



HAL
open science

L'anticipation d'arrondissement vocalique et le Modèle d'Expansion du MouvementDeux extensions: du français à l'anglais et de l'adulte à l'enfant

Aude Noiray

► **To cite this version:**

Aude Noiray. L'anticipation d'arrondissement vocalique et le Modèle d'Expansion du MouvementDeux extensions: du français à l'anglais et de l'adulte à l'enfant. Linguistique. Université Stendhal - Grenoble III, 2007. Français. NNT: . tel-00180059

HAL Id: tel-00180059

<https://theses.hal.science/tel-00180059>

Submitted on 17 Oct 2007

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

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

**UNIVERSITE GRENOBLE 3 – STENDHAL
U.F.R. DES SCIENCES DU LANGAGE**

N°attribué par la bibliothèque

--	--	--	--	--	--	--	--	--	--

THESE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE STENDHAL

Discipline : « Sciences du Langage »

Préparée à l'Institut de la Communication Parlée dans le cadre de
l'**Ecole Doctorale « Langues littéraires et sciences humaines »**

Présentée et soutenue publiquement

Par

Aude Noiray

Le 29 juin 2007

Titre :

L'anticipation d'arrondissement vocalique et
le *Modèle d'Expansion du Mouvement*
Deux extensions : du français à l'anglais et de l'adulte à l'enfant

Sous la direction de :

Christian ABRY & Marie-Agnès CATHIARD

JURY

M. BONNOT Jean-François	Professeur	Rapporteur
M. ORLIAGUET Jean-Pierre	Professeur	Rapporteur
Mme. CATHIARD Marie-Agnès	MCF	Co-directrice
M. ABRY Christian	Professeur	Directeur

**UNIVERSITE GRENOBLE 3 – STENDHAL
U.F.R. DES SCIENCES DU LANGAGE**

N°attribué par la bibliothèque

--	--	--	--	--	--	--	--	--	--

THESE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE STENDHAL

Discipline : « Sciences du Langage »

Préparée à l'Institut de la Communication Parlée dans le cadre de
l'**Ecole Doctorale « Langues littératures et sciences humaines »**

Présentée et soutenue publiquement

Par

Aude Noiray

Le 29 juin 2007

Titre :

L'anticipation d'arrondissement vocalique et
le *Modèle d'Expansion du Mouvement*
Deux extensions : du français à l'anglais et de l'adulte à l'enfant

Sous la direction de :

Christian ABRY & Marie-Agnès CATHIARD

JURY

M. BONNOT Jean-François	Professeur	Rapporteur
M. ORLIAGUET Jean-Pierre	Professeur	Rapporteur
Mme. CATHIARD Marie-Agnès	MCF	Co-directrice
M. ABRY Christian	Professeur	Directeur

*A Jacques,
Un pas devant l'autre,
Un projet, l'un après l'autre...*

Remerciements

Je me suis régulièrement dit au cours de la rédaction « Quand tu en seras à la page des remerciements... ». Eh bien non ! Arrivée à ce point du chemin, nombreux sont ceux que l'on a envie de remercier et l'on ne veut évidemment oublier personne...

Je voudrais commencer par remercier mon Directeur de thèse, Christian Abry, pour m'avoir permis de continuer dans la recherche et d'apporter un éclairage sur un sujet aussi complexe que l'anticipation. Je voudrais également le remercier de m'avoir appris à gérer la pression de l'arène et préparé à l'exercice tant redouté des questions... Bien du temps s'est écoulé depuis ma première communication orale où le trac me paralysait tellement que j'en oubliais presque mon nom !

Merci pour m'avoir transmis la rigueur, encore et toujours, le souci de la vérification et du détail ; de m'avoir aidé à diminuer le nombre de mes idées reçues. Merci pour les discussions scientifiques, nombreuses et fructueuses, parsemées de digressions déroutantes ! Ces discussions où je ne comprenais rarement plus de 10% de leur contenu mais qui, quelques semaines plus tard, retentissaient, tels des déclics, ouvrant de nouvelles portes à ma réflexion.

