et ingestives (Ferrari *et al.*, 2003).

• Architecture expérimentale

Deux macaques adultes *ayant une vie sociale* au sein d'une "colonie" strasbourgeoise ont pour tâche d'exécuter et d'observer des actions buccales et manuelles variées. Des actions orofaciales liées à la fonction alimentaire sont testées en perception et en production¹⁹⁶. Elles comprennent: la préhension dentale, la préhension labiale (*LP-ing*), casser avec les dents, aspirer le liquide contenu dans une seringue (*téter*), mâcher, la protrusion-rétraction de la langue (*TP-ing*), tenir avec les dents et retirer d'un support. Des actions orofaciales faisant partie du *répertoire communicatif des sujets* sont testées en perception uniquement, sauf cas exceptionnels. L'on y trouve les gestes suivants: le **lip-smacking**¹⁹⁷ (*LS*), le **pucker**¹⁹⁸ (*LP-*

¹⁹⁴ Il sont au nombre de 19: 6 sont strictement sélectifs (3 pour prendre, 1 pour tenir et 2 pour placer) et 13 sont globalement sélectifs (8 pour prendre et tenir, 1 pour prendre et manipuler, 4 pour approcher et manipuler).

¹⁹⁵ Cette expérience pourrait rappeler aux adeptes de la perception de parole celles de dévoilement progressif (ou *gating*) où les sujets ont pour tâche l'identification de stimuli de parole amputés d'un extrait plus ou moins long en fin de leur réalisation. Nous nous interrogeons donc sur la possibilité que des résultats de même ordre qu'ici, en condition de vision partielle, puissent être obtenus en substituant au cache une interruption des gestes d'action et de pantomime à l'endroit du début de la césure dans la présente étude.

¹⁹⁶ Ici, toutes les actions présentées ont été exécutées par des humains. Les auteurs nous apprennent que l'utilisation de vidéo 2D et 3D de congénères exécutant des actions manuelles ne permettent pas de mettre en évidence les propriétés perceptives des neurones de F5 qui restent insensibles à l'artifice visuel.

¹⁹⁷ Le **lip-smacking** (ou bisou) est utilisée, par les macaques, au moins en situation d'apaisement. Ce geste a au moins pour composante une cyclicité mandibulaire (MacNeilage, 1998).

¹⁹⁸ Le **pucker** (ou LEN pour Lips-Ears-Neck) est une expression faciale, unique aux macaques à queue de cochon, où les lèvres sont protruses, tandis que le cou est étendu et que les oreilles sont en arrière. Cette expression a lieu dans plusieurs contextes sociaux. Elle est donnée par des adultes (i) aux juvéniles retrouvés et est renvoyée par ceux/celles-ci à ceux/celles-là, (ii) à d'autres adultes avant leur monte, ou avant d'autres interactions sociales amicales, telles que le toilettage, ou encore (iii) aux humains. L'on considère que sa fonction est de réduire la distance entre les "inter-puckerer" (elle fait l'objet de "turn-takings") et qu'elle joue un rôle important dans la coordination des activités sociales.

com, Figure 3-7), le **teeth-chatter**¹⁹⁹ (TC), la **bouche ouverte**²⁰⁰ (MO-com) et la **protrusion de la langue** (TP-com).

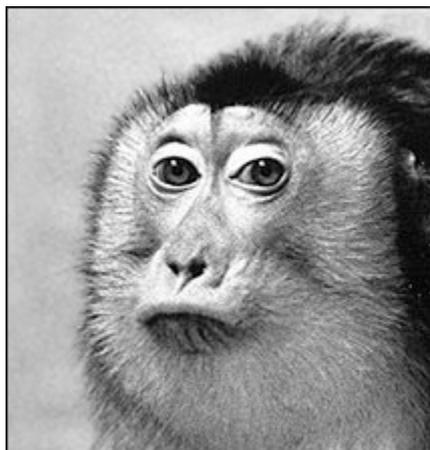


Figure 3-7: Un pucker produit par une macaque à queue de cochon.

- **Propriétés caractéristiques**

Sur 485 neurones enregistrés dans F5, 130 sont des cellules miroirs VM d'actions orofaciales ou à la fois orofaciales et manuelles²⁰¹: 79 des neurones miroirs buccaux sont étudiés plus en détail. Parmi ceux-ci, 67 sont visuellement spécifiques d'actions sur des objets (85%) et 12 le sont d'actions corporelles. Les premiers, nommés *miroirs buccaux ingestifs* par les auteurs, en comprennent 41 strictement sélectifs en perception pour la préhension dentale (19), casser avec les dents (8), téter (6), et 27 sensibles à plusieurs actions où la préhension dentale figure de manière presque systématique²⁰². Les seconds, ou *miroirs buccaux communicatifs*, sont en majorité (10/12) strictement sélectifs en perception pour LS (2), LP-com (3), TP-com (4) et TCh (1); les deux restants s'activent autant pour la vision de LP-com que de TP-com.

¹⁹⁹ **Teeth-chatter** (ou claquement des dents): Nous n'avons pas trouvé la signification de ce geste dans le répertoire des macaques à queue de cochon, mais il est très probable que ce soit le même que chez les macaques des forêts (*Macaca sylvianus*), chez qui le teeth-chatter a lieu avant des interactions agonistes (Fischer et Hammerschmidt, 2002).

²⁰⁰ L'expression faciale de **bouche ouverte**, chez les macaques, est un signe de menace.

²⁰¹ Plus précisément, les 485 cellules se répartissent en 412 dans l'hémisphère droit d'un sujet et 73 dans l'hémisphère gauche de l'autre. De plus, 15 d'entre elles n'ont pas de propriété motrice manuelle ou buccale. Parmi les 470 actives pour la production d'actions, 101 sont spécifiques d'actions manuelles, 171 d'actions manuelles et buccales ingestives et communicatives, 198 d'actions orofaciale ingestives et communicatives. Dans les 369 neurones actifs pour l'exécution d'actions buccales ou buccales et manuelles, 180 sont actifs pour la perception de stimuli visuels, dont 50 pour la présentation d'objet fixes ou en mouvement, et 130 pour des actions buccales, ou manuelles et orofaciales. L'on peut remarquer que l'effet de la perception auditive n'est pas contrôlé dans cette étude.

²⁰² Les 8 miroirs buccaux VM strictement spécifiques restants déchargent pour la perception de l'action d'enlever (4) et celle de tenir avec les dents, TP-ing ou mâcher (4). Les 26 miroirs orofaciaux globalement spécifiques le sont pour les combinaisons d'actions suivantes: la préhension dentale et casser avec les dents (7), la préhension dentale et mâcher (6), la préhension dentale et téter (5), la préhension dentale et tenir avec les dents (4), et d'autres jeux d'actions (4).

Pour ce qui est de leur degré de congruence perceptuo-motrice, les miroirs buccaux ingestifs se répartissent en 37% strictement et 58% globalement congruents²⁰³. Bien que les miroirs orofaciaux communicatifs soient insensibles à la vision d'actions ingestives et qu'ils soient pour la plupart strictement sélectifs en perception, sauf exception²⁰⁴, ils n'en déchargent pas moins pour l'exécution d'actions ingestives spécifiques liées à leur-s action-s communicative-s effective-s en perception. Ce lien ne semble pas résider dans leur but fonctionnel: les actions effectives en production sont ingestives et s'appliquent donc à des objets, tandis que celles effectives en perception sont communicatives et donc corporelles.

Congruence articulaire des neurones miroirs communicatifs

Dans les limites de ce qu'il peut être proposé vu le faible effectif des miroirs communicatifs rapportés à ce jour, si la similarité entre action-s effective-s en vision et en exécution est cherchée au niveau du mouvement, un tableau intéressant peut émerger. En effet, d'après nous, leurs actions effectives en perception et en production semblent partager un but commun en termes de *configuration articulaire*, et cela selon une gradation, allant de stricte à ambiguë, de même ordre que celle mise en évidence pour la congruence perceptuo-motrice des neurones miroirs VM manuels et buccaux ingestifs. Plus précisément, l'on peut distinguer 3 degrés de *congruence articulaire*: strict, global et difficile à définir. Les neurones de congruence articulaire stricte comprennent ceux strictement spécifiques de TP-com (4) en perception qui déchargent uniquement pour l'exécution de TP-ing, ceux qui ne répondent qu'à la vision de LS (2) qui s'activent pour l'exécution de téter et LS, dans un cas, et un neurone sensible à LP-com dont la seule action effective en production est LP-ing. Le rapprochement entre LS et téter se comprend en considérant que le mouvement buccal de téter d'une seringue (comme d'une paille) peut vraisemblablement être décomposé en une séquence d'ouverture-fermeture de la mâchoire comme une succession de "bisou", ou LS, bien que le fluide en jeu ne soit pas de même nature dans les deux cas (voir aussi Fogassi et Ferrari, 2005). Avec un degré de congruence articulaire moins stricte, un neurone qui répond à la fois à la perception de LP-com et de TP-com, s'active spécifiquement pour l'exécution de la version ingestive de chacune de ces actions (LP-ing et TP-ing). De même, la cellule strictement sélective en perception pour TCh l'est aussi en production pour la préhension dentale. Par contre, la non-congruence articulaire est illustrée par un neurone qui ne répond qu'à la présentation de LP-com et qui s'active en production pour la préhension dentale et la mastication, ou encore par une cellule spécifique de LP-ing et TP-ing en perception mais uniquement de la préhension dentale en production.

Toutefois, un neurone nous rappelle que les fonctions de communication *vs.* alimentaire des miroirs orofaciaux ont certainement leur importance dans la compréhension du lien entre action-s effective-s en perception et celle-s en production des miroirs communicatifs. En effet, cette cellule est strictement sensible à la vision de LP-com, mais, à la différence de ses semblables, elle ne décharge pour l'exécution d'aucune action ingestive et ne s'active que pour la production de LS. Cela indique que le tableau ci-dessus s'inscrit probablement dans un ensemble plus large, où il pourrait figurer comme l'échantillon d'un lien entre actions buccales ingestives et communicatives, en termes de congruence globale *stricto sensu*, notamment via la communauté de leurs mouvements et de leurs effecteurs (langue et lèvres), à la manière dont des actions buccales et manuelles ingestives partagent un même substrat neuronal miroir.

²⁰³ 5% des miroirs buccaux ingestifs sont non-congruents. Sinon, nous n'avons pas plus d'information sur la congruence des miroirs buccaux ingestifs.

²⁰⁴ Un neurone miroir orofacial communicatif est strictement sélectif en perception pour LP-com, et en production pour LS. C'est un des 2 rares miroirs au cours de l'enregistrement desquels le sujet a produit une action communicative effective pour la cellule, et ce spontanément en réponse aux gestes communicatifs de l'humain (ici, LP-com).

En résumé, les macaques à queue de cochon disposent d'au moins quelques neurones miroirs VM capables de donner une équivalence, en termes de configuration articulaire, entre actions orofaciales perçues et exécutables, semblables à celles que sont capables d'imiter à partir de leur vision des nouveaux-nés humains et chimpanzés (voir plus bas).

Toutefois, l'intérêt des adeptes du développement de la parole pour les neurones miroirs ne devrait pas s'arrêter au fait qu'ils puissent faire le lien entre la production d'une action et sa vision sur autrui, et concerner des actions orofaciales corporelles apparentées à celles de parole²⁰⁵. En effet, ils peuvent implémenter des équivalences audio-motrices et audio-visuo-motrices, en témoignant d'une coordination audio-visuelle pour la perception d'actions.

c. Miroirs audio-visuo-moteurs et audio-moteurs d'actions manuelles sur des objets

Les mouvements biologiques, qu'ils soient produits par soi ou par autrui, sont perceptibles par tous les canaux sensoriels à disposition, notamment le visuel et l'auditif, et le proprioceptif pour les mouvements propres. Les actions n'échappent pas à la règle, et cette loi de multi-modalité de l'action se retrouve dans des neurones miroirs. En effet, deux études parmesanes sont relatives à des neurones miroirs d'actions manuelles *bruyantes* dans F5: l'une est celle de mise en évidence (Kohler, Keysers, Umiltà, Fogassi, Gallese et Rizzolatti, 2002), l'autre reprend les données de la première pour préciser la contribution de l'audition et de la vision dans la sensibilité des miroirs audio-visuels (Keyser, Kohler, Umiltà, Nanetti, Fogassi et Gallese, 2003).

Des neurones F5 de trois macaques adultes, vivant en laboratoire, sont testés sur leur activité potentielle lors de la perception auditive, de la vision et de la production d'actions manuelles sur des objets (Kohler *et al.*, 2002): casser la coque d'une cacahuète, déchirer une feuille de papier, secouer du papier, froisser un sac en plastique, laisser tomber un bâton à terre, manipuler de la nourriture sèche, taper un métal contre un autre métal et prendre un anneau qui émet un son au moment du contact²⁰⁶.

➤ Equivalence audio-motrice

Sur 497 cellules enregistrées²⁰⁷, 63 déchargent pour la production et l'audition d'actions manuelles sur des objets qui sont majoritairement celles de casser (27) et de déchirer (21)²⁰⁸. Les auteurs précisent que ce sont les plus fréquentes dans le répertoire de comportements manuels bruyants des sujets.

²⁰⁵ De plus, la perception de parole peut s'interroger sur la possibilité que des neurones miroirs VM d'actions orofaciales corporelles puissent s'activer pour la perception d'actions communicatives dont l'atteinte du but, qui reste à définir dans ce contexte, ne serait qu'inférée par le sujet, comme ce qui a été mis en évidence sur des miroirs VM d'actions manuelles sur des objets.

²⁰⁶ Des sons non-liés à des actions manuelles sur des objets ont été testés sur des cellules pour vérifier la spécificité des réponses auditives. Ces sons pré-enregistrés comprennent: du bruit blanc, des tons purs, des clics, des cris de macaques et d'autres animaux.

²⁰⁷ En réalité, cet ensemble de cellules est constitué de 2 groupes. Le premier (211) a été testé avec les sons non-liés à des actions manuelles et le son *réel* des actions manuelles exécutées par un/une expérimentateur/trice. Le second (286) est testé avec le son pré-enregistré des actions manuelles. La majorité (61) des cellules ayant des propriétés motrices et auditives semblent être issues du second groupe (286) (voir plus bas dans Keysers *et al.*, 2003).

²⁰⁸ Il en reste 5 pour froisser, 3 pour taper, 3 pour secouer, 2 pour manipuler et 2 pour laisser tomber un bâton. De plus, l'activité de 32 des 63 cellules a aussi été suivie lors de l'audition de sons non liés à des actions manuelles sur des objets: 5 seulement présentent une réponse relativement faible à ce type de stimuli (sans avoir plus de précision quant à leur nature).

33 de ces miroirs *a priori* AM ont été testés sur leur capacité de discriminer 2 actions manuelles, présentées en ordre aléatoire, en vision seule (*Condition V*), en audition seule (*Condition A*), et en perception audio-visuelle (*Condition AV*)²⁰⁹. Une quatrième condition expérimentale est celle de l'exécution de ces actions (*Condition M*) pour 28 cellules. Kohler *et al.* (2002) distinguent des neurones "sélectifs" et des "non-sélectifs": les premiers ne déchargent que pour la perception de leur action la plus effective (*sélectifs+*), les seconds s'activent en plus pour une de leur actions moins efficaces (*globalement sélectifs*).

Il en résulte que 29 cellules sont sélectives+ en audition, dont 22 le sont aussi en vision (*neurones miroirs AVM*) et 4 n'ont pas de réponse visuelle (*neurones miroirs AM*)²¹⁰. Parmi les neurones sélectifs+ sur le son, (i) 16 ont un taux de décharge similaire quelle que soit la condition perceptive ($AV=A=V$), (ii) 10 s'activent plus en condition AV que A ($AV>A$), et (iii) 3 sont plus sensibles au son seul qu'aux deux autres stimuli (*A-max*).

Pour le premier type de cellules, le son seul est aussi informatif que la vue seule, ou que l'audio-vision de l'action. Il vient que pour la moitié des neurones sélectifs+ en audition, quelle que soit la modalité en jeu un même équivalent moteur est évoqué, et les canaux auditif et visuel se valent. D'après les auteurs, le second groupe de cellules a besoin de l'information issue des deux modalités pour "décrire" l'action, qui en condition écologique est toujours perçue de manière bimodale. De plus, l'importance de l'audition dans l'activation des cellules A-max et $AV>A$ pourrait s'expliquer par le fait que le son soit souvent un indice d'accomplissement de l'action, ie. de *l'atteinte de son but*.

Globalement, sur l'ensemble de l'échantillon de 33 neurones relatifs à l'audition et l'exécution d'actions manuelles sur des objets, les cellules déchargent en réponse aux sons des actions de manière spécifique: elles montrent une *discrimination auditive* entre les sons d'actions différentes. De plus, les actions "préférées" en audition sont aussi celles qui induisent les réponses les plus élevées en exécution et en vision seule.

Autrement dit, ces cellules miroirs sont susceptibles de traduire en des termes moteurs (potentiels) le son d'actions manuelles sur des objets, et donc d'implémenter une *équivalence audio-motrice*, menant vraisemblablement à une capacité de *discrimination auditive*. Toutefois, il est important de noter que le son des actions manuelles sur des objets permet probablement de distinguer les actions bruyantes plus sur la base de leur *but* que sur celle des moyens précis pour les atteindre²¹¹.

➤ **Équivalence audio-visuo-motrice et coordination intermodale**

Dans l'approfondissement de cette étude, Keyssers *et al.* (2003) ajoutent que sur 286 neurones de F5, 130 répondent à la production et à la perception auditive et/ou visuelle d'actions manuelles sur des objets, dont 61 à l'audition d'actions. Cette publication précise l'interaction entre audition et vision au niveau des 22 neurones miroirs AV sus-cités dont 14 ont été testés en condition M pour leur action la plus efficace et pour celle moins efficace.

²⁰⁹ Pour chaque cellule, les deux actions correspondent à la plus efficace en audition et une des actions moins efficaces. En condition A et AV, le son est toujours pré-enregistré et, pour chaque action, 10 exemplaires sonores légèrement différents ont été diffusés. De plus, en condition V, les objets auxquels s'appliquent les actions sont modifiés de manière à ne plus émettre de son: la coque de la cacahuète est déjà cassée, le papier est mouillé, la feuille à secouer est en mousse, le sac en plastique est remplacé par de la mousse, le bâton tombe sur de la mousse, l'anneau n'émet plus de son. Taper un métal contre un autre métal n'est pas utilisé.

²¹⁰ Il reste 3 cellules sélectives+ en audition qui sont globalement sélectives en vision. Nous n'avons pas de précision sur la nature des actions en jeu, mais elle correspondent probablement à casser et déchirer.

²¹¹ Il semblerait que Kohler *et al.* (2002) n'aient pas cherché à définir la structuration potentielle des neurones miroirs spécifique de l'audition en termes de but et de moyen, comme ce qui a pu être fait dans le cas de la vision.

Il en ressort que 3 catégories de miroirs AVM peuvent être distinguées. Le premier groupe de cellules (11) sont de type $AV=A=V$ (voir plus haut). Les neurones de la seconde catégorie (8) sont plus sensibles en perception bi-modale que uni-modale (*AV-max*). Pour 5 d'entre eux, l'activité en condition audio-visuelle est la somme des activités auditive et visuelle (*miroirs AVM additifs*), tandis que 3 ne répondent pas à la vision seule ($AV>A$). Les 3 cellules du troisième groupe sont de type A-max.

Sur l'ensemble de l'échantillon de 22 neurones, une interaction audition-vision significative est mise en évidence. D'après les auteurs, cette dépendance inter-modale s'explique par le fait que, quand le son est absent, la vision de l'action la plus effective contribue fortement au taux de décharge, mais pas si le son est présent. De plus, si un seul des neurones miroirs AVM était utilisé pour discriminer les actions en jeu, les performances des sujets seraient en moyenne de 97% en condition bi-modale, contre 90% environ en condition uni-modale (88% pour la vision et 89% pour l'audition). Chose curieuse, sur les 14 cellules testées en condition M, si le taux de décharge en exécution était la base de discrimination des sujets, leurs performances seraient de 81% (vs. 87, 86 et 97%, avec l'audition, la vision et l'audio-vision, respectivement).

En d'autres termes, l'on retrouve au niveau des neurones miroirs AVM une *coordination audio-visuelle* dans la perception d'actions comme en témoigne la possibilité (i) de compenser et de renchérir une modalité par une autre, et (ii) de donner une équivalence entre l'audition et la vision d'actions en plus de celle entre perception et production. Leur capacité de discrimination audio-visuelle d'actions est supérieure aux discriminations uni-modales, voire à celle en production.

D'après Keyssers *et al.* (2003) les neurones miroirs AVM pourraient se développer par un apprentissage associatif de type Hebbien liant l'audition, la vision et la production des actions propres par contiguïté temporelle, dans le même ordre d'idées que la proposition avancée par Rizzolatti et Luppino (2001, voir section IV.A.2.a.).

d. Un circuit de neurones miroirs PFG/PF-F5c

Selon le groupe de Parme, F5c fait partie d'un circuit voué à la reconnaissance d'actions d'autrui. Ce circuit est formé des connections réciproques que partagent F5c avec la région du lobule pariétal inférieur PFG/PF, où se trouvent aussi des neurones miroirs (Fogassi *et al.*, 2005)²¹². Plus précisément, ce circuit permettrait de convertir la description perceptive de l'action d'autrui en des termes moteurs potentiels (ou prémoteurs). De là, l'action perçue serait

²¹² Ajoutons que jusqu'à présent, les publications du groupe de Parme se sont référées uniquement à PFG, mais des données d'anatomie récentes défendent une nouvelle nomenclature (Luppino *et al.*, 2004). "A propos de nomenclature, note que dans nos futures publications, nous nous référerons à PF, PFG comme étant les aires strictement liées au cortex prémoteur ventral (aires F4 et F5) et que, en utilisant cette nomenclature, les neurones miroirs semblent être principalement localisés dans l'aire PFG, où la représentation de la main est prépondérante. Cette nouvelle attribution est due au fait que le neuroanatomiste du laboratoire de Luppino dans notre Institut a effectué une parcellarisation du lobule pariétal inférieur et a redéfini les aires rapportées auparavant par Pandya et Seltzer. Dans la nouvelle parcellarisation de Luppino, PF est bien moins étendue qu'avant et PFG est une grande région, bien définie. En comparant nos cartes fonctionnelles avec leurs aires cytoarchitectoniques, nous pensons désormais que la majorité des neurones miroirs tombent dans PFG" (Fogassi, com. pers., 21/01/2005, traduction de "[a]bout the nomenclature, note that in our future publications on parietal neurons we will refer to PF, PFG as the areas strictly linked with ventral premotor cortex (areas F4 and F5) and that, using this nomenclature, mirror neurons appear to be located mainly in area PFG, where the hand representation prevails. This new attribution depends on the fact that the neuroanatomists of Luppino's lab in our Institute made a cytoarchitectonic parcellation of the inferior parietal lobule and redefined the areas previously reported by Pandya and Seltzer. In the new Luppino's parcellation PF is much smaller than before and PFG is well defined, large region. Comparing our functional maps with their cytoarchitectonic areas, we now believe that most parietal mirror neurons fall in PFG.").

(directement) située par rapport au répertoire comportemental du sujet et donc reconnue. Que sait-on de PFG/PF?

PFG/PF

Cette région à dominante sensorielle a une représentation de la face et de la main et est multimodale. En effet, elle comprend en grande majorité des neurones qui répondent aux stimuli visuels, *somesthésiques* (proprioceptifs inclus), et *visuo-somesthésiques*. Parmi les cellules ayant des propriétés visuelles, une partie d'entre elles décharge spécifiquement pour la perception de mouvements biologiques et, en particulier d'actions (Rizzolatti et Luppino, 2001, Rizzolatti et Craighero, 2004). Certaines sont des neurones miroirs VM. En plus d'être connectée de manière réciproque à F5c, PFG/PF est liée (i) à l'aire somesthésique secondaire (SII), et (ii) par des connections réciproques, au Sillon Temporal Supérieur (STS).

STS

STS est une zone sensorielle qui comprend une représentation somatotopique, comme F5c et PFG/PF. L'on y trouve des neurones visuels sensibles à des mouvements biologiques complexes tels que la marche, la grimpe et bouger les bras, dont une petite partie déchargent pour la vision d'actions manuelles sur des objets, comme des cellules de PFG/PF (voir pour revue Carey, Perrett et Oram, 1997, cité par Iacoboni *et al.*, 2001). Toutefois contrairement à PFG/PF, et *a fortiori* F5c, STS ne semble pas avoir de neurones actifs lors de la production (Rizzolatti et Craighero, 2004).

L'idée est que STS enverrait une description visuelle de l'action au circuit PFG/PF-F5c qui "projetterait" cette information dans la représentation motrice du répertoire comportemental du sujet, de manière à la reconnaître (Rizzolatti et Craighero, 2004; Rizzolatti *et al.*, 2000). Nous reviendrons plus bas sur l'utilité de ce modèle fonctionnel en tant que base cérébrale de l'imitation visuelle chez l'adulte humain.

3. Une cerise sur le gâteau?

Les neurones miroirs de F5c sont susceptibles de traduire en des termes moteurs potentiels des actions perçues visuellement, auditivement et de manière bimodale. Nous allons tracer des éléments de liaison possible entre propriétés de ces systèmes miroir observés chez les macaques et architecture de buts et de moyen postulée pour les mécanismes d'imitation au cours du développement précoce chez les humains.

a. Neurones miroirs et comportements perceptuo-moteurs précoces

Au niveau moteur, des neurones miroirs VM concernent des actions *orofaciales* (i) qui figurent dans le répertoire *communicatif* des macaques à queue de cochon adultes et (ii) dont certaines font l'objet d'*imitation visuelle* par des *nouveaux-nés humains* (LP et TP²¹³, section III.A.1.a.) et *chimpanzés* (TP, section III.B.2.d.). Cela n'implique pas qu'il y aurait (i) un déterminisme génétique, ou une "innéité", des neurones ou d'aires miroirs, ni d'ailleurs (ii) la présence de neurones ou d'aires miroirs dans le lobe frontal de nouveaux-nés humains et chimpanzés, ou pis (iii) un récapitulonnisme du développement cognitif, mais indique simplement que les actions et la mise en équivalence perceptuo-motrice sur lesquelles portent des neurones miroirs de macaques à queue de cochon peuvent avoir une pertinence vis-à-vis

²¹³ LP: protrusion des lèvres, TP: protrusion-retraction de la langue.

du développement de la communication oro-faciale d'autres primates et, en particulier, par rapport au *développement de la parole*. En effet, dans le cas des macaques à queue de cochon, au moins, les neurones miroirs ne concernent *a priori* que des actions faisant partie du *répertoire comportemental* individuel de sujets *adultes*. En particulier, les neurones miroirs communicatifs ont été mis en évidence sur des individus vivants en *colonie*²¹⁴, et les actions sur des objets pour agir sur d'autres objets ne semblent pas activer de neurones miroirs, en accord avec la non-utilisation d'outils par des macaques.

De plus, l'architecture de buts et de moyens décrite par le groupe de Parme sur des macaques nous semble avoir des similitudes avec l'espace des actes, proposé par Meltzoff et Moore (1997), dans leur modèle AIM. L'on pourra remarquer l'idée commune d'une cartographie de moyens multiples pour atteindre un même but.

Un autre phénomène trouvé dans les neurones miroirs VM d'actions manuelles sur des objets intéresse l'imitation visuelle au cours du développement. En effet, certains sont sensibles à des actions dont le but n'est qu'inféré par les sujets. De là, ils sont potentiellement impliqués dans des processus de conversion visuo-motrice de même type que ceux vraisemblablement sous-jacents à l'*imitation visuelle* d'actions manuelles de *but inféré* que montrent des humains de 18 mois (section III.A.1.c).

Dans le même ordre d'idée, les neurones miroirs AVM pourraient participer à des mécanismes d'inversion²¹⁵ audio-visuo-motrice, voire audio-motrice, de nature similaire à ceux vraisemblablement nécessaires à l'*imitation audio-visuelle précoce* (section III.A.3.), si ce n'est à l'*appariement audio-moteur néonatal* (section III.A.2.).

En effet, au niveau perceptif, ces cellules sont susceptibles de jouer un rôle dans des processus de *coordination audio-visuelle* et de discrimination bimodale d'actions. Bien que leurs actions effectives soient manuelles et appliquées à des objets, les miroirs AVM n'en présentent pas moins des propriétés d'intermodalité observée pour la *parole*, et ce dès l'âge de 2 mois (section III.A.2.).

De plus, les données publiées dans Keysser *et al.* (2003) suggèrent que les neurones miroirs VM d'actions bruyantes décrits section IV.A.2.a. sont très probablement pour une moitié d'entre eux des miroirs AVM. Par ailleurs, Ferrari *et al.* (2003) trouvent des neurones miroirs VM d'actions orofaciales ingestives et communicatives. Pour mémoire, les macaques à queue de cochon ont une communication principalement basée sur la modalité visuelle. Or, dans l'étude de Kohler *et al.* (2002), des sons non-liés à des actions manuelles sur des objets ont été testés sans résultat: ils comprennent du bruit blanc, des tons purs, des clicks, des cris de macaques et d'autres animaux, mais ne semblent pas contenir de "bruits de bouche". Ainsi est-il plausible que des neurones miroirs AVM, voire AM, sensibles à des *sons d'actions ingestives* existent dans F5c. Quel en serait l'intérêt? La réponse réside dans des arguments théoriques sur l'émergence de la parole au cours de l'évolution.

b. Neurones miroirs visuo-moteurs orofaciaux et Théorie Frame/Content

Section IV.A.2.b., Ferrari *et al.* (2003) mettent en évidence un substrat neuronal commun entre actions communicatives et ingestives au niveau des cellules miroirs d'actions orofaciales de macaques à queue de cochon. En effet, la plupart des neurones miroirs communicatifs montrent une congruence en termes de *configuration articulaire* entre actions ingestives effectives en production et actions communicatives effectives en perception. Ces résultats font

²¹⁴ Nous pensons que c'est leur répertoire comportemental qui justifie que les participants de cette expérience ont été amenés de Strasbourg à Parme pour participer à l'étude, alors que des macaques de la même espèce, vivants en laboratoire, sont présents sur place.

²¹⁵ Schématiquement, l'inversion est la conversion d'une information sensorielle en des termes moteurs *potentiels*.

écho à la théorie Frame/Content selon laquelle les systèmes de communication orofaciale des primates, dont celui de la parole, pourrait être "issus" de comportements ingestifs s'exprimant dans des interactions sociales de type affiliatif, telles que l'épouillage (MacNeilage, 1998; Fogassi et Ferrari, 2005; voir aussi chapitre 1, section III).

Un autre élément va dans le sens de la théorie: les neurones miroirs communicatifs ont été trouvés pour des gestes affiliatifs du répertoire communicatif des macaques à queue de cochon, notamment LP-com et LS²¹⁶. Notons que MacNeilage (1998) compte LS dans les précurseurs potentiels de la cyclicité mandibulaire caractéristique de la parole, qui émerge sous la forme du babillage canonique au cours du développement.

Un dernier point en appui à la théorie est le suivant. D'après des données d'embryologie comparative, de cytoarchitectonique et fonctionnelles, *F5 est l'homologue putative de l'aire de Broca* (Rizzolatti, Luppino et Matelli, 1998, Figure 3-8). Rappelons que, pour MacNeilage, l'aire de Broca fait partie du système en charge de l'apprentissage du contenu du cadre syllabique, et ce par imitation. Chose marquante, F5 comprend des neurones capables de faire le lien entre les actions d'autrui et celles de soi, bien que l'effecteur en jeu, le plus fréquemment mis en évidence, à ce jour, soit la main plutôt que la bouche.

Toutefois, sans rentrer dans la polémique de "l'origine manuelle vs. orofaciale du langage", l'on peut remarquer que les systèmes de communication brachio-manuelle ne semblent exister que chez des Hominés (eg. Chimpanzé, Humain) au sein des Primates (Cf. Figure 3-5). Par ailleurs, le répertoire communicatif des macaques à queue de cochon ne paraît pas doué de gestes brachio-manuels. Il se pourrait donc que les systèmes de communication brachio-manuelle, défendus comme primordiaux *stricto sensu* par certains (eg. Rizzolatti et Arbib, 1998), n'aient émergé, au niveau phylogénétique, qu'*après* des systèmes de communication orofaciale tels que ceux des Macaques ou encore des Ouistitis à toupet blanc ("marmosets" pour les anglophones). Tout cela n'enlève rien au fait que l'émergence de la communication brachio-manuelle hominée pourrait, comme l'orofaciale, utiliser un système miroir du même type que celui présent pour des actions orofaciales et manuelles chez des macaques à queue de cochon. Les lecteurs/trices intéressé/ées pourront trouver une remise en question du bien fondé de "l'hypothèse brachio-manuelle" dans Dale, Richardson et Owren (2003). Pour eux, l'effecteur (manuel) des actions effectives des neurones miroirs n'est pas central dans l'explication de l'émergence évolutive de la parole: c'est *l'intégration multimodale perceptuo-motrice* qu'ils implémentent qui importe le plus (voir aussi MacNeilage, 1998, voir p. 511).

²¹⁶ LP-com: pucker, LS: lip-smacking.

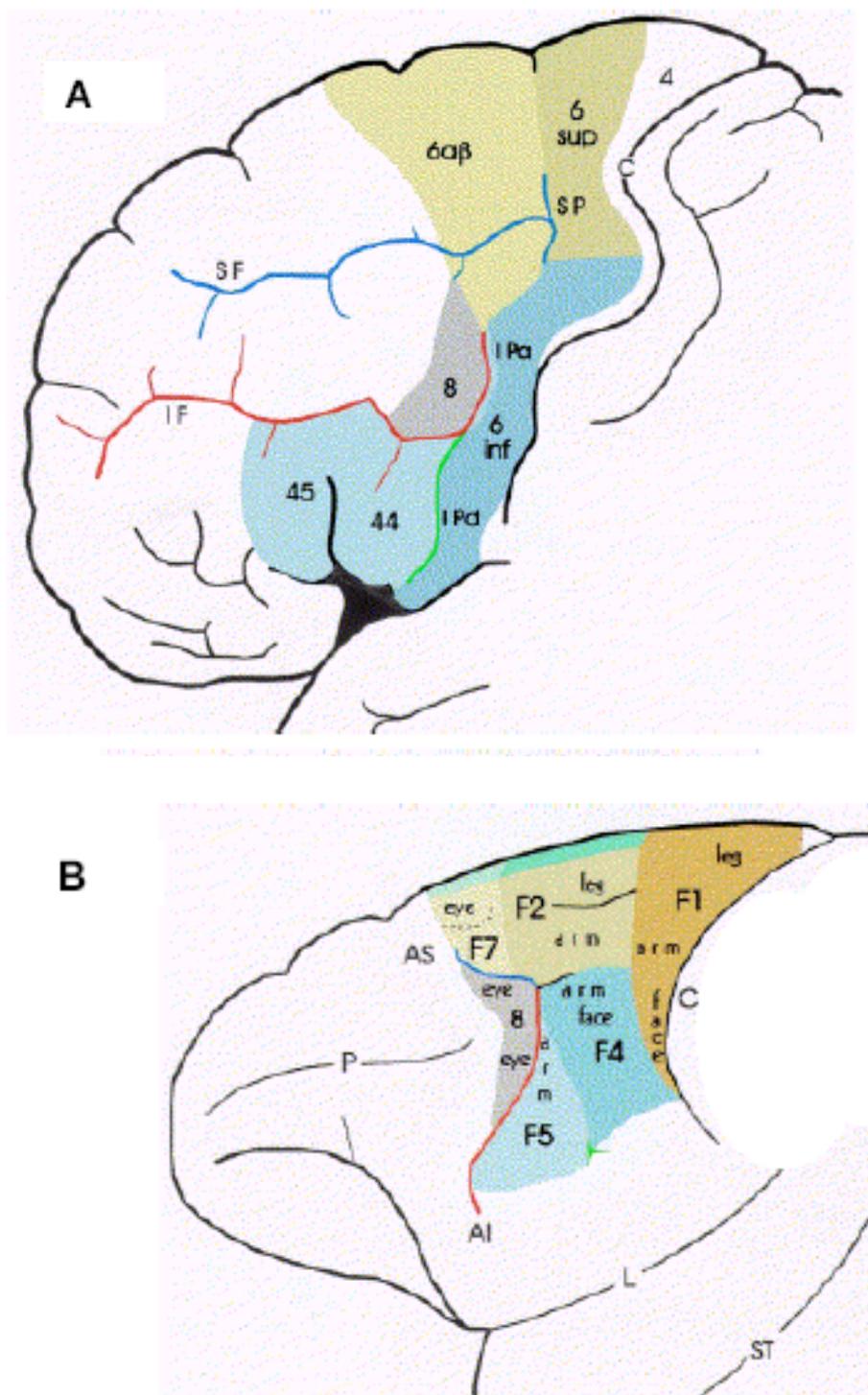


Figure 3-8: Les divisions du système cortical moteur humain (A) en fonction de leur homologie avec celles des macaques à queue de cochon (B). Le code de couleur indique l'homologie qui a été proposée entre les deux espèces. Les aires corticales F2-F7 du macaque sont étiquetées selon Matelli et al. (1985), celles de l'humain sont définies selon Brodman avec des sous-divisions supplémentaires de l'aire 6 d'après Vogt et Vogt. Les cissures sont étiquetées comme suit. (A) pour l'humain, C: Centrale, SF: Frontale Supérieure, IF: Frontale Inférieure, SP: Précentrale Supérieure, et (B) pour le Macaque, C: Centrale, AS: Arquée Supérieure, AI: Arquée Inférieure, L: Latérale, ST: Temporale Supérieure. Figure adaptée de Rizzolatti *et al.* (1998).

De "l'évolution des gestes communicatifs"

Précisons que nous ne défendons pas ci-dessus l'existence de "gènes de gestes" de communication. En effet, concrètement, ces gestes, comme tout autre mouvement contrôlé, se *développent*, et nécessitent en plus d'apprentissages individuels des apprentissages sociaux. Pour rappel, les informations résultant d'apprentissage au cours de la vie des individus ne peuvent être transmises via des gènes aux générations suivantes.

Notre point de vue est le suivant. Les ancêtres putatifs communs aux macaques et aux humains ont pu leur transmettre un ensemble de contraintes génétiques permettant, notamment, l'émergence de systèmes de communication orofaciale. De plus, le chemin allant de ce nœud ancestral à la feuille qu'est l'espèce humaine est vraisemblablement jonché de modifications génétiques, eg. des mutations, qui marquent la différence entre une espèce-fille et son espèce-mère. Or, il est plus que probable que l'information contenue dans le pool génétique de la fille soit *conditionnée* par celle contenue dans le génome de la mère et donc de ses ascendantes, par le simple jeu de leur lien de filiation.

De là, le chemin évolutif par lequel "est passé le génome d'une espèce" représente en soi une contrainte pesant sur l'information qu'il met à disposition pour le développement des individus. Et c'est dans les limites de cet univers des possibles génético-environnementaux²¹⁷, conditionné par celui de ces ancêtres, qu'une espèce donnée développe ses comportements. Pour illustration de la contrainte de l'ascendance, en partant du génome d'une espèce affiliée aux Vertébrés, si une mutation²¹⁸ empêche la formation de la corde au niveau embryonnaire, il en résulte notamment que le tube neural et donc le système nerveux dans son ensemble, ne peuvent se former, vu que leur développement dépend de la présence de cette corde. De là, l'individu au sens génétique ne peut voir le jour, au moins parce que le reste de son génome n'est pas celui d'un Invertébré, et la mutation qu'elle porte ne pourra pas passer à la génération suivante. Globalement, c'est *parce que* l'émergence d'un tube neural dans l'histoire évolutive des Vertébrés s'est faite *à partir* d'un génome Chordé que ce donné représente une *contrainte* qui pèse sur ce que peuvent et ne peuvent vraisemblablement pas devenir des Vertébrés. En d'autres termes, le profil génétique des Vertébrés est *conditionné par leur ascendance* Chordée. Cet exemple illustre aussi le poids qu'a *a fortiori* le *développement* sur l'évolution des êtres vivants.

En résumé, si les primates n'utilisaient pas de mouvements ingestifs buccaux, il est plausible que des systèmes de communication orofaciale, tel que celui de la parole, n'auraient pas émergé à ce jour dans cet embranchement. Mais cela ne signifie pas que l'information relative à ces gestes serait codée génétiquement.

²¹⁷ L'évolution des gènes ne peut être pensée sans l'environnement. A notre connaissance, l'environnement *contre-sélectionne* les modifications génétiques apportées, notamment, par les mutations: il dicte quelles nouveautés *ne* peuvent pas être maintenues dans le génome des espèces qui l'ont "choisi" pour niche (et non celles qui "doivent" être maintenues). Le génome d'une espèce est donc, plus précisément, conditionné par l'histoire génético-environnementale de celui de ses ascendantes.

²¹⁸ Cette mutation peut avoir lieu lors d'une gamétogenèse (formation des spermatozoïdes et des ovules) chez les parents biologiques.

B. Système d'appariement perception-production d'actions d'Humains adultes

Comme mentionné ci-dessus, le lien entre l'aire F5 des macaques à queue de cochon et l'aire de Broca (BA44²¹⁹) ne s'arrête pas à des données d'embryologie comparative et de cytoarchitectonique, elles partagent des propriétés fonctionnelles, d'après des études de neuroimagerie.

1. Broca et F5 au niveau fonctionnel

Si BA44 est connue pour son implication dans la production de parole, l'on peut citer, notamment, les travaux de Buccino et collègues qui montrent sa sensibilité à la perception visuelle d'actions orofaciales corporelles (communicatives) et d'actions orofaciales sur des objets (ingestives), et ce, même si elles sont produites par les membres d'espèces non-humaines²²⁰ (Buccino, Canessa, Patteri, Lagravinese, Benuzzi, Porro et Rizzolatti, 2004, cité par Buccino, Binkofski et Riggio, 2003). En particulier, les auteurs montrent que cette activation est fonction du statut concevable *vs.* inconcevable de l'action vis-à-vis du répertoire du sujet. Plus précisément, si le geste est inconnu en production (e.g. aboiement) et que la similarité entre le corps propre et celui de l'autre (e.g. espèce canine) est relativement faible, BA44 reste silencieuse. D'après Rizzolatti et Craighero, l'activation de BA44 lors de l'observation d'actions reflète une "projection" de ces actions perçues sur le *répertoire moteur effectif* des sujets, mécanisme qui ne peut avoir lieu pour des actions inconnues. Dans ce dernier cas, les actes d'autrui ne seraient reconnus que sur la base de leurs propriétés en termes d'analyse visuelle et non en termes moteurs.

De plus, dans une étude antérieure, Buccino et co-auteurs retrouvent, en accord avec d'autres, que cette aire est organisée selon une somatotopie analogue à celle de F5, et s'active notamment pour la perception visuelle d'actions manuelles *corporelles* et d'actions manuelles sur des objets (Buccino, Binkofski, Fink, Fadiga, Fogassi, Gallese, Seitz, Zilles, Rizzolatti et Freund, 2001, cité par Buccino *et al.*, 2003; voir aussi Rizzolatti et Craighero, 2004). Ce dernier point marque une différence cohérente entre F5 et BA44: les neurones miroirs de F5 des macaques à queue de cochon sont insensibles aux pantomimes, reflétant ainsi l'absence d'actions manuelles corporelles dans leur répertoire, tandis que ce type d'action est présent dans le répertoire d'humains.

Toutefois, une chose bien plus importante pour nous est que l'aire de Broca est impliquée dans le système cortical en charge de *l'imitation* chez des adultes humains.

2. Une architecture fonctionnelle minimale pour l'imitation chez des humains adultes

Des études de neuroimagerie mettent en évidence, chez des adultes humains, un système similaire au circuit miroir PFG/PF-F5c des macaques adultes: il comprend des aires actives pour la perception et la production d'actions. Ce système est aussi recruté lors de l'imitation d'actions faisant partie du répertoire des sujets (eg. Iacoboni, Koski, Brass, Bekkering,

²¹⁹ Au niveau anatomique, l'aire de Broca se trouve dans le gyrus frontal inférieur (*IFG*) de l'hémisphère gauche. Pour information, un gyrus est un "bourrelet cérébral". BA44 est l'aire de Broca au sens strict et correspond à la partie operculaire (*pars opercularis*) de ce gyrus. Des auteurs incluent aussi les aires BA45 et BA47 dans celle de Broca. Elles consistent respectivement en la partie triangulaire (*pars triangularis*) et la partie orbitaire (*pars orbitalis*) de ce même gyrus (voir Tzourio-Mazoyer, 2002, pp. 82 et 90).

²²⁰ Par exemple, Broca était active pour la perception visuelle de parole, mais dans une moindre mesure pour celle de LS (lip-smacking) de macaques.

Woods, Dubeau, Mazziotta et Rizzolatti, 2001; Iacoboni, Woods, Brass, Bekkering, Mazziotta et Rizzolatti, 1999; voir pour revue Rizzolatti et Craighero, 2004, et Iacoboni, 2003).

En rendant compte d'expériences sur des humains et en s'inspirant des travaux relatifs au circuit miroir des macaques, Iacoboni (2003) propose une architecture minimale pour l'imitation. Elle comprend (i) la partie rostrale du lobule pariétal inférieur (*PC*), (ii) la pars opercularis du gyrus frontal inférieur (*IFG*), où se trouve BA44, et (iii) le sillon temporal supérieur (*STS*)²²¹. *PC* et *IFG* constituent le circuit miroir humain, où *PC* est prise pour homologue fonctionnelle de *PFG/PF* (Rizzolatti et Craighero, 2004).

L'idée de Iacoboni est la suivante (Iacoboni *et al.*, 2001). Si la production de mouvements laisse une trace de leurs conséquences sensorielles, comme James (1890) l'a avancé, et que la représentation motrice d'une action volontaire évoque une représentation sensorielle de ses conséquences (ou "copie sensorielle interne"), l'imitation peut être sous-tendue par un mécanisme faisant le lien entre la représentation perceptive de l'action propre et celle de l'action à imiter, lien qui active en retour des représentations motrices pertinentes sachant l'acte cible.

Sans rentrer dans le détail, le modèle de Iacoboni, qui est en accord avec celui de Rizzolatti *et al.* (2000) pour la reconnaissance d'actions des macaques (voir section IV.A.2.d), peut s'énoncer comme suit.

STS, région sensible aux mouvements biologiques chez les humains autant que chez les macaques, fournit une description visuelle de l'action à imiter au circuit *PC-IFG*.

PC enrichit cette description d'informations somesthésiques sur l'action perçue²²². La partie dorsale d'*IFG* donne le but de cette action et produit un ensemble de commandes motrices potentielles *a priori* nécessaires à son exécution, ou **plan moteur**²²³. Ce plan, qui peut être conçu comme la sortie d'un **modèle inverse**²²⁴, est envoyé au secteur ventral d'*IFG* qui en produit une **copie d'efférence**. Cette copie est, par définition, une prédiction des conséquences sensorielles de l'action planifiée: elle peut être vue comme la sortie d'un **modèle direct**^{225 226}.

La copie d'efférence de l'action imitative est renvoyée à *STS* où il y a comparaison entre la description perceptive de l'action perçue et celle de l'action planifiée. Si la différence est relativement grande, le plan moteur est corrigé jusqu'à convergence. Si l'appariement est correct, l'action planifiée est initiée par le sujet, et dans le cas où les conséquences sensorielles

²²¹ Pour la situation anatomique de ces zones du cortex cérébral, voir Tzourio-Mazoyer, 2002a, pp. 82 et 90, pour le gyrus frontal inférieur dans le lobe frontal; Petit, 2002, pp. 93, 97 et 99, pour le lobule pariétal inférieur dans le lobe pariétal, et Tzourio-Mazoyer, 2002b, pp. 111 et 122 pour le sillon temporal supérieur dans le lobe temporal. Pour information, *PC* est la zone où se trouve le gyrus supra-marginal (BA40) que des auteurs soupçonnent être impliqué dans l'encodage de buts phonologiques (Sato, Baciù, Lœvenbruck, Schwartz, Cathiard, Segebarth et Abry, 2004; Paulesu, Frith et Frackowiak, 1993).

²²² Rappelons que *PFG/PF* macaque, homologue putative de *PC*, comprend des neurones visuels, somesthésiques et visuo-somesthésiques. Ces derniers sont susceptibles de donner une équivalence visuo-somesthésique. De plus, du côté humain, *PC* est une région corticale associative, qui se trouve à la convergence de flux d'informations issus de plusieurs modalités sensorielles, tandis que *STS* est proposé comme lieu majeur des interactions audio-visuelles (Calvert et Campbell, 2003).

²²³ Pour Iacoboni (2003), la partie dorsale d'*IFG* est de type *miroir*, tandis que sa partie ventrale est plutôt *prémotrice*.

²²⁴ Un **modèle inverse** trouve les commandes motrices nécessaires à l'atteinte d'un état sensoriel donné dont il prend la description en entrée. Ce type de modèle entre en jeu dans le contrôle moteur.

²²⁵ Un **modèle direct** simule le comportement du système moteur périphérique: il donne les conséquences sensorielles d'un plan moteur, comme s'il avait effectivement été exécuté. Ce type de modèle est le résultat d'un apprentissage où des plans moteurs sont associés à leurs conséquences sensorielles.

²²⁶ Pour information, Miall (2003) propose que *STS* n'est pas nécessaire à l'implémentation des modèles directes et internes du circuit miroir, sachant que le cervelet pourrait tout aussi bien s'acquiescer de cette tâche et qu'il est lié, lui aussi, à *PC* et *IFG*.

réelles de cette action confirment la prédiction, le lien entre les modèles direct et inverse qui lui sont associés est renforcé.

Conclusion

Nous avons pris dans ce chapitre le parti de nous appuyer sur une analyse aussi exhaustive et fine que possible des données expérimentales sur l'imitation et sur certains de ses corrélats neurophysiologiques potentiels proposés dans la littérature. Cette démarche a été adoptée pour les raisons suivantes. Tout d'abord, ce travail n'avait pas, à notre connaissance, été fait *de l'intérieur* du domaine de la parole. Or, la thèse que nous défendons désigne l'imitation comme étant un mécanisme essentiel à l'apprentissage de comportements nouveaux et arbitraires et donc à l'*ajustement* des productions de soi à celles de la langue ambiante (ou *tuning*) au cours du développement de la parole.

Les données montrent sans conteste la variété et la complexité des mécanismes d'imitation. En particulier, ceux-ci sont présents dès les premiers jours de vie et ne disparaissent pas pour, au contraire, se raffiner et se structurer, permettant ainsi la capture, la représentation et le contrôle de comportements de plus en plus complexes. *Nous proposons que ces capacités d'imitation, que l'on retrouve vraisemblablement chez tous les Hominiés, sont une pièce maîtresse de l'émergence des "schèmes sensori-moteurs" de la parole dans les premiers mois et les premières années de vie.* Nous considérons de plus que le noyau fonctionnel de ces mécanismes implique une composante fondamentale, celle de la *mise en correspondance de représentations sensorielles multiples*, avec l'hypothèse supplémentaire qu'une *relation du système somesthésique aux modalités externes* pourrait être au cœur de cette mise en correspondance, en fournissant le germe opérationnel de l'appariement entre extéro-perception et motricité.

Les données de neurosciences relatives aux adultes macaques et humains proposent un cadre solide – et régulièrement confirmé et affiné par les publications successives – pour une architecture neuronale sous-jacente à des mécanismes d'appariement perception-production, dont l'imitation chez les humains. Bien que ces modèles fonctionnels ne soient pas directement transposables au cas juvénile, sachant notre manque de connaissances sur le fonctionnement cérébral précoce, ils nous informent au moins sur l'état final du/des processus de développement qui y mène/ent.

L'étape suivante du travail que rapporte ce mémoire consiste à tenter de nous doter de concept-outils pouvant nous aider à rendre compte de l'apprentissage de représentations sensori-motrices susceptibles de servir, à la fois, de *support* au phénomène d'imitation et de *réceptable* des résultats et nouveaux contenus d'apprentissage issus de l'expérience imitative de l'individu.

Ceci nous conduira à un travail de modélisation autour de deux composantes principales: (i) un *modèle direct* des relations sensori-motrices, qui pourrait être appris par une exploration progressive de l'espace sensori-moteur de l'agent (humain, animal ou artificiel), et (ii) un *modèle inverse* conduisant à estimer une commande adéquate capable de reproduire un percept, donc à l'imiter. La dernière partie de ce mémoire rendra compte de ce travail de modélisation.



PARTIE III:
**Modélisation des mécanismes
d'exploration et d'imitation**

Introduction à la troisième partie

Notre projet veut faire passer un agent sensori-moteur en croissance par des étapes de développement observées sur des nourrissons humains (voir chapitre 2, Introduction), afin, notamment, (i) d'éclairer des zones d'ombre de l'état de nos connaissances sur le phénomène et (ii) d'interroger les représentations cognitives susceptibles d'entrer en jeu dans le développement de la production de la parole. Les deux premières parties de ce mémoire voulaient présenter la toile de fond de notre démarche en donnant, d'une part, la situation de notre point de vue sur le développement de la production verbale et, d'autre part, les modèles mis à disposition par nos prédécesseurs. Ici, se trouve le chaudron où nous concevons et construisons nos propres outils de modélisation pour simuler et interroger des facettes du phénomène. Des hypothèses sous-jacentes à la construction de l'androïde pourront être affinées, par la confrontation des prédictions de nos outils à des données réelles.

A ce jour, les ingrédients principaux mis en œuvre dans l'élaboration du premier avatar sont les suivants: (i) le modèle articulatoire-acoustique de production de la parole VLAM, qui intègre la croissance non-linéaire et non-uniforme du conduit vocal (chapitre 2, section I), (ii) des systèmes auditif, visuel et tactile simplifiés (chapitre 2, section II.4.) (iii) un mécanisme d'apprentissage et d'inversion de type bayésien qui tient compte de relations de dépendance entre des variables articulatoires et perceptives (chapitre 2, section III et IV), (iv) un scénario de développement de la coordination du système articulatoire, inspiré de la théorie Frame/Content (chapitre 1, section III), et (v) des données acoustiques et de transcription mesurées auprès de vrais nourrissons publiées par d'autres auteurs.

Dans ce cadre, deux notions, référant chacune à un phénomène vraisemblablement à l'œuvre au cours du développement de la parole, nous servent d'axes de modélisation: l'exploration (vocale) et l'imitation (orofaciale).

Deux processus de base: l'exploration et l'imitation

Dans notre vision du développement de la parole, le mécanisme d'**exploration** désigne l'expérimentation par l'individu d'un ensemble de gestes du conduit vocal (ou configurations articulatoires) ayant notamment des effets auditifs et tactiles. L'idée est que la répétition de la concomitance geste-conséquences perceptives fait vraisemblablement l'objet de l'apprentissage d'une correspondance production-perception multimodale¹ par contiguïté temporelle. Cette *carte cognitive* de l'espace sensori-moteur des possibles en cours sous-tend le développement du contrôle des mouvements audibles de l'appareil vocal et, par là, celui de la parole en devenir. En d'autres termes, la finesse et l'étendue de cette représentation contraignent, de fait, la production volontaire, et ce en tout point du décours développemental. Aussi, *le domaine d'exploration articulatoire-acoustique observé, ou estimé, est pris pour une approximation de l'espace sensori-moteur potentiellement appris par le sujet.*

Toutefois, la contribution de cet apprentissage individuel dans l'émergence et la mise à jour de la carte perceptuo-motrice des propriétés effectives du conduit vocal se double d'un apprentissage social: l'**imitation** oro-faciales de modèles de parole. Rappelons qu'en psychologie, l'imitation correspond à l'appariement de son propre comportement à celui d'un autre individu (le modèle), ou mieux, à donner un équivalent moteur à cette cible perceptive (voir chapitre 3, section I). L'imitation en acte, i.e. les instances d'imitation, s'inscrivent au même titre que les vocalisations spontanées dans l'espace sensori-moteur d'exploration. Or, les capacités humaines d'imitation, ce dès la naissance, montrent que l'individu en

¹ Dans les termes de l'éthologie, il s'agit de la formation d'une carte cognitive (voir chapitre 3, section I.1.).

développement ne fait pas que produire: il perçoit des cibles potentielles "pour" l'imitation (chapitre 3, section III). De là, l'imitation permet d'orienter l'exploration articulatoire-acoustique *vers* les objets (e.g. réalisations phonémiques) de la/des langue-s ambiante-s, menant ainsi à une inter-structuration cohérente et dynamique des paysages articulatoire et perceptif, entre eux, via la carte cognitive qui représente leurs liens, et vis-à-vis de l'environnement.

Schématiquement, l'exploration est vue comme le support à l'émergence du contrôle, et l'imitation dirige l'exploration sur les productions de la parole ambiante. Bien que ces deux phénomènes ne soient observables en conditions écologiques que dans la résultante de leur interaction, i.e. dans ce que les nourrissons produisent effectivement aux niveaux acoustique et articulatoire, nos travaux de modélisation ont proposé d'en rendre compte de manière relativement indépendante.

Notre objectif est de simuler ces deux processus avant et au début du babillage canonique, en se basant sur des données acoustiques réelles, recueillies à ces deux périodes de développement, aux âges respectifs de 4 et 7 mois. Le quatrième chapitre veut estimer le domaine d'exploration articulatoire-acoustique à ces deux périodes, tandis que le cinquième propose de rendre compte de performances en imitation orofaciale précoce des voyelles [i a u] établies au cours d'une expérience vue au chapitre 3.

Chapitre 4: Modélisation de l'exploration avant et au début du babillage

INTRODUCTION

Afin de réfléchir sur les caractéristiques de l'information potentiellement apprise au cours du développement de la parole et menant probablement à sa forme adulte, le point de départ de la conception de notre agent se trouve dans l'ajustement des capacités de production du modèle VLAM au domaine d'exploration articulatoire-acoustique de (vrais) nourrissons. En effet, les paramètres moteurs d'entrée de VLAM sont issus de l'analyse statistique de coupes radiographiques d'un conduit vocal adulte en action (chapitre 2, section I). De là, ils reflètent ce qu'un *adulte* est capable de produire en termes de configurations articulatoires, tandis que les nourrissons n'exploitent vraisemblablement qu'un sous-ensemble de ce qui peut être effectué avec le système sensori-moteur dont ils/elles héritent phylogénétiquement (voir tout au long du chapitre 1, les remises en question du postulat *Ji*, de Jakobson, section I.1.a.).

Toutefois, une recherche bibliographique nous a permis de constater un manque de caractérisation non-ambiguë du développement de la *production* de la parole. En effet, les techniques utilisées comprennent essentiellement des méthodes de transcription et la mesure de variables de phonétique acoustique. Hormis la classification sensori-motrice de Koopman-van Beinum et van der Stelt (chapitre 1, section I.3.a.), toutes ont été conçues pour l'étude de la *parole* adulte, non pour celle de son développement: elles ne distinguent pas de manière contrôlée les effets de la croissance (chapitre I, section I.2.) de ceux du mouvement des articulateurs sur les patrons acoustiques juvéniles car aucune n'observe directement l'intérieur du conduit vocal.

Pour pallier le manque de caractérisation proprement articulatoire des productions précoces, nous avons donc construit des outils d'analyse par modélisation, en prenant VLAM comme référence dictant l'univers des *possibles* articulatoire-acoustiques des productions vocaliques. Le premier but est de traduire des données acoustiques réelles relevées à 4 et 7 mois, soit avant et au début du babillage canonique, en termes d'*exploration articulatoire simulée* à partir des paramètres de commande de VLAM, en tentant de tenir compte des effets acoustiques dus à la croissance non-linéaire et non-uniforme du conduit vocal. Autrement dit, nous proposons de nouveaux moyens techniques pour estimer les domaines articulatoires à l'origine des espaces d'exploration observés au niveau acoustique dans le cas de la parole en devenir.

Au sortir de l'analyse, il sera possible d'apprécier la validité des résultats obtenus par rapport à ceux de la littérature. En particulier, dans le cas du babillage, nous pourrons confronter des prédictions basées sur le domaine d'exploration articulatoire inféré à 7 mois aux descriptions symbolique et acoustique de ce comportement, au départ du développement du contrôle du cadre syllabique. Plus précisément, ces descriptions s'expriment respectivement (i) par l'observation de *cooccurrences closant-vocant* dans le formalisme transcriptionnel (chapitre 1, section II.1.a.) et (ii) par une répartition différentielle du second formant à l'ouverture de la mâchoire en fonction de sa valeur dans la portion vocalique de la chaîne préverbale (Sussman, *et al.*, 1999; voir aussi chapitre 1, section III, point *iii*).

Dans un premier temps, nous ferons un état des lieux synthétique des méthodes classiques utilisées pour la caractérisation de la production pré-verbale afin de préciser leurs limites au regard de l'étude du développement et de proposer une approche alternative dans le cadre de la modélisation des mécanismes de production périphériques (section I). Comprenant donc la

modélisation comme un outil d'analyse, nous présenterons les données réelles de 4 et 7 mois qui servent de référence à la "paramétrisation" de VLAM pour doter notre agent d'un conduit vocal aux capacités réalistes (section II). Suivra une description des méthodes que nous avons développées pour estimer l'espace d'exploration articulatoire-acoustique avant et au début du babillage avec les résultats auxquels elles mènent (section III). De là, nous essaierons d'en savoir plus sur l'exploration observée au cours du babillage canonique par la confrontation de prédictions de nos outils aux résultats fournis par d'autres méthodes (section IV). Globalement, le but sera de tirer des enseignements de l'écart entre nos prédictions et la réalité, avec le souci à tout moment d'apprécier la cohérence de notre démarche vis-à-vis d'elle-même mais aussi de la réalité décrite par d'autres auteurs avec d'autres approches (section V)

I. COMMENT CARACTERISER LA PRODUCTION DE LA PAROLE EN DEVENIR?

1. Par des méthodes traditionnelles: un manque d'information articulatoire

Traditionnellement, l'étude de la production de la parole *en devenir* se base sur les outils conçus pour l'étude de la parole *adulte*. Trois niveaux d'analyse peuvent être distingués dans l'ordre temporel des phénomènes qui se succèdent dans le processus de production: celui de la physiologie, celui du traitement du signal acoustique et celui de la linguistique. A l'ordre temporel se superpose un ordre spatial du niveau d'observation qui va du cerveau de la personne qui produit à celui de l'expérimentateur/trice, en passant par le milieu matériel qui les sépare. Notamment, le long de ce continuum spatio-temporel, la mesure est d'interprétation de plus en plus complexe, les phénomènes d'aval étant conditionnés par le cumul de ceux d'amont. De plus, si l'on écarte les mécanismes ayant lieu avant la contraction musculaire vu leur accès difficile quelque soit l'âge des sujets humains, la donnée peut être considérée de plus en plus dépendante de l'expérience individuelle de l'observateur/trice.

a. Mesures physiologiques

Les mesures physiologiques vont du suivi de l'activité cérébrale (antérieure à la commande de contraction) à celle de la forme du conduit vocal en passant par l'enregistrement de l'activité musculaire (pour plus de détails, voir Kent, 1997, pp. 303-327). La première ne concernant pas les outils à notre disposition¹, force est de constater le peu d'information publié sur l'exploration précoce réelle en des termes proprement articulatoires.

De nos jours, des raisons éthiques protègent les nourrissons humains de l'exposition aux rayons X ou tout autre procédé invasif, tel que l'électromyographie (EMG) avec insertion d'électrodes ou l'imagerie par résonance magnétique (IRM)². Des raisons pratiques rendent difficiles l'enregistrement des mouvements linguaux via des tétines à capteurs (Maeda communication personnelle). Les seules mesures articulatoires disponibles avant l'âge de 2 ans portent sur des mouvements labiaux et/ou mandibulaires, en utilisant l'enregistrement vidéo de diodes à émission infrarouge (Munhall et Jones, 1998) ou de marqueurs réfléchissants (Green *et al.*, 2000; Green *et al.*, 2002), et sur l'activité de muscles mandibulaires, à l'aide d'électrodes électromyographiques (EMG) de surface (Moore et Ruark,

¹ En particulier, VLAM est un modèle *articulatoire*-acoustique, donc de mécanismes périphériques, et non un modèle simulant des processus ayant lieu dans le système nerveux central.

² De même, l'innocuité de l'utilisation des ultra-sons, qui permettent d'observer l'intérieur du conduit, à des âges précoces reste débattue.

1996). En particulier, les méthodes courantes ne permettent donc pas de mesurer les *configurations linguales* associées aux vocalisations infantiles.

Globalement, l'étude des capacités motrices précoces se voit restreinte à une investigation en aval des activités articulatoires par le biais de méthodes de transcription et de mesures acoustiques (Kent et Miolo, 1995).

a. Mesures perceptives: les transcriptions

Les études de transcription (principalement centrées sur le babillage canonique) utilisent par définition une grille de symboles, le plus souvent l'Alphabet Phonétique International (API), pour coder les productions perçues et étiquetées par des transcrip-teurs/trices adultes.

Cependant, le problème des transcriptions phonétiques est que "l'utilisation de symboles de consonnes et voyelles impliquent le *contrôle* indépendant [du cadre] des segments"³, comme l'ont remarqué Davis et MacNeilage (2002b) (voir aussi Kent et Murray, 1982), tandis que les vocalisations sont plutôt le sous-produit d'un comportement moteur immature qui n'a pas d'interprétation linguistique à proprement parler, ni même d'intention linguistique non-ambiguë. De plus, la perception des catégories phonémiques représentées par l'API peut être biaisée par l'expérience linguistique personnelle de l'auditeur/trice. Par exemple, les anglophones natifs/ves peinent à catégoriser les sons vocaliques nasalisés, par ailleurs fréquents dans les productions infantiles (Beddor et Strange, 1982; Matyear, 1997), vu qu'aucun contraste vocalique n'est basé sur cette caractéristique en Anglais.

Une proposition qui rend la transcription plus indépendante des systèmes phonémiques adultes se trouve dans la classification sensori-motrice de Koopman-van Beinum et van der Stelt (chapitre 1, section I.3.a.), où les vocalisations sont réparties selon leur type de phonation et leur type d'articulation, avec une grille d'encodage des sons vocaliques basée sur des catégories relativement grossières définies le long des dimensions avant-arrière et haut-bas de la cavité orale⁴. Ce système manque néanmoins de précision pour la description des activités articulatoires liées à la production préverbale.

a. Mesures acoustiques

Un moyen d'éviter les limites inhérentes aux méthodes de transcription dans l'étude de la production de la parole en devenir est de centrer l'analyse sur l'onde sonore (eg. Buhr, 1980; Kent et Murray, 1982; Lieberman, 1980; Matyear *et al.*, 1998). La tradition de la phonétique acoustique a permis de penser les configurations articulatoires de la parole en termes de fréquences de résonances (ou formants) extraites de sonagrammes. Les formants sont intrinsèquement liés à la forme du conduit vocal, et leur estimation peut être presque automatisée grâce à l'utilisation d'outils de traitement du signal, tels que le Codage Linéaire Prédicatif (LPC) ou la méthode des cepstres⁵. Ces techniques ont beau être d'une grande utilité pour l'analyse de la parole adulte, elles s'avèrent difficiles d'utilisation sur la parole en devenir⁶.

En effet, la fonction de transfert du conduit vocal est dans ce cas sous-échantillonnée par la fréquence fondamentale élevée des voix de nourrissons, rendant le recueil de telles données relativement imprécis et pauvre en termes de nombre de mesures (voir à ce sujet une proposition d'analyse automatisée de données acoustiques par analyse en composantes

³ "[U]se of consonant and vowel symbols implies "independent control of segments" (Davis et MacNeilage, 2002b, p. 138).

⁴ Rappelons que ces variables n'ont à notre connaissance pas encore été exploitées dans le cadre leur méthode de transcription.

⁵ Voir par exemple Johnson (2003) pour les techniques d'analyse de phonétique acoustique.

⁶ En particulier, la méthode des cepstres est inutilisable.

principales dans van der Stelt, Wempe et Pols, 2003). De plus, l'analyse acoustique ne peut pas mener à une interprétation directement articulatoire des sons, vu que (i) il n'est pas possible de séparer l'effet acoustique de la croissance du conduit vocal (chapitre 1, section I.2.) de celui des mouvements des articulateurs et que (ii) l'estimation de la forme du conduit vocal ou de sa configuration motrice d'entrée en fonction de la sortie acoustique est un **problème mal-posé**⁷. Ainsi, l'étude acoustique seule ne peut-elle pas fournir assez d'information pour délimiter les habiletés articulatoires précoces.

a. Mesures perceptives et acoustiques

Comme l'ont fait remarquer Kent et Murray (1982), la procédure la plus adéquate pour l'étude des capacités articulatoires précoces, à ce jour, est d'exploiter dans le même temps l'analyse acoustique et l'analyse transcriptionnelle, de manière (i) à faire une validation réciproque de la transcription perceptive et de la description acoustique, minimisant ainsi les biais liés à la perception, et (ii) à tirer profit des relations entre l'articulatoire et l'acoustique issues d'investigations chez l'adulte (voir par exemple Buhr, 1980; Lieberman, 1980; Matyear *et al.*, 1998).

Cependant, cette caractérisation reste pour le moins imprécise, puisque le conduit vocal adulte n'est pas le résultat d'une transformation linéaire du conduit du nourrisson. Cette différence soulève le problème de la normalisation vocalique que de nombreux/euses investigateurs/trices tentent de pallier en cherchant une procédure de conversion des mesures acoustiques issues de conduits vocaux juvéniles en équivalents adultes, dans l'espoir de situer les vocalisations précoces dans des espaces vocaliques normalisés de manière cohérente (voir une revue et quelques propositions pour traiter le problème de la normalisation dans Ménard, Schwartz, Boë, Kandel et Vallée, 2002; pour des biais dus à l'influence entre mesures acoustiques et perceptives voir van der Stelt *et al.*, 2003).

1. Par la modélisation articulatoire: proposition d'une stratégie alternative

Les méthodes de mesure traditionnelles ne peuvent contrôler deux particularités de la parole en développement: (i) la production *préverbale* est articulatoire sans s'inscrire dans un système phonémique ou phonétique adulte et (ii) son domaine d'exploration acoustique, voire articulatoire, est conditionné par les effets de la croissance du conduit vocal. Notre projet défend qu'il est possible d'obtenir des informations sur la forme du conduit vocal des productions juvéniles à partir de leur description acoustique tout en respectant au mieux ces deux points fondamentaux, et ce grâce à l'utilisation du modèle articulatoire VLAM. Pour rappel, VLAM est un modèle de production vocalique de morphologie et de capacités articulatoire-acoustiques compatibles avec celles d'un conduit vocal réel qui intègre la croissance non-linéaire et non-uniforme du conduit: son utilisation permet une normalisation des vocalisations précoces.

Tirer une information articulatoire d'une information acoustique met, par définition, face à la difficulté des processus d'inversion propre aux problèmes *many-to-one*⁸. Toutefois, nous verrons que l'on peut inférer un certain nombre de tendances articulatoires importantes à la compréhension de l'émergence de la production de la parole, par des analyses basées sur VLAM et des données formantiques.

⁷ Un **problème mal posé** est un problème qui n'admet pas de solution unique, ou de solution dépendant continuellement de l'entrée (Hadamard, 1923). L'on y trouve les problèmes du type many-to-one (voir section III.2.).

⁸ Voir section III.2.

Plus précisément, notre démarche consiste à *situer* des vocalisations de nourrissons par rapport aux capacités acoustiques et articulatoires du modèle. De là, il est possible d'ajuster globalement les simulations du modèle aux données acoustiques réelles, et de déterminer quelques tendances articulatoires de base, voire leur évolution au cours du développement. Ce problème est posé en termes de probabilité conditionnelle: *quels sont les corrélats articulatoires les plus vraisemblables des sons vocaliques de vrais nourrissons, sachant la valeur de leurs deux premiers formants?* Nous ne prétendons pas donner une réponse tranchée à cette question mais proposer des outils pour extraire certaines connaissances phonétiques, et permettre la formulation d'hypothèses, sur l'exploration articulatoire précoce. Les outils que nous avons construits établissent une analyse globale basée sur VLAM: ils ont été développés pour estimer des patrons majeurs de cette exploration et ont pris les noms de *cadrage acoustique*, *cadrage articulatoire* et *cadrage géométrique*. En attendant de revenir sur le détail de ces méthodes d'analyse par modélisation articulatoire, la section suivante présente les données réelles que nous avons exploitées.

II. AVEC DES DONNEES ACOUSTIQUES DE REFERENCE⁹

Pour jeter quelques lumières sur les configurations articulatoires sous-jacentes aux productions acoustiques pré-verbales, notre attention s'est focalisée sur des vocalisations infantiles produites *avant* et au début du babillage canonique, respectivement aux âges de 4 et 7 mois. En effet, bien que la plupart des études développementales traitent exclusivement des patrons de babillage canonique, les individus sont tout aussi susceptibles d'acquérir une expérience sensori-motrice *avant* le babillage qu'au cours de celui-ci, comme en témoigne le développement de la coordination orolaryngée (chapitre 1, section I.3.b.). Ainsi avons-nous choisi d'interpréter, par rapport aux simulations du modèle, des données réelles de nourrissons recueillies au début du babillage mais aussi avant cette période.

Un corpus pour chaque âge a été sélectionné pour fournir, dans le même temps, une description acoustique (les deux premiers formants) et la transcription phonétique des vocalisations recueillies. Notons que, si nos analyses ont essentiellement besoin de valeurs acoustiques, la cohérence des résultats obtenus doit être évaluée par rapport aux observations de nos prédécesseurs qui ont en grande majorité utilisé des méthodes de transcription. Aussi, le choix de ces données ne peut être basé sur un critère de représentativité statistique. En effet, la littérature dans le domaine accuse la rareté des corpus alliant acoustique et transcription, et nous contraint donc à utiliser le matériel disponible.

Toutefois, ces données proviennent de publications relatives à des études indépendantes de la nôtre, et l'une de l'autre, donnant ainsi une certaine diversité de condition de recueil et une certaine impartialité pour l'évaluation de l'efficacité de nos outils d'estimation du domaine d'exploration précoce au cours du développement de la parole. En effet, la première étude traite de l'imitation vocale précoce notamment à 4 mois; elle vous a été présentée dans la deuxième partie de ce mémoire (voir Kuhl et Meltzoff, 1996, chapitre 3, section III.A.3). La seconde est de Matyear et collègues (Matyear, 1997; Matyear *et al.*, 1998): elle teste la dominance du cadre sur la contribution potentielle du velum dans l'explication des modifications acoustico-phonétiques à l'intérieur des séquences CVC de babillage (pour la dominance du cadre voir chapitre 1, section II.1.a.). Dans le premier cas, la production est plutôt stimulus-induite, ou exogène, tandis que dans le second, elle est plutôt spontanée, ou endogène¹⁰.

⁹ Le titre de cette section ne veut pas être trompeur: le terme de "référence" signifie juste que nous les utilisons comme référence et non qu'elles représentent une référence dans l'absolu.

¹⁰ Nous discuterons des conséquences potentielles de cette différence de contexte expérimental sur nos résultats de simulation dans la dernière section de ce chapitre.

1. Le corpus de pré-babillage

Le premier jeu de données est tiré d'une figure représentant la valeur des deux premiers formants de 45 sons vocaliques produits par 24 sujets âgés de 20 semaines, vivant en milieu anglophone. Ces données ont été publiées dans le *Journal of the Acoustical Society of America* (figure 3 dans Kuhl et Meltzoff, 1996). Les fréquences formantiques ont été extraites d'une copie de l'article à l'aide d'un scanner. Rappelons que l'étude originale correspond à une expérience menée en laboratoire. Elle avait pour objet l'imitation vocale précoce des voyelles /i/, /a/ et /u/, présentées en modalité audio-visuelle sur un écran sous la forme d'un visage et d'une voix adulte¹¹ aux sujets dont les productions subséquentes ont été analysées aux niveaux phonétique et acoustique (à chaque fois que la qualité du signal le permettait). Un aspect important de cette expérience est que le protocole pourrait avoir encouragé les nourrissons à produire des vocalisations au dépens d'autres types de sons, et à explorer dans les limites de leurs capacités, leur répertoire articulatoire-acoustique vu que les 3 cibles correspondent aux extrêmes des configurations articulatoires du conduit vocal¹² (voir chapitre 2, figure 2-4). Le système de transcription utilisé était le système vocalique de l'Anglais mais les auteurs ont regroupé les éléments transcrits en trois catégories : pseudo-/a/, qui comprend [a æ ʌ], pseudo-/i/, correspondant aux voyelles [i ɪ ε], et pseudo-/u/ pour les sons [ʊ u]¹³. Les signaux acoustiques ont été numérisés à une fréquence de 20 kHz, et des valeurs formantiques ont été estimées au croisement d'une analyse par spectrogramme à bande étroite (114 Hz), par une transformée de Fourier rapide (256 points) et par un codage prédictif linéaire (fenêtre de 10 ms et un filtre d'ordre 12). La valeur fréquentielle des deux premiers formants correspond à la moyenne des mesures effectuées sur 5 instants de chaque vocalisation (le début, 1/4, 1/2, 3/4 de la durée et la fin), sachant qu'aucune différence significative n'a été trouvée entre les mesures des points successifs.

1. Le corpus de début de babillage

Les données de Matyear et collègues consistent en des vocalisations spontanées produites par trois nourrissons issus de milieux anglophones. Ces vocalisations ont été audio-enregistrées sur le lieu de vie des sujets du début de leur babillage canonique au début de la période de leurs premiers mots. Ces productions ont été transcrites selon le système de l'API. Le corpus analysé est composé de proto-syllabes CVC, constituées des closants [b d m n g ŋ] redupliqués de part et d'autre des vocants [i ɪ e ε æ ə ʌ a u ʊ o ɔ]. Les portions vocaliques ont été analysées acoustiquement dès que deux transcrip-teurs/trices ont proposé le même étiquetage. Chaque item a été numérisé à une fréquence de 16 kHz. Les fréquences formantiques ont été estimées grâce à un spectrogramme (100 Hz de largeur de bande) qui a permis la sélection d'une portion de 200 ms de la partie stable du vocant à partir de laquelle un

¹¹ Le modèle est dans cette expérience un individu de sexe féminin, mais le genre du visage et de la voix du personnage-stimuli n'est vraisemblablement pas perçu à cet âge: cette caractéristique ne semble pas détectée avant l'âge de 8 mois (Patterson et Werker, 2003).

¹² Rappelons aussi que chaque sujet n'est exposé qu'à une seule cible. Il est difficile de savoir l'incidence de ce point sur notre méthode d'analyse. Toutefois, le paradigme de psychologie sous-jacent à l'étude de Kuhl et Meltzoff a pour principe implicite de considérer que leurs données représentent ce qu'un sujet théorique pourrait avoir produit à lui seul. De plus, la difficulté que représente l'analyse acoustique des vocalisations de cet âge indique qu'un seul sujet aurait eu du mal à produire 45 vocalisations analysables dans le temps imparti à l'expérience pour chaque individu. Nous ne pouvons donc que nous associer à l'idée selon laquelle les données issues des 24 sujets réels fournissent le portrait du comportement d'un seul sujet théorique sur une durée égale à la somme des temps impartis à chaque individu.

¹³ Les données de la publication ne présentent pas le détail de la transcription de départ: nous ne disposons que de l'étiquetage [i] vs. [a] vs. [u].

codage prédictif linéaire à des intervalles de 50 ms a généré un spectre moyenné sur la zone stable.

De ces données acoustiques, nous avons sélectionné toutes les productions relevées à 7 mois sur les deux sujets dont le babillage a commencé à cet âge¹⁴, avec 98 items en tout.

Seules les 7 catégories vocaliques suivantes y figurent: [i ɪ ε æ ə a o]¹⁵.

III. PAR UN AJUSTEMENT DU MODELE ARTICULATORI-ACOUSTIQUE AUX DONNEES ACOUSTIQUES

L'analyse par modélisation articulatoire que nous proposons se décline en trois temps. Le premier projette le domaine d'exploration acoustique observé dans l'espace acoustique complet de VLAM : il s'agit du cadrage acoustique des données. De là, une procédure cherche le sous-espace du domaine d'exploration *articulatoire* totale de VLAM le plus probable sachant les données : on parle de cadrage articulatoire. La troisième étape utilise ce sous-espace pour donner à chaque vocalisation une description plausible de la forme de conduit vocal qui l'a émise : c'est le cadrage géométrique. Schématiquement, le long de ce plan d'analyse, les données acoustiques réelles sont traduites successivement en termes d'acoustique simulée, de distribution de variables articulatoires puis de distribution de coupes sagittales, permettant le passage d'un niveau acoustique à une description phonétique sans faire appel à un système de transcription.

1. Cadrage acoustique

Si l'on veut interpréter les productions acoustiques des nourrissons en des termes acoustiques et articulatoires adultes, l'étape préliminaire est de *normaliser* ces sons, c'est-à-dire de trouver leur position dans un système acoustique de référence équivalent à ce qui est nommé « espace vocalique » chez l'adulte, en prenant en compte les différences morphologiques entre le conduit vocal infantile, à un âge donné, et le conduit adulte, indépendamment de tout système linguistique. Cette normalisation peut être obtenue en se focalisant sur les contraintes physiques par lesquelles les configurations vocaliques possibles du conduit sont limitées. Ces contraintes comprennent au moins les faits suivants: (i) les axes de mouvement des articulateurs trouvent leurs bornes aux parois rigides du conduit vocal, (ii) les structures articulatoires ont des limites biomécaniques, telles que le degré avec lequel la mandibule peut être abaissée, et (iii) l'aire de la plus petite constriction le long du conduit vocal (lèvres incluses) doit être supérieure à une certaine valeur pour vérifier la/les condition/s de vibration des cordes vocales et/ou d'écoulement laminaire (voir Pélorsson, 2001). L'espace acoustique résultant de ces contraintes correspond à la partie vocalique des *sons possibles de parole*, notion qui a été proposée par Lindblom et collègues (Liljencrants et Lindblom, 1972; Lindblom, 1986; Lindblom, 1990).

a. Méthode

Dans la présente étude, VLAM est utilisé en tant que modèle de production vocalique oral qui simule les propriétés acoustiques et biologiques d'un conduit vocal humain pris comme standard¹⁶. Pour un âge spécifié, toutes les configurations articulatoires de VLAM donnent

¹⁴ Le troisième individu a commencé à babiller à 9 mois.

¹⁵ [i] : *fr.* vie; [ɪ]: *ang.* because; [æ] : *ang* cat ; [a] : *fr.* capable; [ε] : *fr.* bêlement; [ø] : *fr.* demi ; [o] : *fr.* repos. Sont considérées **antérieures**: [i ɪ ε æ]. Sont **centrales**: [ə a]. Est **postérieure**: [O].

¹⁶ Ici, "standard" est à prendre au sens où bien que le modèle soit basé sur la morphologie d'une locutrice donnée, le modèle est pris comme représentant par défaut les capacités du conduit vocal d'un sujet théorique.

lieu à des sons qui s'inscrivent tous dans un espace acoustique qui a été nommé **Espace Vocalique Maximal (EVM)** (Boë, Perrier, Guérin et Schwartz, 1989). L'EVM correspond à ce qu'un-e enfant de même âge serait capable d'articuler si il/elle utilisait le jeu complet de commandes articuloire de VLAM. L'EVM représente donc tous les sons vocaliques oraux de parole projetés dans l'espace des formants ($F_i, i=1..n$). Le plan (F_1, F_2) décrit le triangle vocalique que des phonéticien-ne-s utilisent dans l'étude de la parole adulte: ses coins comprennent les voyelles [i a u] (voir chapitre 2, figure 2-4).

Le cadrage acoustique consiste en la superposition du jeu de vocalisations réelles produit à 4 et 7 mois sur l'EVM de 4 et 7 mois, respectivement¹⁷. Pour générer ces deux EVM, la valeur des paramètres de commandes P_i est tirée aléatoirement d'une distribution uniforme ($-3 < P_i < +3$ écart-types ($e-t$), pour LP, LH, J, TB, TD, TT, et $-1 < L_x < +1 e-t$) et fournie en entrée de VLAM aux âges d'intérêt. Les aires minimales intra-orale et inter-labiale choisies sont de 0.1 cm^2 et 0.01 cm^2 , respectivement: ces seuils ont été utilisés par Ménard *et al.* (2002) pour la simulation de la production néonatale. Un total d'environ 25 000 couples de valeurs formantiques, correspondant à des simulations satisfaisantes, ont été obtenues pour chacun des âges testés¹⁸.

Il est à noter que cette procédure permet non seulement d'effectuer une normalisation graphique des productions précoces, mais aussi de vérifier que les vocalisations à notre disposition appartiennent bien aux capacités acoustiques de VLAM au même âge. Ceci est un pré-requis à l'utilisation de VLAM comme base à l'inférence de "modèles" de production préverbale plausibles. Pour chaque période développementale, si cette condition obligatoire est vérifiée, l'exploration acoustique du groupe de vocalisations est caractérisée et le cadrage articuloire peut être effectué.

b. Résultats

Chaque jeu de vocalisations réelles s'inscrit dans l'EVM de même âge¹⁹ (Figure 4-1 et 4-2). De plus, les données réelles ne couvrent pas toute l'aire acoustique qu'elles auraient si les individus avaient utilisé toutes les configurations articuloires possibles. Plus précisément, avant le babillage canonique, comme le montre la Figure 4-1, les vocalisations tendent à se grouper au cœur du triangle, autour de la position neutre, définie par la configuration dont tous les paramètres articuloires de VLAM sont de valeur nulle. Dans la terminologie phonétique, par rapport au triangle vocalique acoustique de VLAM, les données de 4 mois correspondent à des réalisations plutôt centrales, et mi-hautes à mi-basses: les productions les plus antérieures, les plus postérieures et les plus ouvertes ne sont pas exploitées²⁰. Dans le cas du début de babillage, illustré Figure 4-2, les productions vocaliques explorent plus la partie inférieure de l'espace acoustique qu'à la période développementale précédente: ici, la dimension verticale (le long de F_1) semble prédominante dans l'explication de la dispersion acoustique.

¹⁷ La taille du conduit vocal des sujets réels n'étant pas accessible, nous ne pouvons que prendre en compte leur âge, bien qu'il existe en réalité une variabilité interindividuelle de la croissance du conduit vocal.

¹⁸ Plus précisément, l'EVM à 4 mois comprend 24660 simulations, celui à 7 mois 24971.

¹⁹ La/le lectrice/teur remarquera qu'une vocalisation se trouve légèrement en dehors de l'EVM à 7 mois. Cela peut être dû à des imprécisions dans la mesure des formants et/ou à nos choix heuristiques de modélisation. Dans ce dernier cas, l'on retiendra que le son aurait pu être inclus dans l'EVM si le domaine de variation des variables articuloires prises en compte avait été légèrement plus étendu que l'intervalle $[-3, +3] e-t$.

²⁰ Ces zones correspondent donc aux trois coins du triangle formé par l'EVM.

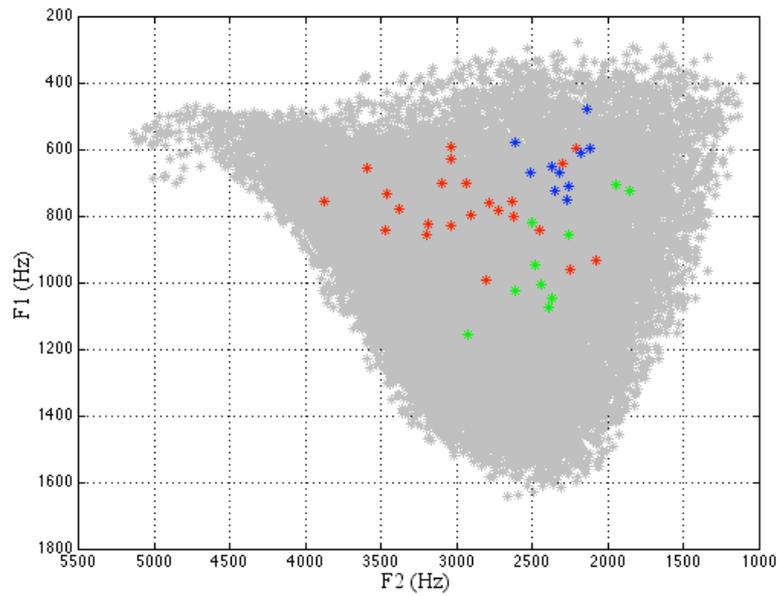


Figure 4-1: Représentation graphique du résultat du cadrage acoustique des données de pré-babillage (4 mois) dans le plan (F1, F2), exprimé en Hertz. Les points gris correspondent à l'EVM à 4 mois, les points de couleurs correspondent aux vocalisations réelles en fonction de leur étiquetage ([i] en rouge, [a], en vert, et [u], en bleu).

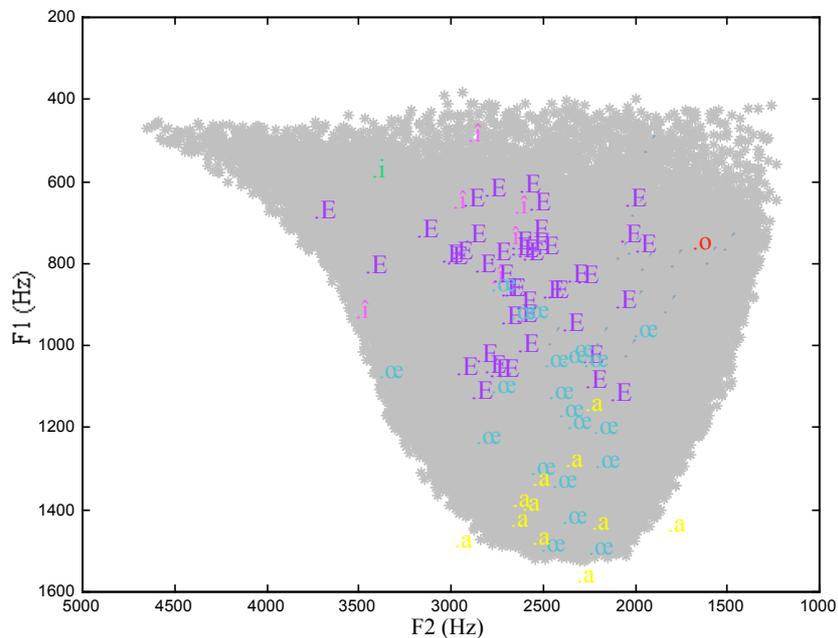


Figure 4-2: Représentation graphique du résultat du cadrage acoustique des données de début de babillage dans le plan (F1, F2), exprimé en Hertz. Les points gris correspondent à l'EVM à 7 mois, les points de couleurs correspondent aux vocalisations réelles en fonction de leur étiquetage, avec une couleur pour chaque classe phonétique.

2. Cadrage articulatoire

Une manière de décrire l'espace articulatoire qui pourrait rendre compte de l'exploration acoustique des sujets est de chercher la combinaison de paramètres de commande motrice de VLAM qui génère un domaine d'exploration acoustique aussi proche que possible de la zone acoustique occupée par les vocalisations réelles dans l'EVM. Cette technique essaie de dépasser un problème *a priori* insoluble, celui de l'inversion de l'acoustique à l'articulatoire dû aux relations $n:1$ (ou **problème many-to-one** ou **inverse**) entre les domaines articulatoire et acoustique en production de la parole, où *un* son peut être issu de n configurations articulaires, ou encore, de n combinaisons de valeurs motrices possibles (voir par exemple Atal, Chang, Mathews, et Tukey, 1978; Boë, Perrier et Bailly, 1992; Ouni et Laprie, 2005). Cependant, l'on peut essayer de caractériser de manière globale le domaine articulatoire sous-jacent à un corpus de vocalisations infantiles par rapport aux capacités de VLAM.

Dans un premier temps, la méthode et les résultats obtenus à 4 et 7 mois vous seront présentés. Ensuite, VLAM étant un modèle de production de sons oraux alors que des données du corpus de babillage sont potentiellement nasalisées, nous tenterons d'apprécier l'incidence de la non-prise en compte du phénomène de nasalisation dans les productions de babillage sur les résultats de la première analyse par une série de simulations subsidiaires.

a. Méthode

D'un point de vue probabiliste, le cadrage articulatoire cherche le jeu de paramètres articulaires (ou **sous-modèle**) le plus probable d'après la distribution des réalisations acoustiques simulées qui lui sont associées (distribution théorique), sachant la distribution des vocalisations réelles des sujets au même âge (distribution réelle) dans le plan des deux premiers formants. En d'autres termes, le but est de trouver le sous-modèle qui maximise la probabilité $P(M_i/D)$, où M_i est le $i^{\text{ème}}$ sous-modèle, caractérisé par la distribution des sorties acoustiques qu'il a générées, tandis que D représente la distribution des données réelles dans l'espace formantique²¹. Cette technique nécessite (i) la définition des sous-modèles articulaires qui entrent en compétition les uns avec les autres, (ii) la génération de leur espace acoustique de réalisation à chaque âge ainsi que le calcul de la distribution statistique correspondante dans le plan (F1, F2), et (iii) la sélection du meilleur sous-modèle sachant la distribution statistique des vocalisations infantiles de même âge.

Quatre paramètres articulaires ont été sélectionnés pour la mise en œuvre de la procédure: LH qui contrôle l'ouverture-fermeture des lèvres, J qui commande l'ouverture de la mandibule, et TB et TD qui déterminent ensemble la configuration de la langue selon les dimensions verticale et horizontale du conduit. Ces paramètres sont les plus pertinents vis-à-vis de l'acoustique de la parole pour des raisons liées. En premier lieu, J, TB et TD rendent compte de l'essentiel de la variance observée dans les profils de langue ayant servi de base de données à l'analyse statistique dont VLAM est issu: ces axes factoriels ont expliqué 81% de l'inertie totale (Tableau 4-I). De plus, ce sont les commandes majeures de la détermination de la position, X_c , et de l'aire, A_c , de la constriction intra-orale minimale : ces variables dépendent de la forme de la langue. La valeur de l'aire aux lèvres, A_l , résulte quant à elle des effets combinés de LH et J. Or, X_c , A_c , et A_l sont les corrélats principaux des fréquences formantiques²² (Fant, 1960; Boë *et al.*, 1995b). Trois paramètres de commande ne sont pas pris en compte dans le cadrage articulatoire que nous avons mis en œuvre : la variable de protrusion labiale, LP, celle de la pointe de la langue, TT, et celle de la hauteur du larynx, Lx.

²¹ La démarche utilisée ici pour traiter le problème inverse emprunte donc un cadre probabiliste de type Bayésien.

²² Des auteurs les proposent pour paramètres de *contrôle* des productions vocaliques (voir Boë *et al.*, 1992).

Cette restriction s'appuie sur le fait qu'elles jouent un rôle mineur dans l'acoustique des voyelles en particulier, et que leur absence rend l'interprétation des résultats plus claire.

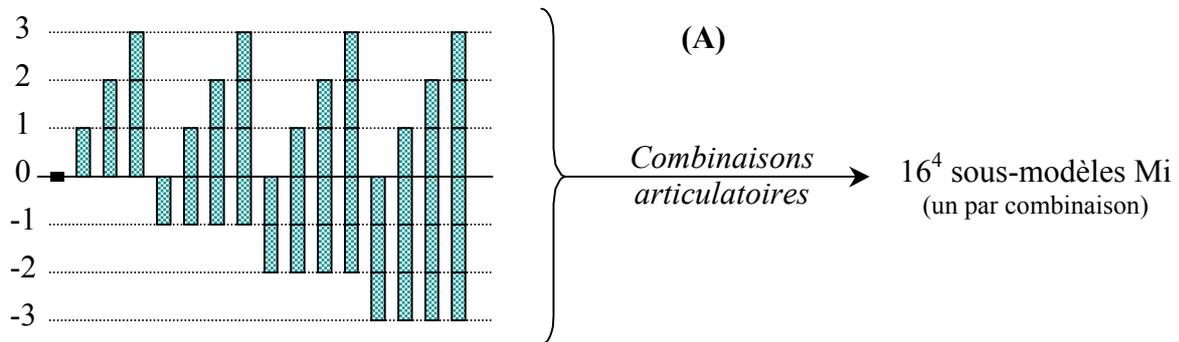
Tableau 4-I: Pourcentage de l'inertie totale expliquée par chaque facteur dans l'ACP de Maeda (données extraites de Maeda, 1990, sujet

Facteur	J	TB	TD	TT	Lx
Pourcentage de l'inertie totale	15	43	23	7	4

Pour chacun des quatre paramètres entrant en jeu (LH, J, TB et TD), l'intervalle de variation va de nul à $[-3 ; +3] e^{-t^{23}}$. Cet intervalle inclut toujours la valeur 0, i.e. celle de la position neutre des articulateurs de VLAM, et ses bornes varient par pas de un écart-type. Ainsi les domaines de variation utilisés sont de la forme $[-a ; +b] e^{-t}$, où a et b sont des entiers positifs ou nuls et inférieurs ou égaux à 3 e^{-t} . La valeur des paramètres articulatoires non-sélectionnés (LP, TT et Lx) est fixée à 0. L'on obtient donc en tout 16^4 , soit 65536, sous-modèles: pour chacune des 4 commandes, 16 intervalles sont testés avec 4 valeurs possibles pour a et 4 pour b^{24} . De là, pour chaque corpus de sons vocaliques infantiles, D, le sous-modèle Mi qui maximise la probabilité $P(\text{Mi}/D)$ est sélectionné parmi les 65536 possibles (voir annexe A1 pour plus de détails sur le calcul de $P(\text{Mi}/D)$). L'organigramme complet du cadrage articulatoire est fourni Figure 4-3.

²³ L'intervalle $[-3 ; +3] e^{-t}$ correspond à plus de 99% des valeurs possibles du paramètre dans la base de données dont SMIP est issu.

²⁴ En clair, les intervalles de valeur sont les 16 suivants: $[-3 ; 3]$, $[-3 ; 2]$, $[-3 ; 1]$, $[-3 ; 0]$, $[-2 ; 3]$, $[-2 ; 2]$, $[-2 ; 1]$, $[-2 ; 0]$, $[-1 ; 3]$, $[-1 ; 2]$, $[-1 ; 1]$, $[-1 ; 0]$, $[0 ; 3]$, $[0 ; 2]$, $[0 ; 1]$, $[0 ; 0] e^{-t}$. De plus, à l'intérieur d'un intervalle donné, la valeur de chaque paramètre de commande varie par pas de 0.2 e^{-t} . Cela revient à échantillonner chaque dimension articulatoire de manière régulière, pour contrôler l'effet de la distribution des variables motrices sur le classement des sous-modèles. De plus, comme dans le cas de la génération de l'EVM, la valeur minimale de Ac est de 0.1 cm^2 , et celle de Al de 0.01 cm^2 .



Domaines de variation de J, TB, TD et LH
(LP et Lx fixé à 0)

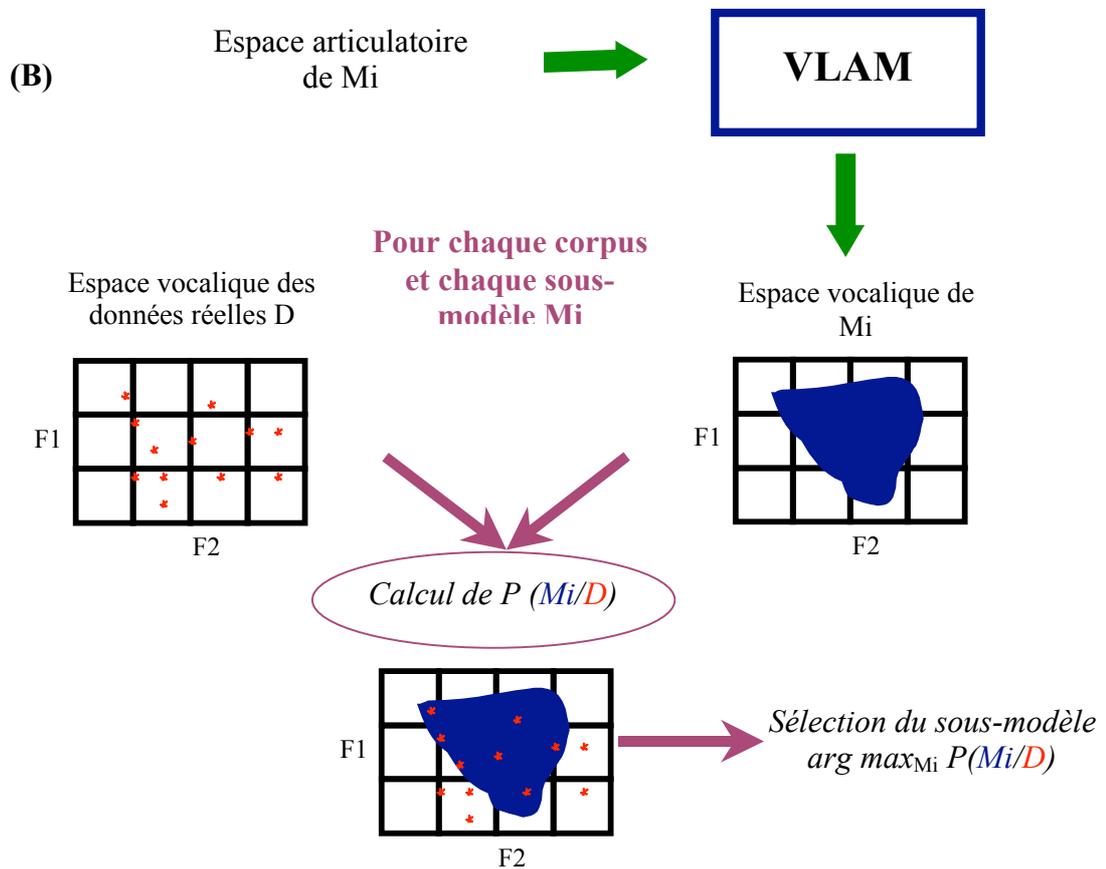


Figure 4-3: Organigramme du cadrage articulatoire. (A) Génération des espaces articulatoires de chaque sous-modèle M_i . (B) Génération de l'espace acoustique correspondant à 4 et 7 mois, calcul de la probabilité de chaque sous-modèle sachant les données réelles, puis sélection du meilleur sous-modèle.

b. Résultats

Considérons tout d'abord les résultats obtenus sur le corpus de 4 mois. Les figures 4-4 et 4-5 montrent la superposition des données réelles sur l'espace d'exploration acoustique des modèles à 1 et 2 paramètres non-nuls ayant pour intervalle de valeur $[-3 ; +3]$ *e-t*. Les graphiques correspondant aux cas où une seule dimension articulaire varie (Figure 4-4) donnent la sensibilité des paramètres acoustiques aux variations de la valeur de la commande autour de la configuration neutre (Cf. notion de *macrosensibilité*, Majid, 1986). Ces macrosensibilités sont d'amplitude globalement équivalente d'un paramètre à l'autre, bien qu'elles se répartissent de manière différente entre les formants F1 et F2 (voir aussi Ménard *et al.*, 2004). En considérant l'ensemble des graphiques des figures 4-4 et 4-5, il apparaît que deux paramètres articulatoires ne suffisent pas à reproduire, ou *cadrer*, l'étendue de l'exploration acoustique des données réelles à 4 mois.

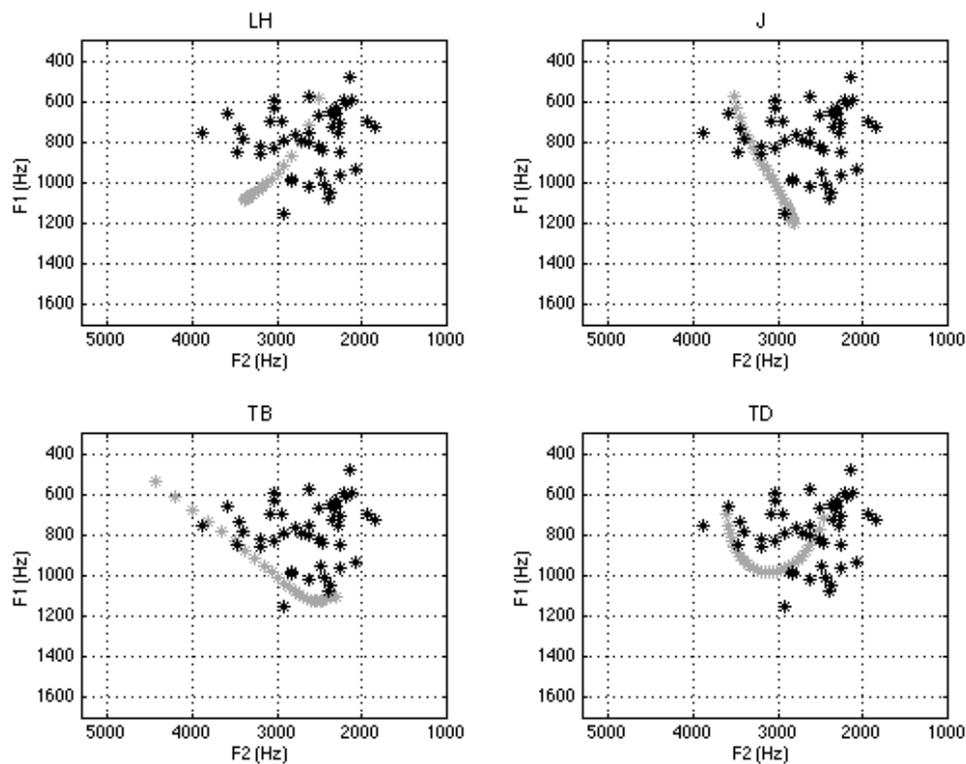


Figure 4-4: Représentation graphique du cadrage articulaire des vocalisations de 4 mois, avec les sous-modèles à 1 paramètre articulatoire non-nul variant de -3 à 3 *e-t*, par pas de $0,2$ *e-t*.

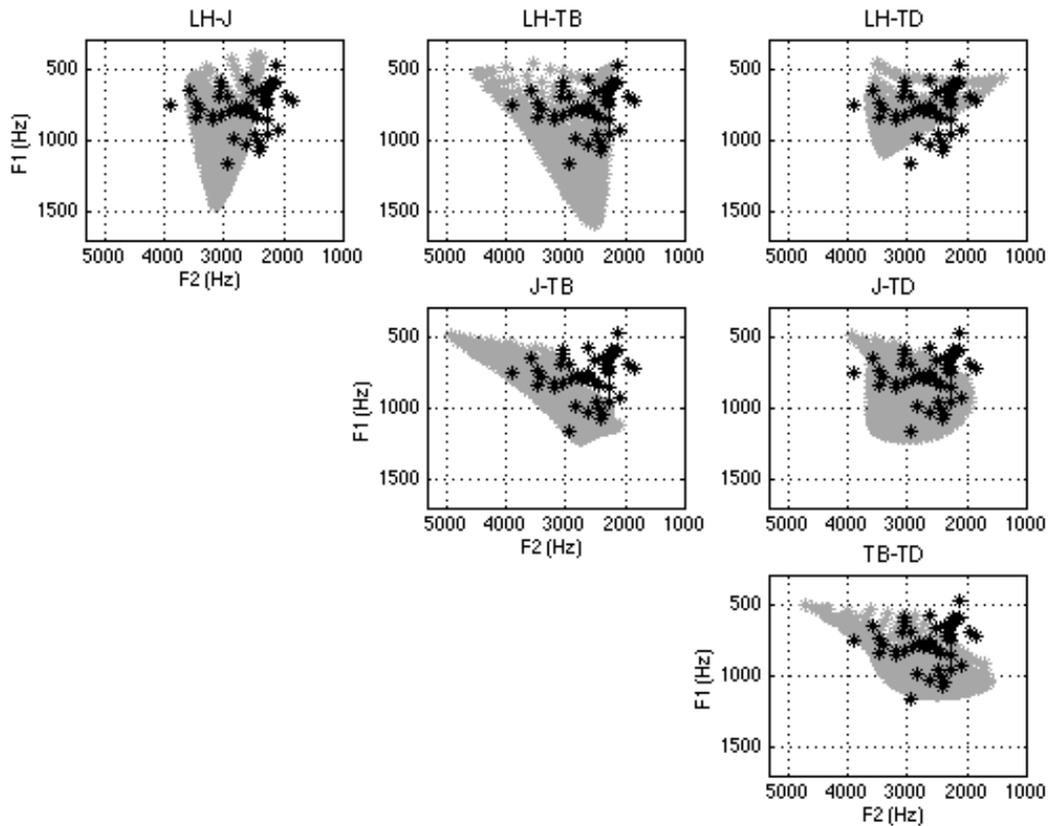


Figure 4-5: Représentation graphique du cadrage articulatoire des vocalisations de 4 mois, avec les sous-modèles à 2 paramètres articulatoires non-nuls variant de -3 à 3 $e-t$, par pas de $0,2$ $e-t$.

Les figures 4-6 et 4-7 illustrent le cadrage de ces données par le meilleur sous-modèle, en termes de maximisation de $P(Mi/D)$, parmi ceux ayant respectivement 3 et 4 paramètres. Les résultats montrent que les données de 4 mois nécessitent un modèle d'au minimum 3 dimensions articulatoires non-nulles pour être simulées par VLAM, dont au moins une commande pour la langue (TB ou TD, ou les deux ; voir annexe 2). Plus précisément, quand TB et TD sont tout deux présents²⁵, leur exploration favorise les valeurs négatives de TB et positives de TD²⁶. Le meilleur sous-modèle à 3 commandes non-nulles montre des domaines de variation de LH, J TB et TD respectivement de $[-2; 1]$, $[0; 0]$, $[-3; 1]$ et $[-1; 3]$ $e-t$: il ne comprend pas de variation sur le paramètre mandibulaire (J).

En effet, parmi les sous-modèles à 3 dimensions, ceux où J est non-nulle ont au moins une vocalisation à l'extérieur de leur domaine d'exploration acoustique, menant à des valeurs de la probabilité $P(Mi/D)$ relativement faibles²⁷. Lorsque les 4 dimensions sont en jeu, le domaine de variation des commandes est très variable. Dans ce cas, les compensations

²⁵ Cela concerne les quelques 150 premiers modèles à 3 dimensions non-nulles.

²⁶ Schématiquement, les valeurs négatives de TB correspondent à un recul de la masse de la langue en direction de la paroi dorsale du pharynx inférieur, tandis que les valeurs positives de TD arquent la masse de la langue en rapprochant son milieu de l'arrière du palais dur.

²⁷ Ils apparaissent dans le classement par ordre croissant de $P(Mi/D)$ à des positions supérieures à 150.

articulatoires ne permettent pas de dégager un patron d'exploration aux tendances aussi stables que celles des modèles à 3 commandes non-nulles.

Pour chaque sous-modèle, il est possible d'estimer l'étendue globale de l'exploration articulatoire en calculant le rapport entre le nombre de configurations motrices générées par le sous-modèle et le nombre de configurations produites par le jeu à 4 dimensions complet, i.e. si chacun des 4 paramètres articulatoires inclus dans l'analyse varie systématiquement entre -3 et +3 *e-t*. Dans l'ensemble, il s'avère que les meilleurs sous-modèles à 3 et 4 commandes non-nulles exploitent de 8 à 13% de l'espace articulatoire possible²⁸. En d'autres termes, le volume total d'exploration articulatoire du corpus de 4 mois prend de 8 à 13% du volume à 4 dimensions disponible quand J, LH, TB, TD varient selon leur intervalle maximum [-3 ; +3] *e-t*.