Enfin, je voudrais le remercier de m'avoir permis de découvrir le versant phylogénétique du langage. Même si mes connaissances dans ce domaine sont encore très rudimentaires, j'ai beaucoup apprécié me plonger dans les lectures sur l'imitation chez les primates non-humains, sur le système de neurones miroirs, sur le paradigme de l'émulation vs. imitation avec ses implications pour le langage chez l'enfant. Ces lectures ont été pour moi passionnantes et ont apporté une nouvelle dimension à mon travail de thèse.

En second lieu, je voudrais remercier « ma chef », Marie-Agnès Cathiard. Que te dire d'autre qu'un grand merci ! J'ai eu grand plaisir à travailler avec toi durant ces trois années. Merci de m'avoir fait confiance en m'encadrant sur un sujet dont j'étais totalement ignorante au début du voyage. Merci pour ton implication à tous les stades de la thèse, pour ta disponibilité sans faille et pour ta grande tolérance à mes étourderies !

Merci pour ton encadrement quotidien, ta grande rigueur, ton appui expérimental et statistique. Merci également pour ton soutien, ton écoute, tes connaissances tout au long du chemin et pour m'avoir posé les bonnes questions. Tes encouragements réguliers m'ont vraiment aidé à avancer, même dans les moments où je me sentais un peu perdue.

J'ai appris à te connaître au fil des années et j'espère que nos échanges s'ancreront dans le temps... Nos discussions scientifiques ou pas autour d'un thé, nos grignotages de sucreries et nos petits mails du soir vont beaucoup me manquer !

Je vous remercie donc tous les deux pour m'avoir appris tant de choses. Votre encadrement est de loin le meilleur que j'ai reçu.

Je souhaiterais également remercier les membres de mon jury, J.-F. Bonnot et J.-P. Orliaguet, d'avoir pris le temps d'examiner mon travail. C'est un honneur pour moi de bénéficier de leurs retours scientifiques sur ces trois années de recherche. Leurs expériences dans les domaines de la coarticulation et de l'anticipation m'ouvriront de nouvelles pistes de réflexions.

Mes pensées se tournent ensuite vers la rue Sainte Catherine à Montréal. Au cours de ma thèse, j'ai eu la chance de bénéficier d'une collaboration de longue date entre Christian Abry, Marie-Agnès Cathiard et Lucie Ménard, ce qui nous a permis de mettre en place l'une des deux études décrites dans ce manuscrit.

Je remercie donc Lucie Ménard pour m'avoir invité au « Laboratoire de Phonétique » qu'elle dirige à l'Université du Québec à Montréal (UQAM). Ce séjour dans la Belle Province reste pour moi un excellent souvenir et ce n'est pas sans nostalgie que je me remémore ces quelques mois passés là bas. Je voudrais la remercier pour son accueil chaleureux et son enthousiasme scientifique qui a également nourri ma motivation de poursuivre dans la recherche. Je la remercie de m'avoir impliqué dans la vie de son laboratoire, et pour m'avoir donné l'opportunité de présenter nos résultats à plusieurs reprises. Ses encouragements et son entrain ont été pour moi d'un grand support dans les moments difficiles. De cette collaboration est née une amitié et j'espère avoir le plaisir prochainement de développer nos échanges.

*J'en profite également pour remercier toute la « gagne » du Laboratoire de Phonétique qui m'a si bien accueillie : Annie, Anthony, Caroline, Jérôme, Johana et Vincent. Merci de m'avoir fais découvrir les 5 à 7 et de m'avoir initié aux expressions québécoises. Je sais maintenant qu'il ne faut pas confondre 'tongues' avec 'gougounes' !
Merci à ceux qui ont bien voulu se prêter au jeu des lèvres bleues !
Merci à Johana d'avoir participé sans retenue au déroulement des enregistrements et d'avoir été avec Vincent mon guide lors de notre passage à Québec. Je garde un très bon souvenir de cette journée.*

Un grand merci également à Jérôme pour avoir répondu si promptement à toutes mes questions techniques, pratiques, parfois même un peu farfelues ! Ton aide et tes encouragements ont été très précieux.