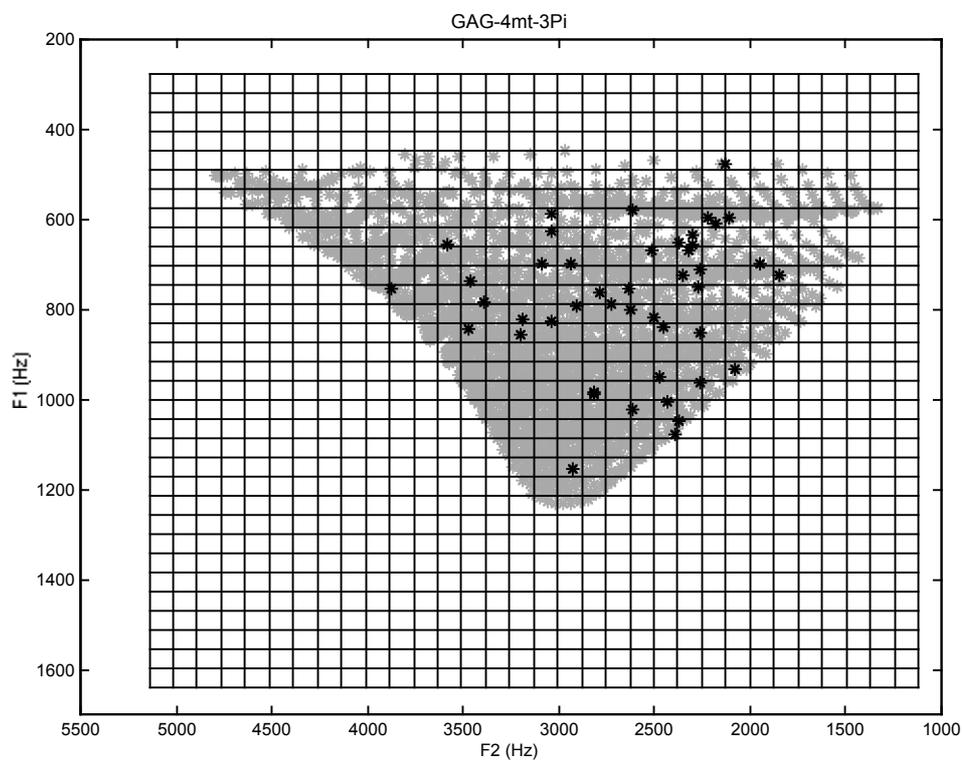


Figure 4-6: Représentation graphique du cadrage articulatoire des vocalisations de 4 mois, avec le meilleur sous-modèle à 3 paramètres articulatoires non-nuls.

²⁸ Cette proportion a été estimée sur le sous-modèle à 4 paramètres car le nombre de simulations réalisées par la combinaison de 3 commandes n'est pas comparable à celui de la combinaison de 4 dimensions.

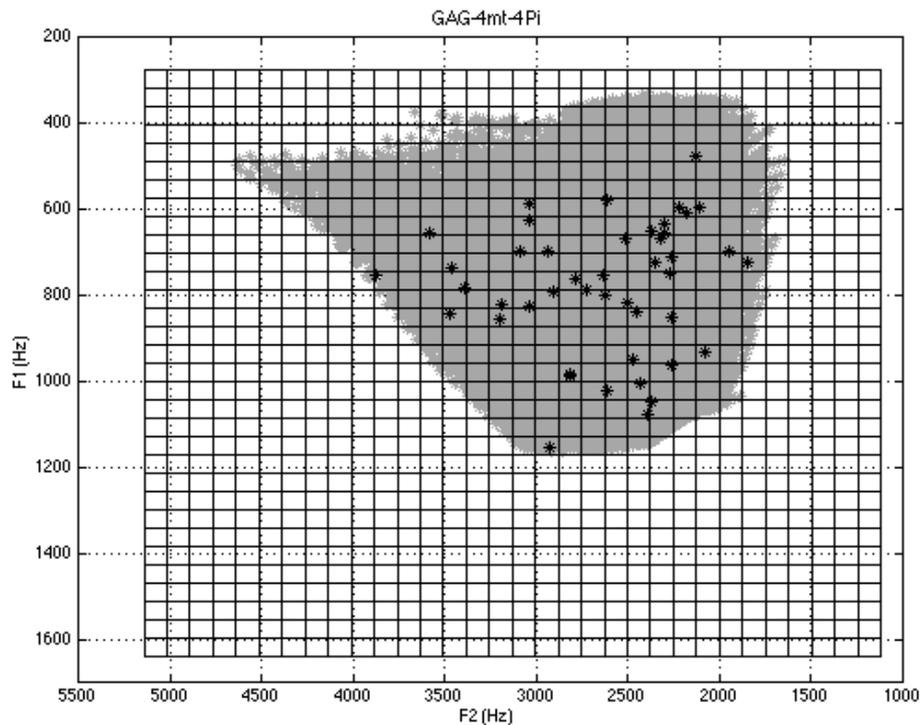


Figure 4-7: Représentation graphique du cadrage articulatoire des vocalisations de 4 mois, avec le meilleur sous-modèle à 4 paramètres articulatoires non-nuls.

Dans le cas du corpus de 7 mois, vu que l'exploration acoustique est plus étendue que celle à 4 mois, l'on s'attend à ce que plus de commandes et/ou des intervalles de variation plus larges soient nécessaires à leur cadrage articulatoire. En effet, les simulations indiquent que 3 dimensions ne suffisent pas à reproduire le domaine d'exploration acoustique des données de babillage (voir annexe A3 pour des figures et annexe A4 pour plus de détails). La figure 4-8 illustre le cadrage de ces données par le meilleur sous-modèle, en termes de maximisation de $P(\text{Mi}/D)$, donc parmi ceux à 4 dimensions non-nulles. Dans ce sous-modèle, la mandibule est la seule commande qui exploite toute l'étendue de l'intervalle $[-3 ; +3]$ *e-t*. De plus, les domaines de variation de LH, TB et TD sont respectivement de $[-1; 2]$, $[-1; 3]$ et $[-1; 3]$.

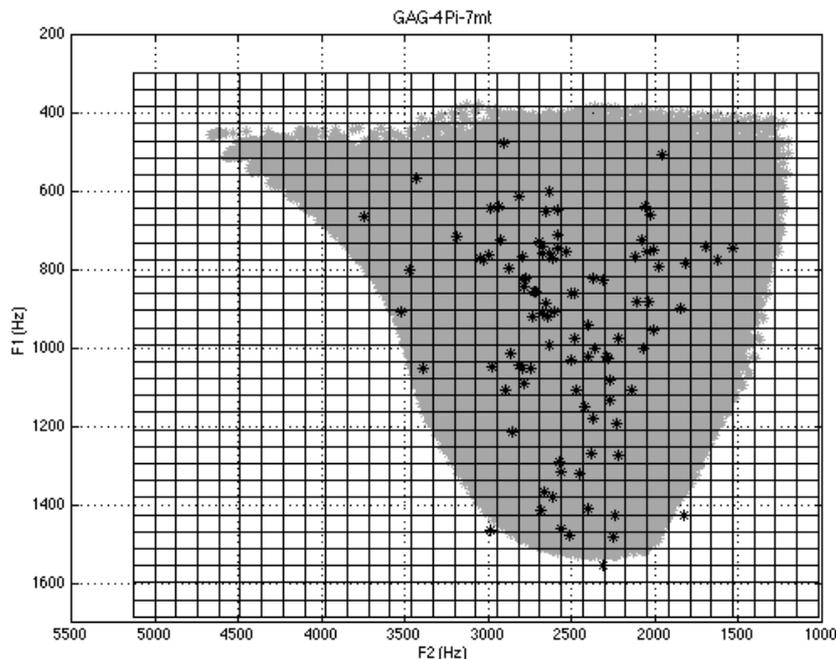


Figure 4-8: Représentation graphique du cadrage articulaire des vocalisations de 7 mois, avec le meilleur sous-modèle qui est de 4 paramètres articulatoires non-nuls.

Ainsi nos simulations montrent-elles que les paramètres utilisés par les meilleurs sous-modèles à 3 dimensions de 4 mois tendent à être aussi utilisés par ceux à 4 dimensions de 7 mois, bien qu'avec des domaines de variation quelque peu différents²⁹, et que la modification radicale de l'exploration est introduite par une exploitation maximale du domaine de variation de la commande mandibulaire.

Dans l'ensemble, les meilleurs sous-modèles à 4 dimensions de la période de babillage explorent de 36 à 58% du volume articulaire qui leur est possible, soit sensiblement plus qu'à 4 mois.

c. Quelques propositions pour traiter le cas de la nasalisation à 7 mois

Nous avons vu au chapitre 1 section III.1.a., qu'une séquence de babillage donnée peut être vue comme une forme statique, ou presetting, du conduit vocal à laquelle (seule) l'oscillation mandibulaire se superpose. En particulier, la position du velum restant statique au cours du cycle mandibulaire, l'on prévoit que les vocants produits en contexte de closant nasal sont eux-mêmes nasalisés: la position basse du velum serait un élément constitutif du presetting. Matyear et collègues (Matyear, 1997; Matyear *et al.*, 1998) ont trouvé des éléments en faveur de cette hypothèse sur la totalité du corpus qu'elles ont recueilli du début du babillage aux premiers mots (Section I.2.), vu que les sons vocaliques tendent à avoir "une fréquence de premier formant plus faible pour les voyelles ouvertes et une fréquence de second formant

²⁹ En particulier, la langue semble explorer des configurations plus antérieures à 7 mois qu'à 4 mois, avec des valeurs de TB qui tendent à être positives.

plus faible pour les voyelles antérieures³⁰ (Matyear *et al.*, 1998, p. 1) quand ces vocants se trouvent en environnement nasal que lorsque leur environnement est oral. Jusque là, notre étude traite les sons vocaliques de 7 mois, issus de la même base de données que Matyear et collègues, comme étant purement oraux, en accord avec la transcription qu'elles en ont faite, bien qu'environ 68% de ces productions aient eu lieu en contexte nasal. Or, la variable motrice dont l'absence du jeu de paramètres articulatoires de VLAM peut être remarquée est celle de la position du vélum.

La présente analyse tente d'estimer l'incidence d'une nasalisation potentielle sur les résultats de notre analyse par cadrage articulatoire.

Principe

Une première analyse aurait pu consister à introduire explicitement le velum dans nos simulation. Hélas, ceci est techniquement impossible. En effet, la cavité nasale n'est pas représentée dans VLAM à cause (i) du manque de données sur la croissance de cette partie du conduit vocal et (ii) des problèmes soulevés par la nasalisation au niveau de sa simulation acoustique. En ce qui concerne le second point, l'on notera que l'introduction d'une paire pôle-zéro, qui est une caractéristique acoustique des voyelles nasalisées, dans la fonction de transfert d'une voyelle par ailleurs orale rend la définition des fréquences formantiques difficile (Rossato, 2000). Un moyen de contourner ces limites de modélisation acoustique aurait été d'analyser uniquement les vocalisations ayant lieu en environnement oral. Cette solution a été écartée en particulier parce que [ə] n'est jamais entouré de closants oraux dans les données de 7 mois: le cadrage articulatoire n'aurait pas pris en compte les configurations articulatoires correspondantes. Toutefois, vu que [ə] est la classe vocalique qui montre le moins de différences acoustiques entre les contextes oral et nasal (Matyear *et al.*, 1998), les profils formantiques des sons qui lui sont affiliés en environnement nasal peuvent être pris comme une estimation satisfaisante de leur homologue en environnement oral. Pour apprécier l'effet de l'absence de commande vélaire sur les résultats inférés à partir des données de babillage à 7 mois, nous avons décidé de générer des corpus constitués soit de vocants uniquement oraux, et de [ə], soit de vocants "oralisés", c'est-à-dire pour lesquels nous avons entrepris de remplacer la valeur éventuellement nasalisée par une valeur corrigée, susceptible de prendre en compte les écarts formantiques dus à la position basse du velum.

Méthode

Deux jeux de données ont été constitués. Le premier, ou *corpus restreint*, comprend tous les sons vocaliques trouvés en environnement oral dans le corpus de base ainsi que les sons étiquetés [ə]. Le second, ou *corpus modifié*, inclut tous les sons vocaliques trouvés en environnement oral dans le corpus de base ainsi que les réalisations acoustiques ayant lieu en contexte nasal après qu'ils ont été modifiés de manière à ce que la moyenne de chacun des deux premiers formants de leur classe phonétique en situation nasale soit la même que celle de cette classe en situation orale³¹.

³⁰ "[A] lower frequency first formant region in low vowels and a lower frequency second formant region in front vowels".

³¹ Cette procédure est une sorte d'"oralisation" des sons vocaliques potentiellement nasalisés. Nous faisons ici l'hypothèse que, pour chaque classe phonétique, la variance des données acoustiques est la même en environnement nasal qu'en environnement oral. De plus, quand des catégories phonétiques n'ont pas lieu en contexte oral, leurs sons ne sont pas transformés, comme dans le cas du schwa. Les calculs de l'"oralisation" sont effectués sur les fréquences formantiques exprimées en Hertz.

Toutefois, la durée de l'analyse par cadrage articulatoire est relativement élevée³². Pour écourter ce temps de calcul, les cadrages articulatoires exposés ici sont tels que seuls les domaines de variation symétriques des paramètres de commandes non-nuls sont pris en compte, i.e. [-3; 3], [-2; 2], [-1; 1] *e-t*, au lieu des 16 servant à la spécification de l'androïde (Section III.2.a.). L'appréciation de l'incidence de la non-prise en compte de la nasalisation par VLAM s'appuie sur la comparaison des résultats obtenus par ce cadrage articulatoire partiel sur le corpus de base, le corpus restreint et le corpus modifié.

Résultats

La figure 4-9 montre l'issue de ces expériences subsidiaires. Les panneaux 4-9A, 4-9B et 4-9C correspondent respectivement aux jeux de vocalisations de base, restreint et modifié. Le Tableau 4-II fournit les attributs des gagnants correspondants. En comparaison avec les résultats obtenus sur les données originales (4-9A), les deux alternatives (4-9B et 4-9C) mènent à une utilisation légèrement plus large des paramètres articulatoires pour recouvrir les sons vocaliques "réels". Ceci est dû au fait que la proportion de vocalisations périphériques³³ dans ces deux derniers jeux de données est plus élevée que dans le jeu original, favorisant ainsi les sous-modèles ayant, au sein de leur espace d'exploration acoustique, les fréquences de simulations périphériques les plus importantes. Toutefois, l'exploration articulatoire des sous-modèles pour les jeux restreint et modifié reste proche de celle trouvée pour le jeu de base. De plus, celui-ci s'avère faire partie des cinq meilleurs sous-modèles dans le cadrage articulatoire des jeux de données restreint et modifié. En somme, d'après ce test, l'absence de commande vélaire dans VLAM ne semble pas avoir de conséquence importante sur notre estimation des corrélats articulatoires des vocalisations de babillage à 7 mois.

Tableau 4-II: Domaines de variation des paramètres articulatoires des meilleurs sous-modèles issus du cadrage articulatoire partiel des corpus de base, restreint et modifié.

Paramètre articulatoire	LH	J	TB	TD
Corpus de base	[-2; 2]	[-3; 3]	[-3; 3]	[-2; 2]
Corpus restreint	[-3; 3]	[-3; 3]	[-3; 3]	[-2; 2]
Corpus modifié	[-3; 3]	[-3; 3]	[-3; 3]	[-3; 3]

³² Le temps de calcul nécessaire au cadrage articulatoire de chaque corpus peut être d'environ 100 heures, pour 16 intervalles de variation possibles.

³³ C'est-à-dire vers les côtés gauche et droit du triangle du domaine d'exploration acoustique.

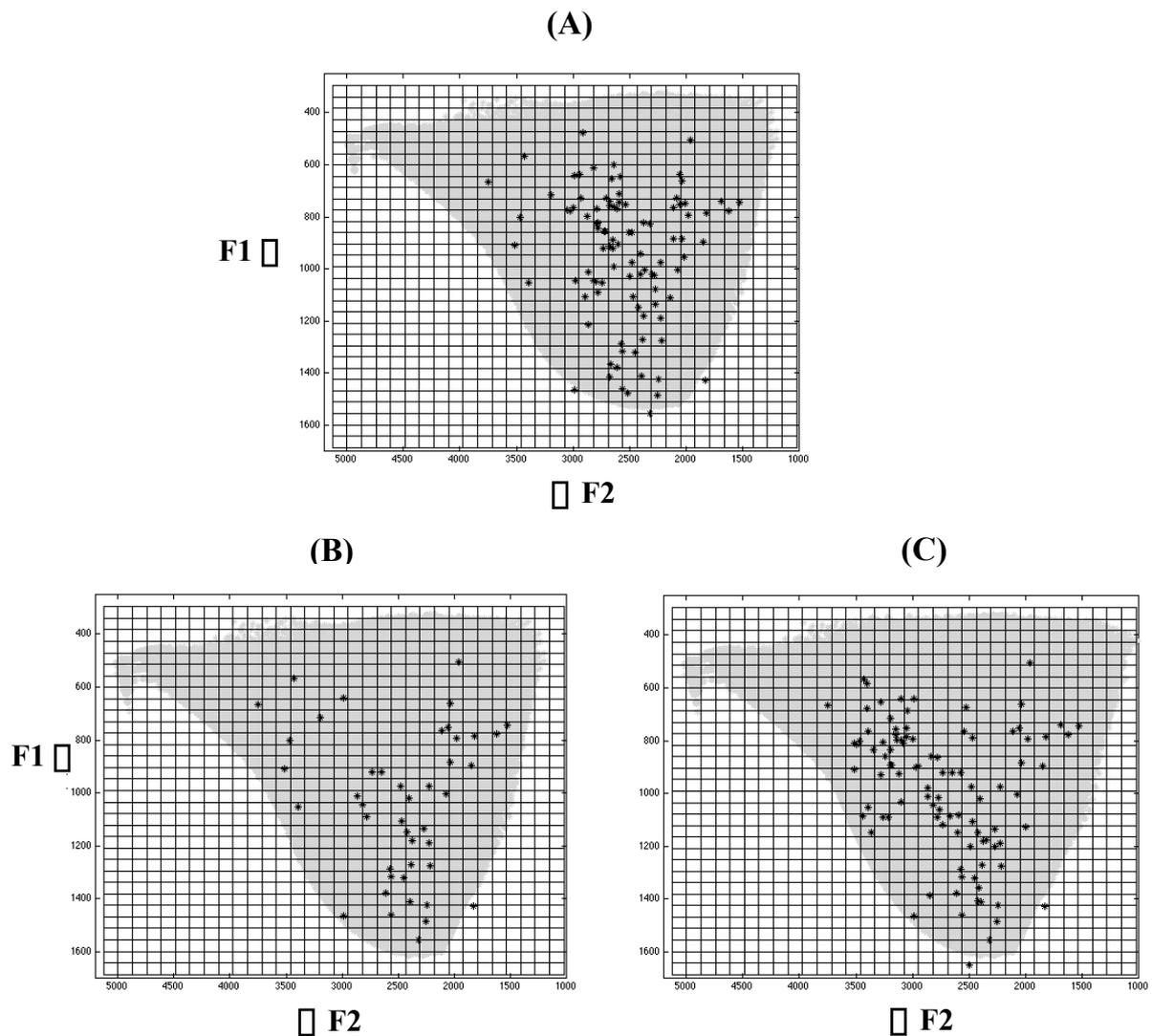


Figure 4-9: Représentation graphique du cadrage articulatoire des corpus de vocalisations de 7 mois de base (A), restreint (B) et modifié (C), avec le meilleur sous-modèle parmi ceux ayant des domaines de variation symétriques des paramètres de commandes non-nuls (voir le texte).

3. Cadrage géométrique

Les meilleurs sous-modèles issus du cadrage articulatoire sont vus comme les "modèles" articulatoire-acoustiques simulant le mieux les productions infantiles avant et au début du babillage canonique. Le cadrage géométrique est une méthode pour inférer la configuration géométrique (ou *forme*) de conduit vocal qui pourrait avoir mené à chaque vocalisation enregistrée à 4 et 7 mois, à partir du meilleur sous-modèle à 3 dimensions articulatoires non-nulles dans le premier cas et à 4 dimensions dans le second. En particulier, le but est d'obtenir des informations sur les formes potentielles de langue dans les vocalisations infantiles, de manière à caractériser plus précisément l'exploration articulatoire des nourrissons. Ce procédé nous permettra de revenir sur une question soulevée par des prédécesseurs au sujet de l'origine développementale de la forme de conduit la plus fréquente de la voyelle [u] adulte.

a. Méthode

Les caractéristiques suivantes ont été choisies pour décrire la forme du conduit vocal à partir de sa coupe sagittale: (i) les coordonnées (X_h , Y_h) du point le plus haut de la langue dans un système de référence fixe, (ii) l'aire (A_c) et la distance à la glotte (X_c) de la constriction intra-orale principale le long du conduit vocal, ainsi que l'aire aux lèvres (A_l) (Figure 4-10). Ces jeux de variables fournissent une caractérisation synthétique de la production orale au niveau géométrique (Boë *et al.*, 1995b)³⁴.

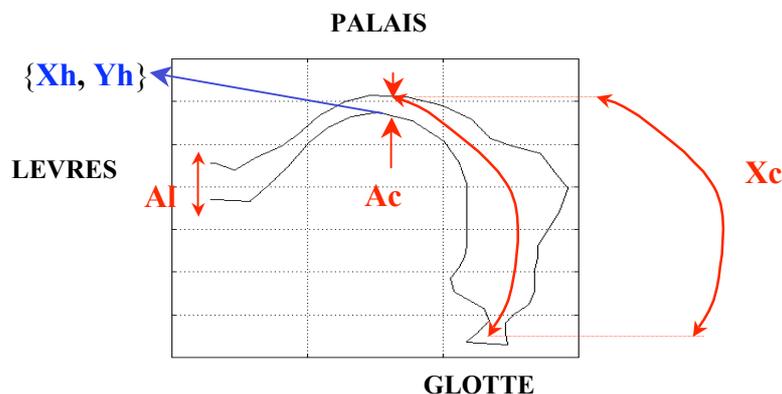


Figure 4-10: Représentation graphique des deux systèmes de variables géométriques choisis pour décrire la forme du conduit vocal, sur la coupe sagittale d'une configuration articuloire prototypique de [i] (contour gris), générée par VLAM à 4 mois. X_c : lieu de constriction, A_c : aire de constriction, A_l : aire interlabiale. X_h et Y_h : abscisse et ordonnée du point le plus haut de la langue.

Le cadrage géométrique est basé sur l'"**inversion exhaustive**" de chaque son vocalique infantile de l'espace acoustique vers l'espace géométrique du sous-modèle de même âge sélectionné à l'issu de l'analyse par cadrage articuloire. L'inversion exhaustive d'une vocalisation consiste en la sélection de *toutes* les simulations du sous-modèle qui se trouvent dans le voisinage auditif de la vocalisation réelle analysée³⁵. Les valeurs de X_h , Y_h , X_c , A_c et A_l qui correspondent aux configurations articuloires sélectionnées sont vues comme acoustiquement plausibles sachant la donnée réelle. La moyenne de chacun de ces paramètres est estimée, et associée à la vocalisation en cours de traitement. En effet, les phénomènes compensatoires nous ont poussés à associer chaque vocalisation réelle à un paramètre statistique capable de rendre compte de manière globale de la distribution des variables géométriques dans les simulations de son voisinage.

Le résultat de cette analyse prend la forme d'une représentation dans les espaces $\{X_h, Y_h\}$ et $\{X_c, A_c, A_l\}$ des configurations de conduit vocal potentiellement explorées par les

³⁴ L'on peut remarquer que ces paramètres correspondent aux variables du système somesthésique simplifié de l'androïde (chapitre II, section II.4.).

³⁵ Dans les présentes simulations, ce voisinage correspond à la région délimitée par un cercle de rayon $0.4 B_k$, centré sur la vocalisation à inverser (pour plus d'information sur l'échelle des Bark, voir chapitre I, section II.4.). L'utilisation d'une échelle perceptive plutôt que strictement acoustique a pour but de reproduire l'assimilation des sons perceptivement proches à une même catégorie. De plus, l'espace acoustique du sous-modèle sur lequel se base l'inversion est généré en tirant des configurations articuloires selon une loi uniforme dans les limites définies par les domaines de variation des commandes de ce sous-modèle, de manière à obtenir un total de 60 000 simulations recevables (i.e. avec des valeurs minimum de A_c et de A_l de 0.1 cm^2 et 0.01 cm^2 , respectivement).

nourrissons de 4 et 7 mois. Toutefois, l'interprétation de ces résultats nécessite des points de référence. Ces repères correspondent à l'inversion exhaustive des prototypes acoustiques des voyelles [i a u ə] à 4 et 7 mois en utilisant l'espace d'exploration de VLAM comme base d'inversion (voir Ménard 2000, pour la notion de prototype acoustique). Les sons choisis pour les prototypes [i a u] sont tels que leur position sur les EVM de 4 et 7 mois soit équivalente à celle qu'ils occupent sur l'EVM adulte dans le plan (F1, F2)³⁶. Le son du prototype schwa est le produit de la configuration articuloire neutre à chacun des deux âges.

En somme, le cadrage géométrique de chaque corpus de données réelles informe sur leur espace d'exploration géométrique et peut donc nous éclairer sur les formes de langues qui pourraient être impliquées dans les productions vocaliques réelles avant et au début du babillage canonique.

b. Résultats

La figure 4-11 présente les résultats obtenus à 4 mois. En accord avec notre compréhension du domaine d'exploration articuloire du meilleur sous-modèle pour les données de pré-babillage, les formes moyennes de langue tourneraient autour de la configuration neutre avec une tendance à s'étendre vers celles du type [u], quelque soit l'espace considéré. Les lieux moyens de la constriction intra-orale sont surtout antérieurs (4-11C), la zone du [u] semblant marquer une limite postérieure du recul de la langue, tandis que l'aire de cette constriction est relativement large. Globalement, les valeurs moyennes d'aire aux lèvres sont petites (4-11D, voir plus bas).

À 7 mois (Figure 4-12), les positions de langue inférées (4-12B) montrent une exploration plus étendue des dimensions haut-bas et avant-arrière qu'à la période développementale précédente. Ce résultat trouve écho dans les valeurs moyennes des attributs, Xc et Ac, de la constriction intra-orale (4-12C).

Aux deux âges, les constrictions labiales moyennes correspondent à des valeurs d'aire, Al, plutôt faibles. Cependant, à 7 mois, la distribution des aires aux lèvres moyennes (4-12D) représentent des valeurs plus élevées qu'à 4 mois (4-11D), ce qui va dans le sens d'une exploration géométrique plus importante des configurations de conduit vocal ouvertes. Il est probable qu'une sous-estimation des aires aux lèvres dans nos simulations soit due aux paramètres articuloires mis à l'écart de l'analyse dans le cadrage articuloire (section III.2.), en particulier LP (protrusion labiale) et potentiellement Lx (hauteur du larynx). En effet, les valeurs d'aire aux lèvres les plus petites sont obtenues dans la région de faibles valeurs formantiques F1 et F2, qui se produisent typiquement avec deux résonances de Helmholtz associées aux cavités avant et arrière couplées (Badin, Perrier, Boë et Abry., 1991). L'allongement de la cavité avant par la protrusion des lèvres (LP) ou l'allongement de la cavité arrière par abaissement du larynx (Lx) aurait résulté en une diminution des valeurs de F1 et F2. Ainsi les valeurs faibles de F1 et F2 réelles auraient-elles pu être obtenues avec des valeurs plus élevées d'aire aux lèvres si LP et/ou Lx avaient été pris en considération dans le cadrage articuloire. En d'autres termes, les formes moyennes de conduit vocal inférées à partir des productions infantiles qui sont acoustiquement proches du [u] équivalent-adulte à 4 et 7 mois pourraient avoir des valeurs d'aire aux lèvres supérieures à celles trouvées par ce cadrage géométrique.

Globalement, bien que les résultats obtenus portent sur des moyennes, appelant donc à prendre des précautions quant à ce qui peut être inféré sur les formes de conduit vocal sous-

³⁶ À 4 mois, les valeurs formantiques des prototypes sont les suivantes: pour [i], F1=568 Hz, F2=4514 Hz; pour [a], F1= 1498 Hz, F2=2496 Hz; pour [u], F1=588 Hz, F2=1533 Hz; et pour [ə], F1=982 Hz, F2=3066 Hz. À 7 mois, ces valeurs sont : pour [i], F1=558 Hz, F2=4378 Hz; pour [a], F1= 1473 Hz, F2=2542 Hz; pour [u], F1=537 Hz, F2=1533 Hz; et pour [ə], F1=928 Hz, F2=2971 Hz.

jaçentes aux vocalisations réelles, les tendances trouvées sont cohérentes avec celles des cadrages acoustique et articuloire.

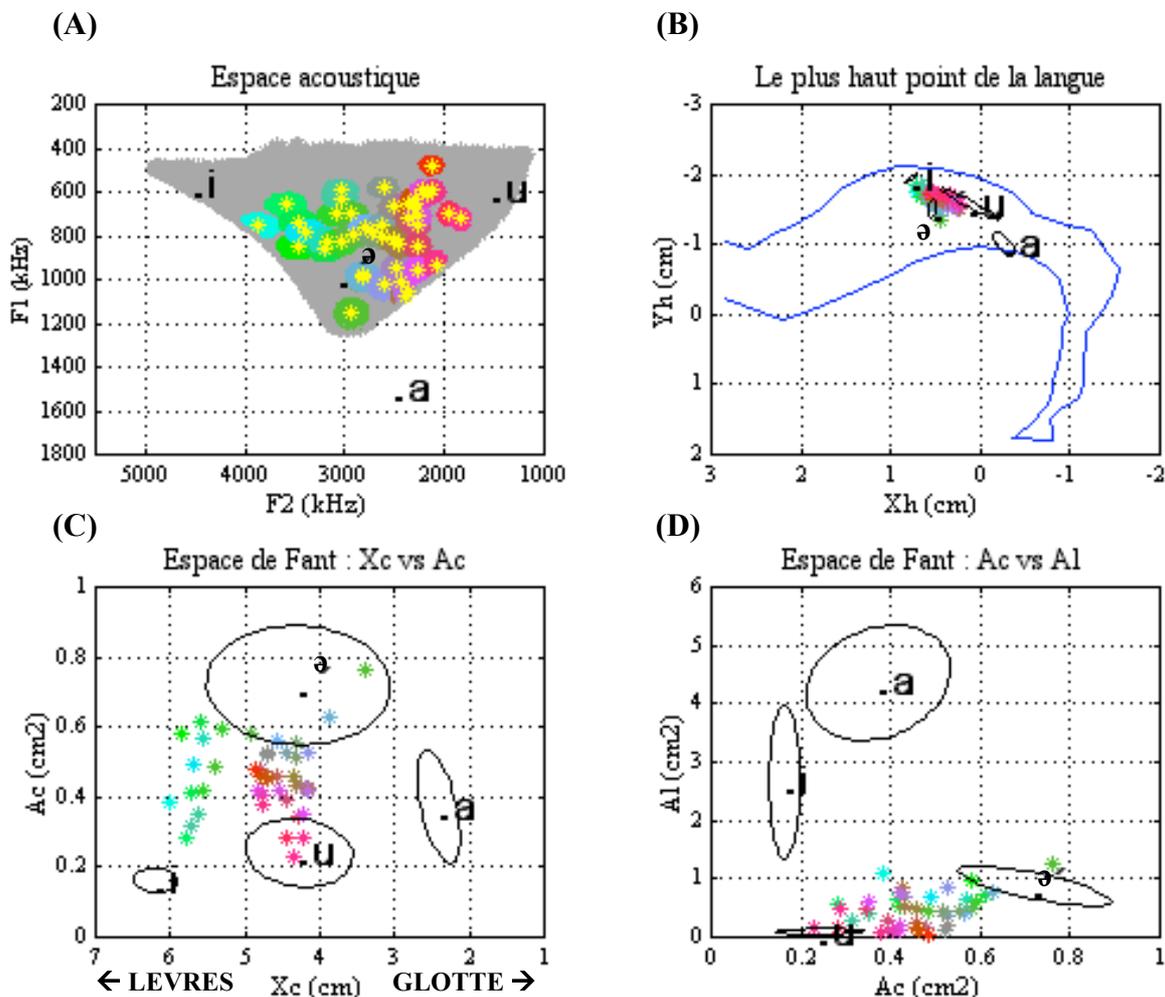


Figure 4-11 : Représentation graphique du cadrage géométrique des vocalisations de 4 mois, avec le meilleur sous-modèle à 3 paramètres articulatoires non-nuls, dans trois plans des espaces géométriques {Xc, Ac, Al} et {Xh, Yh}.

LEGENDE :

- * : Données réelles
- ● ● : Éléments sélectionnés pour l'inversion et * * * : moyennes résultantes
- i a u ə** : Prototypes ajoutés aux données acoustiques réelles et moyennes résultant de leur inversion
- : Ellipses de dispersion des valeurs issues de l'inversion des prototypes ajoutés
- : Coupe sagittale du conduit de VLAM pour la configuration articuloire du prototype [a] de VLAM
- : Domaine d'exploration acoustique du meilleur sous-modèle à 3 paramètres articulatoires non-nuls à 4 mois (60000 simulations)

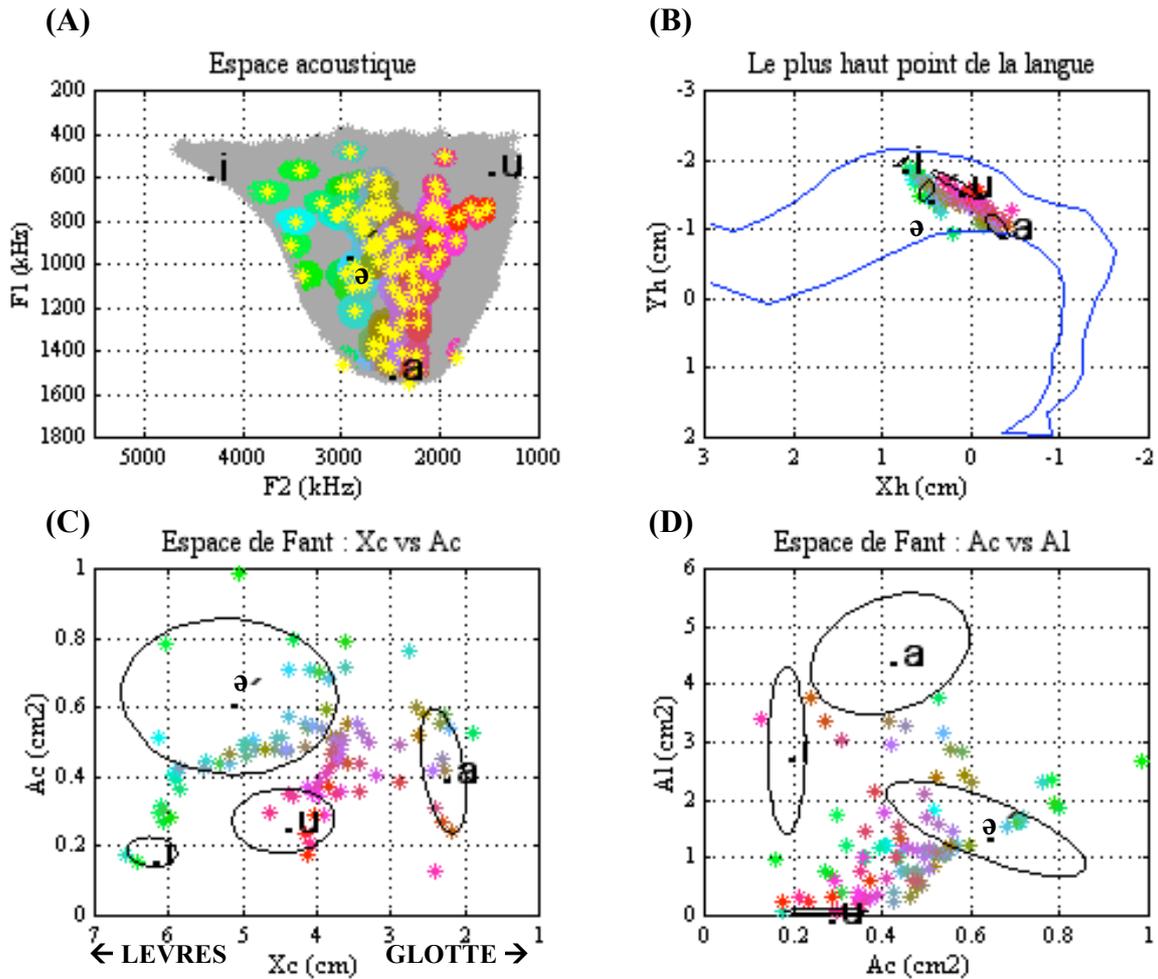


Figure 4-12: Représentation graphique du cadrage géométrique des vocalisations de 7 mois, avec le meilleur sous-modèle à 4 paramètres articulatoires non-nuls, dans trois plans des espaces géométriques $\{Xc, Ac, Al\}$ et $\{Xh, Yh\}$.

LEGENDE :

- * : Données réelles
- : Éléments sélectionnés pour l'inversion et * * * : moyennes résultantes
- i a u ə : Prototypes ajoutés aux données acoustiques réelles et moyennes résultant de leur inversion
- : Ellipses de dispersion des valeurs issues de l'inversion des prototypes ajoutés
- ~ : Coupe sagittale du conduit de VLAM pour la configuration articulatoire du prototype [a] de VLAM
- : Domaine d'exploration acoustique du meilleur sous-modèle à 4 paramètres articulatoires non-nuls à 7 mois (60000 simulations)

Le [u] adulte est-il palatal pour des raisons développementales?

Théoriquement, les phénomènes compensatoires permettent au conduit vocal humain de produire les trois premiers formants du [u] grâce à trois types de configurations articulatoires (Boë, Abry, Beautemps, Schwartz et Laboissière, 2000): on parle de [u] vélo-palatal, vélo-pharyngal et pharyngal, dont la constriction intra-orale principale est, respectivement, palatale, dans le haut pharynx et pharyngale. Toutefois, des études radiographiques montrent que les locuteurs/trices (adultes) de toutes les langues testées ne produisent que des [u] de type palatal (Wood, 1979). Le [u] pharyngal n'a jamais été relevé. Le [u] vélo-pharyngal n'a été observé que sur de rares sujets dont la tâche était de prononcer la voyelle [u] avec un tube entre les lèvres, les forçant à compenser au niveau acoustique la perturbation que représente l'ouverture labiale (paradigme du *lip-tube*, Savariaux, Perrieret Orliaguet, 1995). La difficulté de la tâche a mené Abry et Badin (1996) à la proposition selon laquelle le [u] palatal pourrait être la première stratégie de production de [u] au cours du développement de la production de la parole. Autrement dit, sa prépondérance à l'âge adulte s'expliquerait par sa cartographie sensori-motrice précoce, empêchant ainsi les individus ayant un tube entre les lèvres de "réellement abandonner le lien acquis entre la cible acoustique du [u] et la position [vélo-palatale de la langue]"³⁷ (Abry et Badin, 1996, p. 181). Notre but est ici de tester si nos simulations des capacités de productions les plus précoces pourraient s'accorder avec cette hypothèse, en effectuant l'inversion exhaustive de la vocalisation la plus proche (en termes de Bark et de Hertz) du prototype acoustique de [u] à 4 mois.

Il en résulte que les formes de conduit vocal susceptibles de produire ce [u] juvénile³⁸ montrent: (i) des plus hauts points de la langue palataux à pharyngaux, si le domaine d'exploration (complet) de VLAM à 4 mois est pris pour base de l'inversion (Condition contrôle, Figure 4-13), mais (ii) des configurations de langue *uniquement de type palatal*, si l'inversion est basée sur le sous-modèle à 3 paramètres articulatoires non-nuls qui simule le mieux les capacités d'exploration de cette période développementale (Condition test, Figure 4-14).

Il vient que le pseudo-[u] aux alentours de 4 mois serait de configuration palatale comme chez l'adulte, et pourrait donc vraisemblablement être appris précocement au cours du développement, fournissant ainsi un précurseur potentiel de cette voyelle dans la parole adulte. Ce résultat va dans le sens de l'hypothèse d'Abry et Badin.

³⁷ "[T]o really abandon their acquired link between the [u] acoustic target and [the velo-palatal] position [of the tongue]".

³⁸ Les valeurs formantique du [u] juvénile sont les suivantes: F1= 722 Hz, F2= 1846 Hz.

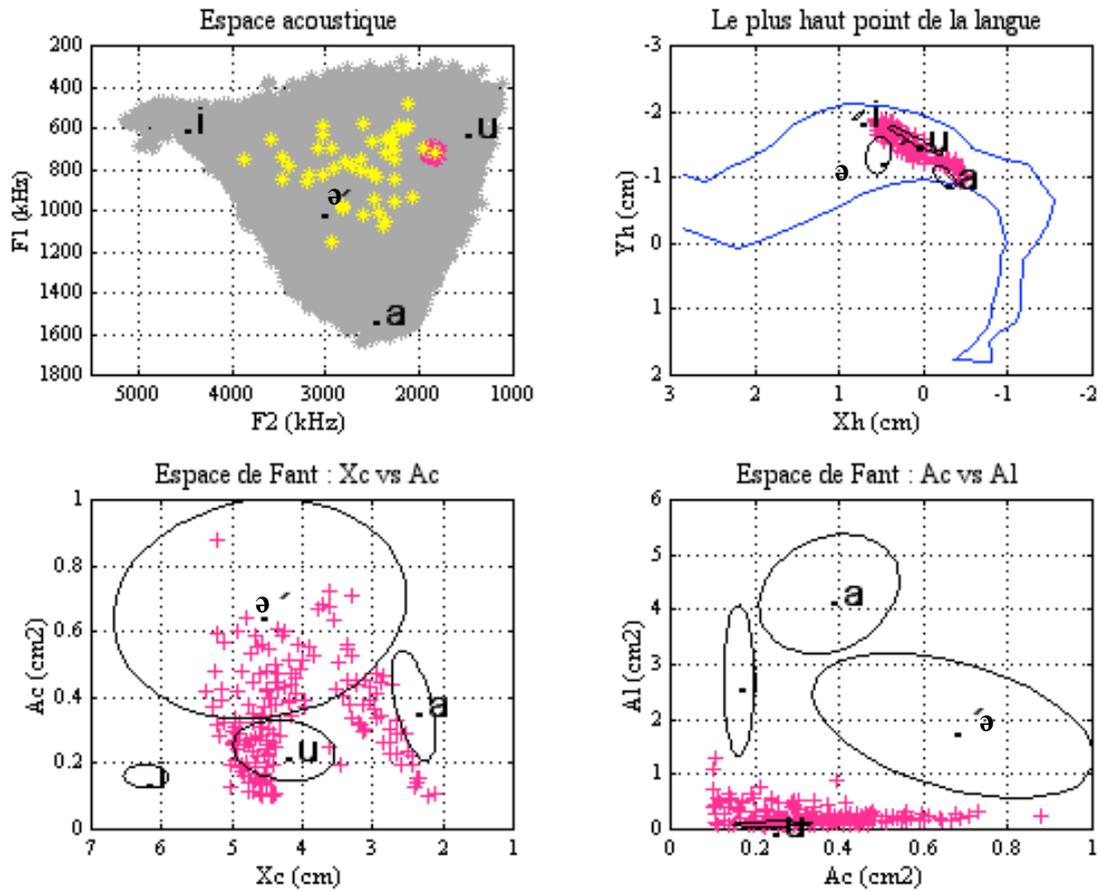


Figure 4-13: Représentation graphique du cadrage géométrique du [u] juvénile avec VLAM à 4 mois comme base d'inversion (voir Figure 4-11, pour la légende).

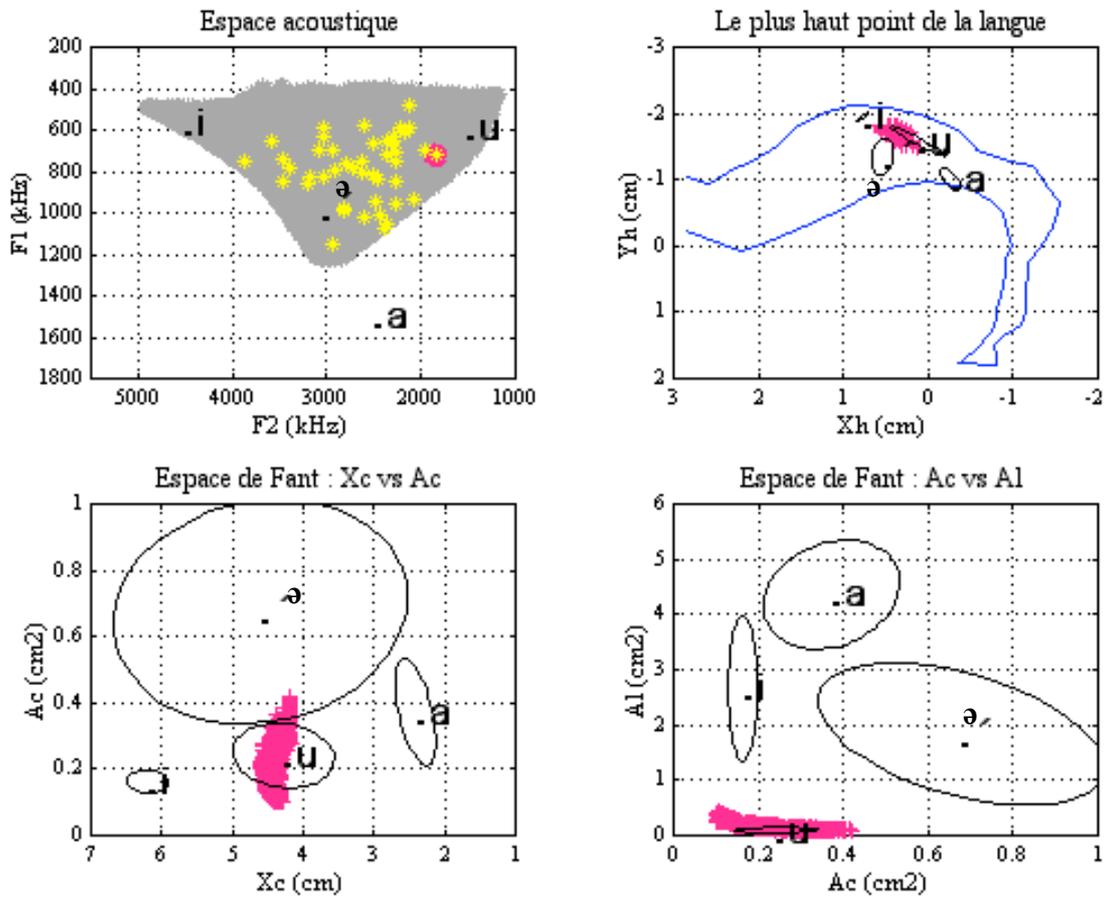


Figure 4-14: Représentation graphique du cadrage géométrique du [u] juvénile avec le meilleur sous-modèle à 3 paramètres moteurs non-nuls à 4 mois comme base d'inversion (voir Figure 4-11, pour la légende).

IV. PREDICTIONS DU MODELE A 7 MOIS ET CONFRONTATION A LA REALITE

1. Le cas des cooccurrences dans le formalisme de la théorie Frame/Content

Dans la théorie Frame/Content, la configuration articulatoire d'un vocant donné est conçue comme une pré-configuration (*presetting*) du conduit vocal à laquelle la cyclicité mandibulaire se superpose, produisant de fait une alternance de vocants et de closants (chapitre 1, section III.1.a.). Ici, le but est de simuler ce phénomène afin de prédire les patrons de cooccurrence qui en résulteraient et de comparer les prédictions de la modélisation à celles de la théorie ainsi qu'aux patrons réels établis à partir de l'analyse transcriptionnelle de Matyear et collègues (section II.2.).

a. Méthode

Le principe de ces simulations est successivement (i) d'associer à *chaque* vocalisation du corpus à 7 mois *plusieurs* configuration-s articulatoire-s adéquate-s sur le meilleur sous-modèle de début de babillage, (ii) pour chaque configuration inférée de monter la mâchoire jusqu'à fermeture du conduit vocal, et (iii) de déterminer le lieu de closion, afin d'établir l'association vocant-closant qui en résulterait. Cela permettra de comparer ces cooccurrences simulées, à celles trouvées dans le corpus réel et à celles prédites par la théorie (section III.1.a.).

Plus précisément, pour chaque vocalisation du corpus de 7 mois, les configurations articulatoires potentiellement sous-jacentes à sa description acoustique ont été estimées par l'inversion exhaustive de l'espace acoustique vers l'espace *articulatoire* dans le meilleur sous-modèle à 4 dimensions non-nulles (section III.2.b.). En d'autres termes, il y a sélection de toutes les configurations articulatoires, générées par le sous-modèle de production le plus vraisemblable à 7 mois, qui produisent un résultat acoustique voisin du son de la vocalisation réelle³⁹.

La suite de la procédure est empruntée à Vilain *et al.* (1999, voir chapitre 1, section II.1.b.). La mandibule des configurations inférées est levée jusqu'à fermeture du conduit vocal, quel que soit son lieu, i.e. entre les lèvres, ou entre la langue et la paroi dorsale du conduit⁴⁰. Les closants correspondants ont été classifiés comme étant labiaux, coronaux ou palato-vélaires d'après la position de la première closion. Plus précisément, si la fermeture a d'abord lieu au niveau des lèvres, i.e. avec une aire aux lèvres nulle ($A_l=0$), le closant est considéré labial. Si la closion initiale se trouve à l'intérieur du conduit vocal, i.e. avec une aire de constriction nulle ($A_c=0$), l'étiquetage est soit coronal soit palatal, en fonction de la position de la fermeture. En effet, une ligne de démarcation entre les deux types de lieux

³⁹ Dans les présentes simulations, l'espace acoustique du sous-modèle et le voisinage des vocalisations sont les mêmes que pour le cadrage géométrique. Cependant, pour chaque vocalisation seules 33 simulations ont été tirées au hasard au sein du voisinage acoustico-perceptif pour mener le reste de la procédure, pour des raisons de temps de calcul: la levée de la mandibule se fait par pas de 0.1 *e-t*, jusqu'à ce que A_l ou A_c soit inférieure ou égale à 0.01, avec à chaque itération le calcul de toutes les variables de sortie de VLAM. L'on remarquera que l'utilisation d'un même nombre de simulations pour toutes les vocalisations à inverser présente l'avantage de respecter la représentativité de chaque catégorie de vocants dans le corpus de données réelles de départ, évitant ainsi la sur-représentation des catégories dont les réalisations ont une forte densité de simulations dans leur voisinage acoustico-perceptif (notamment celles qui se trouvent au centre de l'espace d'exploration acoustique simulé).

⁴⁰ Autrement dit, hormis les variations de commande mandibulaire J, la configuration articulatoire reste constante: l'on retrouve la notion de *pré-configuration* du conduit vocal proposée par la théorie. De plus, la valeur de J est augmentée jusque contact, même si elle dépasse la valeur de +3 *e-t*, en accord avec Vilain *et al.* (1999).

d'articulation a été définie à 6 cm de la glotte ($X_c = 6$ cm), avec les coronales en aval ($X_c \geq 6$ cm) et les palato-vélaires en amont ($X_c < 6$ cm) de cette frontière. Le choix de la valeur-limite de X_c se base sur l'étude de séquences CV menée sur le modèle de production adulte SMIP (Berrah, 1994). Cette valeur est telle qu'elle sépare les closions coronales des palatales en partant de la configuration articulaire d'un [i] prototypique dans le modèle VLAM à 7 mois (voir annexe A5).

De là, la proportion de labiales, de coronales et de palato-vélaires générées à partir des configurations articulaires sélectionnées par l'inversion des vocants de chaque catégorie vocalique est calculée, dans le but de comparer cette distribution simulée à celle prédite par la théorie et celle établie par transcription.

b. Résultats

➤ Un point de vue en surface

La figure 4-15 présente la distribution en pourcentage des cooccurrences entre types de closants et types de vocants sur le total des simulations pour chacune des 7 catégories de vocants dans le corpus de 7 mois. L'on trouve que les sons des vocants antérieurs ainsi que le [a] central sont majoritairement associés à des closants coronaux (dans 60 à 80% des cas pour les antérieures et 87% des cas pour [a]), et à une proportion non-négligeable de palato-vélaires (pour 17 à 25% des simulations dans les antérieures et 13% des simulations pour [a]). La catégorie centrale [ə] et celle postérieure [o] sont le plus souvent associées à des closants labiaux (avec 44 et 67% des simulations respectivement) et dans une moindre mesure à des palato-vélaires (respectivement 37 et 33% des cas).

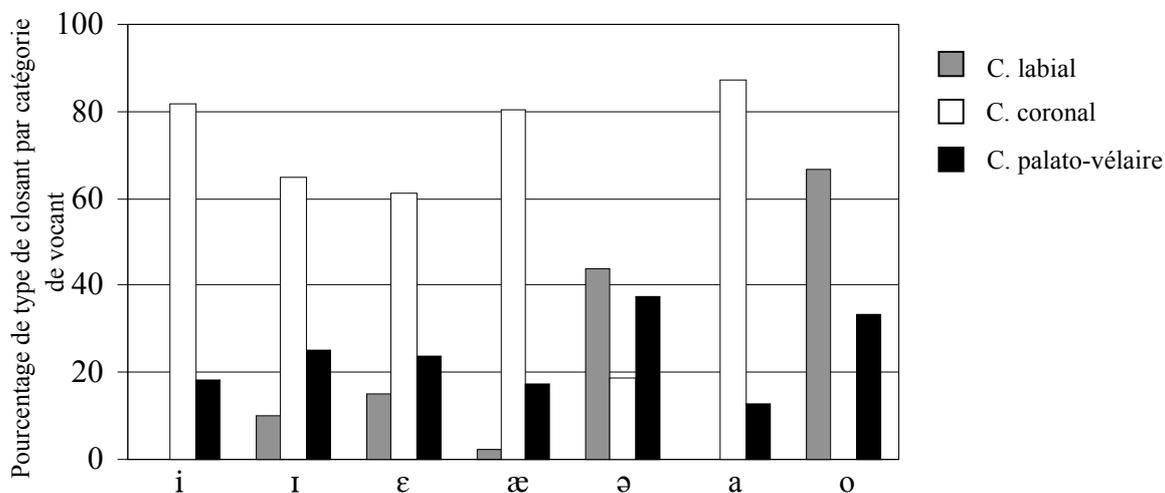


Figure 4-15: Distribution en pourcentage des cooccurrences entre types de closant (labial vs. coronal vs. palato-vélaire) et types de vocant sur le total des simulations pour chacune des 7 catégories de vocants du corpus de 7 mois.

La distribution des lieux de closion en fonction de la catégorie vocalique prédite par notre procédure est plus ou moins en accord avec les prédictions de la théorie Frame/Content. Si la prépondérance des coronales avec les vocants antérieurs va dans son sens, un des écarts les plus flagrants réside dans le cas de [a]. En effet, cette catégorie vocalique est supposée centrale: la théorie prévoit une prédominance de son occurrence avec des closants bilabiaux et non coronaux comme le proposent nos simulations de babillage. Hormis les imprécisions

potentielles sur les mesures acoustiques et/ou perceptives que l'on ne peut contrôler, cet écart à la prédiction théorique pourrait s'expliquer (i) par le domaine d'exploration articulaire du sous-modèle⁴¹ et/ou (ii) par la morphologie particulière du modèle VLAM que l'on sait, d'après les simulations menées par Vilain *et al.* (1999) sur SMIP, avoir tendance à produire des cadres de types coronaux pour une fermeture mandibulaire à partir du prototype [a] dans le cas du conduit vocal d'âge adulte.