Je remercie David Ostry du « Motor Control Lab » (Université McGill, Montréal) d'avoir eu la gentillesse de nous prêter les locaux ainsi que le système Optotrak pour l'enregistrement de nos locuteurs.

De retour en France, il me faut saluer les membres de l'Institut de la Communication Parlée. En premier lieu, je remercie Christophe Savariaux, Alain Arnal et Coriandre Vilain pour leur apport technique et expérimental. Merci à Christophe de m'avoir initié aux secrets de TACLE. Ce fût un apprentissage douloureux au début mais qui s'est avéré très fructueux lorsque l'on regarde les beaux résultats que nous avons obtenus. Je n'oublierai pas toutes ces heures passées à décrypter les lèvres bleues !

Merci à Alain pour son aide lors des petits « bricolages » pour l'enregistrement de nos jeunes locuteurs. Je le remercie de m'avoir fait partager ses connaissances mais aussi pour son enthousiasme et sa pédagogie.

Un grand merci à Coriandre, Virginie Attina, Christophe et de nouveau Lucie pour avoir eu la patience de m'aider à déchiffrer le langage de Matlab ou encore de Trap.

Merci à tout le « campus » pour sa bonne humeur, pour les pauses café, les pauses gourmandes, les fous rire, pour l'entraide dans les moments de rush.

Merci à Dominique, Marie-Thé et Marie-Ange pour leur disponibilité, leur gentillesse et leur aide administrative tout au long de ces années. Vous êtes extra !

Je remercie Jean-Luc Schwartz et Pierre Badin pour m'avoir donné l'opportunité de présenter régulièrement mon travail lors de conférences. Ces expériences ont toutes été enrichissantes. Elles m'ont permis de rencontrer des spécialistes de mon domaine et de recueillir de précieux conseils mais aussi d'aborder mon sujet sous différents angles.

Merci à « l'équipe pédagogique » avec laquelle j'ai fait mes premiers pas en tant qu'enseignante : Françoise Boch, Marie-Agnès Cathiard, Solange Rossato, Anne Vilain. Ces trois années d'échanges et d'enseignement ont été très enrichissantes tant sur le plan

professionnel que personnel. C'est à vos côtés que je me suis découverte une réelle passion pour l'enseignement – contrairement à mes appréhensions du début – et j'espère vivement pouvoir renouveler ces expériences.

Je me tourne maintenant vers « les filles » qui se reconnaîtront sous ce nom. J'aurais tellement de choses à vous dire en ces instants où nos routes prennent toutes des tournants mémorables. Cela fait plus de 7 ans que l'on partage les trois quarts de nos journées ensemble. Il y en a eu des « groomings » du matin, des pauses thés de quatre heures, des piques niques, des anniversaires, des départs, des retours, des coups de blues, des crises, des fous rires nerveux, des mariages, et même des naissances !

Bien plus que des copines de boulot, vous êtes devenues des amies. Des amies particulières car les seules à si bien comprendre cette période particulière de la thèse...J'espère que notre lien et notre complicité – qui n'est plus à consolider – franchira le cap des kilomètres et des mois à venir. Un grand merci à vous les filles pour votre soutien, votre présence, votre écoute, vos coups de pouces, vos cadeaux et petites surprises.

Enfin, je veux vraiment remercier toute la bande de Grenoble : le trio des copines Amandine, Fanny, Johana, devenu quatre avec Isa ; les gars : Jey, Rem, Steph, Duc, Mathieu. Je ne leur dirai à tous qu'un profond merci, pour tout. Les savoyards, Aurèle et Eric... Les amis de Normandie, Jeanne et Philippe, Marie et Jean-Charles, Servanne et Brice, les amis du Québec Sophie et Seb, ma petite sœur Clémence, sans compter la troupe des Chapon qui a su me bichonner au moment où j'en avais le plus besoin... Vous avez été des perles !

Un grand merci également à Nicole...

Merci à ma mère, qui m'a épaulé, écouté patiemment ronchonner, qui a stressé autant que moi sans le montrer et qui m'a fait de bons petits plats pendant le rush final !