En effet, en 2000, Vilain soutenait, modélisation à l'appui, que le patron de cooccurrence en partant d'une langue en position de repos (vocant central) ne donne pas obligatoirement une closure labiale (cadre pur), et en particulier que ce patron est fonction de la morphologie du conduit vocal (Vilain, 2000, et voir Tableau IV, chapitre 1, section III.1.b.).

En réalité, dans les limites de ce qui peut être tiré du nombre relativement faible de données réelles à notre disposition, la distribution des lieux de closure en fonction de la catégorie vocalique, établie d'après la transcription de Matyear et collègues, montre que la majorité des closants du corpus sont coronaux. Plus précisément un sujet sur deux ne produit que ce type de closant, tandis que seul l'autre sujet produit des bilabiales. De plus, le second réalise les sons étiquetés [ə] en majorité avec une fermeture aux lèvres. En d'autres termes, le schwa mène à des cadres plutôt coronaux ou plutôt labiaux en fonction de l'individu. Ce patron de cooccurrence n'est pas rigoureusement en accord avec les prédictions de la théorie Frame/Content, sans pour autant remettre en question de manière fondamentale son axiome de départ sur la prépondérance de la cyclicité mandibulaire dans l'explication de la variabilité des types vocaliques et consonantiques à l'intérieur des séquences de babillage canonique (voir Section V pour une confrontation plus générale aux données avancées par la théorie).

➤ Un point de vue en profondeur

A l'argument de la variabilité inter-individuelle dans le type de closure en partant d'une même configuration articulaire de départ due à la morphologie du conduit vocal, nous voulons soulever la question de cette variabilité au niveau *intra*-individuel. En effet, le détail de nos simulations montre que, dans de nombreux cas, la détermination d'un *unique* lieu de closure ne coule pas de source.

Prenons l'exemple de la simulation de la fermeture mandibulaire à partir d'une configuration résultant de l'inversion d'un son étiqueté [ɪ]. D'après notre procédure, la première closure est coronale (en accord avec la prédiction de cadres antériorisés par la théorie Frame/Content). Toutefois, un suivi de la coupe sagittale nous montre que ce lieu précède de peu une closure *labiale* (Figure 4-16). Si la langue et les lèvres de VLAM n'étaient pas rigides, mais *molles* comme dans la réalité, le closant produit à l'ouverture aurait très bien pu être [bd]. De même, dans les simulations à partir du son transcrit [o], les fermetures bilabiale et vélaire peuvent "cacher" des closures, respectivement, vélaire et bilabiale (Figure 4-17), menant potentiellement à la production de [bg] dans la réalité.

⁴¹ En effet, nous avons vu que le meilleur sous-modèle issu du cadrage articulaire des données de 7 mois (section III.2.b.) correspond à une exploration des configurations de langue plutôt antérieure, avec des valeurs de TB (avancement-rétraction du corps de la langue) le plus fréquemment positives. La distribution des paramètres articulaires étant uniforme dans les simulations présentes, la prépondérance de configurations linguales antérieures, lors de la sélection des pré-configurations voisines des vocants réels, pourrait avoir favorisé l'occurrence des closures intra-orales situées vers l'avant du conduit vocal (C. coronal).

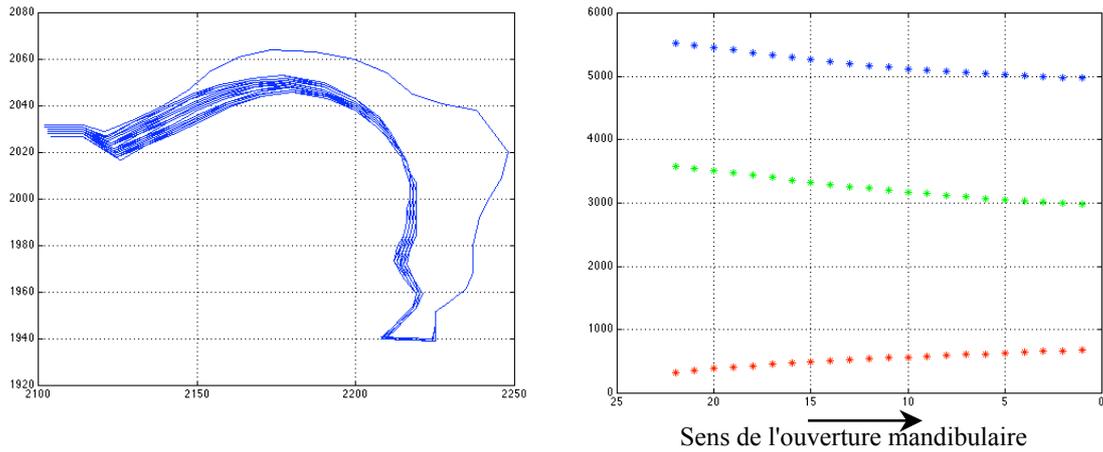


Figure 4-16: Suivis de la coupe sagittale du conduit vocal (à gauche) et des trois premiers formants exprimés en Hertz (à droite, dans le sens CV) au cours de la fermeture mandibulaire d'une pré-configuration simulée résultant de l'inversion d'un vocant (réel) étiqueté [ɪ] par Matyear et collègues. Le lieu de closion initial est coronal mais précède de peu une closion bilabiale comme le montre le suivi de contour sagittal.

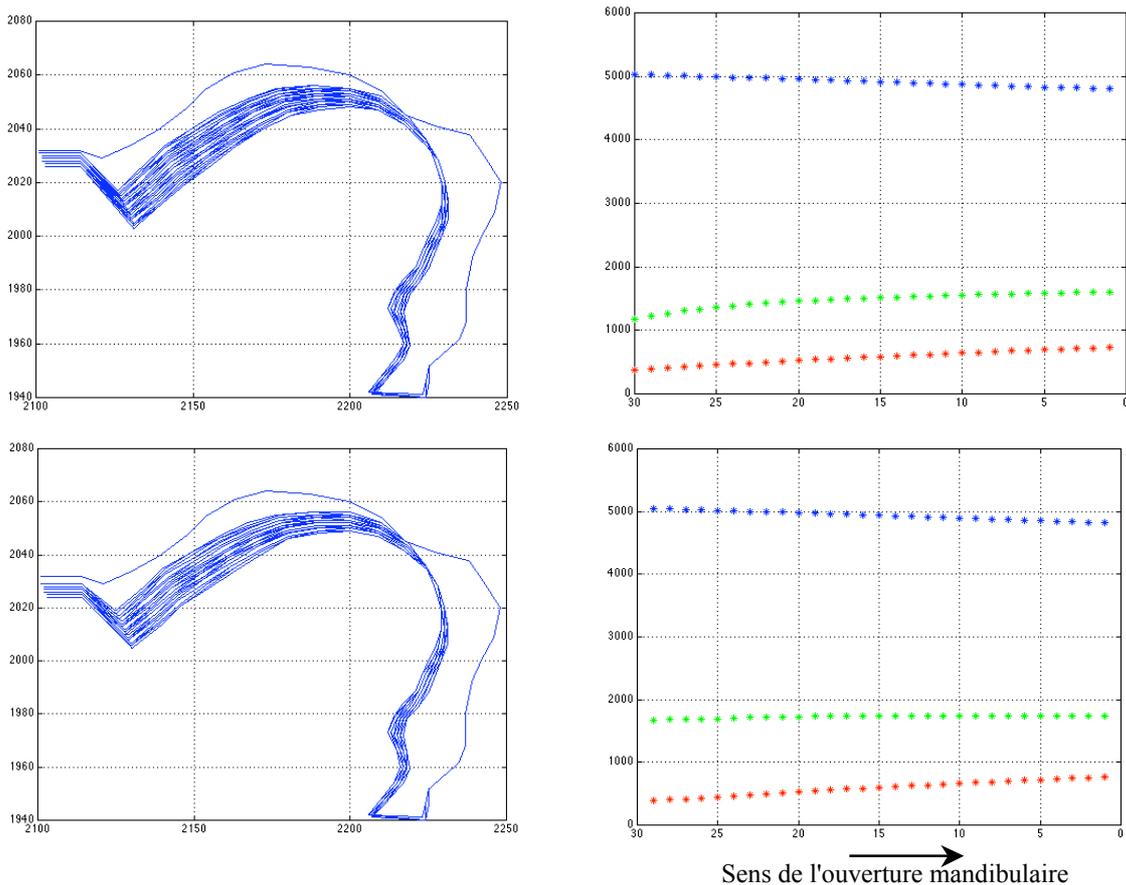


Figure 4-17: Suivis de la coupe sagittale du conduit vocal (à gauche) et des trois premiers formants exprimés en Hertz (à droite, dans le sens CV) au cours de la fermeture mandibulaire de deux pré-configurations simulées résultant de l'inversion d'un vocant (réel) étiqueté [o] par Matyear et collègues. Le lieu de closion initial est bilabial mais presque vélaire dans un cas (en haut) et vélaire mais presque bilabiale dans l'autre (en bas).

Nous ne faisons pas là un procès de la théorie: ses prédictions de base sont tout à fait vraisemblables, comme en témoignent des simulations à partir d'un vocant étiqueté [ε] (e.g. Figure 4-18). Cependant, si la closion peut être coronale pour certaines réalisations de cette catégorie vocalique, un suivi de coupe sagittale montre que la même closion peut cacher des réalités différentes et plus ambiguës pour d'autres réalisations. Comment juger de l'*unique* lieu d'articulation dans le cas d'un écrasement de la langue contre le palais, situation qui peut être doublée d'une fermeture labiale quasi-concomitante (Figure 4-19)?

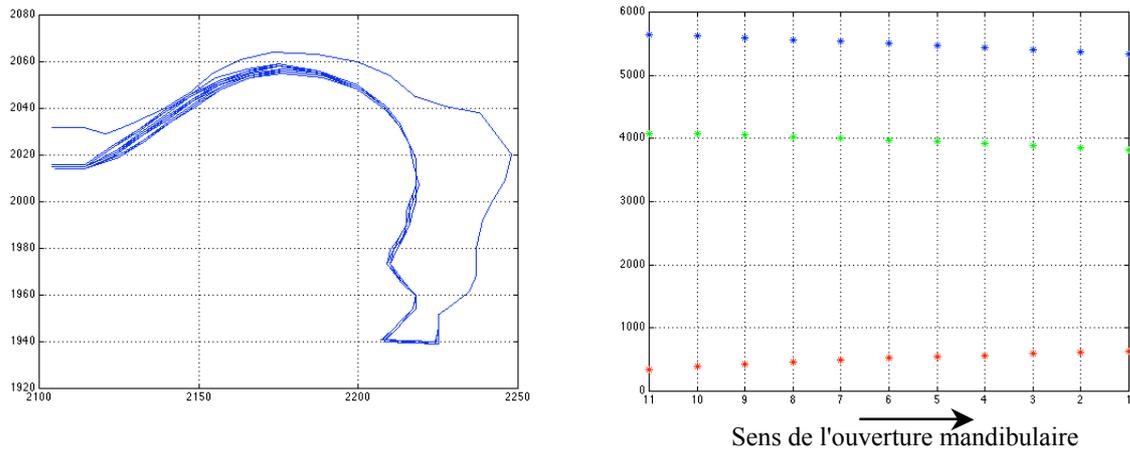


Figure 4-18: Suivis de la coupe sagittale du conduit vocal (à gauche) et des trois premiers formants exprimés en Hertz (à droite, dans le sens CV) au cours de la fermeture mandibulaire d'une pré-configuration simulée résultant de l'inversion d'un vocant (réel) étiqueté [ε] par Matyear et collègues. Le lieu de closion initial est coronal en accord avec la prédiction d'un cadre antérieurisé par la théorie Frame/Content.

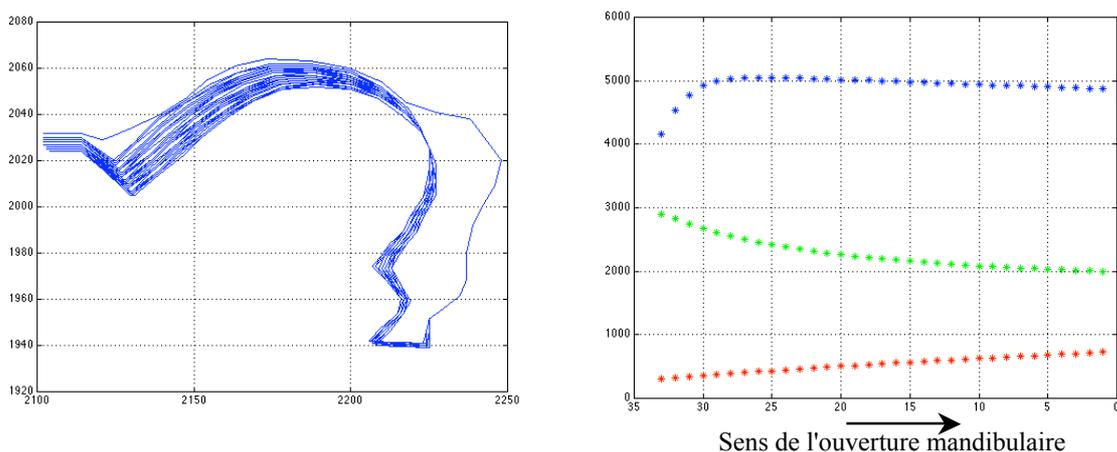


Figure 4-19: Suivis de la coupe sagittale du conduit vocal (à gauche) et des trois premiers formants exprimés en Hertz (à droite, dans le sens CV) au cours de la fermeture mandibulaire d'une pré-configuration simulée résultant de l'inversion d'un vocant (réel) étiqueté [ε] par Matyear et collègues. Le lieu de closion initial est coronal, mais le suivi de contour montre qu'il y a écrasement de la langue contre le palais et qu'une closion bilabiale est imminente.

Ainsi, comme une pierre à l'édifice, nous ajouterons à l'argumentaire de Vilain *et al.* (1999) sur la variabilité inter-individuelle que la variabilité du cadre dit "pur" se trouve aussi au niveau intra-individuelle, comme l'illustrent les simulations à partir de sons étiquetés [ə]. En accord avec la notion de cadre pur, la closure peut être labiale, avec un suivi de coupe sagittale sans ambiguïté (Figure 4-20). Toutefois, en fonction des réalisations, cette fermeture peut tout aussi bien être précédée de peu d'un contact vélaire (Figure 4-21). Un autre exemple, est fourni pour un son [a] qui donne des closures coronales mais aussi vélares (Figure 4-22, dans le second cas, l'on notera l'écrasement de la langue contre le palais).

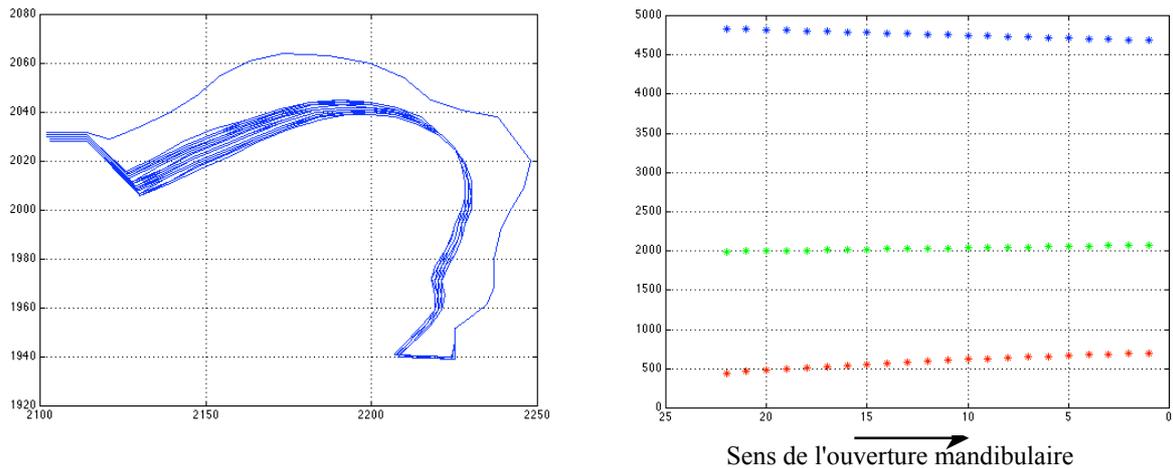


Figure 4-20: Suivis de la coupe sagittale du conduit vocal (à gauche) et des trois premiers formants exprimés en Hertz (à droite, dans le sens CV) au cours de la fermeture mandibulaire d'une pré-configuration simulée résultant de l'inversion d'un vocant (réel) étiqueté [ə] par Matyear et collègues. Le lieu de closure initial est clairement bilabial en accord avec la prédiction d'un cadre pur par la théorie

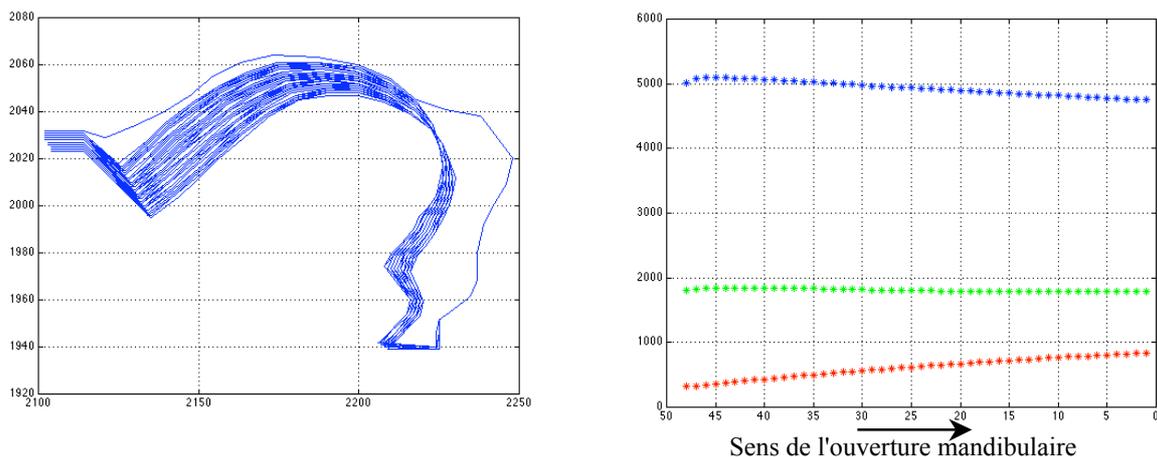


Figure 4-21: Suivis de la coupe sagittale du conduit vocal (à gauche) et des trois premiers formants exprimés en Hertz (à droite, dans le sens CV) au cours de la fermeture mandibulaire d'une pré-configuration simulée résultant de l'inversion d'un vocant (réel) étiqueté [ə] par Matyear et collègues. Le lieu de closure initial est vélaire, mais presque labial.

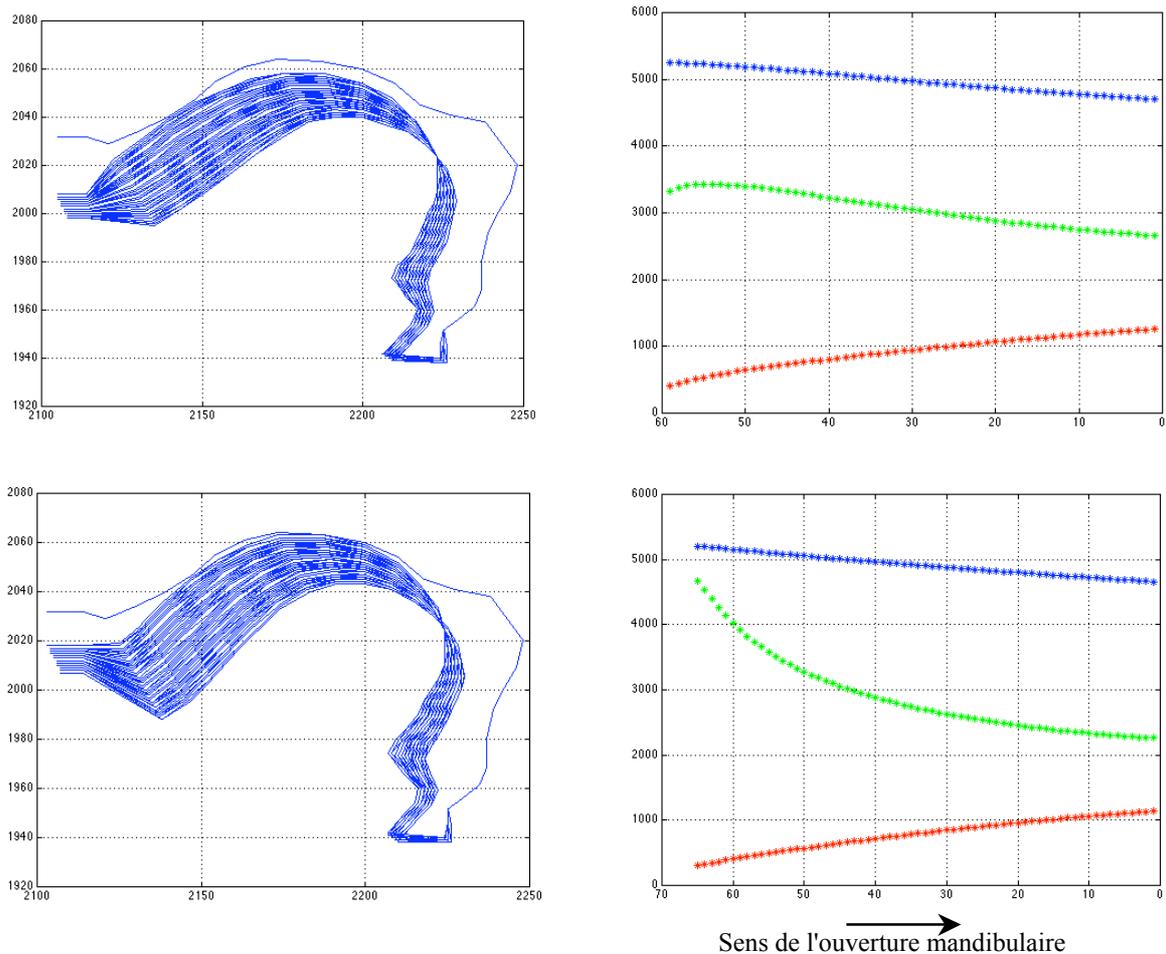


Figure 4-22: Suivis de la coupe sagittale du conduit vocal (à gauche) et des trois premiers formants exprimés en Hertz (à droite, dans le sens CV) au cours de la fermeture mandibulaire de deux pré-configurations simulées résultant de l'inversion d'un vocant (réel) étiqueté [a] par Matyear et collègues. Le lieu de closion initial est coronal dans un cas (en haut) et vélaire mais presque coronal par écrasement de la langue contre le palais dans l'autre (en bas).

En conclusion, vu que les lois de production des conduits vocaux juvéniles ont peu de chance d'échapper aux phénomènes de compensation articulatoire, un même jeu de résonance du conduit vocal peut cacher des configurations articulatoires différentes. De là, la superposition de la cyclicité mandibulaire peut mener à des patrons de cooccurrence plus diversifiés que ne le prévoit la théorie Frame/Content: ils seraient fonction de la morphologie propre du conduit vocal mais aussi de la réalisation en cours, i.e. d'un **bruit biologique**⁴² (Brosda, 1999). Aussi, une association closant-vocant établie par transcription peut cacher des réalités différentes en termes articulatoires.

Ainsi, la définition des trois types de cadre doit-elle être *nuancée* à cause de la **variabilité interindividuelle** de la morphologie du conduit vocal et de la **variabilité intra-individuelle** dans la production de vocalisations similaires.

⁴² Ce **bruit biologique** s'ajoute aux phénomènes compensatoires et peut correspondre à de petites variations au niveau des commandes mais aussi au niveau du résultat de l'exécution de ces commandes par le système effecteur.

Toutefois, cette nuance n'enlève rien à l'idée fondamentale que la pré-configuration du conduit vocal et le type de cadre d'une même proto-syllabe puissent être liés par le truchement du mouvement mandibulaire.

2. Le cas de l'émergence du contrôle du système articulateur au niveau acoustique dans le paradigme du locus

Un "point info" sur le paradigme du locus

Dans la lignée ouverte par les travaux de Delattre et collègues sur la théorie du locus, développée aux laboratoires Haskins depuis les années 50s, Sussman mène depuis 15 ans une série d'études dont le but est de caractériser les patrons de coarticulation anticipatrice dans les productions de type son consonantique-son vocalique (*CV*) sur la base de l'analyse des relations entre les fréquences de résonance du début et de la fin des transitions formantiques dans ces séquences (Sussman, Fruchter, Hilbert et Sirosch, 1998).

L'idée-phare de Sussman est la proposition selon laquelle la coarticulation entre C et V est:

- (i) maximale, si la relation, supposée linéaire, entre le second formant au départ de la transition CV ($F2_initial$), valeur que l'on associe à la réalisation de C, et le second formant au climax de V ($F2_centre$) est de pente voisine de 1, puisque les configurations articulaires consonantique et vocalique sont proches⁴³ (Figure 4-23A).
- (ii) absente, si les valeurs de $F2_initial$ et $F2_centre$ sont indépendantes l'une de l'autre, i.e. la droite qui caractérise leur relation serait de pente nulle, puisque les configurations articulaires consonantique et vocalique ne s'influenceraient pas l'une l'autre (Figure 4-23B).

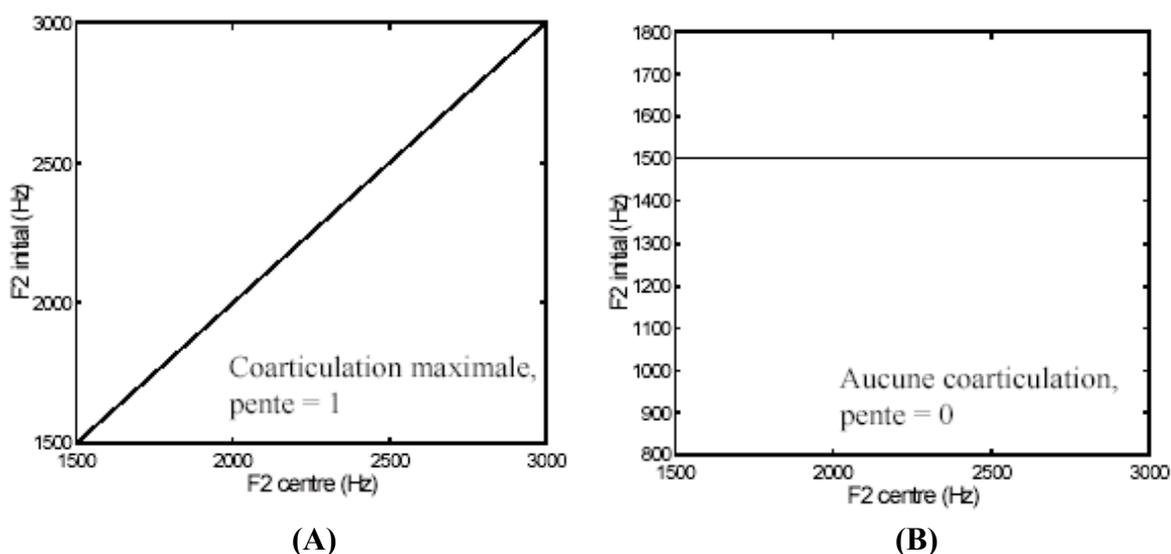


Figure 4-23: Représentation schématique du principe des équations du locus d'après Sussman *et al.* (graphiques tirées de Ménard, 2000).

⁴³ Précisons d'emblée que même " dans le cas le plus clair, [c'est-à-dire] celui des bilabiales, même si la forme globale du conduit vocal pour la voyelle était complètement anticipée dans la phase de closure, la vitesse d'ouverture des lèvres ne pourrait clairement pas être aussi rapide que ce qui est théoriquement nécessaire pour produire les mêmes valeurs de second formant dans les phases de départ et cible de la voyelle (pente de 1). " (Abry *et al.*, 2005: " in the most clear case, that of bilabials, even if the overall vocal tract shape for the vowel was fully anticipated into the closure phase, the rate of opening of the lips obviously could not be as fast as required theoretically for producing the same 2nd formant values at the onset and target phases of the vowel (slope 1)"). La pente de 1 est donc une abstraction théorique.

a. Des éléments pour une modélisation de l'émergence du contrôle du système articulatoire dans le formalisme du locus

Sur la base du principe des équations du locus, Sussman *et al.* (1999) ont réalisé une étude de la dispersion de F2 dans les productions de type syllabique d'un sujet âgé de 7 à 40 mois⁴⁴ pour caractériser au niveau acoustique l'émergence du contrôle du système articulatoire (voir chapitre 1, section III, point *iii*). Ils ont montré comment les pentes de coarticulation évoluaient progressivement, en fonction du lieu d'articulation, pour converger sur des valeurs adultes, dès 12 mois pour le cas des labiales, mais plus tard pour les coronales et les vélares.

Les tendances de cooccurrence de la théorie Frame/Content dans le formalisme du locus

Partant de la conviction qu'un phénomène doit vraisemblablement s'exprimer dans différents formalismes, nous avons cherché à reformuler les prédictions de tendance à la cooccurrence des séquences de babillage de la théorie Frame/Content dans les termes d'une répartition structurée et cohérente des valeurs de second formant dans les transitions CV. Nos prédictions sont donc les suivantes (Figure 4-24).

Tout d'abord, les patrons de cooccurrence devraient se traduire par *l'association préférentielle de zones de valeurs de F2_centre à un type majoritaire de closant*: les F2_centre les plus faibles (vocants postérieurs) liés aux palato-vélares, les F2_centre médians (vocants centraux) aux labiaux, et les F2_centre les plus élevés (vocants antérieurs) aux coronaux.

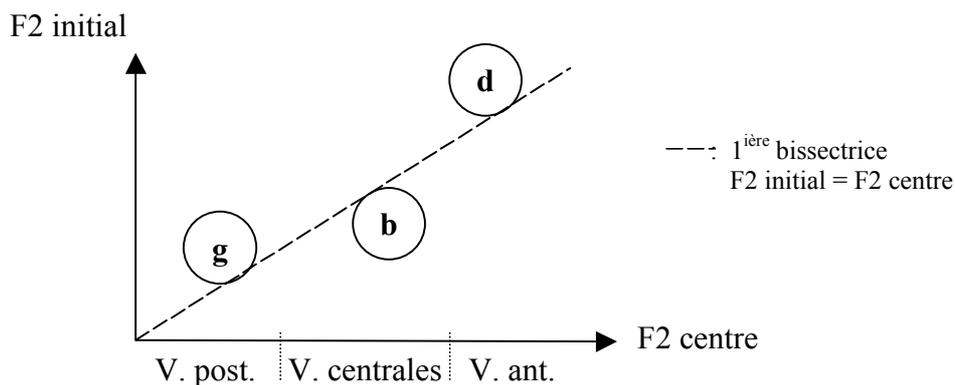


Figure 4-24: Représentation schématique des tendances attendues dans la répartition des valeurs de second formant au cours des transitions CV de babillage d'après les prédictions de la théorie Frame/Content. **b:** C. labial, **d:** C. coronal, **g:** C. palato-vélaire.

De plus, le fait que seule la mâchoire soit mobile au début du babillage canonique devrait avoir pour conséquence que les vocants et closants contigus soient relativement semblables en terme de configuration articulatoire, la pré-configuration étant considérée stable, en particulier, le long de la dimension avant-arrière au cours du cycle. Il vient que la simulation de séquences closant-vocant par geste mandibulaire pur devrait produire des *données autour de la première bissectrice* ($F2_initial = F2_centre$), réparties en trois amas résultants des effets de cooccurrence sus-cités.

Un troisième point concerne l'évolution de la valeur de F2 au cours de la transition en fonction du lieu d'articulation du son consonantique. Classiquement, l'on considère que cette

⁴⁴ Les données brutes sont issues des enregistrements (sonores) et de la transcription effectués par Davis et MacNeilage (1995). Les productions de ce sujet n'apparaissent pas dans le corpus de babillage que nous avons étudié jusque là.

transition est plutôt montante, i.e. $F2_centre \geq F2_initial$, pour les coronaux et les palato-vélaires, mais plutôt descendante, i.e. $F2_centre < F2_initial$, pour les labiaux. L'on en tire que la représentation graphique de $F2_initial$ en fonction de $F2_centre$, en fonction des types de closant, donnerait les *bilabiaux plutôt sous la première bissectrice*, et les *coronaux et les palato-vélaires plutôt au-dessus* de cette ligne.

Nous avons donc décidé de comparer nos résultats acoustiques de la simulation du comportement de babillage à partir du corpus à 7 mois de Matyear et collègues (section IV.1.b.) aux données de dispersion formantique des vocalisations de babillage à 10 mois publiées par Sussman et coauteurs⁴⁵.

De l'émergence du contrôle local des contacts dans le formalisme du locus

D'après Abry *et al.* (2003), dans l'émergence du contrôle du système articulaire, la maîtrise du cycle mandibulaire est suivie d'un début d'émancipation des articulateurs portés par la mâchoire pour le contrôle local des contacts, permettant ainsi la production d'un premier contenu segmental de type consonantique (chapitre 1, section III, point *iii*). Bien que cette perspective dépasse le cadre de cette thèse, nous voulons proposer quelques éléments de modélisation pour la simulation de l'indépendance des mouvements de la pointe de la langue et des lèvres, dans le formalisme du locus.

Nous avançons l'hypothèse globale que la *cible* du contrôle consonantique est la situation où l'*articulateur caractéristique* de chaque type de contact, i.e. les lèvres pour le labial, la pointe de la langue pour le coronal et la masse de la langue pour le palato-vélaire, s'est complètement affranchi de l'*articulateur porteur* (la mandibule), le contact tendant à devenir le produit *exclusif* du mouvement de son articulateur spécifique. Précisons que la notion de cible est, ici, à prendre au sens d'*objectif idéal jamais atteint*, sachant que, en réalité, la mâchoire reste partie prenante de la production des consonnes.

En des termes très simplifiés, l'on peut supposer que l'individu passe donc d'une production des lieux labial, coronal et palato-vélaire par un geste mandibulaire pur, à la différenciation de la production de ces contacts par l'action exclusive de leur articulateur caractéristique respectif (sans intervention de la mandibule) quelle que soit la configuration articulaire vocale.

Afin de simuler l'atteinte de cette cible, dans le cas des contacts bilabiaux et coronaux, le pur geste mandibulaire de la modélisation du début du babillage canonique a été remplacé par un *pur geste labial* et un *pur geste apical* pour la production de contacts, respectivement, labiaux et coronaux, à partir des configurations articulaires résultant de l'inversion des vocants de 7 mois de Matyear et collègues.

b. Un mouvement exclusivement mandibulaire pour le babillage canonique

A partir des simulations de la Section IV.1.b., nous avons estimé la valeur du second formant juste avant closure, i.e. avec des valeurs de $A1$ et/ou Ac égale-s à 0.01 cm^2 . La Figure 4-25A donne la répartition de $F2_initial$ en fonction de $F2_centre$ simulée pour chaque type de closant⁴⁶. Les données de Sussman *et al.* sont présentées sur la même figure pour comparaison (Figure 4-25B).

⁴⁵ Les premières données obtenues pour les lieux bilabial, coronal et vélaire disponibles dans la publication de Sussman *et al.* correspondent aux vocalisations de leur sujet à 10 mois.

⁴⁶ Sur la figure, seule 11 simulations par vocalisation (tirée au hasard) sont représentées (au lieu de 33) pour rendre les résultats plus lisibles.

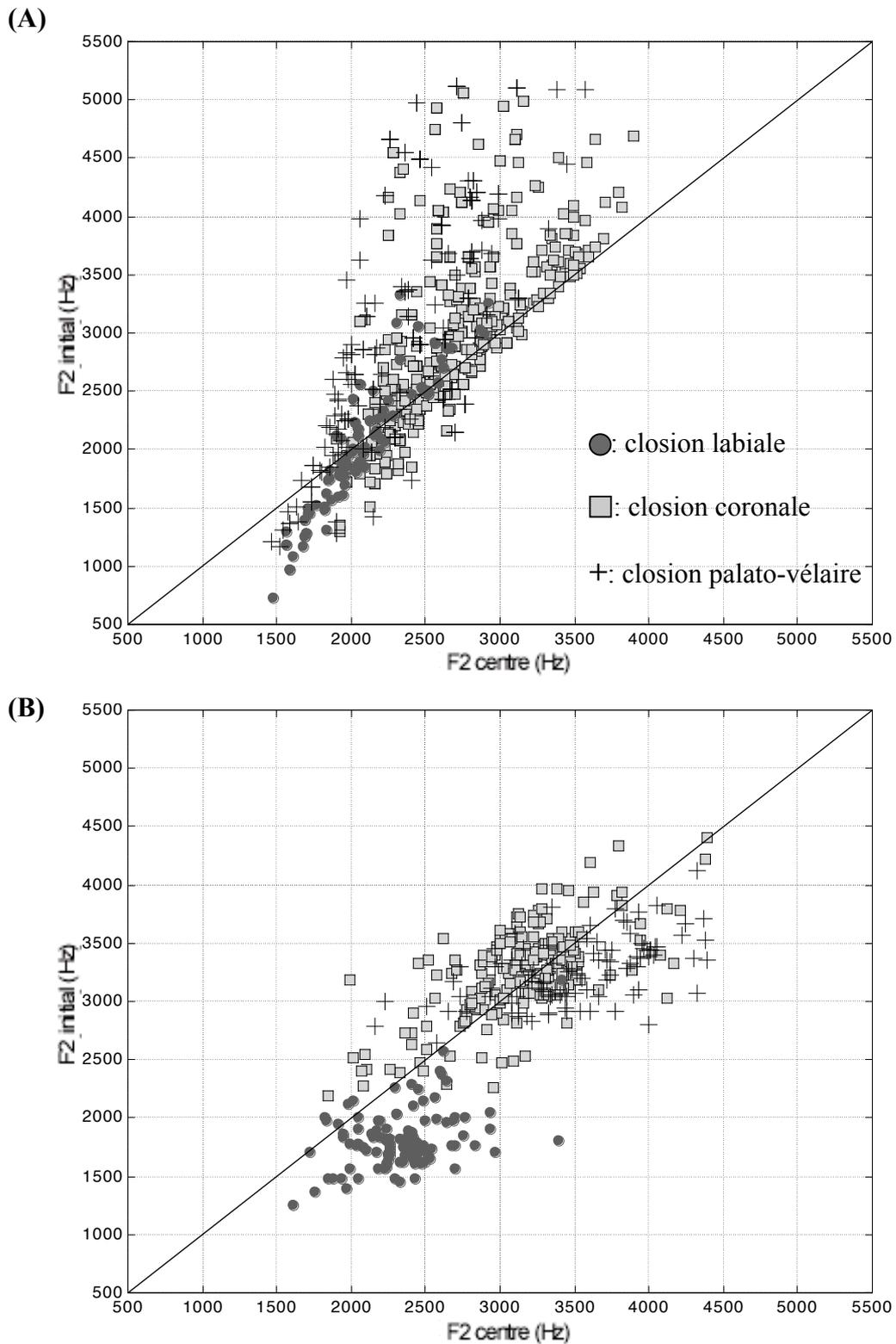


Figure 4-25: Répartition de F2_initial en fonction de F2_centre pour chaque type de closure (A) dans le cas de la simulation de la cyclicité mandibulaire à partir des données de Matyear et collègues à 7 mois, et (B) dans les données de Sussman *et al.* à 10 mois.

Les simulations sont raisonnablement cohérentes en termes de tendance avec les données de ces auteurs pour les closants des types labial et coronal. Le groupe des labiaux est associé à des valeurs de F2 basses au niveau du closant et du vocant, avec des valeurs à l'ouverture généralement inférieures aux valeurs vocaliques, i.e. les données se situent principalement en-dessous de la première bissectrice sur la figure. La situation est radicalement inverse pour les F2 consonantiques et vocaliques de valeurs plus élevées, i.e. la majorité de ces points se trouvent au-dessus de la diagonale. Ces deux groupes de données sont compatibles avec les connaissances classiques sur les articulations labiales et coronales, pour lesquelles l'on s'attend effectivement à observer une valeur de F2 à l'ouverture, respectivement, plus faible et plus élevée que la valeur de F2 vocalique correspondante.

Cependant, la situation est plus complexe pour le groupe des palato-vélaires. En accord avec les statistiques présentées Figure 4-15 (section IV.1.b.), les simulations mènent à ce type de closure à partir de la région des vocants postérieurs ayant des valeurs faibles de F2, comme prédit par la théorie Frame/Content, mais aussi à partir de la région des vocants de F2 élevés, contrairement aux attentes théoriques. Cela est vraisemblablement dû au fait que l'élévation de la mandibule des configurations vocaliques antérieures mène à un rapprochement entre le palais et toute la partie buccale de la langue: il peut en résulter des closures autant coronales que palatales, en fonction de la position de la constriction Xc la plus faible juste avant fermeture par rapport à la valeur limite de 6 cm à la glotte, choisie comme frontière entre ces deux catégories de closants dans nos simulations⁴⁷. Toutefois, ce résultat n'est pas en contradiction avec les données (réelles) de Sussman *et al.* puisque, dans leur corpus, les closants de type [g] ont lieu pour les vocants (antérieurs) de F2 élevés et non pour les postérieurs.

Dans l'ensemble, les résultats de nos simulations ne sont pas très éloignés des données de la réalité. En particulier, l'on observe bien les trois principaux points attendus: (i) la répartition des données par groupes de valeurs de F2_centre selon le lieu, ce qui est caractéristique de la notion de cooccurrence, (ii) le regroupement des données autour de la première bissectrice, et (iii) des valeurs de F2_initial plutôt inférieures à F2_centre pour les contacts labiaux, et plutôt supérieures pour les contacts coronaux. Cela montre que la dispersion des données de Sussman *et al.* (1999) à 10 mois est plus ou moins compatible avec ce à quoi peut mener un pur mouvement vertical de la mandibule, i.e. ce que la théorie Frame/Content postule être la première étape de la production de couples closant-vocant. L'écart entre, d'une part, les prédictions de la théorie Frame/Content et, d'autre part, les données de Sussman *et al.* et nos résultats de simulations pourrait s'expliquer, justement, par l'existence de lieux d'articulation ambigus et de variabilités inter- et intra-individuelles dans les productions réelles⁴⁸.

c. Le contrôle de contacts consonantiques par un pur geste labial et un pur geste apical

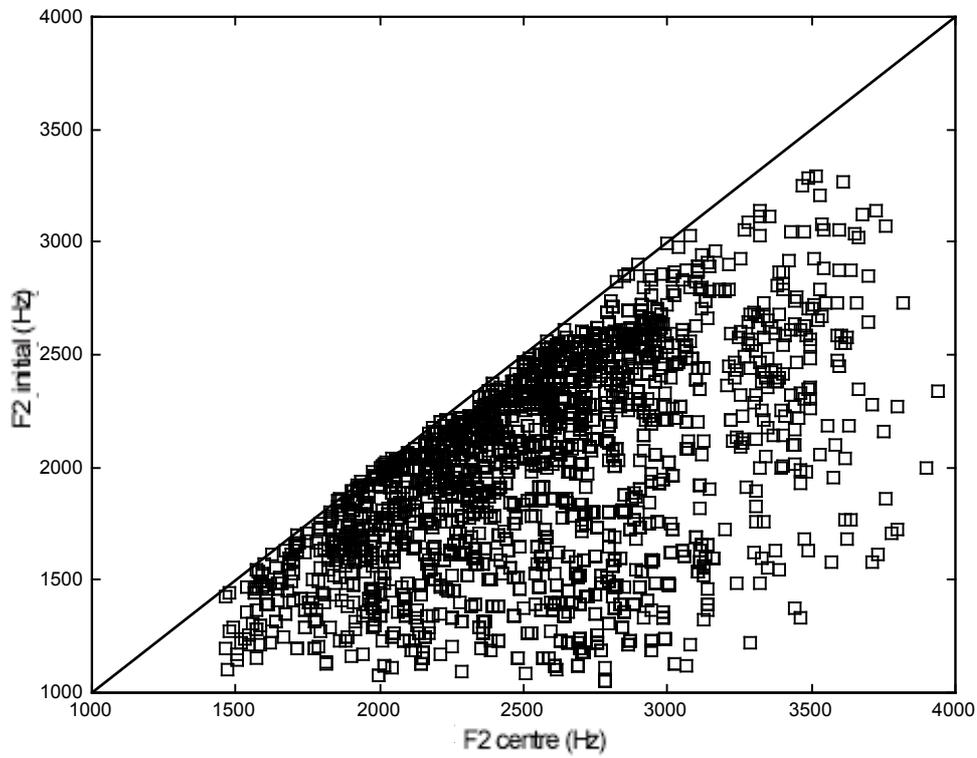
Nous avons monté la pointe de la langue (par la commande TT) seule et fermé les lèvres (par la commande LH) seules de configurations articulatoires résultant de l'inversion des vocants de 7 mois de Matyear *et al.* (section IV.1.b), pour modéliser le cas idéal de la production contrôlée des contacts de types, respectivement, coronal et labial. Comme dans le cas du pur geste mandibulaire, nous avons estimé la valeur du second formant juste avant closure, i.e.

⁴⁷ La Figure 4-19 (section IV.1.b.) donne un exemple du même ordre. Ce vocant est antérieur d'après sa transcription. Bien que la montée mandibulaire mène à un premier contact coronal, le suivi de la coupe sagittale révèle un écrasement de la langue contre le palais, et le suivi des transitions formantiques montre un patron de type palato-vélaire, avec un F2 descendant et un F3 montant.

⁴⁸ Rappelons que nos simulations ont montré qu'une même étiquette transcriptionnelle pouvait cacher des réalités différentes au niveau de la forme du conduit vocal (section IV.1.b., un point de vue en profondeur).

avec une valeur de A_l , pour les contacts labiaux, et une valeur de A_c , pour les contacts coronaux, égale à 0.01 cm^2 . La Figure 4-26 donne la répartition de $F2_initial$ en fonction de $F2_centre$ simulée pour chacun de ces deux types de closants.

(A)



(B)

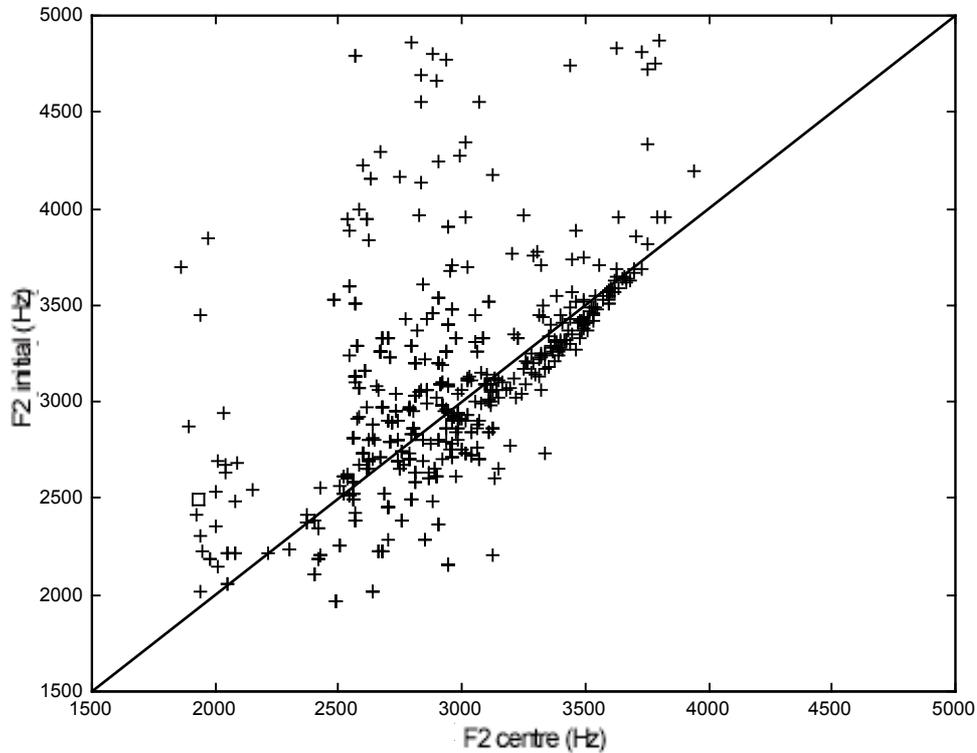


Figure 4-26: Répartition de $F2_initial$ en fonction de $F2_centre$ pour les contacts labiaux par une fermeture des lèvres seules (A) et pour les contacts coronaux par une levée de la pointe de la langue seule (B) à partir des configurations articulatoires résultant de l'inversion des données de Matyear et collègues à 7 mois.

Les patrons de dispersion obtenus ici sont radicalement différents de celui résultant de la simulation d'un geste mandibulaire pur. En effet, le regroupement des points autour de la diagonale en amas relativement liés au type de vocant (Figure 4-25) laisse la place à deux groupes de points très séparés, respectivement en-dessous (4-26A) et au-dessus (4-26B) de la diagonale, chacun d'eux couvrant l'essentiel de la gamme des valeurs de F2_{centre}. Il vient qu'il y a effectivement possibilité d'articuler en association avec toute voyelle soit une consonne labiale, soit une coronale, et que ces deux lieux d'articulation consonantique sont bien différenciés.

La réalité est bien sûr éloignée de l'atteinte de ces cibles (voir Sussman *et al.*, 1991). Néanmoins, le chemin du développement du contrôle des contacts consiste à passer progressivement d'un geste purement mandibulaire, qui ne permet pas de contrôler séparément les closants et vocants contigus, à des gestes *contrastés*, qui impliquent séparément les lèvres ou la pointe de la langue en fonction du contact désiré. De plus, la "dispersion maximale" représentée Figure 4-25, et interprétée par Abry (2004) dans le cadre de la Théorie de la Dispersion-Focalisation (TDF), est remplacée en situation réelle par une "dispersion suffisante", vraisemblablement moins efficace au niveau perceptif mais moins coûteuse au niveau articuloire, comme l'a proposé Lindblom dans sa théorie de la variabilité adaptative (Lindblom, 1990).

V. DISCUSSION

Dans l'ensemble, ces analyses suggèrent une exploration articuloire qui s'étend de 4 à 7 mois, indépendamment de l'effet de la croissance du conduit vocal sur l'acoustique des productions, avec une implication plus marquée de la mandibule en début de babillage (7 mois) qu'à la période de développement antérieure (4 mois). Toutefois, l'ensemble de nos inférences sont à prendre avec précaution. En effet, elles se basent sur l'appariement entre des corpus acoustiques de vocalisations de nourrissons et les productions d'un modèle articuloire du conduit vocal en croissance, VLAM. Ainsi doivent-ils être considérés comme plutôt préliminaires, sachant la difficulté que représente l'ajustement de productions réelles à celles d'un modèle. Dans ce qui suit, ces résultats seront d'abord mis en lien avec les connaissances disponibles dans le domaine du développement de la production de la parole, pour apprécier leur validité. Puis, les intérêts et les limites spécifiques à notre approche d'analyse par modélisation des vocalisations infantiles seront discutés.

1. Le portrait cohérent d'une exploration articuloire croissante

Les résultats des cadrages acoustiques et articuloires indiquent que l'exploration à 4 mois est de relativement antérieure à centrale, restant voisine de la configuration articuloire neutre, qui est généralement considérée comme étant la forme du conduit vocal au repos. Cette exploration impliquerait au moins trois dimensions articuloires avec au minimum un paramètre de langue, tandis que la mandibule semble jouer un rôle mineur. En effet, la commande de mâchoire ne paraît pas être indispensable pour expliquer une grande partie de la dispersion acoustique des données à 4 mois. Globalement, le volume d'exploration articuloire à 4 mois représente de 8 à 13% de l'espace total à 4 dimensions articuloires disponible quand les paramètres LH, J, TB et TD varient selon leur domaine de valeur le plus grand ([-3; +3] *e-t*). A 7 mois, l'exploration est plus étendue qu'auparavant, occupant de 36 à 58% de l'espace complet à 4 dimensions, et la mandibule joue désormais un rôle prédominant, menant ainsi à une exploitation marquée des contrastes ouvert-fermé dont les corrélats acoustiques se lisent sur l'axe du premier formant de l'espace vocalique. Ce portrait est cohérent avec les données de transcription de la littérature (chapitre 1, section I.1.c.) et vient

renforcer la remise en question des postulats d'exploration aléatoire et d'emblée maximale de Jakobson que de nombreuses approches de modélisation de l'acquisition du langage, antérieures, ont repris (e.g. Laboissière, Schwartz et Bailly, 1991; Markey, 1994; Guenther, 1995; Bailly, 1997). Nous reviendrons en discussion générale sur l'intérêt d'une exploration croissante des configurations possibles du système articulatoire au cours du développement de la production de la parole.

Il est important de rappeler, à ce point de l'exposé, que les deux corpus étudiés correspondent à des conditions expérimentales assez différentes: les vocalisations sont plutôt exogènes dans le cas des données à 4 mois, tandis qu'elles sont plutôt endogènes à 7 mois. Toutefois, l'imitation vocale (dans le corpus de 4 mois) pourrait avoir encouragé les individus à explorer de la manière la plus large possible leur répertoire articulatoire-acoustique vu que les cibles auxquelles ils ont été exposés correspondent aux configurations articulatoires extrêmes du conduit vocal humain. Il vient que l'exploration inférée à partir du corpus de pré-babillage (i) pourrait être une sur-estimation de ce que les mêmes sujets produisent spontanément, et donc (ii) doit être vue comme une approximation de l'espace maximal explorable à 4 mois. De là, le degré d'expansion du domaine d'exploration entre les corpus de 4 et 7 mois de la présente étude est vraisemblablement une *sous-estimation* de ce qu'il est "spontanément" dans la réalité. De plus, la seconde partie de ce mémoire montre que l'imitation orofaciale est présente très précocement au cours du développement. Il est très probable que les productions de babillage, supposées purement spontanées, soient en réalité le résultat de la combinaison d'une cyclicité mandibulaire endogène et de l'apprentissage sensori-moteur d'une information extraite de la parole ambiante avant même la période du babillage. En effet, les voyelles /i a u/ sont fréquentes en anglais (milieu linguistique des sujets de Matyear et collègues) et l'imitation peut prendre une forme différée.

En ce qui concerne le corpus de 7 mois, le fait qu'il soit dédié au comportement de babillage explique vraisemblablement l'importance du paramètre mandibulaire pour en rendre compte dans le cadrage articulatoire. Or, le mouvement de la mâchoire est justement la caractéristique essentielle de cette période développementale. Ainsi nos résultats s'accordent-ils assez bien avec les données de transcription qui montrent que la dimension "verticale" explique une plus grande part de la variance des productions de babillage que la dimension "horizontale" (Davis et MacNeilage, 1995).

La simulation de la cyclicité mandibulaire à partir des données de 7 mois, menant aux patrons de cooccurrence dans le formalisme transcriptionnel de la théorie Frame/Content fournit un portrait plus ou moins en accord avec les prédictions et données avancées par la théorie. L'écart le plus flagrant réside dans le fait que VLAM ne produise pas de closants labiaux en association avec les vocalisations étiquetées [a], catégorie vocalique centrale, bien que ce patron de cooccurrence soit théoriquement attendu. Cependant, une explication qui s'impose est celle de la morphologie du conduit vocal dont les coupes radiographiques ont servi à la construction du modèle. En effet, Vilain *et al.* (1999) ont montré que des modèles articulatoires, ayant des morphologies de palais légèrement distinctes, produisent des patrons de cooccurrence de types distincts, tels que [ba], [da] ou [bda], juste parce que, lorsque la mandibule est levée à partir d'une position centrale, le premier contact peut tout aussi bien avoir lieu entre les lèvres qu'entre la pointe de la langue et le palais antérieur. De plus, nos simulations illustrent l'importance de la prise en compte du bruit biologique, mais aussi qu'une même transcription consonantique peut cacher des réalités différentes.

La répartition des valeurs du second formant à l'ouverture mandibulaire en fonction des valeurs dans la portion vocalique trouvée par Sussman *et al.* (1999) a été simulée sur VLAM. Il en découle que leurs données de 10 mois s'accordent relativement bien avec les simulations des séquences vocant-closant par un geste purement mandibulaire, compatible avec les

prédictions de la théorie Frame/Content. L'étude de Sussman et collègue montre que, plus tard au cours du développement, ces nuages de dispersion s'étendent progressivement selon l'axe du second formant de la portion vocalique pour toutes les catégories de vocants ([b d g]). Cette tendance indique une autonomisation progressive des contenus vocaliques et consonantiques vis-à-vis du cycle mandibulaire, avec l'implication d'articulateurs spécifiques à chaque catégorie d'occlusives, tels que LH pour les closions labiales, TT pour l'élévation de la pointe de la langue dans les coronales.

2. Les intérêts et les limites de notre analyse par modélisation

Sachant la difficulté que représente l'acquisition de données articulatoires sur des nourrissons, voire l'estimation de paramètres de phonétique acoustique au cours du développement précoce de la production de la parole (Van der Stelt, Wempe et Pols, 2003), il est incontestablement utile d'être capable d'apparier des données réelles aux simulations d'un modèle articulatoire-acoustique du conduit vocal, en particulier si l'on considère l'importance des mécanismes de croissance et le problème que pose leur prise en compte dans l'analyse phonétique (voir Ménard *et al.*, 2004).

Toutefois, il est raisonnable de ne pas perdre de vue que VLAM présente les limitations classiques de bien d'autres modèles de production de parole. Il ignore la variabilité inter-individuelle de la morphologie du conduit vocal, ce qui s'avère important notamment dans la compréhension de l'écart entre les patrons de cooccurrence prédits par la théorie Frame/Content et ceux obtenus par la simulation du comportement de babillage (Vilain *et al.*, 1999; Vilain, 2000). Il lisse des détails morphologiques et linéarise ou simplifie les changements tridimensionnels dus à la croissance dans les structures articulatoires que sont la mandibule, la langue, les lèvres et le larynx (Ménard, 2000). Néanmoins, cette étude ainsi que celles qui l'ont précédée montrent que ces simplifications n'ont pas d'effets drastiques sur son réalisme, vu qu'elles mènent à des patrons acoustiques congruents avec ceux de données réelles (Ménard, 2002; Ménard *et al.*, 2004). L'utilisation de ce modèle, qui est, à ce jour, le seul modèle statistique capable de traiter des productions de la naissance à l'âge adulte, reste donc pertinente.

Une des critiques qui pourraient être faites à cette étude est l'utilisation de degrés de liberté mis en évidence chez l'adulte pour analyser des vocalisations de nourrissons. Bien sûr, ce procédé est discutable: somme toute, personne ne connaît exactement les degrés de liberté du conduit vocal à la naissance, et la manière dont ils évoluent au cours du développement. L'intérêt de notre analyse est de ne faire aucune hypothèse *a priori* sur ces données articulatoires absentes de la littérature, et de considérer des données acoustiques dans les termes d'une exploration articulatoire progressive d'un conduit vocal en croissance. Un moyen de simuler au mieux les propriétés acoustiques et articulatoires de la parole est de tirer profit d'un modèle computationnel basé sur des analyses statistiques de coupes sagittales d'un conduit vocal réel. Cela nécessite des techniques multidimensionnelles telles que les analyses factorielles dont fait partie l'analyse en composantes principales (ACP). Toutefois, même s'il était possible de suivre le même protocole que celui dont est issu le SMIP avec un nourrisson pour producteur de profils de conduit vocal (chapitre 2, section I), l'on risquerait d'obtenir un nombre de facteurs articulatoires inférieur à celui trouvé chez l'adulte, puisque seule une partie des configurations possibles du conduit vocal humain aurait été représentée dans les données brutes. Par exemple, si le modèle avait été extrait d'un jeu de coupes sagittales de sujets âgés de 4 mois, il est probable que la structure mandibulaire n'aurait pas été correctement représentée. Ainsi, afin de simuler le développement de la parole vu comme un phénomène continu au cours du temps, le modèle de production choisi doit-il impérativement comprendre dès le départ l'ensemble des paramètres de commande qui sont

présents à l'âge adulte. En d'autres termes, le modèle de conduit vocal infantile n'a d'autres choix que de représenter tous les paramètres articulatoires inférés de la parole adulte si l'on veut éviter le postulat irréaliste de l'émergence de nouvelles structures articulatoires au cours de la croissance.

En partant de ce constat, le seul moyen de compenser le fait d'utiliser des paramètres de commande adultes est de simuler leur participation *partielle* à la production en contraignant le domaine de variation des facteurs articulatoires de VLAM, comme le fait le cadrage articulatoire, de manière à sélectionner le jeu de variables motrices le plus probable sachant les vocalisations réelles. En définitive, ce travail plaide en faveur de l'hypothèse selon laquelle les changements observés sur la première année de vie dans les inventaires phonétiques préverbaux ne peuvent s'expliquer uniquement par la croissance du conduit vocal : ils expriment aussi des modifications dans l'*exploitation articulatoire* de l'appareil vocal.

Dans l'ensemble, cette étude nous renforce dans l'idée que VLAM est un modèle de production en croissance *réaliste*, d'après la cohérence que montre (i) la correspondance trouvée entre les simulations acoustiques de VLAM et les données acoustiques réelles dans le cadrage articulatoire, mais aussi (ii) la confrontation des connaissances rencontrées dans la littérature avec les résultats du cadrage articulatoire et de la simulation du comportement de babillage.

Conclusion

Ce chapitre propose un paradigme original: c'est la première fois qu'un modèle articulatoire-acoustique basé sur des propriétés de la production de la parole révélées statistiquement et intégrant la non-linéarité et non-uniformité de la croissance du conduit vocal est exploité dans l'étude de réalisations acoustiques *réelles* de nourrissons pour caractériser leurs habiletés articulatoires. Un aspect important de notre stratégie de modélisation est qu'elle permet de séparer, dans les changements de l'espace acoustique au cours du développement, les variations dues au processus de croissance de celles dues à des modifications de l'exploration articulatoire.

Notre analyse par modélisation suggère que l'exploration articulatoire tend à s'étendre entre 4 à 7 mois, et que la mandibule joue un rôle mineur avant le début du babillage canonique, pour devenir dominant dès le début de cette période développementale. Les prédictions de cooccurrences closant-vocant de la théorie Frame/Content ont été testées par la simulation du comportement de babillage d'après VLAM, en exploitant les configurations de langue inférées à partir des données de 7 mois, fournissant ainsi une estimation des pré-configurations (pressettings) postulées par la théorie. Les simulations articulatoires et acoustiques obtenues sont plus ou moins en accord avec les données réelles de cooccurrence, bien qu'une réplique exacte soit impossible, au moins à cause des variations inter-individuelles dans la morphologie du conduit vocal et intra-individuelles dans la réalisation d'un même son.

Les résultats obtenus ici fournissent les premières spécifications motrices de notre agent. En effet, il devra expérimenter une exploration en deux étapes. A 4 mois, dans une phase de pré-babillage, il exploitera les trois paramètres moteurs (un pour les lèvres et deux pour la langue) les plus aptes à reproduire les vocalisations des individus de 4 mois. A partir de 7 mois, la période de babillage canonique sera principalement caractérisée par la superposition de la cyclicité mandibulaire sur les pré-configurations du conduit vocal dont une partie se trouve dans le répertoire précédent, comme ce qu'il semble se produire dans la réalité. Cette exploration graduelle sous-tendra le processus d'apprentissage d'une représentation sensori-motrice plus ou moins précise des habiletés du robot, c'est-à-dire, la correspondance entre les configurations articulatoires de son conduit vocal et les conséquences perceptives auxquelles elles aboutissent. Cette carte s'adaptera aux changements subis par le robot et lui permettra d'imiter des sons de parole perçus dans son environnement selon ses capacités perceptuo-motrices en cours, permettant ainsi au développement du contrôle de son conduit vocal de se diriger vers ce dont il a besoin pour parler.

Le chapitre suivant propose une implémentation du mécanisme d'imitation pour réfléchir sur les caractéristiques potentielles de ce comportement à 4 mois, tel que rapporté par Kuhl et Meltzoff (1996).

Chapitre 5: Modélisation probabiliste de l'imitation orofaciale avant le babillage

INTRODUCTION

La seconde partie de ce mémoire montre en quoi l'imitation audio-visuelle précoce de modèles orofaciaux est un mécanisme plausible pour l'apprentissage de la parole ambiante, notamment d'après des expériences de psychologie du développement. Ici, nous voulons appuyer cette thèse par la modélisation de ce mécanisme avant le babillage canonique, en montrant, en particulier, que ce comportement ne demande pas nécessairement l'apprentissage d'un grand volume d'information ni une structure de connaissance particulièrement sophistiquée.

L'essentiel de notre argumentaire bibliographique (chapitre 3) s'appuie sur le fait que (i) dès la naissance l'information auditive semble être porteuse d'une information articulatoire, révélant ainsi un précurseur potentiel de l'imitation orofaciale précoce de modèles verbaux (Cf. capacité de donner un équivalent articulatoire en adéquation avec la perception d'un son de parole, section III.A.2.a), (ii) une coordination intermodale dans la perception de configurations orofaciales, dont des éléments de parole, soit vraisemblablement présente dès les deux premiers mois de vie (Cf. capacité de détecter une équivalence visuo-haptique et audio-visuelle, section III.A.2.a et b., respectivement), et (iii) aux environs de 4 mois, des individus montrent une capacité d'imitation de stimuli audio-visuels de voyelles, d'après la transcription phonétique qu'en font des adultes, mais aussi des indices sur une mémoire potentielle des représentations auditive et visuelle de modèles verbaux pour leur imitation "semi-différée" (section III.A.3.).

Dans notre vision du développement de la parole, le comportement d'imitation se base sur une carte de correspondances entre des configurations articulatoires expérimentées par l'individu et leurs conséquences perceptives, avec pour passerelle entre extéroception et motricité le domaine somesthésique : cette représentation cognitive des capacités articulatoires-perceptives du sujet se construit par un apprentissage individuel qui trouve son substrat informationnel dans l'exploration¹. Autrement dit, le domaine d'exploration est l'espace d'où sont tirées les informations sensori-motrices apprises qui servent à la formation et la mise à jour de la carte des liens perception-production sous-jacente à l'imitation orofaciale précoce. Pour interroger les caractéristiques potentielles de la carte cognitive en jeu dans ce comportement, nous proposerons une implémentation du phénomène d'imitation de voyelles à 4 mois dans le formalisme Bayésien (Cf. chapitre 2, section III). En suivant une démarche en accord avec une modélisation réaliste, nous prendrons en considération les capacités de production et les performances rapportées par Kuhl et Meltzoff (1996). Les auteurs ont trouvé que leur sujets âgés de 3 à 4,5 mois produisaient 59% de réponses congruentes avec la cible (voir plus bas, Tableau 5-III). Le chapitre précédent fournit une estimation plausible du domaine d'exploration de leurs sujets de 4 mois en termes de commandes articulatoires, de formes de conduit vocal et de fréquences formantiques. Toutes ou parties de ces données simulées peuvent être utilisées comme base d'apprentissage d'un modèle bayésien de la représentation des correspondances perception-production vraisemblablement à l'œuvre dans l'imitation orofaciale précoce de cibles vocaliques. Nous verrons que deux paramètres sont

¹ L'on retrouve ici les idées de babillage corporel et de représentation production de Meltzoff et Moore (chapitre 3, section II.2.).

supramodale des liens perception-

susceptibles de permettre une évolution de cette structure de connaissance au cours du développement: la quantité d'information à apprendre et la finesse des variables représentées au sein de la carte². A un autre niveau, la question du type de cible sera abordée en fonction (i) des modalités perceptives, auditive *versus* audio-visuelle, de leur présentation, et (ii) de leur appartenance ou non au domaine d'exploration. Mais avant de présenter notre modèle de carte Bayésienne, le plan de simulation que nous avons suivi et les résultats qui en découlent, nous voulons revenir sur un problème qui a été soulevé dans la littérature avant la publication de l'étude de Chen *et al.* (2004, voir chapitre 3, section III.A.2.a). Cette question est relative à la possibilité que les performances en imitation décrites par Kuhl et Meltzoff soient basées sur la capacité d'imitation visuelle néonatale (Studdert-Kennedy, 1993), vue au chapitre 3 (section III.A.1.).

Au sujet de l'estimation des capacités articulatoires à 4 mois

Un élément important de l'ensemble des simulations présentées ici est que les capacités articulatoires à 4 mois prises pour une approximation de l'espace d'exploration ont été estimées par un cadrage articulatoire *partiel* des données de pré-babillage, vues au chapitre présent, où seuls les domaines de variation symétriques des paramètres de commandes non-nuls sont pris en compte, i.e. [-3; 3], [-2; 2], [-1; 1] *e-t*, au lieu des 16 servant à la spécification de l'androïde (chapitre 4, section III.2.a.). Le sous-modèle correspondant montre les domaines de variation suivants: [-2; 2], [-1; 1] et [-2; 2] pour LH, TB et TD, respectivement (J étant de valeur nulle).

La prise en compte du meilleur sous-modèle à 4 mois sur la base d'un cadrage articulatoire avec des domaines de variation symétriques au lieu du meilleur sous-modèle proposé au chapitre 4 s'explique par le fait qu'en réalité les premières analyses par cadrage articulatoire ainsi que la modélisation du phénomène d'imitation ont d'abord été effectuées avec des domaines de variation symétriques uniquement et que ce n'est que suite aux critiques d'un/une reviewer, lors de la soumission d'un article relatant nos travaux d'estimation des capacités articulatoires avant et au début du babillage, que le cadrage articulatoire a été amélioré en incluant les domaines de variation articulatoires asymétriques. Mais, nous n'avons hélas pas eu le temps de refaire les simulations d'imitation avec le meilleur sous-modèle à 4 mois le plus récent dont les domaines de variation pour LH, TB et TD sont respectivement [-2; 1], [-3; 1] et [-1; 3] *e-t* (J étant de valeur nulle). Toutefois, les informations qui seront tirées des présentes simulations semblent, nous le verrons, assez générales et indépendantes du domaine de variation des paramètres articulatoires.

Il vient que, dans ce chapitre, toute référence au meilleur sous-modèle à 4 mois est à comprendre au sens de celui issu de l'analyse par cadrage articulatoire partiel, à bornes symétriques.

² Nous n'aborderons pas la question de l'émergence de l'utilisation de nouvelles commandes articulatoires (TT, LP et Lx): dans la présente étude, le jeu de variables articulatoires pris en considération sera fixe.

I. TEST DE L'IMITATION VISUELLE PURE

Dès la naissance, des individus sont capables d'imiter des gestes spécifiques de la face, i.e. d'apparier stimuli visuels et proprioceptifs (chapitre 3, section III.A.1.). L'imitation de voyelles à 4 mois peut-elle être basée sur des capacités d'imitation purement visuo-motrices?

1. Méthode

Pour tester si l'imitation visuo-motrice pourrait rendre compte des performances de l'expérience de Kuhl et Meltzoff, les valeurs de l'aire aux lèvres (AI) de prototypes [i a u] dans VLAM à 4 mois ont été utilisées comme cibles (Tableau 5-I): elles ont été inversées, de l'espace géométrique à l'espace articulatoire, de manière exhaustive via le modèle résultant du cadrage articulatoire (partiel) des données de pré-babillage. Ces cibles sont donc un équivalent adulte des trois voyelles. Notons que nous supposons résolu, ici comme tout au long de ce chapitre, le problème de la *normalisation visuelle*, c'est-à-dire que nous supposons que les nourrissons savent convertir une cible géométrique adulte en une cible infantile. Bien qu'à notre connaissance le problème de la normalisation n'ait guère été étudié dans le cas de la lecture labiale, l'on peut supposer que la taille du visage est une information disponible et pertinente pour s'acquitter de cette tâche³.

L'inversion des cibles procède comme suit. Une série de simulations (60 000) est générée de manière aléatoire à l'intérieur du domaine d'exploration inféré à 4 mois. Pour chacune des cibles [i a u], les configurations dont la valeur de AI se trouve dans le voisinage⁴ de l'aire aux lèvres de la cible sont sélectionnées⁵. Les sons correspondant à ces simulations, qui ont une aire aux lèvres adéquate mais une forme de langue arbitraire, sont calculés puis étiquetés [i], [a] ou [u] en fonction de la cible la plus proche dans le plan (F1, F2) en terme de distance Euclidienne⁶.

Tableau 5-I: Caractéristiques des cibles utilisées pour la simulation de l'imitation purement visuelle.

	AI (cm ²)	F1 (Hz)	F2 (Hz)
[i]	1.28	580	4243
[a]	2.15	1415	2509
[u]	0.07	568	1519

1. Résultats

Globalement, cette expérience de modélisation mène à environ 51% de réponses correctes, ce qui n'est pas très inférieur au score réel de 59%. Cela suggère que l'imitation faciale pourrait en partie expliquer les résultats de l'imitation vocale précoce obtenus par Kuhl et Meltzoff.

³ Plus précisément, l'estimation de la taille du visage de l'autre pourrait, par exemple, se baser sur la distance séparant les yeux, sachant que les yeux sont un point de focalisation de l'attention de nouveau-nés.

⁴ Ce voisinage est défini comme étant les simulations dont l'aire aux lèvres est à une distance inférieure ou égale à 0.05 cm² de l'aire aux lèvres de la cible.

⁵ Il s'avère que, bien que ces cibles soient extérieures aux domaines d'exploration articulatoire et acoustique du sous-modèle servant de base à l'inversion, leur aire aux lèvres est incluse dans le domaine d'exploration géométrique

⁶ Le calcul de cette distance se base sur les valeurs formantiques exprimées en Bark (voir chapitre 2, section II.4.).

Cependant, vu qu'il n'y a aucune raison de supposer que les nourrissons effectuent une imitation visuelle *parfaite* de l'aire aux lèvres, le score de cette expérience a de fortes chances d'être inférieur à 51%. De plus, la matrice de confusion de ce test fournit un profil typique de confusion visuelle (Tableau 5-II): de nombreuses confusions ont lieu entre [i] et [a] et entre [i] et [u], mais aucune n'a lieu entre [a] et [u] dont les aires intero-labiales sont les plus distantes l'une de l'autre. Or, ce profil est assez différent du patron réel (Tableau 5-III). Donc, l'information visuelle seule et l'imitation visuo-motrice ne semblent pas suffisantes pour expliquer les données d'imitation vocale précoce observée à 4 mois.

Tableau 5-II: Matrice de confusion résultant de la simulation de l'imitation purement visuelle. Les effectifs sont normalisés de manière à ce que les totaux par colonne correspondent aux nombres de présentation par cible de Kuhl et Meltzoff.

Cibles Réponses	[i]	[a]	[u]	Total
[i]	28	40	28	96
[a]	39	55	0	94
[u]	0	0	34	34
Total	67	95	62	224

Tableau 5-III: Matrice de confusion résultant de l'expérience d'imitation menée par Kuhl et Meltzoff (1996). Ces données sont produites par l'ensemble de leurs sujets, âgés de 3 à 4,5 mois, et sont issues de la Figure 5 de la publication (p. 2433, dans Kuhl et Meltzoff, 1996).

Cibles Réponses	[i]	[a]	[u]	Total
[i]	22	11	4	37
[a]	25	66	14	105
[u]	20	18	44	82
Total	67	95	62	224

II. IMPLEMENTATION DE LA CARTE DES CORRESPONDANCES PERCEPTION-PRODUCTION

Il est probable que, par les vocalisations qu'ils expérimentent depuis leur naissance, les sujets de 4 mois aient commencé à se construire une carte perceptuo-motrice du comportement de leur conduit vocal. Une telle connaissance acquise pourrait sous-tendre l'imitation orofaciale de modèles vocaliques, puisqu'elle permet d'inférer la configuration motrice qui peut générer l'état sensoriel équivalent à la cible perçue. Telle est l'hypothèse de travail de ce qui suit.

La programmation Bayésienne a été utilisée dans la modélisation de l'imitation précoce de cibles auditives et audiovisuelles de type [i a u]. Dans ce cadre de travail, l'agent apprend une carte sensori-motrice du comportement de son conduit vocal qui correspond à une description probabiliste des liens observables entre ses variables motrices et perceptives. Puis, l'imitation se rapporte à l'inversion d'un état sensoriel en un équivalent articulo-motriciel. Pour exposer la structure de cette carte, nous reprenons globalement la méthodologie de son élaboration (voir chapitre 2, section III). Celle-ci s'effectue en deux phases. Dans un premier temps, une phase de description permet de spécifier les connaissances préalables du système (section II.1.) et d'identifier, par apprentissage, les paramètres libres (section II.2.) de la structure de dépendance issue de la spécification : l'on obtient une description effective du système. Dans un second temps, la phase d'utilisation permet la mise en œuvre de la description pour la simulation des processus d'imitation auditive et d'imitation audio-visuelle (section III).

1. Spécification des connaissances préalables à l'apprentissage

a. Choix des variables motrices et perceptives

Les paramètres moteurs ont été sélectionnés pour leur capacité à rendre compte des vocalisations recueillies à 4 mois: ce sont ceux du meilleur sous-modèle de pré-babillage⁷. Le choix des variables motrices est donc la séparation des lèvres (LH), et les commandes du corps (TB) et du dos (TD) de la langue.

Les variables auditives sont les deux premières fréquences formantiques (F1, F2) exprimées en Bark (échelle de perception des fréquences, voir chapitre 2, section II.4.).

Une hypothèse importante dans la présente implémentation est le rôle de pivot joué par les variables géométriques, qui sont (Xh, Yh), i.e. les coordonnées du point le plus haut de la langue dans un référentiel fixe sur la coupe sagittale du conduit vocal⁸, et A1, l'aire aux lèvres. Ces variables présentent à nos yeux plusieurs intérêts. D'abord, elles sont des produits possibles du système somesthésique comme nous l'avons vu au chapitre 2, section II.4.. Ensuite, elles fournissent des descripteurs compacts entre variables articulo-motricielles et variables auditives, en étant bien contrôlées par l'articulation, et ce sont de bons prédicteurs des formants du spectre (Boë *et al.* 1992). Elles devraient permettre à ce titre de passer à l'avenir à des modélisations articulo-motricielles plus complexes, à 4, 5, 6 ou 7 paramètres de VLAM. Enfin, comme nous avons pu le voir dans la seconde partie de ce mémoire, l'information somesthésique est susceptible d'entrer en jeu dans le phénomène d'imitation orofaciale de modèles de parole par le fait qu'elle constitue un intermédiaire entre les domaines extéro-perceptif et articulo-motriciel. Notons d'ailleurs que des expériences de perturbation par bite-block menées par Lindblom, Lubker et Gay (1979) aboutissent précisément au rôle clé de ces variables dans les processus de contrôle.

⁷ Voir Introduction de ce chapitre, *Au sujet de l'estimation des capacités articulo-motricielles à 4 mois.*

⁸ Voir chapitre 4, Figure 4-10. Nous avons choisi Xh et Yh plutôt que Xc et Ac parce que la courbe de Xc en fonction de la configuration de la langue présente des discontinuités, contrairement aux courbes de Xh et Yh (voir Boë *et al.*, 1995b).

Ainsi notre implémentation comporte-t-elle 3 variables articulatoires LH, TB, TD; trois variables géométriques (somesthésiques) Xh, Yh, Al; et deux variables auditives F1, F2. Notons que (F1, F2) sont fournies par les production vocaliques de soi et d'autrui. Et que, dans le cas de l'imitation visuelle, Al est également susceptible d'être extraite de l'observation d'une cible labiale.

Le domaine de variation de chaque variable correspond à l'espace exploré par le sous-modèle sélectionné à l'issue du cadrage articulatoire partiel. De plus, toutes les variables sont supposées discrètes.

Plus précisément, F1, F2, LH, TB et TD ont 16 états possibles, avec un pas de discrétisation constant⁹. Xh, Yh et Al sont, elles aussi, discrétisées de manière régulière, mais Al est discrétisées selon son *logarithme népérien*¹⁰. Le nombre d'états des variables géométriques correspond à 4 profils de discrétisation possibles: {16, 16, 8}, {8, 8, 4}, {4, 4, 2} et {2, 2, 1} pour le nombre de classes de {Xh, Yh, Al}. Ces profils correspondent à des Degrés de Discrétisation Géométrique (DDG) distincts, qui sont ici classés par ordre décroissant. A chacun correspondra une "carte des correspondances" perception-production se distinguant des autres par la finesse de sa représentation de la forme du conduit vocal. En effet, le domaine de variation des variables géométriques étant fixe, le pas de discrétisation est d'autant plus petit que le nombre d'états possibles de chaque variable est grand. De là, le DDG dicte l'écart minimal nécessaire pour distinguer deux éléments dans le domaine géométrique discrétisé, ou mieux le degré d'acuité somesthésique.

Nous avons donc fait varier les DDG sur une large gamme de précision, correspondant pour le premier profil à un nombre total de 2048 états pour le triplet (Xh, Yh, Al), et pour le dernier profile, à un nombre total de 4 états possibles. Nous avons choisi de prendre un nombre égal d'états sur Xh et Yh et un nombre moitié pour Al, en considérant que cette dernière variable joue un rôle plus faible que les deux autres dans la détermination des valeurs formantiques.

a. Décomposition de la probabilité conjointe¹¹

Pour définir la structure de dépendance sensori-motrice, i.e. la charpente de la carte des correspondances perceptuo-motrices, nous prenons pour point de départ la probabilité de conjonction de toutes les variables à notre disposition :

$$P(LH, TB, TD, Xh, Yh, Al, F1, F2).$$

L'espace des variables géométriques est à la charnière entre l'articulatoire et l'acoustique : ces variables permettent d'une part d'estimer les formants et contraignent d'autre part les commandes articulatoires. Elles peuvent donc prendre le rôle clef de variables indépendantes:

$$\begin{aligned} &P(LH, TB, TD, Xh, Yh, Al, F1, F2) \\ &= P(Xh, Yh, Al) * P(LH, TB, TD, F1, F2 / Xh, Yh, Al) \\ &= P(Xh, Yh, Al) * P(LH, TB, TD / Xh, Yh, Al) * P(F1, F2 / Xh, Yh, Al, LH, TB, TD) \end{aligned}$$

⁹ Toutefois, dans nos simulations des mécanismes d'imitation, la discrétisation de F1 et F2 n'entre pas en jeu.

¹⁰ Le but de cette discrétisation non-linéaire est de donner plus de poids aux variations sur les petites valeurs de l'aire aux lèvres que sur ses valeurs élevées, de manière à rendre compte de la courbe de sensibilité des paramètres acoustiques à ce descripteur de la forme du conduit vocal (Cf. Stevens et House, 1955).

¹¹ Nous rappelons que toutes nos démonstrations utilisent les règles de conjonction et de marginalisation de l'inférence Bayésienne, figurant chapitre 2, section III.

Nous pouvons faire l'hypothèse simplificatrice classique que la dépendance entre les variables acoustiques et les motrices disparaît sous la condition de connaître les valeurs des paramètres géométriques. Il vient que:

$$\begin{aligned} & P(\text{LH, TB, TD, Xh, Yh, Al, F1, F2}) \\ &= P(\text{Xh, Yh, Al}) * P(\text{LH, TB, TD} / \text{Xh, Yh, Al}) * P(\text{F1, F2} / \text{Xh, Yh, Al, LH, TB, TD}) \\ &= P(\text{Xh, Yh, Al}) * P(\text{LH, TB, TD} / \text{Xh, Yh, Al}) * P(\text{F1, F2} / \text{Xh, Yh, Al}). \end{aligned}$$

De plus, nous faisons les hypothèses d'indépendances conditionnelles suivantes :

H_1 : L'aire aux lèvres (Al) est indépendante de la forme de la langue donnée par Xh et Yh, et la position du plus haut point de la langue en ordonnée (Yh) est indépendante de celle en abscisse (Xh) ;

H_2 : La variable auditive F2 est indépendante de F1 sachant les valeurs des variables géométriques Xh, Yh et Al;

H_3 : La commande labiale, LH, est indépendante de la forme de la langue et donc de Xh et Yh;

H_4 : La commande du corps, TB, et du dos, TD, de la langue sont indépendantes de l'aire aux lèvres, Al, et de la commande labiale, LH.

H_1 implique que :

$$\begin{aligned} P(\text{Xh, Yh, Al}) &= P(\text{Xh, Yh}) * P(\text{Al} / \text{Xh, Yh}) = P(\text{Xh}) * P(\text{Yh} / \text{Xh}) * P(\text{Al} / \text{Xh, Yh}) \\ &= P(\text{Xh}) * P(\text{Yh}) * P(\text{Al}). \end{aligned}$$

H_2 entraîne que :

$$\begin{aligned} P(\text{F1, F2} / \text{Xh, Yh, Al}) &= P(\text{F1} / \text{Xh, Yh, Al}) * P(\text{F2} / \text{Xh, Yh, Al, F1}) \\ &= P(\text{F1} / \text{Xh, Yh, Al}) * P(\text{F2} / \text{Xh, Yh, Al}). \end{aligned}$$

D'après H_3 ,

$$\begin{aligned} P(\text{LH, TB, TD} / \text{Xh, Yh, Al}) &= P(\text{LH} / \text{Xh, Yh, Al}) * P(\text{TB, TD} / \text{Xh, Yh, Al, LH}) \\ &= P(\text{LH} / \text{Al}) * P(\text{TB, TD} / \text{Xh, Yh, Al, LH}). \end{aligned}$$

Selon H_4 ,

$$\begin{aligned} P(\text{TB, TD} / \text{Xh, Yh, Al, LH}) &= P(\text{TB} / \text{Xh, Yh, Al, LH}) * P(\text{TD} / \text{Xh, Yh, Al, LH, TB}) \\ &= P(\text{TB} / \text{Xh, Yh}) * P(\text{TD} / \text{Xh, Yh, TB}). \end{aligned}$$

Ainsi, la *structure de dépendance* qui définit le comportement sensorimoteur de l'androïde est-elle :

$$\begin{aligned} & P(\text{LH, TB, TD, Xh, Yh, Al, F1, F2}) \\ &= P(\text{Xh}) * P(\text{Yh}) * P(\text{Al}) \\ &\quad * P(\text{LH} / \text{Al}) * P(\text{TB} / \text{Xh, Yh}) * P(\text{TD} / \text{Xh, Yh, TB}) \\ &\quad * P(\text{F1} / \text{Xh, Yh, Al}) * P(\text{F2} / \text{Xh, Yh, Al}). \end{aligned}$$

a. Paramétrisation des distributions composant la structure de dépendance

Les distributions de Xh, Yh et Al sont considérées uniformes¹². L'on suppose que toutes les variables articulatoires et auditives suivent des lois Gaussiennes conditionnelles¹³ dont les

¹² Soit G une des trois variables géométriques, $P(G) = \frac{1}{\text{Nombre_de_Cas_Possibles}}$.

paramètres libres, i.e. les moyennes et variances, sont à estimer (ou *identifier*) dans la phase d'apprentissage.

1. Identification des paramètres libres de la structure de dépendance par apprentissage

Principe

Pour devenir une *description* sensori-motrice effective du comportement de l'agent, i.e. une carte donnant les correspondances observables, les distributions dont est composée la structure de dépendance nécessitent d'être apprises à partir d'un jeu de données dites "expérimentales". La base d'apprentissage est générée par un tirage aléatoire, selon une loi uniforme, des paramètres de commande de VLAM dans les limites de variation définies par le sous-modèle choisi comme représentatif des vocalisations à 4 mois. Ces configurations articulatoires sont injectées dans VLAM et produisent les valeurs des variables Xh, Yh, Al, F1 et F2, pour les simulations jugées recevables¹⁴. Autrement dit, les données expérimentales sont tirées du domaine articulatoire-géométrico-acoustique d'exploration estimé à 4 mois.

Nous avons testé 15 Tailles de Base d'Apprentissage (TBA) variant de 1 à 60 000 exemplaires¹⁵, pour la structure de dépendance associée à chacun des 4 DDG : l'on obtient un jeu de 60 descriptions possibles (4 DDG x 15 TBA).

Puisque les distributions articulatoires et auditives sont conditionnées par les variables géométriques, les DDG dictent la *taille de l'espace d'apprentissage*, i.e. le nombre de distributions articulatoires et auditives à apprendre pour que la description représente l'ensemble du domaine d'exploration à 4 mois. Par exemple, dans le cas du DDG {16, 16, 8} il faut calculer 2048 moyennes et 2048 variances pour décrire l'ensemble des 2048 distributions de F1 dans l'espace {Xh, Yh, Al}¹⁶ (Figure 5-1). En d'autres termes, la taille de l'espace d'apprentissage suit celle de l'espace géométrique qui comprend 2048, 256, 32 et 4 conditions pour les DDG de {16, 16, 8}, {8, 8, 4}, {4, 4, 2} et {2, 2, 1}, respectivement. Ainsi existe-t-il, de fait, un lien entre la taille de l'espace d'apprentissage, i.e. le nombre de distributions à apprendre, et la finesse de la représentation de l'espace des formes du conduit vocal.

Or, le calcul de la moyenne et de la variance d'une distribution nécessitent au moins deux valeurs. Donc, dans le cas des variables articulatoires et auditives, pour une condition donnée, il faut qu'au moins deux exemplaires de la base d'apprentissage s'inscrivent dans cette condition pour que les paramètres de la distribution qui lui est associée soient effectivement calculés. Il vient que le DDG dicte le nombre de distributions à apprendre pour que la description représente l'ensemble du comportement du système: il contraint la TBA nécessaire au calcul des paramètres de toutes les distributions présentes dans la structure de dépendance. Nous verrons que les performances du système dépendent d'un compromis (*trade-off*) entre le degré de discrétisation des variables géométriques (DDG) et la taille du jeu de données expérimentales (TBA).

¹³ La formule de la gaussienne conditionnelle est:
$$p(V_i = v_i / V_j = v_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}}} \exp\left(-\frac{(v_i - m_{ij})^2}{2\sigma_{ij}^2}\right)$$
, où

m_{ij} , σ_{ij} sont respectivement les moyenne et écart-type de la variable V_i sachant la variable V_j .

¹⁴ Comme dans le cas des analyses par cadrage, les simulations recevables ont des valeurs de Ac et Al supérieures ou égales à 0.1 cm² et 0.01 cm², respectivement, pour assurer un régime acoustique viable.

¹⁵ Plus précisément, le nombre d'exemplaires d'apprentissage est de 1, 5, 10, 25, 50, 100, 250, 500, 5 000, 10 000, 20 000, 30 000, 40 000, 50 000 et 60 000. Nous appellerons *base d'apprentissage totale*, l'ensemble formé par les 60 000 simulations issues du domaine d'exploration estimé à 4 mois.

¹⁶ Ce nombre de distributions correspond au produit des nombres de classes des variables conditionnantes: 16*16*8=2048.

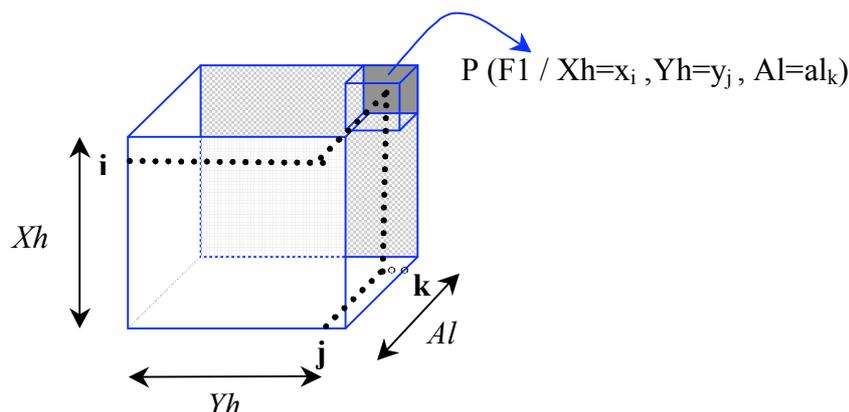


Figure 5-1 : Représentation schématique de l'espace des réalisations de la distribution $P(F1 / Xh=x_i, Yh=y_j, Al=a_k)$. Le petit cube représente l'une des 2048 conditions du type $Xh=x_i$ et $Yh=y_j$ et $Al=a_k$. Les indices i, j et k sont les numéros d'ordre des valeurs discrètes que peuvent prendre respectivement les variables Xh, Yh et Al .

Notons que, dans la présente implémentation du modèle, quand la réalisation d'une condition apparaît moins de deux fois dans la base d'apprentissage, les variables conditionnées par ce cas rare suivent une loi quasi-uniforme¹⁷. Cela équivaut à ôter cette condition du jeu des possibles. Aussi, en théorie, pour avoir une bonne estimation de la variance et de la moyenne de chaque distribution conditionnelle, la TBA doit être au moins 10 fois supérieure au nombre de conditions possibles (Duda et Hart, 1973). Pour illustration, dans le cas des distributions de F1 et de F2, qui sont conditionnée par Xh, Yh et Al , l'on nécessite 40 et 21000 8-uplets ($LH, Tb, Td, Xh, Yh, Al, F1, F2$), respectivement, pour 4 et 2048 conditions possibles.

III. IMPLEMENTATION DE L'IMITATION AUDITIVE ET AUDIOVISUELLE

Globalement, le but est d'apprécier quelle description, parmi les 60 implémentées, est la plus apte à reproduire les performances rapportées par Kuhl et Meltzoff. Après l'apprentissage d'une description définie par un DDG et une TBA donnés, le modèle a été évalué dans sa capacité à simuler l'imitation, i.e. à donner des équivalents articulatoires à des cibles perceptives [i à u], quand ces voyelles sont présentées soit dans la modalité auditive seule (*inversion auditive*), soit dans les modalités auditive et visuelle à la fois (*inversion audiovisuelle*). A chaque type de présentation correspond une question.

1. Des questions Bayésiennes

Dans cette étude, les processus d'imitation sont définis par des questions Bayésiennes de la forme :

$P(LH, TB, TD / Cible_Perceptive)$.

¹⁷ La variance des variables sous cette condition n'étant pas calculable, on la fixe à 10.000: la gaussienne correspondante s'aplatit. Cet acte trouve sa justification dans le fait que la rareté d'une donnée expérimentale la rend suspecte : elle peut être un artefact. Sous la contrainte de représentativité du phénomène modélisé qui pèse sur la base d'apprentissage, uniformiser les distributions sous la condition de cette donnée revient à réduire sa contribution à la description.

➤ Une question pour l'imitation auditive

La question de l'imitation auditive est la suivante: "Quelles configurations articulatoires pourraient avoir pour conséquence le son défini par les valeurs formantiques $F1=f1$ et $F2=f2$ ". Cela revient à tirer de manière aléatoire des valeurs de LH, TB et TD de la distribution suivante¹⁸:

$$P(LH, TB, TD / F1 = f1, F2 = f2).$$

Pour pouvoir répondre à une question, l'inférence Bayésienne a pour principe de reformuler cette question de manière à retrouver la description du système. Il vient:

$$P(LH, TB, TD / F1 = f1, F2 = f2)$$

$$= P(LH, TB, TD, F1=f1, F2=f2) / P(F1=f1, F2=f2)$$

$$= [\sum_{Xh} \sum_{Yh} \sum_{Al} P(LH, TB, TD, Xh, Yh, Al, F1=f1, F2=f2)] / P(F1=f1, F2=f2).$$

où $P(F1=f1, F2=f2)$ est un terme de normalisation, fourni par

$$P(F1=f1, F2=f2) = \sum_{Xh} \sum_{Yh} \sum_{Al} \sum_{LH} \sum_{TB} \sum_{TD} P(LH, TB, TD, Xh, Yh, Al, F1=f1, F2=f2).$$

➤ Une question pour l'imitation audio-visuelle

La question de l'imitation audio-visuelle est la suivante: "Quelles configurations articulatoires pourraient avoir pour conséquence le son défini par les valeurs formantiques $F1=f1$ et $F2=f2$, avec une aire aux lèvres $Al=al$ ". Cela revient à tirer de manière aléatoire des valeurs de LH, TB et TD de la distribution suivante:

$$P(LH, TB, TD / F1 = f1, F2 = f2, Al=al).$$

Cette probabilité conditionnelle se reformule comme suit:

$$P(LH, TB, TD / F1 = f1, F2 = f2, Al=al)$$

$$= P(LH, TB, TD, F1=f1, F2=f2, Al=al) / P(F1=f1, F2=f2, Al=al)$$

$$= [\sum_{Xh} \sum_{Yh} P(LH, TB, TD, Xh, Yh, F1=f1, F2=f2, Al=al)] / P(F1=f1, F2=f2, Al=al).$$

Là encore, $P(F1=f1, F2=f2, Al=al)$ est un terme de normalisation, fourni par

$$P(F1=f1, F2=f2, Al=al)$$

$$= \sum_{Xh} \sum_{Yh} \sum_{LH} \sum_{TB} \sum_{TD} P(LH, TB, TD, Xh, Yh, Al=al, F1=f1, F2=f2).$$

2. Evaluation des capacités d'imitation

Nous avons mené cette évaluation en deux temps. D'abord, nous avons étudié les capacités d'*inversion Bayésienne* proprement dites: pour une cible fournie en entrée et caractérisé par sa paire de formants ($F1, F2$), quel est le son "reproduit" par le modèle à travers une boucle inversion-production, et quel est l'écart de ce son à celui de la cible? Ceci nous a permis d'évaluer le compromis entre le DDG et la TBA, i.e. entre la finesse de la discrétisation

¹⁸ Le tirage aléatoire n'est pas le seul mode possible de sélection de la réponse. En effet, l'on aurait pu choisir les états des variables articulatoires qui maximisent la probabilité de l'articulatoire sachant la cible perceptive. Cette alternative a été écartée parce que les nourrissons ne semblent pas donner une "réponse" identique à elle-même pour les présentations successives de la même cible. Toutefois, lors du tirage aléatoire, les classes de LH, TB et TD les plus probables ont, par définition, une plus grande probabilité d'être sélectionnées que les autres. Notons aussi que, vu que les variables sont discrètes, la valeur qui représente chaque classe de valeur articulatoire, dans la présente étude, est celle du milieu de cette classe.

géométrique (ou somesthésique) et le volume d'apprentissage. Puis, nous avons implémenté un protocole complet de simulation de l'expérience de Kuhl et Meltzoff (1996), afin de comparer simulations et données expérimentales.

a. Etude du compromis entre DDG et TBA pour l'inversion audio-articulatoire

Le lien entre DDG et TBA a été évalué d'après la capacité de l'agent à inverser des simulations issues de la base d'apprentissage totale, pour chacune des 60 descriptions possibles. Ici, la cible est auditive. Pour chaque cible, le tirage sur la distribution correspondant à la question n'est effectué qu'une seule fois. Chaque configuration articulaire résultant de l'inversion est injectée dans VLAM: la sortie acoustique de cette simulation est confrontée aux formants de la cible. L'ensemble des inversions basées sur chaque description a permis de calculer la racine de l'erreur quadratique moyenne respectivement sur F1 et F2 (Figure 5-2)¹⁹.

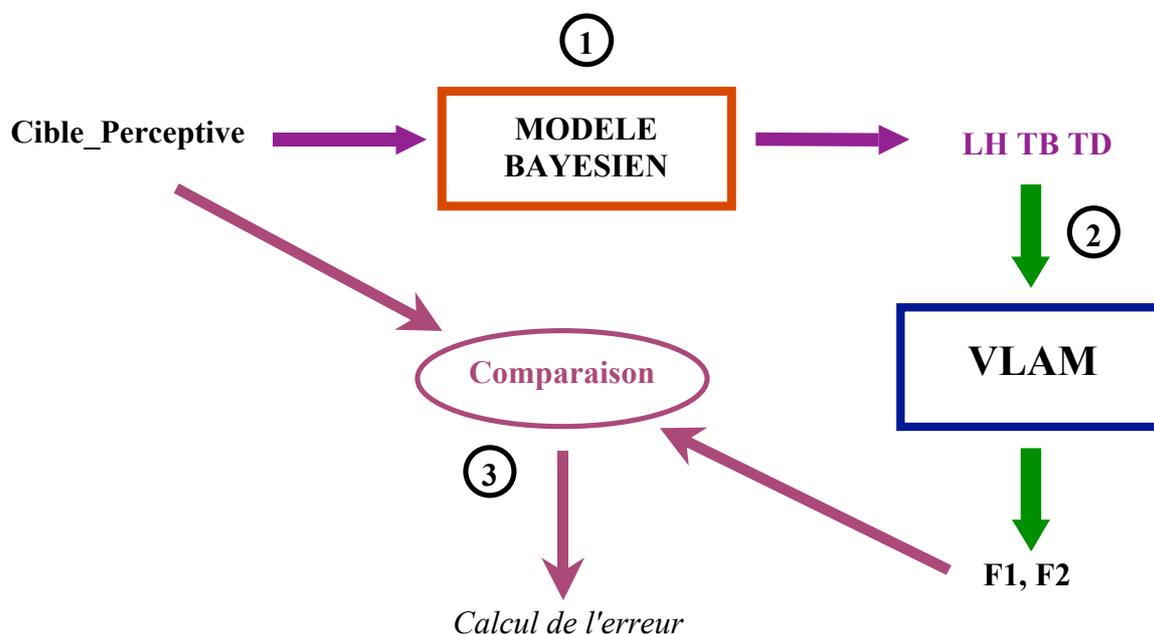


Figure 5-2: Organigramme du processus d'inversion-production. **1.** La cible perceptive est inversée suivant le modèle Bayésien: il en résulte des valeurs de LH, TB et TD. **2.** La configuration articulaire issue de l'inversion Bayésienne est injectée dans VLAM à 4 mois (les valeurs de LP, J, TT et Lx sont nulles). **3.** Les deux premiers formants générés par VLAM sont comparés à ceux de la cible perceptive, permettant le calcul de l'erreur d'inversion.

¹⁹ Toutefois, l'erreur correspondra, ici, à la moyenne des erreurs quadratiques moyennes sur F1 et F2 exprimées en *Bark*, pour donner une dimension perceptive à l'appréciation des « faux-pas » de l'agent. Le choix d'un calcul d'erreur sur les formants plutôt que sur les variables articulatoires trouve sa justification dans les phénomènes compensatoires : une erreur motrice serait difficilement interprétable, vu le nombre de configurations articulatoires pouvant aboutir au même son.

b. Simulation de l'expérience d'imitation de Kuhl et Meltzoff

La simulation complète fonctionne sur le principe suivant. D'abord, l'on se dote d'un ensemble de cibles auditives et audio-visuelles, que nous présenterons plus loin. Pour chacune de ces cibles, le processus d'imitation décrit en III.1. fournit une loi de probabilité adéquate au sens du modèle $P(LH, TB, TD / Cible_Perceptive)$. Nous effectuons alors un tirage de 300 exemplaires selon cette loi, afin de l'exploiter de manière un peu extensive, et d'explorer ses variations articulatoires. Pour chaque tirage, donnant un triplet de valeurs articulatoires (LH, TB, TD), nous calculons par VLAM la paire formantique (F1, F2) qui en résulte: ceci correspond à la simulation des productions infantiles. Il nous reste alors à modéliser le processus de catégorisation phonétique par l'expert: le son est classé [i] vs. [a] vs. [u] selon la cible la plus proche en termes de distance Euclidienne dans l'espace des formants (F1, F2) exprimées en Bark (Figure 5-3). A partir des 3 x 300 réponses obtenues (3 cibles x 300 tirages par cible), l'on peut alors établir des matrices de confusion entre classes d'entrée (celles des cibles données au départ) et classes de sortie (fournies par le processus de catégorisation). Toutefois, nous nous concentrerons surtout sur les scores d'imitation correcte, obtenus par la moyenne des 3 termes diagonaux de chaque matrice.

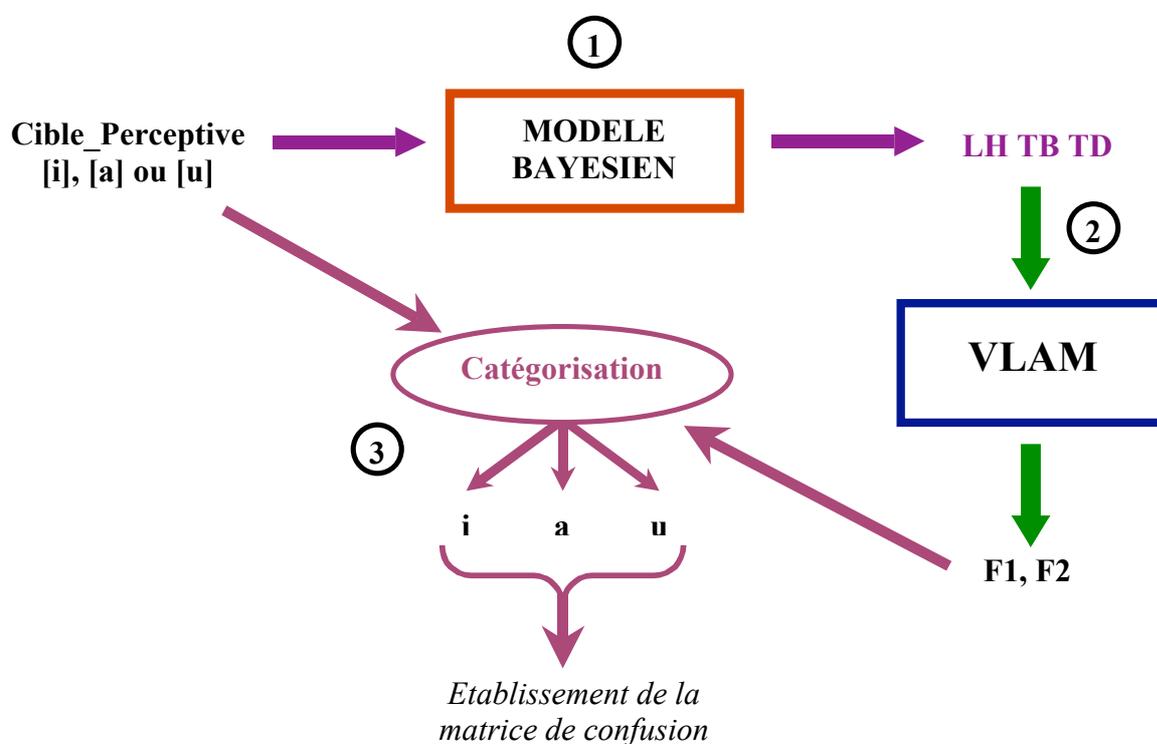


Figure 5-3: Organigramme de la simulation de l'imitation des sons [i], [a] et [u]. 1. La cible perceptive est inversée suivant le modèle Bayésien: il en résulte des valeurs de LH, TB et TD. 2. La configuration articulatoire issue de l'inversion Bayésienne est injectée dans VLAM à 4 mois (les valeurs de LP, J, TT et Lx sont nulles). 3. Les deux premiers formants générés par VLAM sont comparés à ceux de la cible perceptive, pour leur catégorisation en fonction de la cible perceptive la plus proche et l'établissement de la matrice de confusion.

Nous avons utilisé deux types de cibles. Les *cibles externes* correspondent aux prototypes acoustiques de [i a u] dans VLAM à 4 mois qui s'avèrent être en dehors du domaine d'exploration, estimé à 4 mois, aux niveaux articulatoire et acoustique. Les *cibles internes* sont des simulations tirées de chacun des trois coins de l'espace vocalique (F1, F2) du

domaine d'exploration estimé à 4 mois: elles sont vues comme des voyelles [i a u] dans les limites des capacités de 4 mois. Autrement dit, les cibles internes sont des configurations issues de zones potentiellement connues, car incluses dans la base d'apprentissage totale, tandis que les cibles externes sont obligatoirement nouvelles, ou inconnues (Figure 5-4, Tableaux 5-IV et 5-V). La comparaison des performances de l'agent selon qu'il imite des cibles internes ou des cibles externes nous a permis d'apprécier l'effet du paramètre d'appartenance des cibles aux capacités articulatoire-acoustiques, sachant que les sujets de Kuhl et Meltzoff imitent des voyelles qui sont exclues de leurs habiletés motrices d'après notre analyse de cadrage partiel de leurs productions.

Pour les deux types de cibles, nous avons considéré soit une cible auditive pure, associée alors à un processus d'imitation auditive, soit une cible audiovisuelle, associée à un processus d'imitation audiovisuelle. En effet, si nous avons montré dès le départ que l'imitation précoce de voyelle n'est vraisemblablement pas une imitation visuelle pure, la question de savoir si l'imitation est auditive ou audiovisuelle n'est pas clairement tranchée.

Notons que, une fois de plus, nous supposons résolu le problème de la normalisation. Dans le cas de la normalisation auditive, les propositions sont multiples. Elles impliquent en général f_0 . Par exemple, les travaux de Ménard *et al.* (2002) suggèrent que les distances $F1-f_0$ et $F2-F1$ (exprimées en Bark) fournissent des paramètres perceptifs à forte capacité de normalisation. Ainsi, en remplaçant à la fois les formants des cibles adultes et les formants produits par le modèle juvénile par leurs valeurs normalisées, l'on pourrait produire des simulations très acceptables d'imitation infantile de cibles adultes.

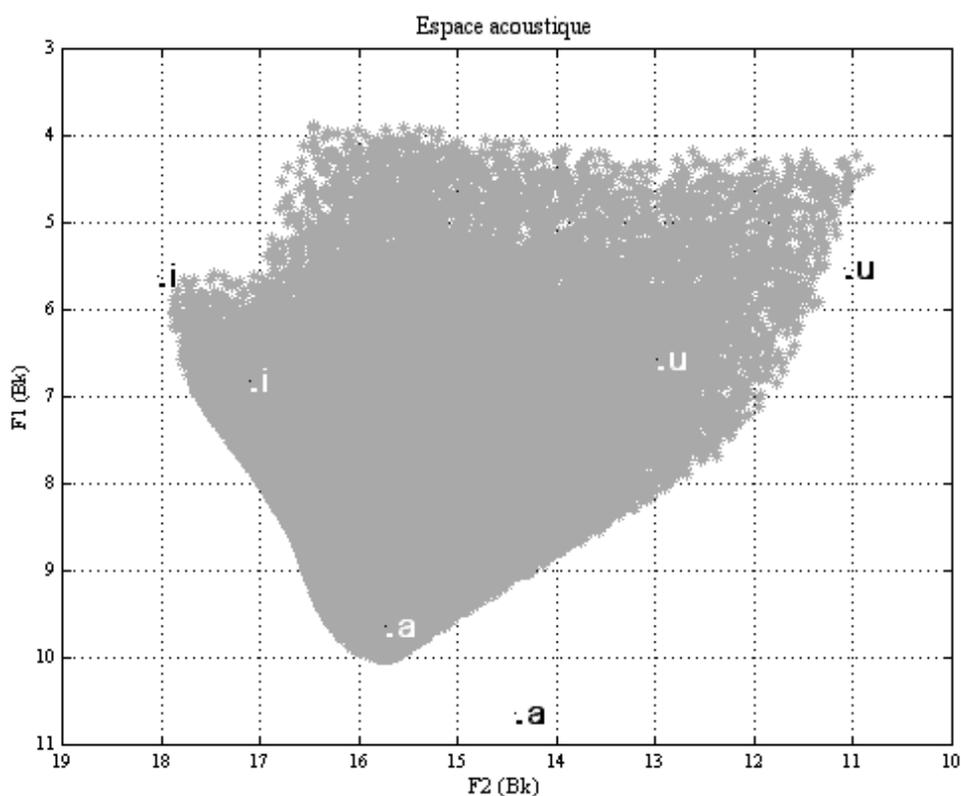


Figure 5-4: Représentation graphique dans le plan des deux premiers formants exprimés en Bark des cibles [i a u] internes (en blanc) et externes (en noir) utilisées pour la simulation de l'imitation par le modèle Bayésien. L'espace acoustique complet d'apprentissage du modèle est représenté en gris. Ces simulations correspondent à l'espace d'exploration du meilleur sous-modèle à 4 mois (voir Introduction).