Merci donc à tous de m'avoir accompagné et soutenu tout le long du chemin.

Cette thèse a bénéficié d'une bourse MENRT. Le projet de recherche mené au « Laboratoire de Phonétique » de l'UQAM à Montréal a bénéficié du soutien du FQRSC et du CRSH.

L'anticipation semble bien être le secret universel pour la réussite des comportements moteurs. Dans la parole, sous l'aspect de la coarticulation, elle permet de souder les composantes de la syllabe, les syllabes entre elles et les mots. Pendant plus de 40 ans, deux principaux modèles rivaux se sont affrontés sur des langues différentes, essentiellement sur l'anticipation visible qu'est l'arrondissement vocalique des lèvres, sans qu'un vainqueur se dégage. Ce qui aboutit au compromis : un modèle *time-locked* pour l'anglais *vs.* un *look-ahead* pour le français, le suédois, etc. *Le Modèle d'Expansion du Mouvement*, élaboré sur le français au début des années 1990, prédit une extension de l'anticipation régulière (*lawful*) en amont de la cible vocalique dans une transition VC_nV , avec un coefficient d'expansion propre au locuteur. La démonstration principale de cette thèse est que les modèles n'ont pas été testés sur le bon paramètre mais sur le paramètre le plus labile pour l'arrondissement : la protrusion. Lorsqu'ils sont testés sur le contrôle de l'aire aux lèvres, les anglophones rejoignent les francophones démontrant un comportement de type *MEM*, régulier et réglé par locuteur. C'est ce que font aussi les enfants, et plutôt tôt : nous n'avons pas trouvé d'enfant qui n'ait pas ce comportement régulier au-delà de 5 ans et demi. En fait, nous avons trouvé que ce comportement des lèvres pouvait être finement « tuné » dans le style *MEM* dès 3 ans et demi. *Le Modèle d'Expansion du Mouvement* semble bien être ainsi le seul à pouvoir réconcilier le principe d'un comportement moteur universel pour la parole avec une paramétrisation spécifique au locuteur.

Mots-clés : Anticipation d'arrondissement, coarticulation, contrôle moteur de la parole, émergence du contrôle vocalique, développement de la parole

Titre en anglais: Vowel rounding anticipation and the *Movement Expansion Model* (MEM)
Two outreaches: from French to English and from adults to children

Anticipatory motor control is a ubiquitous benefit for behavioural efficiency. In speech its coarticulatory aspect is the glue within syllable components, in-between syllables and words. For more than 40 years two main anticipation models have been rivalling on different languages, mainly on the visible lip rounding anticipation of vowels, without a winner. Which led finally to a compromise such as: a *time-locked* model for English *vs.* a *look-ahead* for French, Swedish, etc. The *Movement Expansion Model* elaborated at the beginning of the 90's for French predicts a lawful anticipation extent ahead of the vowel target, along a VC_nV transition, at a speaker-specific rate. The main claim of this thesis is that the testing of the models has focussed on the wrong (the least robust) parameter, namely lip protrusion for rounding. When tested on lip area control, English speakers join French speakers in a regular speaker-specific *MEM* behaviour. Children also do so and quite early, since we found no one who acquired this regular *MEM* behaviour later than 5;6 years, and we could attest the attunement of this lip behaviour as early as 3;6. Hence the *Movement Expansion Model* seems to be the only one to reconcile a universal speech motor behaviour principle with a speaker-specific parameterization.