Tableau 5-IV: Caractéristiques des cibles internes ou connues utilisées dans la simulation Bayésienne de l'imitation

	AI (cm ²)	F1 (Hz)	F1 (Bk)	F2 (Hz)	F2 (Bk)
[i]	1,24	738	6,82	3715	17,11
[a]	2,20	1204	9,63	3039	15,73
[u]	0,09	704	6,57	2025	12,98

Tableau 5-V: Caractéristiques des cibles externes ou inconnues utilisées dans la simulation Bayésienne de l'imitation

	AI (cm ²)	F1 (Hz)	F1 (Bk)	F2 (Hz)	F2 (Bk)
[i]	1,28	580	5,62	4243	18,03
[a]	2,15	1415	10,6405	2509	14,42
[u]	0,07	568	5,5251	1519	11,09

IV. RESULTATS

1. Inversion Bayésienne et compromis DDG-TBA

La Figure 5-5 illustre les résultats obtenus pour l'inversion auditive de 1000 exemplaires choisis de manière aléatoire dans le domaine d'exploration estimé à 4 mois.

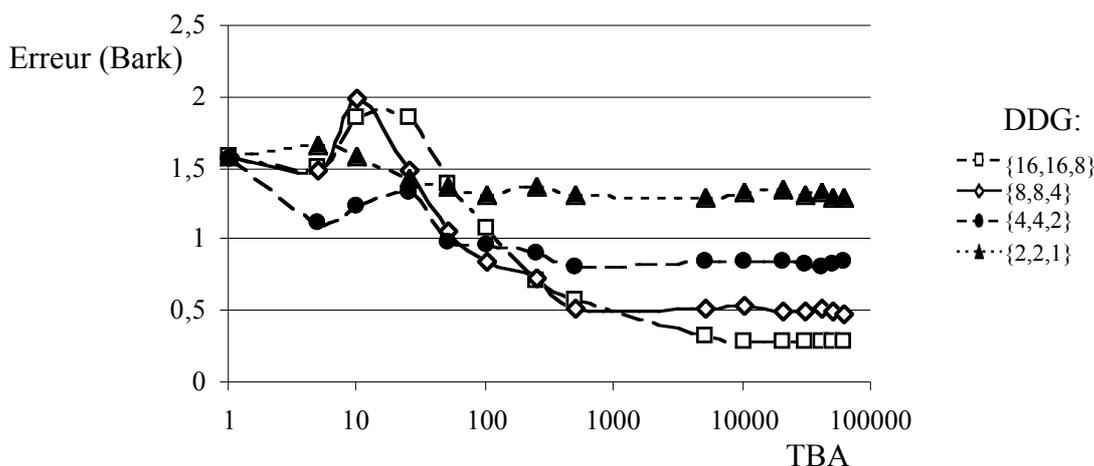


Figure 5-5: Représentation graphique de l'erreur formantique moyenne exprimée en Bark, en fonction de la TBA, pour chaque DDG, à l'issue de l'inversion Bayésienne de 1000 exemplaires tirés aléatoirement de la base d'apprentissage totale.

Quand la valeur de TBA est maximale, plus la discrétisation de l'espace géométrique est élevée (DDG croissant), plus l'erreur est faible : cette erreur atteint des valeurs inférieures à $0,5 Bk^{20}$ pour les deux DDG les plus élevés ($\{8, 8, 4\}$ et $\{16, 16, 8\}$). De plus, pour un DDG donné, l'erreur tend à diminuer avec l'augmentation de la TBA, jusqu'à la limite la plus basse que ce DDG permet à l'agent d'atteindre. Autrement dit, à partir du moment où le nombre de données expérimentales permet à un grand nombre de distributions de la description d'être correctement estimées, l'apprentissage d'exemplaires supplémentaires n'apporte plus de modification majeure à la description.

Toutefois, tous les DDG sauf le plus bas, i.e. $\{2,2,1\}$, mènent à des valeurs erratiques tant que la TBA est inférieure à une certaine valeur qui signe la fin de la *phase de sous-apprentissage* du DDG correspondant. Le sous-apprentissage est dû au nombre relativement faible de distributions articulatoires et auditives effectivement apprises. Dans ce cas, peu de cases dans les espaces de conditions trouvent un nombre de simulations supérieur à 1 nécessaire à estimer les paramètres statistiques des variables qu'elles conditionnent. Il vient que les configurations motrices associées à ces quelques cases apprises sont invariablement choisies par l'agent, indépendamment de leur manque de pertinence vis-à-vis de la cible à inverser et de leur répartition erratique dans l'espace total d'exploration²¹.

Pour chaque DDG, nous avons estimé la TBA minimale nécessaire à l'obtention d'une erreur moyenne 10% supérieure à l'erreur moyenne la plus faible que l'on puisse atteindre (obtenue pour les plus grandes valeurs de TBA). On trouve que cette TBA, qui nous avons associé à la fin du sous-apprentissage, croît de manière relativement linéaire avec la taille de l'espace géométrique en nombre de cases, en ligne avec les recommandations de Duda et Hart (1973) (Figure 5-6).

En d'autres termes, plus les cases sont nombreuses dans l'espace géométrique (plus le DDG est élevé), *plus précises* sont ses variables, et *plus efficace* peut être le processus d'inversion Bayésienne, mais *plus la TBA doit être grande* pour que la carte de l'agent soit représentative de ses capacités sensori-motrices.

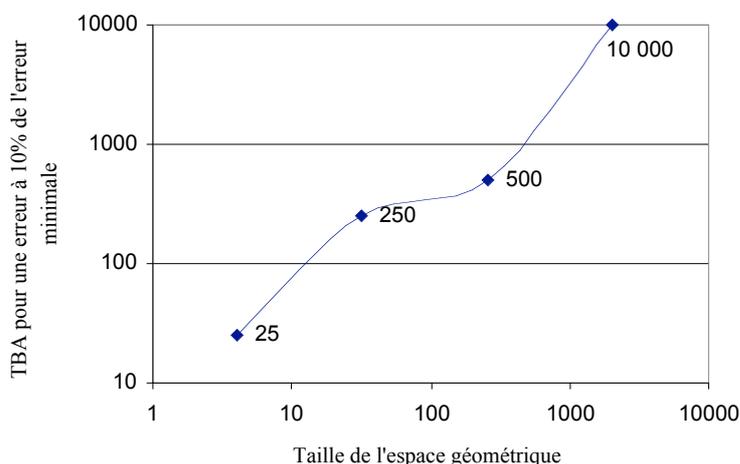


Figure 5-6: Valeur de la TBA minimale pour avoir une erreur à moins de 10% de l'erreur minimale en fonction de la taille de l'espace géométrique.

²⁰ Bien que la capacité de discrimination minimale de l'oreille adulte soit de $0,1 Bk$, les erreurs obtenues ici restent faibles.

²¹ En effet, la variance des distributions qui leur sont associées est inférieure à la valeur donnée par défaut (10 000).

2. Résultats de la simulation de l'expérience de Kuhl et Meltzoff

Nous avons mené des simulations sur 60 descriptions (15 TBA x 4 DDG), avec à chaque fois deux types de cibles (externes *vs.* internes) pour chaque type de processus d'imitation (auditive *vs.* audio-visuelle)²².

Les pourcentages de réponses correctes en fonction du DDG et du type d'inversion auditive *versus* audio-visuelle des cibles [i a u] internes et externes sont présentés Figure 5-7 à 5-10. Les scores pour l'inversion auditive (Figures 5-7 et 5-8) sont globalement inférieurs à ceux de l'inversion audio-visuelle (Figures 5-9 et 5-10). De plus, les tendances suivantes apparaissent.

Compromis DDG-TBA et sous-apprentissage

Comme l'on pouvait s'y attendre, quelque soit le cas, la même tendance au compromis entre DDG et TBA qu'en Figure 5-5 a été trouvée. De plus, tous les DDG, sauf le plus petit ($\{2, 2, 1\}$), imposent à l'agent d'avoir appris un nombre supérieur de données avec les cibles externes qu'avec les cibles internes pour sortir de sa phase de sous-apprentissage, que cela soit en inversion auditive ou audio-visuelle.

Cibles internes vs. externes : le biais du sur-apprentissage

Pour une TBA donnée, les cibles externes ont tendance à mener à des scores inférieurs à ceux obtenus pour des cibles internes : ceci est compréhensible sachant que les premières sont à l'extérieur de l'espace vocalique pris pour l'estimation des capacités de 4 mois contrairement aux secondes. Chose surprenante, dans le cas de l'inversion auditive de cibles externes (Figure 5-8), contrairement à celui de cibles internes (Figure 5-7), le DDG le plus élevé, i.e. la taille d'espace géométrique de 2048 cases, ne conduit jamais à un score de 100% d'imitation correcte, au sens de notre critère, même à TBA maximale (60 000 items). Cela est imputable au problème du *sur-apprentissage*.

On peut l'expliquer de la manière suivante. La phase d'apprentissage consiste à apprendre des moyennes et des variances de lois gaussiennes, qui permettent de "paver" l'espace d'apprentissage. En phase de test, si une donnée fait partie de la région apprise (cas de cibles internes), elle tombe dans la zone d'influence d'une case bien apprise, et conduit à une bonne inversion. Par contre, si elle ne fait pas partie de la région apprise (cas des cibles externes), elle est éloignée de toutes les cases apprises. Or, certaines de ces cases se trouvent avoir des variances plus fortes que d'autres²³. C'est alors ce terme de variance qui se met à jouer le rôle clé dans les calculs de probabilité, et ces cases de forte variance - c'est-à-dire qui tolèrent mieux les écarts à la moyenne - sont sélectionnées, malgré leur manque de pertinence pour l'inversion. En d'autres termes, même lorsque la description est précise et très représentative des capacités sensori-motrices de l'agent, quand les cibles sont inconnues, les configurations articulatoires sont sélectionnées indépendamment de leur pertinence sachant le but perceptif : le mieux est l'ennemi du bien. Ce problème est d'autant plus aigu que les cases sont nombreuses - les écarts de variance sont alors importants - et bien apprises - si elles sont apprises sur peu de données, les variances tendent à mieux s'équilibrer, en reflétant plutôt le hasard des données, que la taille intrinsèque des cases. Ainsi, le sur-apprentissage est réduit lorsque les cases de l'espace des conditions sont "larges", de manière à assurer que la disparité des variances ne prenne pas le pas sur celle des écarts de la cible à la moyenne. En accord avec notre compréhension de ce modèle d'imitation, l'on trouve que les DDG grossiers

²² Cela donne, au final, 240 matrices de confusion, chacune étant établie sur la base de 900 inversions.

²³ Par exemple, sur F1, l'on trouvera pour certaines conditions géométriques des écarts-types de l'ordre de 10⁻² tandis que pour d'autres l'écart-type sera de l'ordre de l'unité. Cela vient notamment des effets de la transformation de l'articulatoire à l'acoustique propre à VLAM.

({4,4,2} et {8,8,4}) ont des scores plus élevés et moins instables que le DDG le plus fin ({16,16,8}) quand les cibles sont inconnues.

Autrement dit, le DDG *doit* donner un nombre assez faible de (grandes cases) pour que l'agent soit capable d'imiter des sons vocaliques qui se trouvent hors de ses habiletés sensori-motrices. Toutefois, le problème du sur-apprentissage est moins aigu si l'information visuelle est elle aussi fournie (Figure 5-10) : comme les valeurs de l'aire aux lèvres pour les voyelles [i a u] de VLAM font partie du domaine d'exploration estimé à 4 mois, l'agent est plus à même de sélectionner des configurations qui produisent des sons proches de la cible.

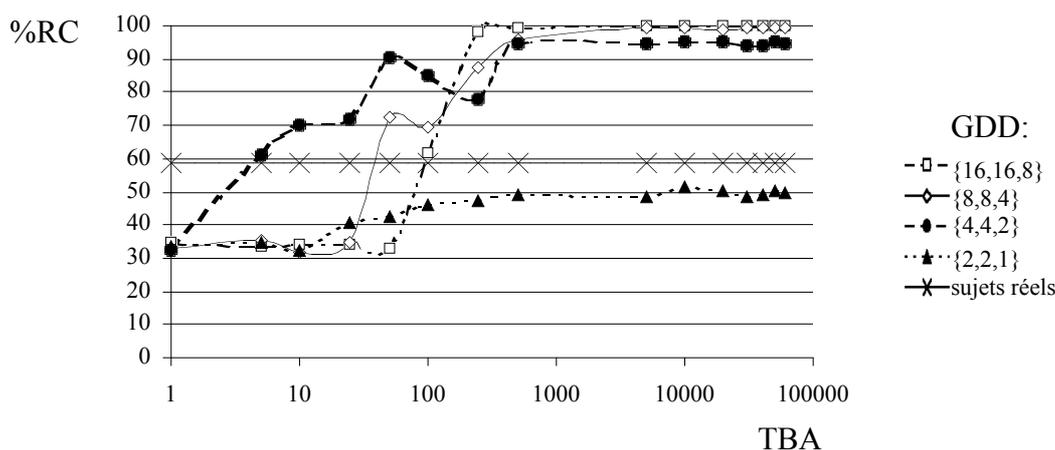


Figure 5-7: Pourcentage de réponses correctes (%RC) pour la simulation de l'imitation auditive des cibles [i a u] internes en fonction de la TBA, pour chaque DDG. La courbe des “sujets réels” correspond aux score obtenus sur l'ensemble des sujets de Kuhl et Meltzoff (1996).

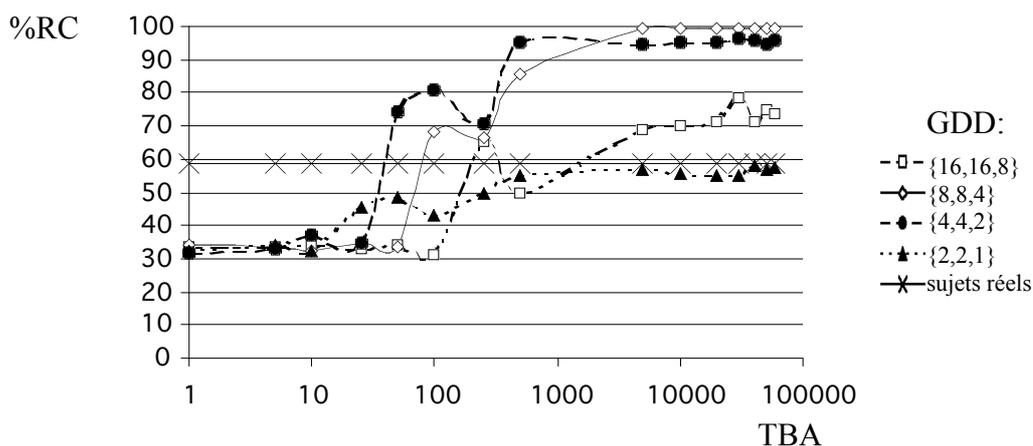


Figure 5-8: Pourcentage de réponses correctes (%RC) pour la simulation de l'imitation auditive des cibles [i a u] externes en fonction de la TBA, pour chaque DDG. La courbe des “sujets réels” correspond aux score obtenus sur l'ensemble des sujets de Kuhl et Meltzoff (1996).

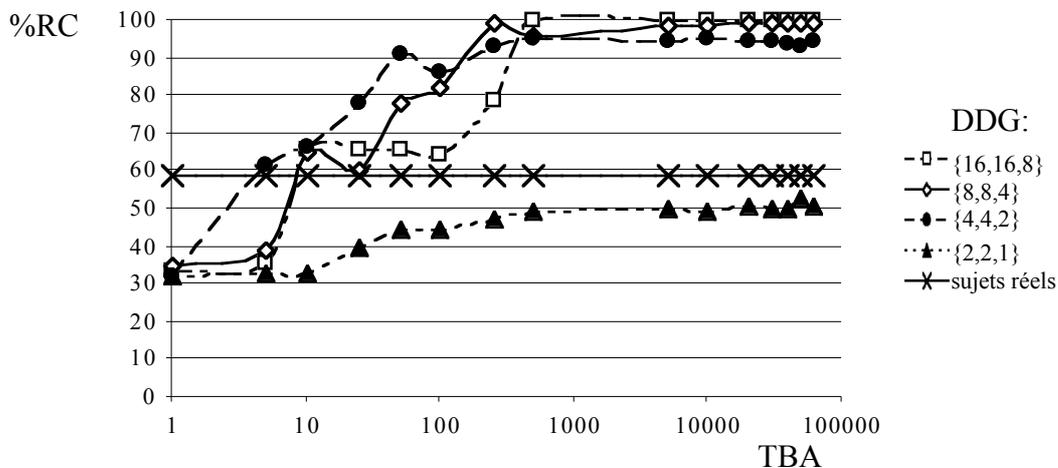


Figure 5-9: Pourcentage de réponses correctes (%RC) pour la simulation de l'imitation audiovisuelle des cibles [i a u] internes en fonction de la TBA, pour chaque DDG. La courbe des “sujets réels” correspond aux score obtenus sur l'ensemble des sujets de Kuhl et Meltzoff (1996).

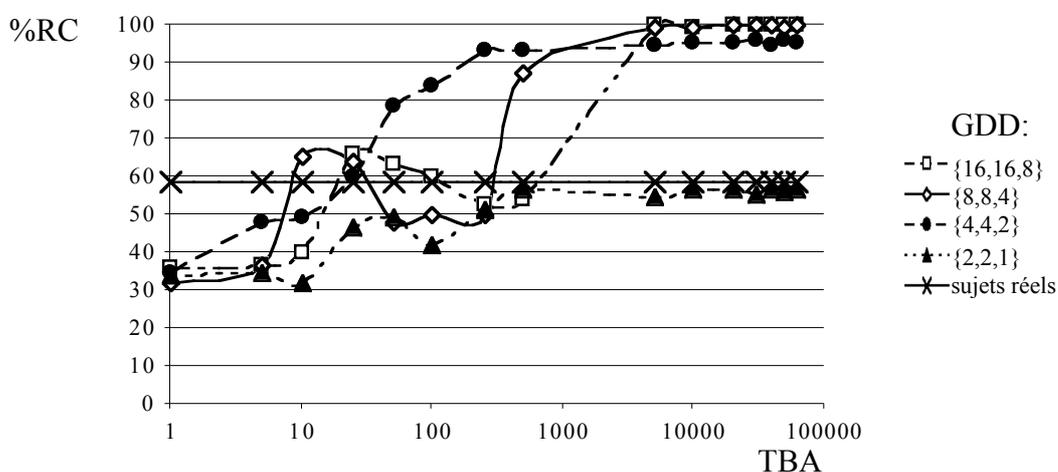


Figure 5-10: Pourcentage de réponses correctes (%RC) pour la simulation de l'imitation audiovisuelle des cibles [i a u] externes en fonction de la TBA, pour chaque DDG. La courbe des “sujets réels” correspond aux score obtenus sur l'ensemble des sujets de Kuhl et Meltzoff (1996).

L'imitation précoce de modèle de parole ne nécessite pas beaucoup d'apprentissage

En définitive, il est important de noter que l'agent n'a besoin ni d'un DDG élevé ni d'une grande TBA, afin de faire aussi bien, si ce n'est mieux, que les nourrissons de Kuhl et Meltzoff en terme de pourcentage de réponses correctes. De plus, la TBA requise est généralement inférieure dans la condition audiovisuelle que dans la condition auditive. Par exemple, dans le cas de cibles externes, qui sont hors des capacités motrice de l'agent comme dans les conditions expérimentales auxquelles les nourrissons semblent avoir été exposés par les auteurs, le score dépasse 60% d'imitation correcte avec un DDG donnant 32 cases et des TBA de 50 et 25 exemplaires pour les inversions auditive et audio-visuelle, respectivement.

Des matrices de confusion

En ce qui concerne les matrices de confusion, elles dépendent beaucoup des diverses configurations de simulation, et bien sûr, aussi du tirage particulier des données d'apprentissage, surtout lorsque leur nombre est faible. Les diverses matrices rencontrées ne reproduisent jamais très fidèlement celle des données réelles. Cependant, dans les différentes configurations testées en condition audio seule, les confusions entre [a] et [i] et entre [a] et [u] s'équilibrent. Ceci s'approche de ce que l'on observe dans les données expérimentales, et qu'il était impossible de reproduire par imitation visuelle pure, pour laquelle les confusions entre [a] et [i] sont beaucoup plus nombreuses que celles entre [a] et [u]. C'était précisément un des enjeux de nos simulations. Les matrices de confusion audiovisuelles sont intermédiaires entre matrices auditives et visuelles, avec une certaine tendance à « écarter » [a] de [u], plus que cela n'est le cas dans les données expérimentales. Cependant, étant donné le grand nombre de variables sur lesquelles nous pourrions jouer pour essayer de nous approcher des données, nous avons jugé plus prudent d'en rester à ces seules conclusions générales.

Conclusion

Comme dans le cas de l'estimation des capacités articulatoires précoces (chapitre 4), cet ensemble de simulations doit être considéré comme une "première", puisqu'il n'existe à notre connaissance aucun travail de modélisation des capacités d'imitation vocale précoce. Il est, à ce titre, riche d'enseignements, et bien-sûr de questions.

Les données vues à travers le modèle

Il permet d'abord de se tourner vers les données elles-mêmes, et apporte à leur sujet deux éléments d'éclairage qui nous semblent très importants.

En premier lieu, l'audition semble indispensable pour rendre compte des performances en imitation vocale des nourrissons de Kuhl et Meltzoff (1996) avec les capacités sensori-motrices inférées à 4 mois, vu que la simulation de l'imitation visuelle pure ne permet de reproduire ni le score ni le profil de réponse rapportés par les auteurs.

Deuxièmement, et c'est à notre sens un élément majeur, très peu de vocalisations (de l'ordre de 25 à 50) sont nécessaires au mécanisme d'apprentissage pour fournir des scores d'imitation au moins aussi élevés que ceux de ces sujets.

D'une certaine manière, il est tout à fait évident que le modèle est en réalité *trop puissant*, et capable à peu de frais de produire des performances d'imitation très élevées. En réalité, il prête sans doute trop aux capacités de contrôle des nourrissons, qui ne sont très probablement pas capables de tirer, même aléatoirement, des vocalisations au sein d'une loi de distribution, fut-elle de variance très élevée.

Le modèle vu à travers les données

Il convient ensuite de revenir sur les caractéristiques du modèle en tant que tel. Il apparaît que la simulation des imitations auditive et audio-visuelle révèle un compromis entre l'acuité somesthésique de la représentation de la forme du conduit vocal (DDG) et la quantité d'information (TBA) à apprendre pour construire une carte sensori-motrice assez représentative des habiletés de l'agent. De plus, nos résultats montrent que la représentation de l'espace géométrique *doit* être assez imprécise (DDG faible) pour que l'agent soit capable d'imiter des sons vocaliques qui se trouvent hors de ses capacités articulatoires-acoustiques. Ceci est du plus grand intérêt puisqu'en fait, les nourrissons doivent acquérir, par imitation, les sons de parole de leur(s) langue(s) ambiante(s) bien qu'ils ne soient pas doués des habiletés motrices requises, dès la naissance.

Enfin, cette étude s'accorde au point de vue selon lequel la formation de la représentation cognitive potentiellement sous-jacente à l'imitation vocale précoce demande aux nourrissons la cartographie d'un nombre plus faible de configurations dans le cas d'une perception audio-visuelle de la parole que sans la vision. Cela fait écho au fait que la perception bimodale facilite le développement phonétique et est congruent avec les légères différences de développement phonétique entre individus voyants et mal-voyants (Mills, 1987 ; Mulford, 1988).

Un point important de ces simulations concerne le caractère arbitraire que peut avoir un choix de modélisation donné. Nous ne reviendrons pas sur la nature même du modèle VLAM, sur laquelle nous avons déjà eu l'occasion de discuter dans le chapitre précédant. Par contre, on peut se demander si nos choix de modélisation Bayésienne, tels qu'ils ont été décrits dans les sections II et III, sont acceptables. En fait, il est important de noter que, quoi que l'on puisse penser de ces choix, que nous avons la faiblesse de trouver judicieux, ils influent vraisemblablement assez peu sur la conclusion majeure de cette étude, qui est le faible volume de données d'apprentissage nécessaires à produire un comportement d'imitation adéquat. En

effet, après tout, le problème posé est un simple problème de distribution à 3 classes, et quel que soit l'outil utilisé pour le résoudre, dans le cadre Bayésien, on comprend bien qu'une dizaine de vocalisations par "classe" suffise à produire un volume d'apprentissage suffisant, pourvu que l'on choisisse un système de discrétisation à gros grains, et quel que soit l'espace dans lequel l'on discrétise.

En résumé, ce travail montre que peu de données sont nécessaires à une reproduction réaliste des scores d'imitation si l'acuité de la perception de la forme du conduit vocal est assez faible. De plus, il fournit le noyau du second mécanisme qui nous semble nécessaire au développement du contrôle de la parole, celui de l'imitation orofaciale précoce. Notre agent sensori-moteur en développement est ainsi doté des deux ingrédients, l'exploration et l'imitation, qui sont vraisemblablement convoqués ensemble dans le parcours de l'émergence des constituant de la parole.

Discussion générale:

Quelques retours sur le moule à connaissances

La construction de notre androïde trouve son germe dans l'emprunt de modèles conceptuels et computationnels. L'état de départ du puzzle relatif au développement précoce de la production verbale comprend les modèles conceptuels (i) de l'émergence de la coordination orolaryngée de Koopman-van Beinum et van der Stelt, (ii) de la cyclicité mandibulaire comme cadre de la future syllabe de Davis et MacNeilage et (iii) du développement du contrôle du système articulatoire d'Abry, Cathiard, Vilain, Laboissière, Loevenbruck, Savariaux et Schwartz. Nous avons proposé que deux mécanismes centraux devaient être explicitement pris en compte et précisément définis pour mieux comprendre le développement de la production verbale: l'exploration et l'imitation orofaciale. Dans le but d'implémenter ces deux mécanismes de manière réaliste, nous avons récupéré des modèles computationnels (i) de chacune des modalités auditive, visuelle et tactile vraisemblablement en jeu dans la communication parlée, (ii) de la production de la parole qui intègre la croissance non-linéaire et non-uniforme du conduit vocal, et (iii) d'apprentissage Bayésien de liens perception-production, et d'inversion Bayésienne de cibles perceptives.

Toutefois, la définition et la défense d'une imitation orofaciale au service du développement de la production de la parole ont nécessité l'élaboration d'une revue de questions relative à l'imitation d'après des recherches en psychologie développementale, en éthologie et en neurosciences: ce travail de réappropriation de connaissances sur l'imitation n'avait, à notre connaissance, jamais été fait de l'intérieur du domaine de la parole. Dit de manière très succincte, l'imitation est la reproduction par l'individu d'une *action* sur la base de sa *perception* dans la conduite d'autrui, et sert vraisemblablement l'*apprentissage* de comportements *nouveaux* et *arbitraires*. Cette recherche bibliographique a suscité, en nous, de nouveaux questionnements et permis de concevoir de nouvelles hypothèses quant à la manière dont les domaines perceptif et moteur pourraient se co-structurer au service du développement de la production verbale dans la/les langue-s ambiantes. Mais avant d'en dire plus sur ces propositions qui nous sont propres, nous désirons revenir sur le bilan de nos expériences de modélisation.

1. Un bilan au niveau computationnel

a. Le principe développemental d'une exploration progressive

Notre analyse par modélisation de données acoustiques sorties de la bouche de vrais nourrissons suggère que l'exploration articulatoire tend à s'étendre entre 4 et 7 mois, et que la mandibule joue un rôle mineur avant le début du babillage canonique, pour devenir dominant dès le début de cette période développementale.

Bien que notre démarche ait emprunté un formalisme très différent des méthodes d'analyse transcriptionnelle, nous avons retrouvé des tendances décrites dans les inventaires de productions vocaliques préverbales. En particulier, nos résultats viennent en renfort à la remise en question du postulat de répertoire infini de Jakobson, et permettent de caractériser la rareté des vocalisations extrêmes du type [i] et [u] dans les répertoires préverbaux, ainsi qu'une extension de l'exploration sur les configurations de conduit vocal ouvertes associée au babillage canonique. Nos résultats à ce niveau nous poussent à suggérer un double jeu de contraintes entre productions infantiles et parole adulte.

D'une part, nous proposons qu'une exploration progressive des capacités du conduit pourrait permettre un développement efficace des capacités de production, au sens où il est *guidé* par l'espace de production adulte, grâce au processus d'imitation. En d'autres termes, plutôt que de perdre une énergie inutile à l'exploration de tout l'espace des possibles, mieux vaut sans doute explorer peu pour développer des germes de relations sensori-motrices, et utiliser ces germes dans le développement d'un espace de production efficace, i.e. fait de ce qui est utile pour parler la/les langue-s de l'environnement, sous l'effet d'un modelage de ce paysage articulatoire par imitation orofaciale. Notons d'ailleurs qu'il est probablement *computationnellement* préférable d'adopter une telle stratégie de déploiement progressif, afin de construire des cartes sensori-motrices incrémentales (voir Diard, 2003, pour des propositions computationnelles originales sur la construction progressive de cartes complexes à partir de cartes plus simples).

D'autre part, nous avons montré que le chemin de développement de la structure du paysage articulatoire pourrait vraisemblablement *contraindre*, en retour, les capacités de productions adultes, comme semble le suggérer, en accord avec l'hypothèse d'Abry et Badin (1996), la caractéristique vraisemblablement palatale du [u]-juvénile, germe du [u] adulte. Il vient que l'adulte conserverait des traces du nourrisson qu'il/elle a été, tout en guidant les pas du nourrisson vers ceux de l'adulte qu'il/elle deviendra. L'on trouve ici instancié, de manière frappante, un thème majeur de la *robotique cognitive*: celui du rôle central des mécanismes développementaux dans l'élaboration de capacités matures et opérationnelles.

b. Des proto-syllabes mieux caractérisées aux niveaux articulatoire et acoustique

Les cooccurrences prédites par la théorie Frame/Content ont été testées par la simulation du cycle mandibulaire à partir des pré-configurations de langue inférées des données de 7 mois. L'établissement des patrons de cooccurrence simulés s'accordent assez bien avec l'idée d'une association préférentielle entre closant et vocant contigus par le truchement du cycle mandibulaire, et donc avec les tendances à la cooccurrence mises en évidence dans le formalisme transcriptionnel. Toutefois, nos simulations ont introduit, en sus de l'importance de la variabilité inter-individuelle de la morphologie du conduit vocal proposée par Vilain *et al.* (1999), l'idée que la variabilité *intra*-individuelle est également à prendre en considération. En effet, un même son vocalique peut conduire à un spectre de contacts possibles relativement varié. De plus, au niveau acoustique, la détermination de la répartition des seconds formants des lieux de contact en fonction de celle des pré-configurations associées, dans le formalisme des équations du locus, a permis de retrouver une tendance au lien closant-vocant proposée par la théorie Frame/Content, et a abouti à des patrons de dispersion compatibles avec ceux des données réelles de Sussman et collègues à 10 mois. Toujours dans le formalisme des équations du locus, nous avons tenté de proposer des pistes pour la modélisation du contrôle des contacts labiaux et apicaux, dans le cas idéal de la seule implication de l'articulateur caractéristique de chacun de ces lieux. L'écart de nos résultats aux données suggère que la réalité de la coarticulation des parties vocalique et consonantique relève vraisemblablement de processus moins rudimentaires, mais ces simulations n'en indiquent pas moins des cibles potentielles pour l'émergence de la coarticulation au cours du développement.

c. Apprendre peu et "grossièrement" pour imiter et mieux apprendre

A partir des capacités articulatoire-acoustiques inférées à 4 mois, nous avons simulé la formation d'une carte cognitive des liens perception-production, où le domaine somesthésique sert de pivot, ainsi que son utilisation dans l'imitation orofaciale de cibles vocaliques [i a u]. Le but était de reproduire les performances en imitation de sujets âgés de 3 à 4.5 mois rapportées dans une publication de la littérature, pour tirer quelques informations sur la carte

cognitive potentiellement en jeu dans ce comportement. Nous avons focalisé notre attention (i) sur deux caractéristiques de la carte des liens, i.e. l'acuité de son espace somesthésique et la quantité d'information exploratoire apprise, et (ii) sur deux caractéristiques de la cible, i.e. son appartenance ou non au domaine d'exploration et les modalités de sa présentation, auditive *vs.* audio-visuelle. Globalement, ce travail montre que peu de données sont nécessaires à une reproduction réaliste des scores d'imitation si l'acuité de la perception de la forme du conduit vocal est assez faible, et que l'imitation de cibles extérieures au domaine d'exploration nécessite une acuité somesthésique faible. De plus, en accord avec l'idée que la multimodalité faciliterait l'imitation (Legerstee, 1990), la quantité d'information nécessaire à la reproduction des performances réelles est inférieure dans le cas des cibles audio-visuelles à celle obtenue pour une perception purement auditive de la cible.

Dans notre implémentation de l'imitation orofaciale, la cible correspond au but de l'acte, il n'y a pas de processus de correction itératif et les réponses de l'agent ne peuvent jamais sortir de son domaine d'apprentissage, empêchant ainsi un développement de l'exploration en fonction des tentatives d'imitation de cibles nouvelles. Cela est en partie dû à notre implémentation du formalisme Bayésien. Toutefois, ce paradigme probabiliste nous a semblé être le plus adapté à l'incomplétude de nos connaissances sur le phénomène de l'imitation orofaciale de modèle de parole. La continuation de ce travail de modélisation consisterait à trouver un moyen d'étendre l'espace d'exploration sous l'effet de l'imitation de cibles nouvelles, et d'étudier l'émergence de la co-structuration des espaces perceptif et articulatoire en fonction des catégories phonétiques de la parole ambiante (nous y reviendrons). Ainsi, là encore, nous sommes devant un processus bidirectionnel, dans lequel l'on peut apprendre peu pour imiter un peu, et imiter même mal, pour mieux apprendre.

2. Un bilan au niveau conceptuel

a. Des organes de la parole émergeant de la co-structuration perceptuo-motrice

Un des apports originaux de cette thèse est la défense et la définition d'une imitation orofaciale au service du développement de la parole. Sans revenir sur les détails, notre réflexion, qui est largement inspirée des travaux et propositions de Meltzoff et collègues, nous a mené à la considération que l'imitation dite "vocale" n'est peut-être pas qu'un problème d'appariement intra-modalité auditive entre la cible et l'acte en cours, comme en témoigne la capacité de donner une équivalence motrice à des cibles de parole dès la naissance (Chen *et al.*, 2004). Un point essentiel sur lequel nous désirons revenir est que la somesthésie pourrait se situer au centre du lien perception-production en jeu autant dans les comportements imitatifs orofaciaux que dans le contrôle du système articulatoire, et que cette passerelle pourrait trouver des bases dans un apprentissage intra-utérin.

En effet, l'on s'accordera sur le fait qu'une configuration articulatoire est une relation-entre-organes et que cet objet est autant descriptible en des termes proprioceptifs que moteurs. Les organes sont ici à comprendre au sens de parties du conduit vocal, dont la définition reste à établir pour le cas de la parole. Nous proposons qu'un organe peut dans le même temps se caractériser par sa "signature cinétique", mais aussi comme surface du corps dont les points ont une grande probabilité d'être tactilement stimulés de manière concomitante lors de l'établissement de contacts avec d'autre surface. Pour illustration, un point de la langue a une probabilité plus élevée d'être tactilement stimulé par un contact avec le palais en même temps qu'un autre point contigu de la langue qu'en même temps qu'un point plus éloigné. Il semble concevable que la répétition d'une stimulation tactile simultanée de micro-surfaces cutanées géographiquement proches au cours de l'activité propre contribue, avec l'établissement d'un modèle interne de leurs propriétés biomécaniques sur la base d'une information

proprioceptive, à l'émergence d'organes au sens de Meltzoff et Moore (1997), et que le développement de ces objets puisse commencer *in utero* (potentiellement avec l'implication du cervelet) en s'inscrivant dans la coordination audio-proprio-visuelle probablement en germe à la même période (voir chapitre 1, section I.3.c., pour des éléments sur l'activité motrice intra-utérine, et chapitre 3, section III.A.2.b., pour une proposition d'un développement intra-utérin d'une coordination audio-proprio-visuelle). L'expérience de Chen *et al.* (2004) suggère que des organes de la parole, au moins basiques, sont présents à la naissance, tels que la mandibule, les lèvres et la langue (Meltzoff et Moore, 1997).

b. Co-émergence des degrés de liberté dans la perception et le contrôle

Cependant, la parole mature demande une différenciation plus fine de ces organes néonataux, en particulier dans le cas de la langue. Nous proposons que la différenciation d'organes linguaux émerge du comportement de babillage dans le même temps que le contrôle des différents "degrés de liberté" de la langue s'émancipe de celui du cycle mandibulaire. Notre idée est la suivante. Les tendances à la cooccurrence idiosyncrasique, observées dans les séquences d'ouverture-fermeture du babillage canonique, nous apprennent qu'une pré-configuration, ou forme du conduit vocal, ou *relation-entre-organes*, donnée en phase d'ouverture mènera à un patron de contact relativement déterminé entre la surface des articulateurs portés (langue et lèvre inférieure) et celle des structures de la paroi dorsale du conduit vocal (palais et lèvre supérieure). Il est très probable que la variabilité de la position de la langue selon l'axe avant-arrière entre les séquences de babillage, associée à la répétition de l'association entre formes de langue à l'ouverture et contacts en fermeture, donne lieu (i) à l'apprentissage de leurs liens par contiguïté temporelle et (ii) à la différenciation de zones de la surface de la langue géographiquement proches en organes de langue permettant ainsi une imitation orofaciale de plus en plus fine des manières arbitraires de produire les sons de la parole ambiante, et dirigeant le développement d'un contrôle du système articulo-moteur qui est posé à la fois par la Théorie Motrice (Liberman et Mattingly, 1985) et par la Théorie de la Perception pour le Contrôle de l'Action (Schwartz *et al.*, 2002). Les développements récents proposés par Studdert-Kennedy et Goldstein (2003) pour inscrire la notion d'organe, au sens de Meltzoff et Moore (1997), dans la Théorie Motrice s'insèrent très bien en base arrière de cette réflexion.

3. Des perspectives pour faire grandir l'androïde

Sur le chemin que nous nous sommes proposés d'emprunter, vers l'élaboration progressive d'un agent capable d'apprendre peu à peu à contrôler son conduit vocal pour produire les sons d'une langue, nous n'en sommes qu'au tout début. Plutôt que de pointer les limitations évidentes de notre approche, liées notamment au caractère contestable de chacun de nos choix – tant des modèles de production et de perceptions, que du cadre théorique de la programmation Bayésienne – nous voudrions suggérer quelques étapes qui nous semblent intéressantes pour faire "grandir" notre androïde. Nous en mentionnerons trois, en nous limitant volontairement à des étapes relativement peu éloignées de notre position actuelle.

a. Comment acquérir le contrôle des contacts?

Dans le schéma développemental proposé par Abry *et al.* (2005), la première étape qui suit la maîtrise du cycle mandibulaire de babillage est le contrôle des contacts, i.e. la capacité d'ajouter à la porteuse-mâchoire une commande réquisitionnant spécifiquement la langue ou les lèvres pour contrôler le contact labial, apical ou vélaire (voir aussi les travaux de Lalevée

et Vilain (2005), sur l'émergence du contrôle du velum à la base du contraste oral-nasal, vers 11-12 mois). Notre hypothèse ici est que la phase de babillage pourrait permettre, dans le cas de cycles mandibulaires de grande amplitude, de découvrir les propriétés de compliance variable de la langue contre le palais ou des lèvres l'une contre l'autre (par écrasement, en quelque sorte), et ainsi d'*augmenter l'espace d'exploration articulatoire-acoustique* par de nouvelles commandes (de la pointe de la langue, TT, par exemple). Cette exploration accrue (voir aussi Brosda, 1999) permettrait de spécifier des cartes élargies, probablement capables, dans un mécanisme d'inversion, de gérer le passage progressif de la proto-syllabe, telle qu'elle apparaît acoustiquement dans nos simulations de la figure 4-25A, vers des syllabes contrôlées en contenu, contrastant geste labial et geste lingual quelle que soit la pré-configuration du conduit. L'idée serait tout simplement de chercher à apprendre des patrons de coarticulation adultes tels que ceux décrits par Sussman et collègues, et de montrer qu'ils passent précisément par un découplage progressif de la partie vocalique et de la partie consonantique, comme le montrent Vilain *et al.* (1999).

b. Comment conquérir ses degrés de liberté?

Cette question est à la fois centrale et délicate (Paviot, 2000). Sans proposer ici de solution miracle, nous voulons revenir sur le point annoncé précédemment. En effet, si le caractère *progressif* de l'exploration articulatoire apparaît comme un résultat important de cette thèse, c'est cette base, selon nous, qu'il faut élaborer. Nous avons proposé qu'exploration et imitation sont des mécanismes complémentaires, et enchâssés. Il nous semble nécessaire de travailler sur cette alternance, en combinant la construction d'un modèle direct, et l'orientation vers de nouvelles régions – de nouveaux degrés de liberté – à travers les processus d'inversion¹. Dans ce contexte, la capacité de construire de manière structurellement pertinente des cartes de complexité croissante, sera essentielle. L'apprentissage du contrôle des contacts, décrit au point précédent, pourrait fournir ici un cadre intéressant. Il permettrait de tester la mise en œuvre de telles cartes, plus complexes que celles que nous avons étudiées: par exemple, en ajoutant un à un de nouveaux paramètres articulatoires, ou encore en introduisant à une étape donnée le *temps* dans les représentations perceptuo-motrices, précisément en relation avec la capacité de contrôler les contenus (contents) des cadres (frames).

c. Que faire des cartes perceptuo-motrices apprises?

Cette dernière question pourrait paraître triviale, dans le contexte où se place ce travail au premier titre, celui du développement du contrôle du système orofacial. Cependant, il nous semble essentiel de retourner, une nouvelle fois, vers le cadre théorique à la base de cette thèse: celui de la Théorie de la Perception pour le Contrôle de l'Action, cadre dans lequel nous défendons une *co-structuration des représentations perceptives et motrices* au cours du développement. En effet, dans ce cadre, les représentations perceptives sont contraintes par l'action. Or, le mécanisme d'exploration que nous avons défini, dans le présent travail, fonctionne précisément ainsi. L'étape qu'il nous faudrait désormais franchir consisterait à *utiliser* ces représentations perceptives contraintes (ou pour être plus précise, ces cartes perceptuo-motrices) pour étudier comment l'enfant (puis l'adulte) peut *traiter l'information perceptive*, prendre des *décisions perceptives*, à partir de ces cartes. En d'autres termes, la question qui se pose est celle de la mise en évidence du rôle de l'action dans la perception – mise en évidence qui surgit dès l'âge de 3 mois, d'après une publication dont nous prenons

¹ Notons que, bien que tout notre travail repose sur un principe d'apprentissage d'un modèle direct suivi d'inversions de ce modèle, un champ considérable à explorer concerne l'hypothèse de processus d'imitation alternatifs, tels que ceux de la correction itérative proposée par Meltzoff et Moore (1997).

connaissance en toute fin de rédaction (voir Sommerville, Woodward et Needham, à paraître). Les travaux ici pourraient concerner notamment l'étude de la perception multimodale de la parole, dans le cadre de l'hypothèse de recodage moteur proposée dans notre laboratoire (Schwartz *et al.*, 1998). On pourrait ainsi se pencher à la fois sur la modélisation des résultats de perception audio-visuelle de type effet McGurk (Burnham et Dodd, 1996, 2004) en référence à ces cartes, mais aussi étudier les relations entre structuration progressive des cartes cognitives et décours développemental des processus de fusion telle qu'elle a été décrite récemment par Sekiyama, Burnham, Tam, et Erdener (2003).

c. Epilogue

L'on comprendra que ce n'est qu'une infime partie du chemin que nous avons parcouru, plus petite certainement que nous ne l'imaginions au départ de cette aventure. Le phénomène à l'étude est complexe, et nous espérons que cette aventure se poursuivra sous une forme ou sous une autre, que cela soit ici ou ailleurs. Les conditions matérielles sont parties prenantes de toute construction: nous tenons à remercier nos soutiens institutionnels venus des projets CNRS STIC Robea, CNRS SHS OHLL, ESF-Eurocore OMLL et SHS Patipapa, mais aussi le Ministère de l'Education Nationale pour l'allocation de recherche à laquelle j'ai eu droit.

BIBLIOGRAPHIE

A

Abry, C. (1997). De la parole articulatoire à la parole narrative : ou du soi vocal à l'autre narratif dans une "théorie de l'esprit". Doctorat d'Etat ès Lettres, Université Stendhal.

Abry, C., et Badin, P. (1996). Speech Mapping as a framework for an integrated approach to the sensori-motor foundations of language. *4th Speech Production Seminar, 1st ESCA Tutorial and Research Workshop on Speech Production Modelling: from control strategies to acoustics*, 175-184, May 21-24, 1996, Autrans, France.

Abry, C., Badin, P., Vilain, A., Stéfanuto, M., et Boë, L.-J. (1997). Une théorie de l'ontogenèse à l'épreuve de trois modèles articulatoires anthropomorphiques. In *Les robots parlent aux robots*, GIS Sciences de la Cognition, Rapport Intermédiaire, pp. 29-37

Abry, C., Benoît, C., Boë, L.J., & Sock, R. (1985). Un choix d'événements pour l'organisation temporelle du signal de parole. *14èmes Journées d'Etudes sur la Parole, Société Française d'Acoustique*, 133-137.

Abry, C. & Boë, L.-J. (1986). Laws for Lips. *Speech Communication*, 5, 97-104.

Abry, C., Boë, L.-J., Laboissière, R., et Schwartz, J.-L. (1998). Emergence of a new puzzle for evolution of speech. *Behavioral and Brain Sciences*, 21 (4), 512-513.

Abry, C., Cathiard, M.-A., Vilain, A., Laboissière, R., Loevenbruck, H., Savariaux, C. & Schwartz, J.-L. (2005). Some insights in bimodal perception given for free by the natural time course of speech production. In *Audiovisual Speech processing, Festschrift Christian Benoît* (E. Vatikiotis-Bateson, G. Bailly, & P. Perrier Eds.). Cambridge, MA, USA: MIT Press (to appear).

Abry, C., Orliaguet, J.P., & Sock, R. (1990). Patterns of speech phasing. Their robustness in the production of a timed linguistic task: single vs. double (abutted) consonants in French. *European Bull. of Cogn. Psych.*, 10, 269-288.

Abry, C., Vilain, A., Schwartz, J.-L., (2004). "Vocalize to Localize"? A call for better crosstalk between auditory and visual communication systems researchers: From meerkats to humans. *Interaction Studies. Social Behaviour and Communication in Biological and Artificial Systems*, Abry, Vilain, J.-L. Schwartz (Eds), 5, 3, 313-325.

Atal, B., Chang, J., Mathews, M., & Tukey, J. (1978). Inversion of articulatory-to-acoustic transformation in the vocal tract by a computer sorting technique. *Journal of the Acoustical Society of America*, 63, 1535-1556.

B

Badin, P., Perrier, P., Boë, L.J., & Abry, C. (1991). Vocalic nomograms: Acoustic and articulatory considerations upon formant convergences. *J. Acoust. Soc. Am.*, 87, 1290-1300.

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

- Bard, K. A., et Russell, C. L. (1999). Evolutionary foundations of imitation: social, cognitive and developmental aspects of imitative processes in non-human primates. In J. Nadel et G. Butterworth (eds), *Imitation in Infancy* (pp. 89-123), Cambridge University Press.
- Barnat, S. B., Klein, P. J., et Meltzoff A. N. (1996). Deferred imitation across changes in context and object: memory and generalization in 14-month-old infants. *Infant Behavior and Development*, 19, 241-251.
- Barr, R., Dowden, A., et Hayne, H. (1996). Developmental changes in deferred imitation by 6- to 24-month-old infants. *Infant Behavior and Development*, 19, 159-170.
- Baudonnière, P.-M. (1997). *Le mimétisme et l'imitation*. Flammarion (Ed.).
- Beardsworth, T., et Buckner, T. (1981). The ability to recognize oneself from a video recording of one's movements without seeing one's body. *Bulletin of the Psychonomic Society*, 18, 19-22.
- Bearko Gleason, J. (2001). The development of language: an overview and a preview. In Bearko Gleason, J. (ed.), *The development of language* (pp. 1-39), Allyn & Bacon.
- Beddor, P. S., & Strange, W. (1982). Cross-language study of perception of the oral-nasal distinction. *Journal of the Acoustical Society of America*, 71, 1551-1561.
- Berrah, A.R. (1994). *L'émergence des structures sonores : les syllabes consonnes-voyelles*. Diplôme d'Etudes Approfondies, Institut National Polytechnique de Grenoble, France.
- Berrah, A.R., Glotin, H., Laboissière, R., Bessière, P., & Boë, L.J. (1996). From form to formation of phonetic structures: an evolutionary computing perspective. In T. Fogarty & G. Venturini (eds.) *ICML '96 Workshop on Evolutionary Computing and Machine Learning* (pp. 23-29). Bari: Italy.
- Bessière, P. (2000). *Vers une théorie probabiliste des systèmes sensori-moteurs*. HDR, Université Joseph Fourier, Grenoble, France.
- Bessière, P., Dedieu, E., et Mazer, E. (1994). Representing robot/environment interactions using probabilities: the "beam in the bin" experiment. ?????
- Bessière, P., Dedieu, E., Lebeltel, O., Mazer, E. et Mekhnacha, K. (1998a). Interprétation ou description (I) : proposition pour une théorie probabiliste des systèmes cognitifs sensori-moteurs. *Intellectica*, 26-27, 257-311.
- Bessière, P., Dedieu, E., Lebeltel, O., Mazer, E. et Mekhnacha, K. (1998b). Interprétation ou Description (II) : Fondements mathématiques de l'approche F+D. *Intellectica*, 26-27, 313-336.
- Bizzi, E. (2001). Motor control. In Wilson, R. A., et Keil, F. C. (eds), *The MIT encyclopedia of the cognitive sciences* (pp. 570-571), MIT Press.
- Blackmore, S. (1998) Imitation and the definition of a meme. *Journal of Memetics – Evolutionary Models of Information Transmission*, 2.
- Bladon, A. (1982). Arguments against formants in the auditory representation of speech. In

The Representation of Speech in the Peripheral Auditory System (R. Carlson & B. Granström, eds.), pp. 95-102. Amsterdam: Elsevier Biomedical.

Boë, L.-J. (1999). Modelling the growth of the vocal tract vowel spaces of newly-born infants and adults. *Proc. XIVth International Congress of Phonetic Sciences*, San Francisco, USA, 2501-2504.

Boë, L.J., Abry, C., Beautemps, D., Schwartz, J.L., et Laboissière, R. (2000). Les sosies vocaliques: inversion et focalisation. XXIII^{ième} Journées d'Etude de la Parole, pp. 257-260.

Boë, L.J., Gabioud, B., & Perrier, P. (1995a). Speech Maps Interactive Plant « SMIP ». *Proc. XIIIth International Congress of Phonetic Sciences*, Stockholm, Sweden, 426-429.

Boë, L.-J, Gabioud, B., Perrier, P., Schwartz, J.-L., et Vallée, N. (1995b). Vers une unification des espaces vocaliques. In C. Sorin et al. (eds.) *Levels in Speech Communication: Relations and Interactions* (pp. 63-71). Elsevier Science B.V.

Boë, L.J., Heim, J.L., Honda, K., & Maeda, S. (2002). The potential of Neandertal vowel space was as large as that of modern humans. *Journal of Phonetics*, 30, 465-484.

Boë, L.J., Perrier, P., et Bailly, G. (1992). The geometric vocal tract variables controlled for vowel production: proposals for constraining acoustic-to-articulatory inversion. *Journal of Phonetics*, 20, 27-38.

Boë, L.-J., Perrier, P., Guérin, B., & Schwartz, J.-L. (1989). Maximal Vowel Space. *Eurospeech89*, 281-284.

Bohannon III, J. N., et Bonvillian, J. D. (2001). Theoretical approaches to language acquisition. In Bearko Gleason, J. (ed.), *The development of language* (pp. 254-314), Allyn & Bacon.

Bonnot, J.F. (1989). Timing extrinsèque et timing intrinsèque : le temps est-il une variable contrôlée ?. *Journal d'Acoustique*, 2, 287-296.

Bonnot, J.F. (1990). Production de la parole et coarticulation : une analyse critique des différents modèles. Travaux de l'Institut de Phonétique de Strasbourg, n°20

de Boysson-Bardies, B., Hallé, P., Sagart, L., et Durant, C. (1989). A crosslinguistic investigation of vowel formants in babbling. *Journal of Child Language*, 16, 1-17.

de Boysson-Bardies, B., Sagart, L. & Durand, C. (1984). Discernible differences in the babbling of infants according to target language. *Journal of Child Language*, 11, 1-15.

de Boysson-Bardies, B., & Vihman, M. (1991). Adaptation to language: Evidence from babbling and first words in four languages. *Language*, 67, 297-319.

Bosma, J.F. (ed.) (1967). *Symposium on oral sensation and perception*. Springfield, Ill.: Charles C. Thomas.

Brooks, R.A., Breazeal, C., Marjanovic, M., Scassellati, B. & Williamson M. (1999). The Cog Project: Building a Humanoid Robot. In C. Nehaniv (ed) *Computation for Metaphors, Analogy, and Agents. Lecture Notes in Artificial Intelligence 1562* (pp. 52-87). New York: Springer.

Brosda, S. (1999). *De la variation dans le babillage canonique: l'apprentissage sensori-moteur*. Mémoire de DEA Sciences Cognitives, ICP, Institut National Polytechnique de Grenoble.

Brosda, S., Laboissière, R. & Abry, C., (2001), Systems, concepts and representations – but where do they come from? From babbling towards phonology. *DECOLAGE 2001* – 2nd Bisontine Conference for Conceptual and Linguistic Development in the Child aged from 1 to 6 years, Besançon, France.

Buccino, G., Binkofski, F., et Riggio, L. (2003). The mirror neuron system and action recognition. *Brain and Language*, à paraître.

Buhr, R.D. (1980). The emergence of vowels in an infant. *Journal of Speech, Language and Hearing Research*, 23, 73-94.

Burnham, D.K. (1998). Language specificity in the development of auditory-visual speech perception. In R. Campbell, B. Dodd & D. Burnham (Eds.) *Hearing by Eye II: Advances in the Psychology of Speechreading and Auditory-visual Speech*. Hove, UK: Psychology Press (27-60).

Burnham, D., & Dodd, B. (1996) Auditory-visual speech perception as a direct process: The McGurk effect in human infants and across languages. In D.G. Stork & M.E. Hennecke (Eds) *Speechreading by humans and machines*. Berlin: Springer-Verlag (103-114).

Burnham, D. & Dodd, B. (2004) Auditory-visual speech integration by pre-linguistic infants: Perception of an emergent consonant in the McGurk effect. *Developmental Psychobiology*, 44,209-220.

Butterworth, G. (1999). Neonatal imitation: existence, mechanisms and motives. In J. Nadel et G. Butterworth (eds), *Imitation in Infancy* (pp. 63-88), Cambridge University Press.