Key-words : Rounding anticipation, coarticulation, speech motor control, emergence of vocalic control, child speech development

Discipline – Spécialité doctorale: Sciences du langage

Intitulé et adresse du laboratoire :

Institut de la Communication Parlée, Dept Parole & Cognition de GIPSA-lab
UMR 5216, CNRS, INPG, UJF, Université. Stendhal
46 avenue Felix Viallet
38031 Grenoble cedex 01 - France –

SOMMAIRE

INTRODUCTION.....	1
PARTIE I. LES MOUVEMENTS ARTICULATOIRES A LARGE EXPANSION CONTROLEE EN ANTICIPATION ET LEURS MODELES.....	5
I.1. Quelques conceptions de la coarticulation.....	7
I.2. L'anticipation comme un effet à courte et longue portée.....	9
I.3. La coproduction comme principe de modélisation du contrôle de la parole.....	13
I.3.1. Rappel de quelques données à l'origine du modèle d'Öhman (1966, 1967).....	14
I.3.2. Le modèle <i>Task Dynamic</i>	16
I.3.2.1. La parole co-produit dynamique des unités phonologiques invariantes de la langue	17
I.3.2.2. Principes d'organisation du modèle <i>Task Dynamic</i>	18
I.3.2.3. Un modèle <i>Task Dynamic</i> pour l'anglais... et un autre pour le français ?.....	20
I.4. Etude d'une des trois dimensions des systèmes vocaliques des langues du monde : l'arrondissement et ses « perturbations » consonantiques.....	22
I.4.1. L'arrondissement vocalique.....	22
I.4.2. L'arrondissement consonantique.....	27
I.4.3. Est-il pertinent d'étudier la lèvre inférieure pour comprendre l'arrondissement ?.....	27
I.4.3.1. Quand des mouvements mandibulaires sont pris pour des mouvements de la lèvre inférieure.....	27
I.4.3.2. Existe-il un mouvement de la lèvre supérieure spécifique pour la réalisation des consonnes coronales non-labialisées ?.....	31
I.4.3.3. Pour ajouter à l'instabilité des gestes labiaux : note sur le phénomène du trough (effet d'auge).....	34
I.5. Les modèles concurrents d'anticipation d'arrondissement vocalique.....	36
I.5.1. Le modèle <i>Time-locked</i> (Bell-Berti & Harris, 1981).....	36
I.5.2. Le modèle <i>Look-ahead</i> (Henke, 1966).....	39
I.5.2.1. Présentation du modèle.....	39
I.5.2.2. L'étude de Benguérel & Cowan (1974) sur le français.....	41
I.5.2.3. L'étude de Sussman & Westbury (1981) sur l'anglais.....	44
I.5.2.4. L'étude de Lubker & Gay (1982) sur le suédois.....	47
I.5.3. Le modèle <i>Hybrid</i> (Perkell & Chiang, 1986).....	50
I.5.3.1. Présentation du modèle.....	50
I.5.3.2. Test des trois modèles classiques de l'anticipation d'arrondissement par Perkell & Matthies (1992).....	51
I.6. L'anticipation vue comme un phénomène 'longue distance' : Le Modèle d'Expansion du Mouvement (MEM) (Abry & Lallouache, 1991).....	59
I.6.1. Le MEM de protrusion (Abry & Lallouache, 1991, 1995a).....	59
I.6.1.1. Présentation du paradigme expérimental.....	59
I.6.1.2. Evénements et phases proposés dans le Modèle MEM pour la description des comportements de protrusion de locuteurs Français.....	60
I.6.1.3. Modélisation du comportement de protrusion.....	61
I.6.2. Adaptation du modèle à la description d'un paramètre plus robuste, l'aire intérolabiale : le MEM de constriction.....	66
I.6.2.1. Description des événements et phases proposés dans le Modèle MEM de constriction.....	66
I.6.2.2. Modélisation du comportement de constriction.....	68

PARTIE II. EXTENSION DU MEM DE PROTRUSION ET DE CONSTRUCTION A DES ADULTES FRANCO-QUEBECOIS ET ANGLLO-AMERICAINS 73

II.1. Protocole expérimental 76

- II.1.1. Sujets 76
- II.1.2. Corpus 76
- II.1.3. Dispositif d'enregistrement des données audiovisuelles 79
- II.1.4. Acquisition des données de la protrusion de la lèvre supérieure 82
- II.1.5. Prétraitements des signaux audiovisuels 83
- II.1.6. Détection et extraction des trajectoires de l'aire aux lèvres 84
- II.1.7. Extraction des trajectoires de protrusion de la lèvre supérieure 88
- II.1.8. Analyse des données articulatoires 89
 - II.1.8.1. Fluctuation des mouvements d'arrondissement : comparaison des séquences [ikstski] vs. [ikstsku] pour la protrusion 89
 - II.1.8.2. Événements cinématiques et phases déterminés sur le signal d'aire aux lèvres 91
 - II.1.8.3. Événements cinématiques et phases déterminés sur le signal de protrusion 93
- II.1.9. Aléas techniques rencontrés au cours de l'étude 94

II.2. L'extension du MEM aux adultes franco-québécois et anglo-américains 95

- II.2.1. L'anticipation du mouvement de protrusion 96
 - II.2.1.1. Locuteurs franco-québécois 96
 - II.2.1.2. Un locuteur anglo-américain protrudeur 105
- II.2.2. L'anticipation du mouvement de constriction labiale 108
 - II.2.2.1. Locuteurs franco-québécois 108
 - II.2.2.2. Locuteurs anglo-américains 114
 - II.2.2.3. Y a-t-il une contribution égale, d'un locuteur à l'autre, des phases de *Time Falling* et *Hold* ? 116

II.3. Discussion générale 120

- II.3.1. De l'influence des consonnes dans les comportements observés 121
- II.3.2. Un paramètre universel : la constriction 121
- II.3.3. Le succès d'un modèle 123

PARTIE III. L'ACQUISITION DU CONTROLE VOCALIQUE 129

III.1. Imitation, émulation dans la production et la perception d'un contrôle labio-laryngé chez le primate non-humain 130

- III.1.1. L'imitation comme transmission de la norme sociale chez les chimpanzés 132
- III.1.2. De l'imitation des gestes oro-faciaux orientés vers la communication chez le macaque rhésus 135
- III.1.3. Le défi du contrôle volontaire oro-laryngé dans la production de vocalisations oro-faciales chez Charlie et Patty de Parme, deux macaques à queue de cochon 137
- III.1.4. Vers un ancrage phylogénétique de la parole ? 140
- III.1.5. L'attunement ou l'appariement progressif de soi avec le monde 148
- III.1.6. La parité entre production bimodale et perception bimodale de la parole : organisation temporelle d'une coordination audiovisuelle 152

III.2. Esquisse des premières habiletés imitatives de l'enfant 162

- III.2.1. Les études pionnières de Meltzoff et Moore et le Mapping Intermodal Actif (AIM) 163
- III.2.2. Sur le chemin de l'imitation des gestes de la parole 169
 - III.2.2.1. Premiers pas vers la parole de l'adulte: l'émulation des buts acoustiques 170
 - III.2.2.2. Modélisation articulatoire des productions acoustiques des bébés 172
 - III.2.2.3. Explication des données de Kuhl & Meltzoff (1996) par la Phonologie Articulatoire 176