C

Calvert, G.A. & Campbell, R. (2003) Reading speech from still and moving faces : the neural substrates of visible speech. *Journal of Cognitive Neuroscience*, 15:57-70.

Campbell, R., Dodd, B., & Burnham D. (eds.) (1998). *Hearing by eye, II. Perspectives and directions in research on audiovisual aspects of language processing*. Hove (UK): Psychology Press.

Chen, X., Striano, T., et Rakoczy, Hannes (2004). Auditory–oral matching behavior in newborns. *Developmental Science* 7(1), pp 42–47.

Chistovich, L.A. (1976). *Physiology of Speech : Human Speech Perception*. Leningrad : Nauka (in Russian).

Chistovich, L.A. (1980). Auditory processing of speech. *Language and Speech*, 23, 67-72.

Crystal, D., (1980/1995). *Dictionary of Linguistics and Phonetics*. Blackwell Publishers (ed.).

Cyrułnik, B. (1989). *Sous le signe du lien*. Hachette (Ed.).

D

Dale, R., Richardson, D. C., et Owren, M. J. (2003). Pumping for gestural origins: The well may be rather dry. *Behavioral and Brain Sciences*, 26 (2), 218-219.

Darlot, C. (2001). Le cervelet. *Pour la Sciences*, 285, 74-82.

Davis, B. L., MacNeilage, P. F., et Matyear, C. L. (2002a). Acquisition of serial complexity in speech production: A Comparison of Phonetic and Phonological Approaches to First Word Production. *Phonetica*, 59, 75-107.

Davis, B. L. et MacNeilage, P. F. (2002b). The Internal Structure of the Syllable: An Ontogenetic Perspective on Origins in Givon, T. et Malle, B.(Eds.), *The Rise of Language out of Pre-language*. (pp. 133-152) Amsterdam Benjamin.

Davis, B. L., et MacNeilage, P. F. (1995). The articulatory basis of babbling. *Journal of Speech, Language and Hearing Research*, 38, 1199-1211.

Dehaene-Lambertz, G., Dehaene, S., et Hertz-Pannier, L. (2002). Functional neuroimaging of speech perception in infants. *Science*, 298, 2013—2015.

Delgutte, B. (1984). Speech coding in the auditory nerve II: Processing schemes for vowel-like sounds. *J. Acoust. Soc. Am.*, 75, 879-886.

Diard, J. (2003). *La carte bayésienne : Un modèle probabiliste hiérarchique pour la navigation en robotique mobile*. Doctorat de l'INPG, Spécialité : « Informatique : Systèmes et Communications ».

Dodd B. , & Campbell, R. (eds.) (1987). *Hearing by eye : the psychology of lipreading*. Lawrence Erlbaum Associates, London.

Dolle, J.-M. (1999). *Pour comprendre Piaget*. Dunod (Ed.).

Duda, R. O. and Hart, P. E., *Pattern Recognition and Scene Analysis*, New York: John Wiley, 1973.

E

Ely, R. (2001). Language and literacy in the school years. *In* Bearko Gleason, J. (ed.), *The development of language* (pp. 409-454), Allyn & Bacon.

F

Fant, G. (1960). *Acoustic Theory of Speech Production*. Mouton: The Hague.

Ferrari, P. F., Gallese, V., Rizzolatti, G., et Fogassi, L. (2003). Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex. *European Journal of Neuroscience*, 17, 1703–1714.

Firestone, S. (1970/2002) *Pour l'abolition de l'enfance*. Editions Tahin Party.

Fischer, J., et Hammerschmidt, K. (2002). An overview of the barbary macaque, *Macaca sylvanus*, vocal repertoire. *Folia Primatologica*, 73, 32-45.

Fogassi, L., et Ferrari, P. F. (2005). Neurones miroir, gestes et évolution du langage. *Primatologie*, 6, 263-286.

Fogassi L., Ferrari P.F., Gesierich, B., Rozzi, S., Chersi, F. et Rizzolatti, G. (2005). Parietal lobe: From action organization to intention understanding. *Science* Vol 308, Issue 5722, 662-667.

Fontaine, R. (1984). Imitative skills between birth and six months. *Infant Behavior and Development*, 7, 323-333.

G

Gallese, V., Fadiga, L., Fogassi, L., et Rizzolatti G. (1996). Action recognition in the premotor cortex. *Brain*, 119, 593-609.

Gallup, Jr., G. G, Anderson, J. R., et Shillito, D. J. (2002). The mirror test. In Bekoff, M., Allen, C, et Burghardt, G. M., *The cognitive animal: empirical and theoretical perspectives on animal cognition* (pp. 325-333), The MIT Press.

Gentil, M. (1990). Organization of the articulatory system: peripheral mechanisms and central coordination. In Hardcastle, W. J., and Marchal, A. (eds.), *Speech production modeling* (pp. 1-22). Kluwer academic Publishers.

Gillis, S., Koopmans-van Beinum, F., et van der Stelt, J. (1988). The prelinguistic stage. In P. Jordens & J. Lalleman (Ed.), *Language Development* (pp. 29-56). Foris Publications Dordrecht.

Goldstein, U.G. (1980). *An articulatory model for the vocal tract of the growing children*. Thèse de doctorat, MIT, Boston (non publiée).

Gordon, R. M. (2001). Simulation vs. Theory-Theory. In Wilson, R. A., et Keil, F. C. (eds), *The MIT encyclopedia of the cognitive sciences* (pp. 765-766), MIT Press.

Graur D., et W-H., Li, (2000) *Fundamentals of Molecular evolution*. Sinauer associates, inc., Publishers, Sunderland, Massachusetts pp. 217-24

Green, J. R., Moore, C. A., Higashikawa, M., et Steeve, R. W. (2000). The physiologic development of motor control: lip and jaw coordination. *Journal of Speech, Language and Hearing Research*, 43, 239-255.

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, 66-79.

Grèzes, J. (2000). *Neuroanatomie Fonctionnelle des Liens entre la Perception et l'Action chez l'Homme*. Thèse de doctorat, Université Claude Bernard, Lyon I.

Guenther, F.H. (1995). Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psychological Review*, 102, 594-621.

H

Hadamard, J. (1923). *Lectures on Cauchy's Problem in Linear Partial Differential Equations*. Yale University Press.

Hanna, E., et Meltzoff, A. N. (1993). Peer imitation by toddlers in laboratory, home, and day-care contexts: implications for social learning and memory. *Developmental Psychology*, 29 (4), 701-710.

Hardcastle, W.J. (1976). *Physiology of speech production*. London: Academic Press.

Harris, M., et Butterworth, G. E. (2002). *Developmental psychology: a student's handbook*. Psychology Press Ltd.

Herman, L. M. (2002). Exploring the cognitive world of the bottlenosed dolphin. In Bekoff, M., Allen, C, et Burghardt, G. M., *The cognitive animal: empirical and theoretical perspectives on animal cognition* (pp. 275-283), The MIT Press.

Hoole, P. (1987). Bite-block speech in the absence of oral sensibility. *Proc. ICPHS, Tallin*, 4, 16-19.

I

Iacoboni, M. (2003). Understanding others: imitation, language, empathy. In Hurley, S., and Chater, N. (Eds), *Perspectives on imitation: from cognitive neuroscience to social science*, Cambridge, MA: MIT Press.

Iacoboni, M., Koski, L. M., Brass, M., Bekkering, H., Woods, R. P., Dubeau, M.C., Mazziotta, J. C., et Rizzolatti, G. (2001). Reafferent copies of imitated actions in the right superior temporal cortex. *Proc. Natl Acad. Sci. U S A*, 98, 13995-13999.

Iacoboni, M., Woods, R. P., Brass, M., Bekkering, H., Mazziotta, J. C., et Rizzolatti, G. (1999). Cortical mechanisms of human imitation. *Science*, 286, 2526-2528.

J

Jakobson, R. (1968/1972). *Child language, aphasia, and phonological universals*. The Hague: Mouton.

Johnson, K. (1997/2003). *Acoustic and auditory phonetics*. Blackwell Publishing.

Jordan, M. I. (1990). Motor learning and the degrees of freedom problem. In Jeannerod, M. (Eds). *Attention and performance* (Vol. 13). Hillsdale, NJ: Lawrence Erlbaum.

Jordan, M. I., & Rumelhart, D. E. (1991). Forward models: supervised learning with a distal teacher. Occasional paper 40. Cambridge, MA: MIT, Center for Cognitive Sciences.

K

Kent, R. D. (1997). *The Speech Sciences*. Singular Publishing Group, Inc.

Kent, R.D., Martin, R.E., & Sufit, R.L. (1990). Oral sensation: a review and clinical prospective. In H. Winitz (ed.), *Human Communication and its Disorders* (pp. 135-191). Norwood, NJ: Ablex Publishing.

Kent, R.D., et Miolo, G. (1995), Phonetic Abilities in the First Year of Life. In Fletcher, P. & MacWinney (Eds.). *The Handbook of Child Language* (pp. 303-334). Blackwell Publishers.

Kent, R. D., & Murray, A. D. (1982). Acoustic features of infant vocalic utterances at 3, 6 and 9 months. *Journal of the Acoustical Society of America*, 72, 353-365.

Keyser, C., Kohler, E, Umiltà, M. A., Nanetti, L., Fogassi, L et Gallese, V. (2003). Audiovisual mirror neurons and action recognition. *Experimental Brain research*, 153, 628-636.

Klein, P. J, et Meltzoff, A. N. (1999). Long-term memory, forgetting, and deferred imitation in 12-month-old infants. *Developmental Science*, 2 (1), 102-113.

Kohler, E., Keyers, C., Umiltà, M. A., Fogassi, L., Gallese, V., et Rizzolatti, G. (2002). Hearing sounds, understanding actions: Action representation in mirror neurons. *Science*, 297, 846-848.

Koopman-van Beinum, F. J., & van der Stelt, J. M. (1998). Early speech development in child's acquiring Dutch mastering general basic elements. In Gillis, S. & de Hower, A. (Ed.). *The Acquisition of Dutch* (pp.101-425). Amsterdam/Philadelphia Benjamins.

Koopmans-Van Beinum, F., & Van Der Stelt, J. (1986). Early stages in the development of speech movements. In B. Lindblom, & Zetterstrom, R. (eds.), *Precursors of Early Speech* (pp. 37-49). New York: Stockton Press.

Kugiumutzakis, G. (1999). Genesis and development of early infant mimesis to facial and vocal models. In J. Nadel et G. Butterworth (eds), *Imitation in Infancy* (pp. 36-59), Cambridge University Press.

Kuhl, P. K.(1991). Human adults and infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics*, 50, 93-107.

Kuhl, P. K., et Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science*, 218, 1138-1141.

Kuhl, P. K., et Meltzoff, A. N. (1984). The intermodal representation of speech in infants. *Infant Behavior and Development*, 7, 361-381.

Kuhl, P. K., & Meltzoff, A. N. (1996). Infant vocalizations in response to speech: vocal imitation and developmental change. *Journal of the Acoustical Society of America*, 100, 2425-2438.

Kuhl, P. K., William, K. A., Lacerda, F., Stevens, K. N., et Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 month of age. *Science*, 255, 606-608.

Kuhl, P. K., William, K. A., et Meltzoff, A. N. (1991). Cross-modal speech perception in adults and infants using nonspeech auditory stimuli. *Journal of Experimental Psychology*, 17 (3), 829-840.

L

Laboissière, R. (1992). *Préliminaires pour une robotique de la communication parlée : inversion et contrôle d'un modèle articulatoire du conduit vocal*. Thèse de Docteur de l'INPG, Signal-Image-Parole, Grenoble, France.

Laboissière, R., Schwartz, J.L., et Bailly, G. (1991). Motor control for speech skills : a connectionist approach. In D.S. Touretzky, J.L. Elman, T.J. Sejnowski, & G.E. Hinton (eds.) *Connectionist Models, Proc. 1990 Summer School* (pp. 319-327). San Mateo CA : Morgan Kaufmann Publishers.

Laboissière, R., Schwartz, J.L., & Bailly, G. (1991). Motor control for speech skills : a connectionist approach. In D.S. Touretzky, J.L. Elman, T.J. Sejnowski, & G.E. Hinton (eds.) *Connectionist Models, Proc. 1990 Summer School* (pp. 319-327). San Mateo CA : Morgan Kaufmann Publishers.

du Lac, S. (2001). Motor Learning. In Wilson, R. A., et Keil, F. C. (eds), *The MIT encyclopedia of the cognitive sciences* (pp. 571-572), MIT Press.

Lalevée, C. & Vilain, A. (2003). Development of speech frame control: a longitudinal study of the oral/nasal control. *Proc. 15th ICPhS*, Barcelona.

Lalevée, C., & Vilain, A. (2005). From the proto-syllabic frame to the syllable: an audio-visual survey of 2 children from 6 to 15 months. *Child Phonology Conference*, Texas Christian University, Fort-Worth.

Landgren, S., & Olsson, K.A. (1982). Oral mechanoreceptors. In S. Grillner (ed.) *Speech Motor Control*. Oxford: Pergamon. Hoole, P. (1987). Bite-block speech in the absence of oral sensibility. *Proc. ICPhS, Tallinn*, 4, 16-19.

Laprie, Y. et Ouni, S. (2002). Introduction of constraints in an acoustic-to-articulatory inversion method based on a hypercubic articulatory table. *International Conference on Spoken Language Processing*, Denver, USA.

Lebeltel, O., Bessière, P., Diard, J., & Mazer, E. (2003). Bayesian robot programming. *Autonomous Robot* (in press), 2003.

Lebib, R., et Baudonnière, P.-M. (1999). L'imitation vocale du nouveau-né au nourrisson de 3-4 mois: les "premiers pas" vers l'acquisition du langage. *Devenir*, 11 (2), 37-52.

Lebib, R., et Baudonnière, P.-M. (2000). Vocal imitation in 3-month-old infants: a releasing stimulus or a releasing context?. *Current Psychology Letters*, 2, 79-93.

Legay, J.-M. (1997). *L'expérience et le modèle*. INRA (Ed.).

Legay, J.-M., et Debouzie, D., (1985). *Introduction à une biologie des populations*. Masson (Ed.).

Legerstee, M. (1990). Infants use multimodal information to imitate speech sounds. *Infant Behavior and Development*, 13, 343-354.

Lieberman, P. (1980). On the development of vowel productions in young infants. In Yeni-Komshian, G.; Kanagh, J.; Fergusson, C. (eds). *Infant Phonology, Vol. 1: Production* (pp. 23-42). New York: Academic Press.

Liljencrants, J., & Lindblom, B. (1972). Numerical simulations of vowel quality systems: The role of perceptual contrast. *Language*, 48, 839-862.

Lieberman, A.M., & Mattingly, I.G. (1985). The motor theory of speech perception revised. *Cognition*, 21, 1-36.

Lindblom, B. (1986). Phonetic universals in vowel systems. In *Experimental Phonology* (J.J. Ohala and J.J. Jaeger, eds.), pp. 13-44. New-York: Academic Press.

Lindblom, B. (1990). On the notion of possible speech sound. *Journal of Phonetics*, 18, 135-152.

Lindblom, B., Lubker, J., & Gay, T. (1979). Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation. *Journal of Phonetics*, 7, 147-161.

Lindblom, B., Lubker, J., & McAllister, R. (1977). Compensatory articulation and the modeling of normal speech production behavior. In R. Carré et al. (eds.) *Articulatory modeling and phonetics* (pp. 147-161). GALF.

Luppino G., Belmalih A., Calzavara R., Matelli M. et Rozzi S. (2004). Afferents from visual, somatosensory and auditory areas to the cortical convexity of the inferior parietal lobule of the macaque monkey. 4th Forum of European Neuroscience.

Luppino, G., et Rizzolatti, G. (2000). The organization of the frontal motor cortex. *News in Physiological Sciences*, 15, 219-224.

Lycan, W. (2001). Intentionality. In Wilson, R. A., et Keil, F. C. (eds), *The MIT encyclopedia of the cognitive sciences* (pp. 413-415), MIT Press.

M

Mackenzie Beck, J. (1997). Organic variation of the vocal apparatus. In W.J. Hardcastle & J. Laver (eds.) *The Handbook of Phonetic Sciences* (pp. 256-297). London: Blackwell Publishers.

MacNeilage, P. F. (1998). The Frame/Content Theory of Evolution of Speech Production. *Behavioral and Brain Sciences*, 21, 499-511.

MacNeilage P. F., et Davis, B.L. (2001). Motor mechanisms in speech ontogeny: phylogenetic, biological and linguistic implications. *Current Opinion in Neurobiology*, 11, 567-569.

- MacNeilage, P. F., et Davis, B. (1990a). Acquisition of Speech Production, Frames then Content. in M. Jannerod (ed), *Attention and Performance, XIII Motor Representation and Control*, (pp.453-476).
- MacNeilage, P. F., et Davis, B. (1990b). Acquisition of Speech Production, the Achievement of Segmental independence in Speech *Production and Speech Modelling* (pp. 55-68), Ed. Kluwer Academic Publishers.
- MacNeilage, P.F., Rootes, T.P., et Chase, R.A. (1967). Speech production and perception in a patient with severe impairment of somesthetic perception and motor control. *Journal of Speech and Hearing Research*, 10, 449-467.
- Maeda, S. (1990). Compensatory articulation during speech: evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model. In W.J. Hardcastle & A. Marchal (eds.). *Speech Production and Modelling* (pp. 131-149). Kluwer: Academic Publishers.
- Maeda, S., & Honda, K. (1994). From EMG to formant patterns of vowels: the implication of vowel systems and spaces. *Phonetica*, 51, 17-29.
- Majid, R. (1986): Modélisation articulatoire du conduit vocal. Exploration et exploitation. Fonctions de macro-sensibilité paramétriques et voyelles du français, Thèse de Doctorat, Grenoble, INP.
- Miall, R. C. (2003). Connecting mirror neurons and forward models. *NeuroReport*, 14 (16), 1-3.
- Markey, K. L. (1994). *The sensorimotor foundations of phonology: a computational model of early childhood articulatory and phonetic development*. Thèse de Doctorat, Université du Colorado. Chicago.
- Martin, J.A.M. (1981). Voice, Speech and Language in the Child.; Development and Disorder. In G.E. Arnold, F. Winckel & B.D. Wyke (eds.) *Disorders of Human Communication, vol. 4*. New-York: Springer-Verlag.
- Matyear, C. L. (1997). *An acoustical study of vowels in babbling*. Thèse de doctorat. Université du Texas. Austin (non publiée).
- Matyear, C. L., MacNeilage, P. F., & Davis, B. L. (1998). Nasalization of vowels in nasal environments in babbling: evidence for frame dominance. *Phonetica*, 55, 1-17.
- Maurer, D. (1993). Neonatal synesthesia: Implications for the processing of speech and faces. In de Boysson-Bardies, B., de Schonen, S., Jusczyk, P., MacNeilage, P., et Morton, J. (eds), *Developmental Neurocognition: speech and face processing in the first year of life* (pp. 109-124). Kluwer Academic Publishers.
- McGurk, H., et MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264, 746-748.
- Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., et Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, 29, 143-178.
- Meltzoff A. N. (1988a). Infant imitation after a 1-week delay: long-term memory for novel acts and multiple stimuli. *Developmental Psychology*, 24 (3), 470-476.

- Meltzoff A. N. (1988b). Infant imitation and memory: nine-month-olds in immediate and deferred tests. *Child Development*, 59, 217-225.
- Meltzoff A. N. (1988c). Imitation of televised models by infants. *Child Development*, 59, 1221-1229.
- Meltzoff, A. N. (1995a). What infant memory tells us about infantile amnesia: long-term recall and deferred imitation. *Journal of Experimental Child Psychology*, 59, 497-515.
- Meltzoff, A. N. (1995b). Understanding the intentions of others: re-enactment of intended acts by 18-month-old children. *Developmental Psychology*, 31 (5), 838-850.
- Meltzoff, A. N. (1999). Origins of the theory of mind, cognition and communication. *Journal of Communication Disorders*, 32, 251-269.
- Meltzoff, A. N. (2002a). Elements of a developmental theory of imitation. In A. N. Meltzoff et W. Prinz (eds), *The Imitative Mind* (pp. 19-41), Cambridge University Press.
- Meltzoff, A. N. (2002b). La théorie du "like me", précurseur de la compréhension sociale chez le bébé: imitation, intention et intersubjectivité. In J. Nadel et J. Decety (eds), *Imiter pour découvrir l'humain* (pp. 33-57), Presses Universitaires de France.
- Meltzoff, A. N., et Borton, R. W. (1979). Intermodal matching by human neonates. *Nature*, 282 (issue n°5737), 403-404.
- Meltzoff, A. N., et Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. *Science*, 198, 75-18.
- Meltzoff, A. N., et Moore, M. K. (1983). Newborn infants imitate adult facial gestures. *Child Development*, 54, 702-709.
- Meltzoff, A. N., et Moore, M. K. (1989). Imitation in newborn infants: exploring the range of gestures imitated and the underlying mechanisms. *Developmental Psychology*, 25 (6), 954-962.
- Meltzoff, A. N., et Moore, M. K. (1992). Early imitation within a functional framework: the importance of person identity, movement and development. *Infant Behavior and Development*, 15, 479-505.
- Meltzoff, A. N., et Moore, M. K. (1994). Imitation, memory, and the representation of persons. *Infant Behavior and Development*, 17, 83-99.
- Meltzoff, A. N., et Moore, M. K. (1995). Infant's understanding of people and things: from body imitation to folk psychology. In J. L. Bermúdez, A. Marcel et N. Eilan (eds), *The Body and the Self* (pp. 43-69), The MIT Press.
- Meltzoff, A. N., et Moore, M. K. (1997). Explaining facial imitation: a theoretical model. *Early Development and Parenting*, 6, 179-192.
- Meltzoff, A. N., et Moore, M. K. (1999). Persons and representation: why infant imitation is important for theories of human development. In J. Nadel et G. Butterworth (eds), *Imitation in Infancy* (pp. 9-35), Cambridge University Press.

Meltzoff, A. N., et Moore, M. K. (2000). Imitation of facial and manual gestures by human neonates. In D. Mind and A. Blater (eds), *Infant development, the essential readings* (pp. 167-181), Blackwell Pub. Ltd.

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.

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. *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 Language, Speech and Hearing Research*, 47, 1059-1080.

Menn, L., et Stoel-Gammon, C. (2001). Phonological development: learning sounds and sound patterns. In Bearko Gleason, J. (ed.), *The development of language* (pp. 70-124), Allyn & Bacon.

Mills, A.E. (1987). The development of phonology in the blind child. In B. Dodd and R. Campbell (Eds.), *Hearing by eye : the psychology of lipreading* (pp. 145-161). London: Lawrence Erlbaum Associates.

Mitchell, R. (2002). Kinesthetic-visual matching, imitation, and self recognition. In Bekoff, M., Allen, C, et Burghardt, G. M., *The cognitive animal: empirical and theoretical perspectives on animal cognition* (pp. 345-351), The MIT Press.

Moore, C. A., et Ruark, J. L. (1996). Does speech emerge from earlier appearing oral motor behaviors? *Journal of Speech, Language and Hearing Research*, 39, 1034-1047.

Mulford, R. (1988). First words of the blind child. In M.D. Smith & J.L. Locke (eds.) *The emergent lexicon: The child's development of a linguistic vocabulary* (pp. 293-338). New-York: Academic Press.

Munhall, K. G., et Jones, J. A. (1998). Articulatory evidence for syllabic structure. *Behavioral and Brain Sciences*, 21, 524-525

N

Nadel, J., et Butterworth, G., (1999). Immediate imitation rehabilitated at last. In J. Nadel et G. Butterworth (eds), *Imitation in Infancy* (pp. 1-5), Cambridge University Press.

Nguyen, D. (2002). La communication des cétacés, *Pour la Science*, Hors-série n°34, 98-102.

O

Ouni, S. et Laprie, Y. (2005). Modeling the articulatory space using a hypercube codebook for acoustic-to-articulatory inversion. *J. Acoust. Soc. Am.*, 118(1), pp. 444-460.

P

- Paulesu, E., Frith, C.D. & Frackowiak, R.S.J. (1993). The neural correlates of the verbal components of working memory. *Nature*, 362: 342-344.
- Patterson, M. L. et Werker, J. F. (2003). Two-month-old infants match phonetic information in lips and voice. *Developmental Science*, 6(2), pp 191–196.
- Paviot, Y. (2000). *Un androïde capable d'apprendre le langage?*. Mémoire de Maîtrise, Université Joseph Fourier (non publiée).
- Payan, Y., et Perrier, P. (1997). Synthesis of V-V sequences with a 2D biomechanical tongue model controlled by the Equilibrium Point Hypothesis. *Speech Communication*, 22, 185-205.
- Pélorson X. (2001). On the meaning and accuracy of the pressure-flow technique to determine constrictions within the vocal tract. *Speech Communication*, 35, 179-190.
- Petit, L. (2002). L'anatomie corticale du lobe pariétal. In Houdé, Mazoyer et Tzourio-Mazoyer (eds.), *Cerveau et psychologie: introduction à l'imagerie cérébrale anatomique et fonctionnelle* (pp. 91-100). Presses Universitaires de France.
- Piaget, J. (1945/1976). *La formation du symbole chez l'enfant*. Ed. Delachaux et Niestlé.
- Piquemal, M., Schwartz, J.L., Berthommier, F., Lallouache, T., et Escudier, P. (1996). Détection et localisation auditive d'explosions consonantiques dans des séquences VCV bruitées. *Actes des XXIèmes Journées d'études sur la parole, SFA*, 143-146.
- Polka, L., et Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, 22 (2), 421-435.
- Pols, L.C.W. (1975). Analysis and synthesis of speech using a broad-band spectral representation. In *Auditory Analysis and Perception of Speech* (G. Fant & M.A.A. Tatham, eds.), pp. 23-36. London: Academic.
- Purves, D., Augustine, G. J., Fitzpatrick, D., Katz, L. C., Lamantia, A. S., et McNamara, J. O. (1997/1999). *Neurosciences*. Ed. DeBoeck Université.

R

- Recasens, D. (1991). An electropalatographic and acoustic study of consonant-to-vowel coarticulation. *Journal of Phonetics*, 19, 177-192.
- Rizzolatti, G., et Arbib, M. A. (1998) Language within our grasp. *Trends Neurosci.* 21, 188–194.
- Rizzolatti, G., et Craighero, L. (2004). The mirror-neuron system. *Annual Review in Neurosciences*, 27:169–92.
- Rizzolatti, R., Fadiga, L., Gallese, V., Fogassi, L. (1996). Premotor cortex and recognition of motor actions. *Cognitive Brain Research*, 3, 131-141.

Rizzolatti, G., Fogassi, L. and Gallese, V. (2000) Cortical mechanisms subserving object grasping and action recognition: a new view on the cortical motor functions. In M. Gazzaniga (ed.), *The new Cognitive Neurosciences* (pp. 539-552), MIT Press.

Rizzolatti, R., Fogassi, F., et Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature reviews*, 2, 661-670.

Rizzolatti, G., et Luppino, G. (2001). The Cortical Motor System. *Neuron*, 31, 889–901.

Rizzolatti, G., Luppino, G., et Matelli, M. (1998). The organization of the cortical motor system: new concepts. *Electroencephalography and clinical Neurophysiology*, 106, 283-296.

Romand, R. (2000). Introduction au fonctionnement du système auditif. In Escudier, P., et Schwartz, J-L., (eds.), *La parole* (pp.101-133). Hermès,

Rossato, S. (2000). Du son au geste, inversion de la parole: le cas des voyelles nasales. Thèse de Doctorat, INP, Grenoble.

Rousset I. (2004). Structures syllabiques et lexicales des langues du monde : données, typologies, tendances universelles et contraintes substantielles. Thèse de doctorat en Sciences du Langage, Université Stendhal, Grenoble.

S

Sato, M., Baciú, M., Løevenbruck, H., Schwartz, J.-L., Cathiard, M.-A., Segebarth, C. & Abry, C. (2004). Multistable representation of speech forms: An fMRI study of verbal transformations. *NeuroImage*, 23(3): 1143-1151.

Savariaux, C., Perrier, P., et 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. *Journal of the Acoustical Society of America*, 98, 2428-2442.

Savariaux, C., Perrier, P., Orliaguet, J.P. & Schwartz, J.L. (1999). Compensation strategies for the perturbation of French [u] using a lip tube. II. Perceptual analysis. *Journal of the Acoustical Society of America*, 106, 381-393.

Schroeder, M.R., Atal, B.S., and Hall, J.L. (1979). Objective measure of certain speech signal degradations based on masking properties of human auditory perception. In B. Lindblom and S. Ohman (eds.) *Frontiers of Speech Communication Research* (pp. 217-229). London: Academic Press.

Schwartz, J.L. (2004) La parole multisensorielle : Plaidoyer, problèmes, perspective . *XXVèmes Journées d'Étude sur la Parole*, xi-xviii.

Schwartz, J.L., Abry, C., Boë, L.J., & Cathiard, M. (2002). Phonology in a theory of perception-for-action-control. In J. Durand, B. Laks (eds.) *Phonetics, Phonology and Cognition* (pp. 255-280). Oxford: Oxford University Press.

Schwartz, Abry et Escudier, (2000) Quelques questions posées par la parole au langage. In Escudier, P., et Schwartz, J-L., (eds.), *La parole* (pp. 391-397). Hermès,

Schwartz, J.L., Arrouas, Y., Beautemps, D., & Escudier, P. (1992). Auditory analysis of

speech gestures. In M.E.H. Schouten (ed.) *The Auditory Processing of Speech – From Sounds to Words* (pp. 239-252). Speech Research, 10, Berlin : Mouton de Gruyter.

Schwartz, J.L., & Boë, L.J. (2000). Predicting palatal contacts from jaw and tongue commands: a new sensory model and its potential use in speech control. *5th Seminar on speech production : Models and data*.

Schwartz, J.L., Boë, L.J., Vallée, N., & Abry, C. (1997). Major trends in vowel system inventories. *Journal of Phonetics*, 25, 233-254.

Schwartz, J.L., Robert-Ribes, J., & Escudier, P. (1998). Ten years after Summerfield ... a taxonomy of models for audiovisual fusion in speech perception. In R. Campbell, B. Dodd & D. Burnham (eds.) *Hearing by eye, II. Perspectives and directions in research on audiovisual aspects of language processing* (pp. 85-108). Hove (UK) : Psychology Press.

Sekiyama, K., Burnham, D., Tam, H., and Erdener, D. (2003) Development of auditory-visual speech perception in Japanese and English speakers. In J-L Schwartz, F. Berthommier, M-A.Cathiard, & D. Sodoyer (Eds) *Proceedings of the 2003 Auditory-visual Speech Processing Workshop* (St. Jorioz , France, Sept 4-7) , Institut de la Communication Parlée, Grenoble, pp. 43-46.

Shumaker, R. W., et Swartz, K. B. (2002). When traditional methodologies fail: cognitive studies of great apes. In Bekoff, M., Allen, C, et Burghardt, G. M., *The cognitive animal: empirical and theoretical perspectives on animal cognition* (pp. 335-343), The MIT Press.

Stevens, K. N., & House, A. S. (1955). Development of a quantitative description of vowel articulation. *Journal of the Acoustical Society of America*, 27, 401-493.

Streri, A. (1998). Action (en psychologie). In Houdé, O., Kayser, D., Koenig, O., Proust, J. et Rastier, F. (eds.), *Vocabulaire de Sciences Cognitives* (pp. 29-34). Presses Universitaire de France.

Streri, A., et Gentaz, E. (2003). Cross-modal recognition of shape from hand to eyes in human newborns. *Somatosensory & Motor Research* 2003; 20(1): 11–16

Studdert-Kennedy, M. (1993). Some theoretical implications of cross-modal research in speech perception. In de Boysson-Bardies, B., de Schonen, S., Jusczyk, P., McNeilage, P., et Morton, J. (eds), *Developmental Neurocognition: speech and face processing in the first year of life* (461-466). Kluwer Academic Publishers.

Studdert-Kennedy, M. et Goldstein, L. (2003) Launching language: The gestural origin of discrete infinity. In M.H. Christiansen et S. Kirby, (eds), *Language Evolution: The States of the Art*. Oxford University Press.

Sommerville, J.A., Woodward, A.L., & Needham, A. (2005). Action experience alters 3-month-old infants' perception of others' actions. *Cognition*, in press.

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, 1080-1096.

Sussman, H. M., Fruchter, D., Hilbert, J., et Sirosh, J. (1998). Linear correlates in the speech

signal: the orderly output constraint. *Behavioral and Brain Sciences*, 21, 241–299.

Sussman *et al.* (1991). Relational invariance for stops. *JASA*, 90 (3).

T

Thelen, E. (1981). Rhythmic behavior in infancy: an ethological perspective. *Developmental Psychology*, 17 (3), 237-257.

Thelen, E., et Smith, L. B. (1995). *A Dynamic systems approach to the development of cognition and action*. MIT Press.

Trevarthen, C. (2001). Intersubjectivity. In Wilson, R. A., et Keil, F. C. (eds), *The MIT encyclopedia of the cognitive sciences* (pp. 415-419), MIT Press.

Tzourio-Mazoyer, N. (2002a). L'anatomie corticale du lobe frontal. In Houdé, Mazoyer et Tzourio-Mazoyer (eds.), *Cerveau et psychologie: introduction à l'imagerie cérébrale anatomique et fonctionnelle* (pp. 81-90). Presses Universitaires de France.

Tzourio-Mazoyer, N. (2002b). L'anatomie corticale du lobe temporal. In Houdé, Mazoyer et Tzourio-Mazoyer (eds.), *Cerveau et psychologie: introduction à l'imagerie cérébrale anatomique et fonctionnelle* (pp. 109-123). Presses Universitaires de France.

U

Umiltà, M. A., Kohler, E., Gallese, V., Fogassi, L., Fadiga, L., Keysers, C., et Rizzolatti, G. (2001). I know what you are doing: A neurophysiological study. *Neuron*, 31, 1-20.

V

Vallée, N. (1994). *Systèmes vocaliques: de la typologie aux predictions*. Thèse de doctorat, Université Stendhal, Grenoble.

van der Stelt, J. M., et Koopman-van Beinum, F. J. (1986). The onset of babbling related to gross motor development. In B. Lindblom, & Zetterstrom, R. (eds.), *Precursors of Early Speech* (pp. 163-173). New York: Stockton Press.

van der Stelt, J. M., Wempe, T. G., et Pols, L.C.W. (2003). Progression in vowel production: comparing deaf and hearing children. Institute of Phonetic Sciences, University of Amsterdam, *Proceedings* 25, 197-206.

Vauclair, J. (2004). Développement du jeune enfant: motricité, perception, cognition. Belin (Ed.)

Vilain, A. (2000). *Apports de la modélisation des degrés de liberté articulatoires à l'étude de la coarticulation et du développement de la parole*. Thèse de doctorat, Université Stendhal, Grenoble.

Vinter, A. (1998). Comment le bébé construit le monde. *Sciences et vie*, Hors-série n°204, 6-15.

W

de Waal, F. (1982): *La politique du chimpanze*. Odile Jacob (Ed.).

Walton, G. E., et Bower, T. G. R. (1993). Amodal representation of speech in infants. *Infant Behavior and Development*, 16, 233-243.

Wood, S. (1979). A radiographic analysis of constriction locations for vowels. *Journal of Phonetics*, 7, 25-43.

Wu, Z.L., Schwartz, J.L., et Escudier, P. (1996). Physiologically plausible modules for the detection of articulatory-acoustic events. In B. Ainsworth (ed.) *Advances in Speech, Hearing and Language Processing, Vol.3 : Cochlear Nucleus* (pp. 479-495). U.K. : JAI Press.



ANNEXES

ANNEXE A1: Des détails sur le calcul de P(Mi/D) (Chapitre 4)

P(Mi / D) se décompose comme suit:

$$P(Mi / D) = P(D / Mi) * P(Mi) / P(D).$$

Les sous-modèles ont *a priori* la même probabilité d'occurrence: P(Mi) est uniforme. De plus, la distribution des données réelles P(D) est la même quelque soit le sous-modèle, et peut donc être ignorée. L'on se centre donc sur P(D/Mi).

$$P(D / Mi) = \sum_j P(D_j / Mi) \quad j \in \{1, \dots, N\}$$

où Dj est la j^{ième} vocalisation (F1, F2) et N est le nombre total de vocalisations dans un jeu de données réelles donné, D. A 4 mois, N=45, et à 7 mois, N=98.

Le calcul de la probabilité ci-dessus nécessite une discrétisation sur ses deux dimensions du plan (F1, F2) qui correspond, ici, à une grille de 32 cases sur 32 de coté constant selon chaque axe formantique, exprimé en Hertz, et cantonnée aux limites acoustiques de l'EVM de même âge. Dans l'EVM de 4 mois, F1 varie dans le domaine [276; 1640] et F2 dans [1121; 5141] Hz. Dans l'EVM de 7 mois, F1 varie dans le domaine [297; 1686] et F2 dans [1022; 5128] Hz.

Puis, pour un sous-modèle donné Mi, la probabilité d'occurrence de chaque case, P(case_k / Mi) est calculée:

$$P(\text{case}_k / Mi) = \text{Nb_sim_case}_k / \text{Nb_sim_tot_Mi} \quad k \in \{1, \dots, 32*32\}$$

où Nb_sim_case_k dénote le nombre de simulations se trouvant dans la k^{ième} case et Nb_sim_tot_Mi est le nombre total de simulations générées par le sous-modèle Mi.

P(Dj / Mi) correspond à la fréquence de la case où se trouve Dj, case_{kj}:

$$P(D_j / Mi) = P(\text{case}_{kj} / Mi)$$

On peut montrer que log P(D/Mi) n'est autre que la distance de Kullback Leibler qui sépare les distributions de D et de Mi sur les 32x32 cases.

Les seules fréquences prises en compte par ce calcul sont celles des cases où se trouve au moins une vocalisation. Cependant, puisque les vocalisations sortant de l'espace acoustique du sous-modèle courant ont une faible probabilité d'être produites par celui-ci, leur P(Dj/Mi) a été fixé à 10⁻²⁰⁰. Par conséquent, moins un modèle arrive à inclure des vocalisations dans son espace de réalisation, plus son score est pénalisé. Inversement, sachant que les paramètres articulatoires sont uniformément distribués, l'augmentation de leur domaine de variation, toutes choses égales par ailleurs, implique celle du nombre total de simulations acoustiques et donc une diminution des valeurs P(case_k/Mi). Ainsi le score des sous-modèle décroît-il quand leur espace acoustique de réalisation tend à déborder des zones de fréquence où les vocalisations réelles se trouvent. En résumé, cette procédure cherche les modèles qui collent le mieux à la distribution acoustique des vocalisations réelles, D.

ANNEXE A2 – Classement des modèles selon la procédure de cadrage articulatoire à 4 mois (Chapitre 4)

Le tableau présente les résultats pour les 250 « meilleurs modèles » avec 4 paramètres, puis pour les 250 « meilleurs modèles » avec 3 paramètres. Chaque ligne du tableau fournit successivement, de gauche à droite, le classement du modèle (sur l'ensemble de tous les modèles à 1, 2, 3 et 4 paramètres testés à 4 mois), le logarithme de la probabilité $p(D/M)$, les régions d'exploration respectives sur les 4 paramètres testés (valeurs minimum et maximum), le nombre de vocalisations du corpus qui sortent de l'espace acoustique produit par le sous-modèle testé, et la taille relative de l'espace d'exploration au sens défini dans le texte.

Nb La valeur -1.80 pour le minimum de Lh correspond en fait à la valeur -2, mais l'exploration est limitée à -1.80 à cause des contraintes acoustiques et articulatoires spécifiées dans le texte.

4Pi

GAG	Score	LHmin	LHmax	Jmin	Jmax	TBmin	TBmax	TDmin	TDmax	NbOut	Rcvt
1	-241,7	-1,80	0	-3	3	-2	3	-1	1	0	11,75
2	-241,7	-1,80	0	-3	3	-2	3	-1	1	0	11,75
3	-241,9	-1,80	0	-3	3	-3	3	-1	1	0	12,73
4	-241,9	-1,80	0	-3	3	-3	3	-1	1	0	12,73
5	-242,1	-1,80	0	-3	1	-2	1	-1	2	0	8,573
6	-242,1	-1,80	0	-3	1	-2	1	-1	2	0	8,573
7	-242,1	-1,80	0	-3	3	-2	1	-1	2	0	9,714
8	-242,1	-1,80	0	-3	3	-2	1	-1	2	0	9,714
9	-242,1	-1,80	0	-3	2	-2	1	-1	2	0	9,388
10	-242,1	-1,80	0	-3	2	-2	1	-1	2	0	9,388
11	-242,5	-1,80	0	-3	0	-2	1	-1	2	0	6,895
12	-242,5	-1,80	0	-3	0	-2	1	-1	2	0	6,895
13	-242,6	-1,80	0	-2	0	-2	1	-1	2	0	4,32
14	-242,6	-1,80	0	-2	0	-2	1	-1	2	0	4,32
15	-242,6	-1,80	0	-3	0	-3	1	-1	2	0	8,026
16	-242,6	-1,80	0	-3	0	-3	1	-1	2	0	8,026
17	-242,7	-1,80	0	-3	1	-3	1	-1	2	0	9,704
18	-242,7	-1,80	0	-3	1	-3	1	-1	2	0	9,704
19	-242,9	-1,80	0	-3	2	-3	1	-1	2	0	10,52
20	-242,9	-1,80	0	-3	2	-3	1	-1	2	0	10,52
21	-242,9	-1,80	0	-3	3	-3	1	-1	2	0	10,84
22	-242,9	-1,80	0	-3	3	-3	1	-1	2	0	10,84
23	-243,3	-1,80	0	-2	3	-2	3	-1	1	0	9,097
24	-243,3	-1,80	0	-2	3	-2	3	-1	1	0	9,097
25	-243,9	-1,80	0	-3	0	-2	0	-1	2	0	4,258
26	-243,9	-1,80	0	-3	0	-2	0	-1	2	0	4,258
27	-243,9	-1,80	0	-2	1	-2	1	-1	2	0	5,998

28	-243,9	-1,80	0	-2	1	-2	1	-1	2	0	5,998
29	-243,9	-1,80	0	-3	1	-3	2	-1	2	0	12,59
30	-243,9	-1,80	0	-3	1	-3	2	-1	2	0	12,59
31	-244	-1,80	0	-3	3	-3	2	-1	2	0	14,57
32	-244	-1,80	0	-3	3	-3	2	-1	2	0	14,57
33	-244	-1,80	0	-3	3	-2	3	-2	1	0	16,48
34	-244	-1,80	0	-3	3	-2	3	-2	1	0	16,48
35	-244	-1,80	0	-3	0	-3	2	-1	2	0	10,22
36	-244	-1,80	0	-3	0	-3	2	-1	2	0	10,22
37	-244	-1,80	0	-3	0	-2	1	-2	2	0	9,051
38	-244	-1,80	0	-3	0	-2	1	-2	2	0	9,051
39	-244	-1,80	0	-3	2	-2	3	-1	1	0	10,98
40	-244	-1,80	0	-3	2	-2	3	-1	1	0	10,98
41	-244,1	-1,80	0	-3	2	-3	2	-1	2	0	13,89
42	-244,1	-1,80	0	-3	2	-3	2	-1	2	0	13,89
43	-244,1	-1,80	0	-3	3	-2	2	-1	2	0	13,44
44	-244,1	-1,80	0	-3	3	-2	2	-1	2	0	13,44
45	-244,2	-1,80	0	-3	2	-3	3	-1	1	0	11,96
46	-244,2	-1,80	0	-3	2	-3	3	-1	1	0	11,96
47	-244,2	-1,80	0	-3	1	-2	1	-2	2	0	11,15
48	-244,2	-1,80	0	-3	1	-2	1	-2	2	0	11,15
49	-244,2	-1,80	0	-3	1	-2	3	-1	1	0	9,726
50	-244,2	-1,80	0	-3	1	-2	3	-1	1	0	9,726
51	-244,2	-1,80	0	-3	1	-3	3	-1	1	0	10,7
52	-244,2	-1,80	0	-3	1	-3	3	-1	1	0	10,7
53	-244,3	-1,80	0	-3	1	-2	3	-2	2	0	18,05
54	-244,3	-1,80	0	-3	1	-2	3	-2	2	0	18,05
55	-244,3	-1,80	0	-3	1	-2	2	-1	2	0	11,46
56	-244,3	-1,80	0	-3	1	-2	2	-1	2	0	11,46
57	-244,3	-1,80	0	-3	2	-2	2	-1	2	0	12,76
58	-244,3	-1,80	0	-3	2	-2	2	-1	2	0	12,76
59	-244,5	-1,80	0	-2	3	-3	3	-1	1	0	9,528
60	-244,5	-1,80	0	-2	3	-3	3	-1	1	0	9,528
61	-244,6	-1,80	0	-3	2	-2	1	-2	2	0	12,09
62	-244,6	-1,80	0	-3	2	-2	1	-2	2	0	12,09
63	-244,6	-1,80	0	-2	0	-2	2	-1	2	0	5,714
64	-244,6	-1,80	0	-2	0	-2	2	-1	2	0	5,714
65	-244,6	-1,80	0	-2	2	-2	1	-1	2	0	6,814
66	-244,6	-1,80	0	-2	2	-2	1	-1	2	0	6,814
67	-244,6	-1,00	0	-2	3	-2	3	-1	1	0	7,33
68	-244,6	-1,80	0	-3	0	-2	3	-2	2	0	14,18

69	-244,6	-1,80	0	-3	0	-2	3	-2	2	0	14,18
70	-244,7	-1,80	0	-3	1	-3	3	-1	2	0	15,1
71	-244,7	-1,80	0	-3	1	-3	3	-1	2	0	15,1
72	-244,7	-1,80	0	-3	3	-2	1	-2	2	0	12,42
73	-244,7	-1,80	0	-3	3	-2	1	-2	2	0	12,42
74	-244,8	-1,80	0	-3	3	-2	3	-2	2	0	21,46
75	-244,8	-1,80	0	-3	3	-2	3	-2	2	0	21,46
76	-244,8	-1,80	0	-3	1	-2	2	-2	2	0	14,93
77	-244,8	-1,80	0	-3	1	-2	2	-2	2	0	14,93
78	-244,9	-1,80	0	-3	2	-2	3	-2	2	0	20,23
79	-244,9	-1,80	0	-3	2	-2	3	-2	2	0	20,23
80	-244,9	-1,80	0	-3	0	-3	3	-1	2	0	12,07
81	-244,9	-1,80	0	-3	0	-3	3	-1	2	0	12,07
82	-244,9	-1,80	0	-2	3	-2	1	-1	2	0	7,14
83	-244,9	-1,80	0	-2	3	-2	1	-1	2	0	7,14
84	-245	-1,80	0	-3	3	-2	2	-2	2	0	17,29
85	-245	-1,80	0	-3	3	-2	2	-2	2	0	17,29
86	-245	-1,80	0	-3	3	-3	3	-1	2	0	17,86
87	-245	-1,80	0	-3	3	-3	3	-1	2	0	17,86
88	-245	-1,80	0	-3	1	-3	3	-2	2	0	19,62
89	-245	-1,80	0	-3	1	-3	3	-2	2	0	19,62
90	-245	-1,80	0	-2	1	-2	3	-1	1	0	7,075
91	-245	-1,80	0	-2	1	-2	3	-1	1	0	7,075
92	-245	-1,80	0	-3	0	-2	2	-2	2	0	11,94
93	-245	-1,80	0	-3	0	-2	2	-2	2	0	11,94
94	-245	-1,80	0	-3	0	-2	2	-1	2	0	9,092
95	-245	-1,80	0	-3	0	-2	2	-1	2	0	9,092
96	-245,1	-1,80	0	-3	2	-3	3	-1	2	0	16,84
97	-245,1	-1,80	0	-3	2	-3	3	-1	2	0	16,84
98	-245,1	-1,80	0	-3	2	-2	2	-2	2	0	16,51
99	-245,1	-1,80	0	-3	2	-2	2	-2	2	0	16,51
100	-245,2	-1,80	0	-3	3	-3	3	-2	1	0	17,9
101	-245,2	-1,80	0	-3	3	-3	3	-2	1	0	17,9
102	-245,3	-1,80	0	-3	3	-3	3	-2	2	0	23,02
103	-245,3	-1,80	0	-3	3	-3	3	-2	2	0	23,02
104	-245,3	-1,80	0	-3	1	-2	0	-1	2	0	5,111
105	-245,3	-1,80	0	-3	1	-2	0	-1	2	0	5,111
106	-245,3	-1,80	0	-3	0	-3	3	-2	2	0	15,74
107	-245,3	-1,80	0	-3	0	-3	3	-2	2	0	15,74
108	-245,4	-1,00	0	-3	3	-2	3	-1	1	0	8,948
109	-245,4	-1,80	0	-2	1	-2	2	-1	2	0	8,078

110	-245,4	-1,80	0	-2	1	-2	2	-1	2	0	8,078
111	-245,4	-1,80	0	-2	2	-2	3	-1	1	0	8,327
112	-245,4	-1,80	0	-2	2	-2	3	-1	1	0	8,327
113	-245,4	-1,80	0	-2	0	-3	1	-1	2	0	4,774
114	-245,4	-1,80	0	-2	0	-3	1	-1	2	0	4,774
115	-245,5	-1,80	0	-3	2	-3	3	-2	2	0	21,79
116	-245,5	-1,80	0	-3	2	-3	3	-2	2	0	21,79
117	-245,6	-1,80	0	-3	1	-2	3	-1	2	0	13,97
118	-245,6	-1,80	0	-3	1	-2	3	-1	2	0	13,97
119	-245,7	-1,80	0	-3	3	-2	3	-1	2	0	16,72
120	-245,7	-1,80	0	-3	3	-2	3	-1	2	0	16,72
121	-245,7	-1,80	0	-3	0	-3	3	-1	1	0	8,552
122	-245,7	-1,80	0	-3	0	-3	3	-1	1	0	8,552
123	-245,8	-1,00	0	-2	3	-3	3	-1	1	0	7,621
124	-245,9	-1,80	0	-3	2	-2	3	-1	2	0	15,71
125	-245,9	-1,80	0	-3	2	-2	3	-1	2	0	15,71
126	-245,9	-1,00	0	-3	3	-3	3	-1	1	0	9,572
127	-245,9	-1,80	0	-3	1	-2	3	-2	1	0	13,81
128	-245,9	-1,80	0	-3	1	-2	3	-2	1	0	13,81
129	-246	-1,80	0	-3	3	-2	1	-1	3	0	11,34
130	-246	-1,80	0	-3	3	-2	1	-1	3	0	11,34
131	-246	-1,80	0	-2	3	-2	2	-1	2	0	10,06
132	-246	-1,80	0	-2	3	-2	2	-1	2	0	10,06
133	-246	-1,80	0	-2	2	-2	2	-1	2	0	9,378
134	-246	-1,80	0	-2	2	-2	2	-1	2	0	9,378
135	-246	-1,80	0	-3	0	-2	0	-1	3	0	5,08
136	-246	-1,80	0	-3	0	-2	0	-1	3	0	5,08
137	-246,1	-1,80	0	-2	0	-2	3	-1	1	0	4,922
138	-246,1	-1,80	0	-2	0	-2	3	-1	1	0	4,922
139	-246,1	-1,80	0	-2	1	-3	1	-1	2	0	6,452
140	-246,1	-1,80	0	-2	1	-3	1	-1	2	0	6,452
141	-246,1	-1,80	0	-3	3	-3	2	-2	2	0	18,85
142	-246,1	-1,80	0	-3	3	-3	2	-2	2	0	18,85
143	-246,1	-1,80	0	-3	1	-3	2	-2	2	0	16,5
144	-246,1	-1,80	0	-3	1	-3	2	-2	2	0	16,5
145	-246,1	-1,80	0	-3	1	-2	3	-3	2	0	21,47
146	-246,1	-1,80	0	-3	1	-2	3	-3	2	0	21,47
147	-246,1	-1,80	0	-3	2	-2	3	-2	1	0	15,49
148	-246,1	-1,80	0	-3	2	-2	3	-2	1	0	15,49
149	-246,2	-1,80	0	-3	2	-2	1	-1	3	0	11,02
150	-246,2	-1,80	0	-3	2	-2	1	-1	3	0	11,02

151	-246,2	-1,80	0	-3	0	-2	3	-1	1	0	7,574
152	-246,2	-1,80	0	-3	0	-2	3	-1	1	0	7,574
153	-246,2	-1,80	0	-3	3	-2	3	-3	2	0	25,3
154	-246,2	-1,80	0	-3	3	-2	3	-3	2	0	25,3
155	-246,2	-1,80	0	-2	0	-2	1	-2	2	0	5,672
156	-246,2	-1,80	0	-2	0	-2	1	-2	2	0	5,672
157	-246,3	-1,80	0	-3	2	-3	2	-2	2	0	18,07
158	-246,3	-1,80	0	-3	2	-3	2	-2	2	0	18,07
159	-246,3	-1,80	0	-3	3	-3	1	-1	3	0	12,48
160	-246,3	-1,80	0	-3	3	-3	1	-1	3	0	12,48
161	-246,3	-1,00	0	-3	3	-2	1	-1	3	0	8,259
162	-246,3	-1,80	0	-3	0	-3	2	-2	2	0	13,5
163	-246,3	-1,80	0	-3	0	-3	2	-2	2	0	13,5
164	-246,3	-1,80	0	-3	2	-3	1	-1	3	0	12,16
165	-246,3	-1,80	0	-3	2	-3	1	-1	3	0	12,16
166	-246,4	-1,80	0	-3	2	-2	0	-1	2	0	5,347
167	-246,4	-1,80	0	-3	2	-2	0	-1	2	0	5,347
168	-246,4	-1,80	0	-3	3	-2	0	-1	2	0	5,357
169	-246,4	-1,80	0	-3	3	-2	0	-1	2	0	5,357
170	-246,4	-1,80	0	-2	3	-2	3	-2	1	0	12,72
171	-246,4	-1,80	0	-2	3	-2	3	-2	1	0	12,72
172	-246,4	-1,00	0	-3	3	-2	1	-1	2	0	7,18
173	-246,5	-1,80	0	-2	0	-3	2	-1	2	0	6,167
174	-246,5	-1,80	0	-2	0	-3	2	-1	2	0	6,167
175	-246,5	-1,80	0	-3	1	-3	1	-2	2	0	12,71
176	-246,5	-1,80	0	-3	1	-3	1	-2	2	0	12,71
177	-246,5	-1,80	0	-3	1	-3	1	-1	3	0	11,34
178	-246,5	-1,80	0	-3	1	-3	1	-1	3	0	11,34
179	-246,5	-1,80	0	-3	2	-2	3	-3	2	0	23,95
180	-246,5	-1,80	0	-3	2	-2	3	-3	2	0	23,95
181	-246,5	-1,80	0	-3	1	-2	1	-1	3	0	10,2
182	-246,5	-1,80	0	-3	1	-2	1	-1	3	0	10,2
183	-246,5	-1,80	0	-3	0	-2	3	-3	2	0	16,94
184	-246,5	-1,80	0	-3	0	-2	3	-3	2	0	16,94
185	-246,5	-1,80	0	-3	0	-3	1	-2	2	0	10,62
186	-246,5	-1,80	0	-3	0	-3	1	-2	2	0	10,62
187	-246,6	-1,80	0	-2	1	-3	3	-1	1	0	7,506
188	-246,6	-1,80	0	-2	1	-3	3	-1	1	0	7,506
189	-246,6	-1,80	0	-3	0	-2	3	-1	2	0	10,94
190	-246,6	-1,80	0	-3	0	-2	3	-1	2	0	10,94
191	-246,6	-1,80	0	-2	2	-3	1	-1	2	0	7,268