III.3. De l'émulation des buts acoustiques à l'imitation d'une coordination articulatoire : le développement du contrôle vocalique.....	179
III.3.1. Ce que nous apportent les études acoustiques.....	179
III.3.2. Quelques clés pour l'acquisition du contrôle moteur.....	182
III.3.2.1. Du contrôle de la mandibule: La théorie <i>Frame then Content</i> de Mac Neilage (1998)	183
III.3.2.2. La théorie <i>Frame then Content</i> serait-elle une ontogénie venue de la phylogénie du langage ?.....	183
III.3.3. Première étape vers l'attunement phonologique: l'émergence de la coordination supra mandibulaire	185
III.3.3.1. L'apport des travaux de Munhall & Jones (1998) sur la coarticulation	185
III.3.3.2. La modélisation des degrés de liberté des articulateurs comme outil pour la compréhension de la coarticulation	187
III.4. Développement de la coarticulation chez l'enfant.....	190
III.4.1. Indices acoustiques de l'acquisition du contrôle de la coarticulation	190
III.4.2. A propos du développement d'un contrôle spécifique: celui des lèvres	191
III.4.2.1. L'indépendance graduelle des articulateurs supra-laryngés	192
III.4.2.2. L'anticipation d'arrondissement vocalique	196
PARTIE IV. EXTENSION DU MEM DE CONSTRICTION A DES ENFANTS DE 3 A 8 ANS : L'ATTUNEMENT A LA LANGUE A TRAVERS L'ANTICIPATION VOCALIQUE.....	203
IV.1. Protocole expérimental	207
IV.1.1. Sujets.....	207
IV.1.2. Corpus.....	208
IV.1.3. Dispositif d'enregistrement des données audiovisuelles.....	208
IV.1.4. Prétraitement des données audiovisuelles	212
IV.1.5. Mesures des événements et des phases sur le signal d'aire aux lèvres	214
IV.1.6. Aléas techniques rencontrés au cours de l'étude.....	215
IV.2. Extension du MEM de constriction à l'enfant francophone.....	216
IV.2.1. Résultats des adultes de référence.....	216
IV.2.2. Résultats des enfants : 1 ^{ère} session d'enregistrement	218
IV.2.2.1. Comportement d'anticipation obtenu chez 6 des 7 de nos petits locuteurs.....	218
IV.2.2.2. Un jeune locuteur qui n'a pas (encore) un comportement d'anticipation lawful	221
IV.2.3. Résultats des enfants : 2 ^{ème} session d'enregistrement un an plus tard.....	222
IV.2.3.1. Mise en évidence de profils d'anticipation vocalique stables	222
IV.2.3.2. Notre 7 ^{ème} petit locuteur sait maintenant anticiper comme les autres enfants.....	225
IV.2.3.3. Contributions respectives des phases de <i>Time Falling</i> et <i>Hold</i>	226
IV.3. Discussion générale	228
CONCLUSIONS ET PERSPECTIVES.....	233
BIBLIOGRAPHIE.....	245
ANNEXES	267

Introduction

Qu'est-ce que l'anticipation ? Le secret du geste comme nous l'affirme le grand spécialiste du contrôle du mouvement, Professeur au Collège de France, Alain Berthoz ? C'est sans doute pour la parole, un contrôle qui permet de donner une certaine unité à la syllabe universelle Consonne-Voyelle, et qui doit être géré pour assurer une liaison fluide entre d'autres unités de taille supérieure, à commencer par les mots. Après plus de 40 ans de tests sur des modèles rivaux, on ne sait toujours pas si ces modèles d'anticipation sont les mêmes pour le français et l'anglais.

Cette thèse devrait apporter des réponses sur le contrôle de l'anticipation des lèvres à la sortie du conduit vocal, qui entre dans la production des voyelles, soit l'*arrondissement*, un geste qui fait passer de la voyelle cardinale extrême [i] à l'autre extrême [u] (avec en plus, en français, la possibilité d'un contraste [i] vs. [y]). Première nouvelle : les résultats divergents pour ces langues viendraient simplement de la mesure qui a été généralement choisie jusqu'à tout récemment : l'avancée ou *protrusion* des lèvres. Pour peu qu'on utilise un descripteur adéquat, celui de la *constriction* ou de l'aire aux lèvres, on verra les anglophones comme les francophones se ranger au même modèle. Il est trop tôt dans cette introduction pour révéler les deux secrets de ce modèle qui n'avait jamais été testé jusque-là sur l'anglais. Mais on peut d'ores et déjà annoncer, au-delà de notre étude sur les adultes, que ce modèle va être acquis très tôt par les enfants : entre 3 et 6 ans.

Notre première partie décrira les modèles concurrents d'anticipation dans la parole qui ont été typiquement testés durant toutes ces années sur le contrôle de l'arrondissement. Nous ferons état des difficultés rencontrées dans l'évaluation expérimentale de ces modèles, que ce soient des difficultés de suivi des mouvements, de contrôle des perturbations des consonnes sur les mouvements propres des lèvres pour les voyelles, de robustesse des événements

choisis et des pièges statistiques auxquels conduisent soit le choix des événements articulatoires et acoustiques, soit une formulation d'invariance reposant sur l'acceptation (*sic !*) de l'hypothèse nulle. Ces problèmes méthodologiques nous amèneront à privilégier un modèle déjà testé sur le français, le *Modèle d'Expansion du Mouvement* ou *MEM*.