192	-246,6	-1,80	0	-2	2	-3	1	-1	2	0	7,268
193	-246,7	-1,80	0	-3	1	-2	0	-1	3	0	5,966
194	-246,7	-1,80	0	-3	1	-2	0	-1	3	0	5,966
195	-246,7	-1,00	0	-2	3	-2	1	-1	3	0	6,236
196	-246,7	-1,80	0	-2	0	-2	1	-1	3	0	5,093
197	-246,7	-1,80	0	-2	0	-2	1	-1	3	0	5,093
198	-246,7	-1,80	0	-2	0	-2	3	-1	2	0	7,058
199	-246,7	-1,80	0	-2	0	-2	3	-1	2	0	7,058
200	-246,7	-1,80	0	-2	0	-2	3	-2	2	0	9,182
201	-246,7	-1,80	0	-2	0	-2	3	-2	2	0	9,182
202	-246,8	-1,80	0	-3	2	-3	1	-2	2	0	13,65
203	-246,8	-1,80	0	-3	2	-3	1	-2	2	0	13,65
204	-246,8	-1,80	0	-2	1	-3	2	-1	2	0	8,532
205	-246,8	-1,80	0	-2	1	-3	2	-1	2	0	8,532
206	-246,8	-1,80	0	-2	2	-3	3	-1	1	0	8,758
207	-246,8	-1,80	0	-2	2	-3	3	-1	1	0	8,758
208	-246,8	-1,00	0	-2	3	-2	1	-1	2	0	5,61
209	-246,8	-1,80	0	-2	0	-2	2	-2	2	0	7,501
210	-246,8	-1,80	0	-2	0	-2	2	-2	2	0	7,501
211	-246,8	-1,80	0	-2	3	-3	1	-1	2	0	7,593
212	-246,8	-1,80	0	-2	3	-3	1	-1	2	0	7,593
213	-246,8	-1,80	0	-3	3	-3	1	-2	2	0	13,99
214	-246,8	-1,80	0	-3	3	-3	1	-2	2	0	13,99
215	-246,8	-1,80	0	-2	1	-2	3	-1	2	0	10,08
216	-246,8	-1,80	0	-2	1	-2	3	-1	2	0	10,08
217	-246,8	-1,80	0	-2	1	-2	3	-2	2	0	13,05
218	-246,8	-1,80	0	-2	1	-2	3	-2	2	0	13,05
219	-246,9	-1,00	0	-2	2	-2	1	-1	3	0	5,91
220	-246,9	-1,00	0	-3	2	-2	1	-1	3	0	7,933
221	-246,9	-1,80	0	-3	1	-2	1	-2	3	0	12,78
222	-246,9	-1,80	0	-3	1	-2	1	-2	3	0	12,78
223	-247	-1,80	0	-2	1	-2	1	-1	3	0	6,881
224	-247	-1,80	0	-2	1	-2	1	-1	3	0	6,881
225	-247	-1,00	0	-3	3	-3	1	-1	3	0	8,984
226	-247,1	-1,80	0	-3	0	-3	1	-1	3	0	9,556
227	-247,1	-1,80	0	-3	0	-3	1	-1	3	0	9,556
228	-247,1	-1,80	0	-3	0	-3	0	-1	2	0	5,389
229	-247,1	-1,80	0	-3	0	-3	0	-1	2	0	5,389
230	-247,1	-1,80	0	-2	3	-3	2	-1	2	0	10,51
231	-247,1	-1,80	0	-2	3	-3	2	-1	2	0	10,51
232	-247,1	-1,80	0	-3	2	-2	1	-2	3	0	13,71

233	-247,1	-1,80	0	-3	2	-2	1	-2	3	0	13,71
234	-247,2	-1,00	0	-2	2	-2	1	-1	2	0	5,284
235	-247,2	-1,80	0	-3	3	-2	1	-2	3	0	14,05
236	-247,2	-1,80	0	-3	3	-2	1	-2	3	0	14,05
237	-247,2	-1,80	0	-2	2	-3	2	-1	2	0	9,832
238	-247,2	-1,80	0	-2	2	-3	2	-1	2	0	9,832
239	-247,2	-1,80	0	-3	3	-2	3	-3	1	0	20,33
240	-247,2	-1,80	0	-3	3	-2	3	-3	1	0	20,33
241	-247,3	-1,80	0	-3	0	-2	3	-2	1	0	10,81
242	-247,3	-1,80	0	-3	0	-2	3	-2	1	0	10,81
243	-247,3	-1,80	0	-2	2	-2	1	-1	3	0	7,696
244	-247,3	-1,80	0	-2	2	-2	1	-1	3	0	7,696
245	-247,3	-1,00	0	-2	3	-2	2	-1	2	0	8,017
246	-247,4	-1,80	0	-3	3	-3	3	-3	2	0	27,13
247	-247,4	-1,80	0	-3	3	-3	3	-3	2	0	27,13
248	-247,4	-1,00	0	-3	2	-2	1	-1	2	0	6,854
249	-247,4	-1,80	0	-2	1	-2	1	-2	2	0	7,769
250	-247,4	-1,80	0	-2	1	-2	1	-2	2	0	7,769

3Pi

GAG	Score	LHmin	LHmax	Jmin	Jmax	TBmin	TBmax	TDmin	TDmax	NbOut	Rcvt
1259	-257,70	-1,80	1	0	0	-3	1	-1	3	0	0,706
1263	-257,70	-1,80	1	0	0	-3	1	-1	3	0	0,706
1267	-257,70	-1,80	1	0	0	-2	1	-1	3	0	0,706
1271	-257,70	-1,80	1	0	0	-2	1	-1	3	0	0,706
1660	-260,75	-1,80	1	0	0	-3	1	-1	2	0	0,65
1664	-260,75	-1,80	1	0	0	-3	1	-1	2	0	0,65
1668	-260,75	-1,80	1	0	0	-2	1	-1	2	0	0,65
1672	-260,75	-1,80	1	0	0	-2	1	-1	2	0	0,65
1883	-262,07	-1,80	1	0	0	-3	2	-1	3	0	0,968
1887	-262,07	-1,80	1	0	0	-3	2	-1	3	0	0,968
1891	-262,07	-1,80	1	0	0	-2	2	-1	3	0	0,968
1895	-262,07	-1,80	1	0	0	-2	2	-1	3	0	0,968
2052	-262,90	-1,80	1	0	0	-3	3	-1	3	0	1,201
2056	-262,90	-1,80	1	0	0	-3	3	-1	3	0	1,201
2060	-262,90	-1,80	1	0	0	-2	3	-1	3	0	1,201
2064	-262,90	-1,80	1	0	0	-2	3	-1	3	0	1,201
2180	-263,46	-1,80	1	0	0	-3	1	-2	3	0	0,886
2184	-263,46	-1,80	1	0	0	-3	1	-2	3	0	0,886
2188	-263,46	-1,80	1	0	0	-2	1	-2	3	0	0,886

2192	-263,46	-1,80	1	0	0	-2	1	-2	3	0	0,886
2309	-264,13	-1,80	1	0	0	-3	3	-1	2	0	1,115
2313	-264,13	-1,80	1	0	0	-3	3	-1	2	0	1,115
2317	-264,13	-1,80	1	0	0	-2	3	-1	2	0	1,115
2321	-264,13	-1,80	1	0	0	-2	3	-1	2	0	1,115
2410	-264,55	-1,80	1	0	0	-3	2	-1	2	0	0,886
2414	-264,55	-1,80	1	0	0	-3	2	-1	2	0	0,886
2418	-264,55	-1,80	1	0	0	-2	2	-1	2	0	0,886
2422	-264,55	-1,80	1	0	0	-2	2	-1	2	0	0,886
2786	-266,55	-1,80	1	0	0	-3	3	-2	3	0	1,527
2790	-266,55	-1,80	1	0	0	-3	3	-2	3	0	1,527
2794	-266,55	-1,80	1	0	0	-2	3	-2	3	0	1,527
2798	-266,55	-1,80	1	0	0	-2	3	-2	3	0	1,527
2816	-266,68	-1,80	1	0	0	-3	2	-2	3	0	1,221
2820	-266,68	-1,80	1	0	0	-3	2	-2	3	0	1,221
2824	-266,68	-1,80	1	0	0	-2	2	-2	3	0	1,221
2828	-266,68	-1,80	1	0	0	-2	2	-2	3	0	1,221
2893	-267,01	-1,80	1	0	0	-3	1	-3	3	0	1,003
2897	-267,01	-1,80	1	0	0	-3	1	-3	3	0	1,003
2901	-267,01	-1,80	1	0	0	-2	1	-3	3	0	1,003
2905	-267,01	-1,80	1	0	0	-2	1	-3	3	0	1,003
2927	-267,14	-1,80	1	0	0	-3	1	-2	2	0	0,83
2931	-267,14	-1,80	1	0	0	-3	1	-2	2	0	0,83
2935	-267,14	-1,80	1	0	0	-2	1	-2	2	0	0,83
2939	-267,14	-1,80	1	0	0	-2	1	-2	2	0	0,83
3117	-268,13	-1,80	1	0	0	-3	3	-2	2	0	1,442
3121	-268,13	-1,80	1	0	0	-3	3	-2	2	0	1,442
3125	-268,13	-1,80	1	0	0	-2	3	-2	2	0	1,442
3129	-268,13	-1,80	1	0	0	-2	3	-2	2	0	1,442
3344	-269,50	-1,80	2	0	0	-3	1	-1	3	0	1,001
3348	-269,50	-1,80	2	0	0	-3	1	-1	3	0	1,001
3352	-269,50	-1,80	2	0	0	-2	1	-1	3	0	1,001
3356	-269,50	-1,80	2	0	0	-2	1	-1	3	0	1,001
3384	-269,68	-1,80	1	0	0	-3	3	-3	3	0	1,78
3388	-269,68	-1,80	1	0	0	-3	3	-3	3	0	1,78
3392	-269,68	-1,80	1	0	0	-2	3	-3	3	0	1,78
3396	-269,68	-1,80	1	0	0	-2	3	-3	3	0	1,78
3422	-269,78	-1,80	1	0	0	-3	2	-2	2	0	1,139
3426	-269,78	-1,80	1	0	0	-3	2	-2	2	0	1,139
3430	-269,78	-1,80	1	0	0	-2	2	-2	2	0	1,139
3434	-269,78	-1,80	1	0	0	-2	2	-2	2	0	1,139

3513	-270,14	-1,80	1	0	0	-3	2	-3	3	0	1,413
3517	-270,14	-1,80	1	0	0	-3	2	-3	3	0	1,413
3521	-270,14	-1,80	1	0	0	-2	2	-3	3	0	1,413
3525	-270,14	-1,80	1	0	0	-2	2	-3	3	0	1,413
3692	-271,04	-1,80	1	0	0	-3	1	-3	2	0	0,948
3696	-271,04	-1,80	1	0	0	-3	1	-3	2	0	0,948
3700	-271,04	-1,80	1	0	0	-2	1	-3	2	0	0,948
3704	-271,04	-1,80	1	0	0	-2	1	-3	2	0	0,948
3804	-271,48	-1,80	1	0	0	-3	3	-3	2	0	1,695
3808	-271,48	-1,80	1	0	0	-3	3	-3	2	0	1,695
3812	-271,48	-1,80	1	0	0	-2	3	-3	2	0	1,695
3816	-271,48	-1,80	1	0	0	-2	3	-3	2	0	1,695
4020	-272,48	-1,80	2	0	0	-3	2	-1	3	0	1,372
4024	-272,48	-1,80	2	0	0	-3	2	-1	3	0	1,372
4028	-272,48	-1,80	2	0	0	-2	2	-1	3	0	1,372
4032	-272,48	-1,80	2	0	0	-2	2	-1	3	0	1,372
4241	-273,55	-1,80	1	0	0	-3	2	-3	2	0	1,33
4245	-273,55	-1,80	1	0	0	-3	2	-3	2	0	1,33
4249	-273,55	-1,80	1	0	0	-2	2	-3	2	0	1,33
4253	-273,55	-1,80	1	0	0	-2	2	-3	2	0	1,33
4292	-273,72	-1,80	2	0	0	-3	3	-1	3	0	1,701
4296	-273,72	-1,80	2	0	0	-3	3	-1	3	0	1,701
4300	-273,72	-1,80	2	0	0	-2	3	-1	3	0	1,701
4304	-273,72	-1,80	2	0	0	-2	3	-1	3	0	1,701
4334	-273,84	-1,80	2	0	0	-3	1	-1	2	0	0,921
4338	-273,84	-1,80	2	0	0	-3	1	-1	2	0	0,921
4342	-273,84	-1,80	2	0	0	-2	1	-1	2	0	0,921
4346	-273,84	-1,80	2	0	0	-2	1	-1	2	0	0,921
4774	-275,90	-1,80	2	0	0	-3	1	-2	3	0	1,255
4778	-275,90	-1,80	2	0	0	-3	1	-2	3	0	1,255
4782	-275,90	-1,80	2	0	0	-2	1	-2	3	0	1,255
4786	-275,90	-1,80	2	0	0	-2	1	-2	3	0	1,255
5117	-277,39	-1,80	2	0	0	-3	3	-1	2	0	1,58
5121	-277,39	-1,80	2	0	0	-3	3	-1	2	0	1,58
5125	-277,39	-1,80	2	0	0	-2	3	-1	2	0	1,58
5129	-277,39	-1,80	2	0	0	-2	3	-1	2	0	1,58
5209	-277,78	-1,80	2	0	0	-3	2	-1	2	0	1,255
5213	-277,78	-1,80	2	0	0	-3	2	-1	2	0	1,255
5217	-277,78	-1,80	2	0	0	-2	2	-1	2	0	1,255
5221	-277,78	-1,80	2	0	0	-2	2	-1	2	0	1,255
5240	-277,85	-1,80	2	0	0	-3	2	-2	3	0	1,73

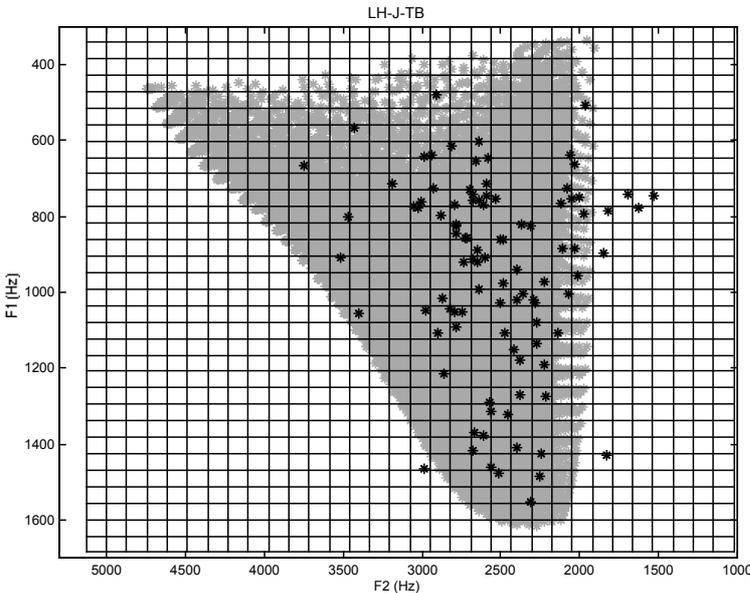
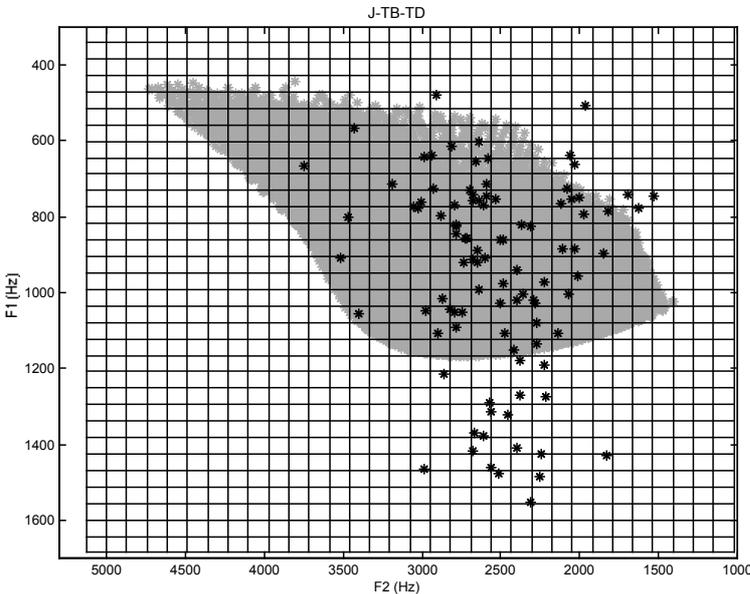
5244	-277,85	-1,80	2	0	0	-3	2	-2	3	0	1,73
5248	-277,85	-1,80	2	0	0	-2	2	-2	3	0	1,73
5252	-277,85	-1,80	2	0	0	-2	2	-2	3	0	1,73
5334	-278,23	-1,80	2	0	0	-3	3	-2	3	0	2,164
5338	-278,23	-1,80	2	0	0	-3	3	-2	3	0	2,164
5342	-278,23	-1,80	2	0	0	-2	3	-2	3	0	2,164
5346	-278,23	-1,80	2	0	0	-2	3	-2	3	0	2,164
5518	-279,00	-1,80	3	0	0	-3	1	-1	3	0	1,294
5522	-279,00	-1,80	3	0	0	-3	1	-1	3	0	1,294
5526	-279,00	-1,80	3	0	0	-2	1	-1	3	0	1,294
5530	-279,00	-1,80	3	0	0	-2	1	-1	3	0	1,294
5702	-279,76	-1,80	2	0	0	-3	1	-3	3	0	1,421
5706	-279,76	-1,80	2	0	0	-3	1	-3	3	0	1,421
5710	-279,76	-1,80	2	0	0	-2	1	-3	3	0	1,421
5714	-279,76	-1,80	2	0	0	-2	1	-3	3	0	1,421
5879	-280,51	-1,80	2	0	0	-3	1	-2	2	0	1,175
5883	-280,51	-1,80	2	0	0	-3	1	-2	2	0	1,175
5887	-280,51	-1,80	2	0	0	-2	1	-2	2	0	1,175
5891	-280,51	-1,80	2	0	0	-2	1	-2	2	0	1,175
6110	-281,65	-1,80	2	0	0	-3	2	-3	3	0	2,001
6114	-281,65	-1,80	2	0	0	-3	2	-3	3	0	2,001
6118	-281,65	-1,80	2	0	0	-2	2	-3	3	0	2,001
6122	-281,65	-1,80	2	0	0	-2	2	-3	3	0	2,001
6153	-281,77	-1,80	2	0	0	-3	3	-3	3	0	2,522
6157	-281,77	-1,80	2	0	0	-3	3	-3	3	0	2,522
6161	-281,77	-1,80	2	0	0	-2	3	-3	3	0	2,522
6165	-281,77	-1,80	2	0	0	-2	3	-3	3	0	2,522
6178	-281,83	-1,80	2	0	0	-3	3	-2	2	0	2,043
6182	-281,83	-1,80	2	0	0	-3	3	-2	2	0	2,043
6186	-281,83	-1,80	2	0	0	-2	3	-2	2	0	2,043
6190	-281,83	-1,80	2	0	0	-2	3	-2	2	0	2,043
6229	-281,98	-1,80	3	0	0	-3	2	-1	3	0	1,774
6233	-281,98	-1,80	3	0	0	-3	2	-1	3	0	1,774
6237	-281,98	-1,80	3	0	0	-2	2	-1	3	0	1,774
6241	-281,98	-1,80	3	0	0	-2	2	-1	3	0	1,774
6488	-283,24	-1,80	3	0	0	-3	3	-1	3	0	2,201
6492	-283,24	-1,80	3	0	0	-3	3	-1	3	0	2,201
6496	-283,24	-1,80	3	0	0	-2	3	-1	3	0	2,201
6500	-283,24	-1,80	3	0	0	-2	3	-1	3	0	2,201
6534	-283,38	-1,80	2	0	0	-3	2	-2	2	0	1,613
6538	-283,38	-1,80	2	0	0	-3	2	-2	2	0	1,613

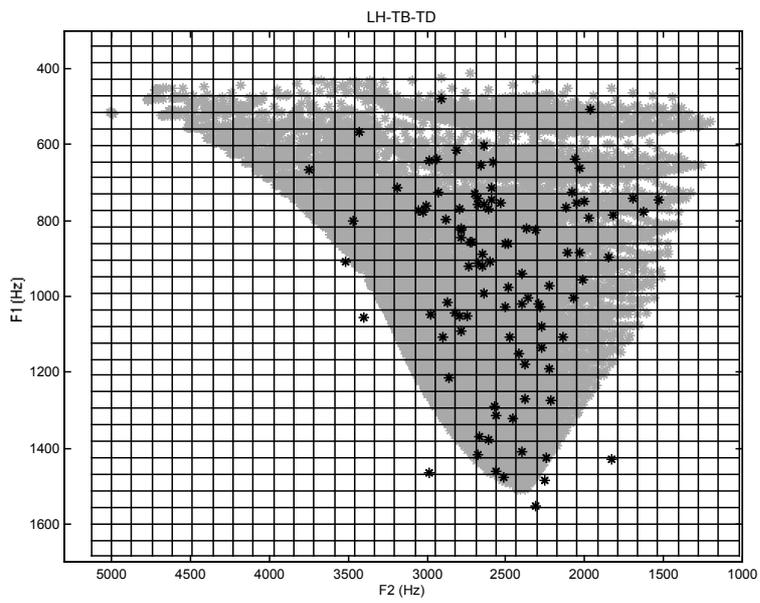
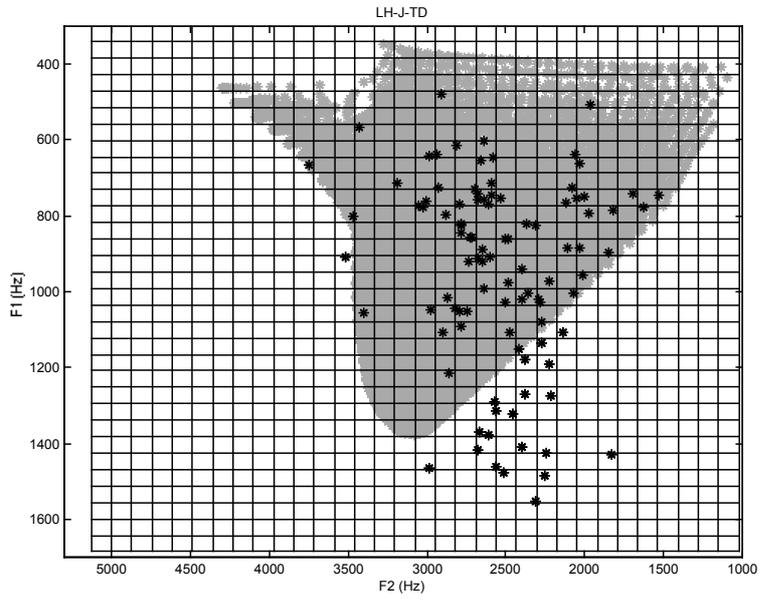
6542	-283,38	-1,80	2	0	0	-2	2	-2	2	0	1,613
6546	-283,38	-1,80	2	0	0	-2	2	-2	2	0	1,613
6756	-284,41	-1,80	3	0	0	-3	1	-1	2	0	1,192
6760	-284,41	-1,80	3	0	0	-3	1	-1	2	0	1,192
6764	-284,41	-1,80	3	0	0	-2	1	-1	2	0	1,192
6768	-284,41	-1,80	3	0	0	-2	1	-1	2	0	1,192
6791	-284,54	-1,80	2	0	0	-3	1	-3	2	0	1,342
6795	-284,54	-1,80	2	0	0	-3	1	-3	2	0	1,342
6799	-284,54	-1,80	2	0	0	-2	1	-3	2	0	1,342
6803	-284,54	-1,80	2	0	0	-2	1	-3	2	0	1,342
6970	-285,38	-1,80	2	0	0	-3	3	-3	2	0	2,401
6974	-285,38	-1,80	2	0	0	-3	3	-3	2	0	2,401
6978	-285,38	-1,80	2	0	0	-2	3	-3	2	0	2,401
6982	-285,38	-1,80	2	0	0	-2	3	-3	2	0	2,401
7039	-285,68	-1,80	3	0	0	-3	1	-2	3	0	1,623
7043	-285,68	-1,80	3	0	0	-3	1	-2	3	0	1,623
7047	-285,68	-1,80	3	0	0	-2	1	-2	3	0	1,623
7051	-285,68	-1,80	3	0	0	-2	1	-2	3	0	1,623
7289	-287,30	-1,80	2	0	0	-3	2	-3	2	0	1,884
7293	-287,30	-1,80	2	0	0	-3	2	-3	2	0	1,884
7297	-287,30	-1,80	2	0	0	-2	2	-3	2	0	1,884
7301	-287,30	-1,80	2	0	0	-2	2	-3	2	0	1,884
7351	-287,68	-1,80	3	0	0	-3	2	-2	3	0	2,238
7355	-287,68	-1,80	3	0	0	-3	2	-2	3	0	2,238
7359	-287,68	-1,80	3	0	0	-2	2	-2	3	0	2,238
7363	-287,68	-1,80	3	0	0	-2	2	-2	3	0	2,238
7407	-288,00	-1,80	3	0	0	-3	3	-1	2	0	2,044
7411	-288,00	-1,80	3	0	0	-3	3	-1	2	0	2,044
7415	-288,00	-1,80	3	0	0	-2	3	-1	2	0	2,044
7419	-288,00	-1,80	3	0	0	-2	3	-1	2	0	2,044
7440	-288,13	-1,80	3	0	0	-3	3	-2	3	0	2,799
7444	-288,13	-1,80	3	0	0	-3	3	-2	3	0	2,799
7448	-288,13	-1,80	3	0	0	-2	3	-2	3	0	2,799
7452	-288,13	-1,80	3	0	0	-2	3	-2	3	0	2,799
7483	-288,39	-1,80	3	0	0	-3	2	-1	2	0	1,623
7487	-288,39	-1,80	3	0	0	-3	2	-1	2	0	1,623
7491	-288,39	-1,80	3	0	0	-2	2	-1	2	0	1,623
7495	-288,39	-1,80	3	0	0	-2	2	-1	2	0	1,623
7657	-289,67	-1,80	3	0	0	-3	1	-3	3	0	1,839
7661	-289,67	-1,80	3	0	0	-3	1	-3	3	0	1,839
7665	-289,67	-1,80	3	0	0	-2	1	-3	3	0	1,839

7669	-289,67	-1,80	3	0	0	-2	1	-3	3	0	1,839
7817	-291,17	-1,80	3	0	0	-3	1	-2	2	0	1,52
7821	-291,17	-1,80	3	0	0	-3	1	-2	2	0	1,52
7825	-291,17	-1,80	3	0	0	-2	1	-2	2	0	1,52
7829	-291,17	-1,80	3	0	0	-2	1	-2	2	0	1,52
7894	-291,63	-1,80	3	0	0	-3	2	-3	3	0	2,588
7898	-291,63	-1,80	3	0	0	-3	2	-3	3	0	2,588
7902	-291,63	-1,80	3	0	0	-2	2	-3	3	0	2,588
7906	-291,63	-1,80	3	0	0	-2	2	-3	3	0	2,588
7927	-291,85	-1,80	3	0	0	-3	3	-3	3	0	3,263
7931	-291,85	-1,80	3	0	0	-3	3	-3	3	0	3,263
7935	-291,85	-1,80	3	0	0	-2	3	-3	3	0	3,263
7939	-291,85	-1,80	3	0	0	-2	3	-3	3	0	3,263
7999	-292,59	-1,80	3	0	0	-3	3	-2	2	0	2,642
8003	-292,59	-1,80	3	0	0	-3	3	-2	2	0	2,642
8007	-292,59	-1,80	3	0	0	-2	3	-2	2	0	2,642
8011	-292,59	-1,80	3	0	0	-2	3	-2	2	0	2,642
8120	-294,11	-1,80	3	0	0	-3	2	-2	2	0	2,087
8124	-294,11	-1,80	3	0	0	-3	2	-2	2	0	2,087
8128	-294,11	-1,80	3	0	0	-2	2	-2	2	0	2,087
8132	-294,11	-1,80	3	0	0	-2	2	-2	2	0	2,087
8213	-295,26	-1,80	3	0	0	-3	1	-3	2	0	1,736
8217	-295,26	-1,80	3	0	0	-3	1	-3	2	0	1,736
8221	-295,26	-1,80	3	0	0	-2	1	-3	2	0	1,736
8225	-295,26	-1,80	3	0	0	-2	1	-3	2	0	1,736
8277	-296,21	-1,80	3	0	0	-3	3	-3	2	0	3,106
8281	-296,21	-1,80	3	0	0	-3	3	-3	2	0	3,106
8285	-296,21	-1,80	3	0	0	-2	3	-3	2	0	3,106
8289	-296,21	-1,80	3	0	0	-2	3	-3	2	0	3,106
8365	-298,09	-1,80	3	0	0	-3	2	-3	2	0	2,437
8369	-298,09	-1,80	3	0	0	-3	2	-3	2	0	2,437
8373	-298,09	-1,80	3	0	0	-2	2	-3	2	0	2,437
8377	-298,09	-1,80	3	0	0	-2	2	-3	2	0	2,437
8625	-703,07	-1,80	0	-3	1	0	0	-3	3	1	1,045
8626	-703,07	-1,80	0	-3	1	0	0	-3	3	1	1,045
8637	-703,19	-1,80	0	-3	2	0	0	-3	3	1	1,148
8638	-703,19	-1,80	0	-3	2	0	0	-3	3	1	1,148
8663	-703,89	-1,80	0	-3	3	0	0	-3	3	1	1,173
8664	-703,89	-1,80	0	-3	3	0	0	-3	3	1	1,173
8737	-706,86	-1,00	0	-3	2	0	0	-3	3	1	0,841
8745	-707,31	-1,00	0	-3	3	0	0	-3	3	1	0,866

8790	-709,82	-1,80	0	-3	0	0	0	-3	3	1	0,83
8791	-709,82	-1,80	0	-3	0	0	0	-3	3	1	0,83
8815	-714,01	-1,80	1	-2	2	0	0	-3	3	1	1,618
8816	-714,01	-1,80	1	-2	2	0	0	-3	3	1	1,618
8824	-714,29	-1,80	1	0	0	-1	1	-3	3	1	0,854
8828	-714,29	-1,80	1	0	0	-1	1	-3	3	1	0,854
8837	-714,56	-1,80	1	-2	3	0	0	-3	3	1	1,679
8838	-714,56	-1,80	1	-2	3	0	0	-3	3	1	1,679
8848	-715,16	-1,80	1	-3	2	0	0	-3	3	1	2,158
8849	-715,16	-1,80	1	-3	2	0	0	-3	3	1	2,158
8856	-715,20	-1,80	1	-2	1	0	0	-3	3	1	1,391
8857	-715,20	-1,80	1	-2	1	0	0	-3	3	1	1,391
8864	-715,32	-1,80	1	-3	3	0	0	-3	3	1	2,219
8865	-715,32	-1,80	1	-3	3	0	0	-3	3	1	2,219
8877	-717,06	-1,80	1	-3	1	0	0	-3	3	1	1,931
8878	-717,06	-1,80	1	-3	1	0	0	-3	3	1	1,931
8884	-718,07	-1,00	1	0	0	-1	1	-3	3	1	0,782
8888	-718,94	-1,80	1	0	0	-1	2	-3	3	1	1,263
8892	-718,94	-1,80	1	0	0	-1	2	-3	3	1	1,263
8896	-719,32	-1,80	1	0	0	-1	3	-3	3	1	1,631
8900	-719,32	-1,80	1	0	0	-1	3	-3	3	1	1,631
8904	-719,37	-1,00	1	-2	2	0	0	-3	3	1	1,433
8908	-719,57	-1,00	1	-2	3	0	0	-3	3	1	1,494
8919	-722,67	-1,00	1	0	0	-3	1	-3	3	1	0,92
8923	-722,67	-1,00	1	0	0	-2	1	-3	3	1	0,92
8929	-722,79	-1,00	1	0	0	-1	2	-3	3	1	1,157

ANNEXE A3 – Figures du cadrage articulatoire à 3 paramètres pour les données de 7 mois (Chapitre 4)





ANNEXE A4 – Classement des modèles selon la procédure de cadrage articulatoire à 7 mois (Chapitre 4)

Le tableau présente les résultats pour les 250 « meilleurs modèles » avec 4 paramètres. Chaque ligne du tableau fournit successivement, de gauche à droite, le classement du modèle (sur l'ensemble de tous les modèles à 1, 2, 3 et 4 paramètres testés à 7 mois), le logarithme de la probabilité $p(D/M)$, les régions d'exploration respectives sur les 4 paramètres testés (valeurs minimum et maximum), le nombre de vocalisations du corpus qui sortent de l'espace acoustique produit par le sous-modèle testé, et la taille relative de l'espace d'exploration au sens défini dans le texte.

Nb La valeur -1.80 pour le minimum de LH correspond en fait à la valeur -2, mais l'exploration est limitée à -1.80 à cause des contraintes acoustiques et articulatoires spécifiées dans le texte.

GAG	Score	LHmin	LHmax	Jmin	Jmax	TBmin	TBmax	TDmin	TDmax	NbOut	Revt
1	-1021,70	-1,00	2	-3	3	-1	3	-1	3	1	40,58
2	-1023,18	-1,00	2	-3	3	-1	3	-2	3	1	50,61
3	-1023,59	-1.8	2	-3	3	-1	3	-2	3	1	55,56
4	-1023,59	-1.8	2	-3	3	-1	3	-2	3	1	55,56
5	-1023,65	-1.8	2	-3	3	-1	3	-1	3	1	44,57
6	-1023,65	-1.8	2	-3	3	-1	3	-1	3	1	44,57
7	-1024,45	-1,00	2	-3	3	-2	3	-1	3	1	46,01
8	-1024,53	-1,00	2	-3	3	-1	3	-1	2	1	36,07
9	-1024,62	-1.8	3	-3	3	-1	3	-1	3	1	57,31
10	-1024,62	-1.8	3	-3	3	-1	3	-1	3	1	57,31
11	-1024,77	-1,00	3	-3	3	-1	3	-1	3	1	53,33
12	-1025,02	-1,00	3	-2	3	-1	3	-1	3	1	43,29
13	-1025,03	-1.8	2	-3	3	-1	3	-1	2	1	39,38
14	-1025,03	-1.8	2	-3	3	-1	3	-1	2	1	39,38
15	-1025,15	-1,00	2	-3	2	-1	3	-1	3	1	36,74
16	-1025,49	-1,00	2	-3	3	-1	2	-1	3	1	31,67
17	-1025,66	-1.8	2	-3	3	-1	3	-2	2	1	50,38
18	-1025,66	-1.8	2	-3	3	-1	3	-2	2	1	50,38
19	-1026,20	-1.8	2	-3	3	-2	3	-1	3	1	50,82
20	-1026,20	-1.8	2	-3	3	-2	3	-1	3	1	50,82
21	-1026,40	-1.8	3	-2	3	-1	3	-1	3	1	45,74
22	-1026,40	-1.8	3	-2	3	-1	3	-1	3	1	45,74
23	-1026,41	-1.8	2	-3	2	-1	3	-2	3	1	50,91
24	-1026,41	-1.8	2	-3	2	-1	3	-2	3	1	50,91
25	-1026,49	-1,00	2	-3	3	-1	3	-2	2	1	46,11
26	-1026,66	-1.8	3	-3	3	-1	3	-2	3	1	71,45
27	-1026,66	-1.8	3	-3	3	-1	3	-2	3	1	71,45
28	-1026,68	-1.8	2	-3	2	-1	3	-1	3	1	40,73
29	-1026,68	-1.8	2	-3	2	-1	3	-1	3	1	40,73

30	-1027,07	-1,00	2	-3	2	-1	3	-2	3	1	45,96
31	-1027,25	-1,00	2	-3	2	-2	3	-1	3	1	42,17
32	-1027,37	-1.8	2	-3	3	-1	2	-1	3	1	34,9
33	-1027,37	-1.8	2	-3	3	-1	2	-1	3	1	34,9
34	-1027,69	-1,00	2	-3	3	-2	3	-2	3	1	57,44
35	-1027,71	-1,00	2	-3	2	-1	2	-1	3	1	29,1
36	-1027,81	-1.8	2	-3	1	-1	3	-2	3	1	43,94
37	-1027,81	-1.8	2	-3	1	-1	3	-2	3	1	43,94
38	-1028,08	-1,00	2	-3	3	-1	3	-3	3	1	58,75
39	-1028,11	-1,00	3	-2	3	-1	3	-2	3	1	54,25
40	-1028,12	-1.8	2	-3	3	-2	3	-2	3	1	63,43
41	-1028,12	-1.8	2	-3	3	-2	3	-2	3	1	63,43
42	-1028,14	-1.8	3	-3	2	-1	3	-1	3	1	52,2
43	-1028,14	-1.8	3	-3	2	-1	3	-1	3	1	52,2
44	-1028,16	-1.8	2	-3	3	-2	3	-1	2	1	44,99
45	-1028,16	-1.8	2	-3	3	-2	3	-1	2	1	44,99
46	-1028,17	-1,00	2	-3	3	-3	3	-1	3	1	48,35
47	-1028,18	-1.8	3	-3	3	-2	3	-1	3	1	65,26
48	-1028,18	-1.8	3	-3	3	-2	3	-1	3	1	65,26
49	-1028,22	-1,00	2	-3	3	-2	3	-1	2	1	40,96
50	-1028,29	-1,00	3	-3	3	-1	3	-2	3	1	66,5
51	-1028,41	-1.8	2	-3	1	-1	3	-1	3	1	35,13
52	-1028,41	-1.8	2	-3	1	-1	3	-1	3	1	35,13
53	-1028,41	-1.8	2	-3	3	-1	3	-3	3	1	64,54
54	-1028,41	-1.8	2	-3	3	-1	3	-3	3	1	64,54
55	-1028,49	-1,00	3	-2	2	-1	3	-1	3	1	38,18
56	-1028,57	-1.8	2	-3	2	-2	3	-1	3	1	46,98
57	-1028,57	-1.8	2	-3	2	-2	3	-1	3	1	46,98
58	-1028,58	-1.8	3	-3	3	-1	3	-1	2	1	50,72
59	-1028,58	-1.8	3	-3	3	-1	3	-1	2	1	50,72
60	-1028,65	-1,00	3	-3	3	-2	3	-1	3	1	60,45
61	-1028,66	-1.8	3	-2	3	-1	3	-2	3	1	57,33
62	-1028,66	-1.8	3	-2	3	-1	3	-2	3	1	57,33
63	-1028,90	-1,00	3	-3	2	-1	3	-1	3	1	48,22
64	-1028,98	-1.8	3	-3	3	-1	2	-1	3	1	44,85
65	-1028,98	-1.8	3	-3	3	-1	2	-1	3	1	44,85
66	-1029,07	-1.8	2	-3	1	-1	2	-1	3	1	28,15
67	-1029,07	-1.8	2	-3	1	-1	2	-1	3	1	28,15
68	-1029,15	-1.8	2	-3	2	-1	2	-1	3	1	32,33
69	-1029,15	-1.8	2	-3	2	-1	2	-1	3	1	32,33
70	-1029,25	-1,00	2	-3	1	-1	3	-1	3	1	31,15
71	-1029,27	-1,00	3	-3	3	-1	2	-1	3	1	41,62

72	-1029,30	-1.8	2	-3	2	-1	3	-2	2	1	45,72
73	-1029,30	-1.8	2	-3	2	-1	3	-2	2	1	45,72
74	-1029,34	-1.8	2	-3	2	-1	3	-1	2	1	35,54
75	-1029,34	-1.8	2	-3	2	-1	3	-1	2	1	35,54
76	-1029,48	-1.8	2	-3	0	-2	3	-1	3	1	32,25
77	-1029,48	-1.8	2	-3	0	-2	3	-1	3	1	32,25
78	-1029,54	-1.8	2	-3	1	-2	3	-1	3	1	41,38
79	-1029,54	-1.8	2	-3	1	-2	3	-1	3	1	41,38
80	-1029,62	-1,00	2	-3	2	-1	3	-1	2	1	32,24
81	-1029,65	-1.8	3	-2	2	-1	3	-1	3	1	40,63
82	-1029,65	-1.8	3	-2	2	-1	3	-1	3	1	40,63
83	-1029,72	-1.8	2	-3	3	-3	3	-1	3	1	53,58
84	-1029,72	-1.8	2	-3	3	-3	3	-1	3	1	53,58
85	-1029,76	-1,00	2	-3	1	-1	2	-1	3	1	24,92
86	-1029,85	-1,00	3	-2	3	-1	3	-1	2	1	39,52
87	-1029,92	-1.8	2	-3	0	-1	3	-2	3	1	32,99
88	-1029,92	-1.8	2	-3	0	-1	3	-2	3	1	32,99
89	-1029,94	-1,00	2	-3	1	-2	3	-1	3	1	36,58
90	-1030,05	-1,00	3	-3	3	-1	3	-1	2	1	47,41
91	-1030,15	-1.8	3	-3	2	-1	3	-2	3	1	65,25
92	-1030,15	-1.8	3	-3	2	-1	3	-2	3	1	65,25
93	-1030,27	-1.8	3	-3	1	-1	3	-1	3	1	44,86
94	-1030,27	-1.8	3	-3	1	-1	3	-1	3	1	44,86
95	-1030,33	-1,00	3	-2	3	-2	3	-1	3	1	47,87
96	-1030,43	-1.8	3	-2	3	-1	3	-1	2	1	41,62
97	-1030,43	-1.8	3	-2	3	-1	3	-1	2	1	41,62
98	-1030,83	-1.8	2	-3	3	-2	3	-2	2	1	57,6
99	-1030,83	-1.8	2	-3	3	-2	3	-2	2	1	57,6
100	-1030,83	-1.8	2	-3	2	-2	3	-2	3	1	58,77
101	-1030,83	-1.8	2	-3	2	-2	3	-2	3	1	58,77
102	-1030,84	-1.8	3	-3	1	-1	2	-1	3	1	35,94
103	-1030,84	-1.8	3	-3	1	-1	2	-1	3	1	35,94
104	-1030,89	-1.8	3	-3	3	-1	3	-2	2	1	64,86
105	-1030,89	-1.8	3	-3	3	-1	3	-2	2	1	64,86
106	-1030,99	-1.8	3	-3	2	-1	2	-1	3	1	41,41
107	-1030,99	-1.8	3	-3	2	-1	2	-1	3	1	41,41
108	-1031,04	-1.8	3	-2	1	-1	3	-1	3	1	33,29
109	-1031,04	-1.8	3	-2	1	-1	3	-1	3	1	33,29
110	-1031,11	-1.8	3	-3	2	-2	3	-1	3	1	60,15
111	-1031,11	-1.8	3	-3	2	-2	3	-1	3	1	60,15
112	-1031,15	-1,00	2	-3	3	-2	2	-1	3	1	37,1
113	-1031,15	-1,00	2	-3	2	-3	3	-1	3	1	44,51

114	-1031,24	-1.8	2	-3	3	-1	3	-3	2	1	59,35
115	-1031,24	-1.8	2	-3	3	-1	3	-3	2	1	59,35
116	-1031,35	-1,00	2	-3	2	-2	3	-2	3	1	52,78
117	-1031,45	-1.8	2	-3	2	-1	3	-3	3	1	59,43
118	-1031,45	-1.8	2	-3	2	-1	3	-3	3	1	59,43
119	-1031,52	-1.8	3	-2	3	-2	3	-1	3	1	50,77
120	-1031,52	-1.8	3	-2	3	-2	3	-1	3	1	50,77
121	-1031,54	-1.8	2	-3	0	-2	3	-2	3	1	40,24
122	-1031,54	-1.8	2	-3	0	-2	3	-2	3	1	40,24
123	-1031,62	-1,00	2	-3	2	-1	3	-2	2	1	41,45
124	-1031,64	-1.8	2	-3	0	-2	2	-1	3	1	27,17
125	-1031,64	-1.8	2	-3	0	-2	2	-1	3	1	27,17
126	-1031,65	-1.8	2	-3	2	-2	3	-1	2	1	41,15
127	-1031,65	-1.8	2	-3	2	-2	3	-1	2	1	41,15
128	-1031,75	-1.8	3	-3	3	-3	3	-1	3	1	68,74
129	-1031,75	-1.8	3	-3	3	-3	3	-1	3	1	68,74
130	-1031,86	-1.8	3	-3	3	-2	3	-2	3	1	81,44
131	-1031,86	-1.8	3	-3	3	-2	3	-2	3	1	81,44
132	-1031,87	-1,00	3	-2	1	-1	3	-1	3	1	30,84
133	-1031,88	-1.8	3	-3	3	-1	3	-3	3	1	82,96
134	-1031,88	-1.8	3	-3	3	-1	3	-3	3	1	82,96
135	-1031,89	-1,00	2	-3	3	-1	2	-1	2	1	27,5
136	-1031,90	-1,00	3	-3	2	-1	2	-1	3	1	38,18
137	-1031,91	-1.8	2	-3	1	-1	3	-2	2	1	38,8
138	-1031,91	-1.8	2	-3	1	-1	3	-2	2	1	38,8
139	-1031,98	-1,00	2	-3	0	-2	3	-1	3	1	27,9
140	-1031,98	-1,00	2	-3	3	-2	3	-2	2	1	52,39
141	-1032,00	-1,00	2	-3	3	-1	3	-3	2	1	54,24
142	-1032,03	-1.8	2	-3	3	-1	2	-1	2	1	30,11
143	-1032,03	-1.8	2	-3	3	-1	2	-1	2	1	30,11
144	-1032,07	-1.8	2	-3	3	-3	3	-1	2	1	47,72
145	-1032,07	-1.8	2	-3	3	-3	3	-1	2	1	47,72
146	-1032,08	-1,00	2	-3	3	-3	3	-2	3	1	60,66
147	-1032,09	-1.8	3	-3	1	-1	3	-2	3	1	56,12
148	-1032,09	-1.8	3	-3	1	-1	3	-2	3	1	56,12
149	-1032,17	-1,00	3	-3	2	-2	3	-1	3	1	55,34
150	-1032,17	-1.8	2	-3	2	-3	3	-1	3	1	49,74
151	-1032,17	-1.8	2	-3	2	-3	3	-1	3	1	49,74
152	-1032,22	-1.8	3	-2	2	-1	3	-2	3	1	51,13
153	-1032,22	-1.8	3	-2	2	-1	3	-2	3	1	51,13
154	-1032,25	-1.8	2	-3	1	-2	3	-2	3	1	51,81
155	-1032,25	-1.8	2	-3	1	-2	3	-2	3	1	51,81

156	-1032,28	-1,00	3	-2	2	-1	3	-2	3	1	48,05
157	-1032,30	-1,00	2	-3	3	-3	3	-1	2	1	43,28
158	-1032,35	-1,00	2	-3	1	-1	3	-2	3	1	39
159	-1032,37	-1,00	2	-3	2	-1	3	-3	3	1	53,65
160	-1032,47	-1,00	3	-3	3	-3	3	-1	3	1	63,51
161	-1032,55	-1.8	2	-3	3	-3	3	-2	3	1	67,23
162	-1032,55	-1.8	2	-3	3	-3	3	-2	3	1	67,23
163	-1032,58	-1,00	2	-3	2	-2	3	-1	2	1	37,13
164	-1032,59	-1.8	2	-3	0	-1	2	-2	3	1	26,85
165	-1032,59	-1.8	2	-3	0	-1	2	-2	3	1	26,85
166	-1032,60	-1.8	3	-3	1	-2	3	-1	3	1	52,81
167	-1032,60	-1.8	3	-3	1	-2	3	-1	3	1	52,81
168	-1032,63	-1.8	3	-3	3	-2	3	-1	2	1	57,85
169	-1032,63	-1.8	3	-3	3	-2	3	-1	2	1	57,85
170	-1032,66	-1.8	2	-3	3	-1	2	-2	3	1	43,29
171	-1032,66	-1.8	2	-3	3	-1	2	-2	3	1	43,29
172	-1032,69	-1.8	3	-2	3	-1	3	-2	2	1	53,22
173	-1032,69	-1.8	3	-2	3	-1	3	-2	2	1	53,22
174	-1032,71	-1,00	2	-3	3	-1	2	-2	3	1	39,26
175	-1032,79	-1,00	2	-3	2	-2	2	-1	3	1	34,53
176	-1032,80	-1,00	3	-2	3	-3	3	-1	3	1	49,19
177	-1032,83	-1,00	3	-2	3	-1	3	-2	2	1	50,48
178	-1032,86	-1.8	2	-3	1	-1	2	-2	3	1	35,3
179	-1032,86	-1.8	2	-3	1	-1	2	-2	3	1	35,3
180	-1032,87	-1.8	2	-3	3	-2	2	-1	3	1	41,16
181	-1032,87	-1.8	2	-3	3	-2	2	-1	3	1	41,16
182	-1033,01	-1.8	2	-3	1	-1	3	-1	2	1	29,99
183	-1033,01	-1.8	2	-3	1	-1	3	-1	2	1	29,99
184	-1033,03	-1,00	3	-3	2	-1	3	-2	3	1	60,3
185	-1033,06	-1.8	2	-3	0	-1	2	-1	3	1	21,48
186	-1033,06	-1.8	2	-3	0	-1	2	-1	3	1	21,48
187	-1033,22	-1.8	2	-3	1	-1	3	-3	3	1	51,5
188	-1033,22	-1.8	2	-3	1	-1	3	-3	3	1	51,5
189	-1033,38	-1,00	3	-2	3	-1	3	-3	3	1	63
190	-1033,39	-1.8	2	-3	1	-3	3	-1	3	1	44,15
191	-1033,39	-1.8	2	-3	1	-3	3	-1	3	1	44,15
192	-1033,45	-1,00	2	-3	3	-2	3	-3	3	1	66,56
193	-1033,46	-1.8	2	-3	1	-2	2	-1	3	1	34,4
194	-1033,46	-1.8	2	-3	1	-2	2	-1	3	1	34,4
195	-1033,49	-1.8	3	-3	0	-2	3	-1	3	1	40,96
196	-1033,49	-1.8	3	-3	0	-2	3	-1	3	1	40,96
197	-1033,57	-1,00	3	-3	3	-1	3	-3	3	1	77,18

198	-1033,57	-1,00	3	-3	3	-2	3	-2	3	1	75,46
199	-1033,67	-1,00	3	-3	3	-1	3	-2	2	1	60,59
200	-1033,72	-1.8	2	-3	2	-1	2	-2	3	1	40,35
201	-1033,72	-1.8	2	-3	2	-1	2	-2	3	1	40,35
202	-1033,73	-1.8	3	-3	0	-1	2	-1	3	1	27,29
203	-1033,73	-1.8	3	-3	0	-1	2	-1	3	1	27,29
204	-1033,73	-1,00	2	-3	1	-2	2	-1	3	1	30,35
205	-1033,75	-1,00	3	-3	1	-1	3	-1	3	1	40,88
206	-1033,79	-1.8	2	-3	0	-3	3	-1	3	1	35,01
207	-1033,79	-1.8	2	-3	0	-3	3	-1	3	1	35,01
208	-1033,84	-1.8	2	-3	3	-2	3	-3	3	1	73,56
209	-1033,84	-1.8	2	-3	3	-2	3	-3	3	1	73,56
210	-1033,89	-1.8	3	-2	1	-1	3	-2	3	1	42
211	-1033,89	-1.8	3	-2	1	-1	3	-2	3	1	42
212	-1033,90	-1,00	3	-2	2	-2	3	-1	3	1	42,77
213	-1033,97	-1.8	3	-2	3	-1	3	-3	3	1	66,63
214	-1033,97	-1.8	3	-2	3	-1	3	-3	3	1	66,63
215	-1033,98	-1.8	2	-3	1	-2	3	-1	2	1	35,6
216	-1033,98	-1.8	2	-3	1	-2	3	-1	2	1	35,6
217	-1033,99	-1,00	3	-2	3	-2	3	-2	3	1	60,04
218	-1034,00	-1.8	3	-2	3	-3	3	-1	3	1	52,23
219	-1034,00	-1.8	3	-2	3	-3	3	-1	3	1	52,23
220	-1034,05	-1.8	2	-3	2	-2	2	-1	3	1	38,58
221	-1034,05	-1.8	2	-3	2	-2	2	-1	3	1	38,58
222	-1034,10	-1,00	2	-3	0	-2	2	-1	3	1	23,49
223	-1034,15	-1,00	2	-3	1	-3	3	-1	3	1	38,92
224	-1034,21	-1,00	3	-3	1	-1	2	-1	3	1	32,71
225	-1034,24	-1.8	3	-3	2	-1	3	-1	2	1	45,61
226	-1034,24	-1.8	3	-3	2	-1	3	-1	2	1	45,61
227	-1034,33	-1.8	3	-3	0	-1	3	-2	3	1	41,91
228	-1034,33	-1.8	3	-3	0	-1	3	-2	3	1	41,91
229	-1034,41	-1.8	2	-3	2	-2	3	-2	2	1	52,94
230	-1034,41	-1.8	2	-3	2	-2	3	-2	2	1	52,94
231	-1034,50	-1.8	2	-3	0	-1	3	-1	3	1	26,56
232	-1034,50	-1.8	2	-3	0	-1	3	-1	3	1	26,56
233	-1034,54	-1.8	2	-3	2	-1	2	-1	2	1	27,53
234	-1034,54	-1.8	2	-3	2	-1	2	-1	2	1	27,53
235	-1034,55	-1.8	3	-2	3	-2	3	-2	3	1	63,68
236	-1034,55	-1.8	3	-2	3	-2	3	-2	3	1	63,68
237	-1034,58	-1.8	3	-3	0	-2	2	-1	3	1	34,5
238	-1034,58	-1.8	3	-3	0	-2	2	-1	3	1	34,5
239	-1034,61	-1.8	2	-3	0	-1	3	-3	3	1	38,66

240	-1034,61	-1.8	2	-3	0	-1	3	-3	3	1	38,66
241	-1034,73	-1.8	2	-3	0	-2	3	-1	2	1	26,95
242	-1034,73	-1.8	2	-3	0	-2	3	-1	2	1	26,95
243	-1034,77	-1.8	3	-3	2	-3	3	-1	3	1	63,63
244	-1034,77	-1.8	3	-3	2	-3	3	-1	3	1	63,63
245	-1034,78	-1,00	3	-3	3	-2	3	-1	2	1	53,83
246	-1034,81	-1.8	3	-2	2	-2	3	-1	3	1	45,66
247	-1034,81	-1.8	3	-2	2	-2	3	-1	3	1	45,66
248	-1034,82	-1.8	3	-2	0	-1	3	-2	3	1	27,79
249	-1034,82	-1.8	3	-2	0	-1	3	-2	3	1	27,79
250	-1034,85	-1,00	2	-3	2	-1	2	-2	3	1	36,32

ANNEXE A5 – Spécification de la frontière entre palatales et vélaires sur VLAM à 7 mois (Chapitre 4)

[i] (coupe sagittale en noir):

paramètres de VLAM : [2 0.14 -1.18 -3 -0.52 -1.76 0.61];

[Xc Ac A1] = [6.1589 0.2879 1.3365]

[di] (coupe sagittale en rouge):

paramètres de VLAM : [-2 0.14 -1.18 -3 -0.52 0.24 0.61]

[Xc Ac A1] = [**6.6946** 0.0008 1.3365]

[gi] (coupe sagittale en bleu):

paramètres de VLAM : [-2 0.14 -1.18 -3 1.48 -1.76 0.61]

[Xc Ac A1] = [**5.8135** 0.0122 1.3365]

La frontière en Xc entre palatales et vélaires sur VLAM à 7 mois a ainsi été fixée à 6, entre les valeurs pour [di] et [gi] (en gras)