Dans la seconde partie, nous avons choisi de tester la validité du modèle *MEM* auprès de locuteurs issus de deux communautés linguistiques, celles du français et de l'anglais d'Amérique du Nord. Le français de France ayant déjà été testé (et il le sera encore avec les locuteurs adultes de référence pour notre partie IV sur l'acquisition du modèle par les enfants), nous avons décidé de faire d'une pierre deux coups. Etant invitée au laboratoire de Phonétique de l'UQAM, nous avons choisi de contraster les deux communautés linguistiques francophone et anglophone présentes à Montréal. L'enjeu principal était, comme nous l'avons déjà dit, de tester le *MEM* de *constriction* sur l'anglais réputé rebelle à ce modèle. Nous avons réalisé une première : intégrer le suivi de *flesh points* (points cutanés) par l'Optotrak du *Motor Control Lab* de McGill, à notre technique originale qui assure un traitement d'images fiables, afin d'obtenir les *formes* et non plus seulement les points, à la sortie du conduit vocal : crucialement l'aire intérolabiale, grâce au *Lip-Shape-Tracker* de l'ICP. Les résultats se sont révélés positifs pour le *MEM*, quelle que soit la communauté, à condition d'utiliser non pas la *protrusion* des lèvres, mesure éminemment fragile ou labile, mais la *constriction*. A cette condition, les anglophones se rangent bien sous le même modèle comportemental d'anticipation que les francophones.

Le second volet de cette thèse s'inscrit dans un cadre développemental. Ainsi, nous proposerons, dans notre partie III, une lecture sélective de la littérature sur l'émergence du contrôle vocalique chez l'enfant. Cette histoire sur l'apparition des différents contrôles nécessaires pour l'acquisition des voyelles est loin d'être simple. Elle nécessite, en premier lieu, une révision profonde des affirmations sur les capacités des enfants de 5 mois à produire des contrastes vocaliques pour les voyelles cardinales extrêmes quasi universelles [i, a, u]. Les seules données disponibles, pour ne pas dire uniques, montrent que les enfants acquièrent le contrôle de la voyelle [a] au plus tôt à la fin de la première année. Quant au contrôle acquis en dernier, celui précisément de l'arrondissement pour [u], il peut bien n'apparaître que vers 3 ou 4 ans. Cette histoire est donc restée particulièrement embrouillée, le consensus étant globalement que les voyelles sont acquises dès 3 ans et, en conséquence (et pour différentes raisons), l'effort a surtout porté sur l'acquisition des consonnes. Les efforts concernant plus

précisément l'évaluation de la maîtrise, non pas des cibles statiques des voyelles, mais de leur coarticulation dans le flux de parole, ont démarré en grand dans le milieu des années 70, de manière significative par l'*acoustique* du phénomène. Et elles sont restées, pour les études d'aussi grande envergure, à un stade plutôt impressionniste (avec transcription *auditive* seulement) jusqu'à ces dernières années, les études *articulatoires* étant malheureusement le parent pauvre de ce domaine de recherche. Cette situation étant donnée, il va de soi que la partie qui va suivre se devait d'apporter des réponses expérimentales claires sur le contrôle articulaire des patrons d'anticipation des voyelles chez l'enfant. Ces patrons comprennent à la fois le but, la constriction de l'aire aux lèvres (*goal*), et le contrôle de son anticipation (*path* ou chemin), soit un patron complet *goal + path*. L'acquisition d'un tel patron est testable dans le cadre du renouveau des études sur l'imitation, tout particulièrement pour celles d'entre elles qui testent, chez le chimpanzé ou l'enfant, la différence entre *émulation* (*goal* seulement) et *imitation* véritable (soit *goal + path*). Une étude toute récente sur la vocalisation volontaire chez le macaque nous montrera que la coordination oro-laryngée, entre la mise en forme du conduit vocal et l'arrivée de l'excitation glottique, donne la base de l'organisation temporelle du contrôle vocalique, une organisation temporelle qui sera réglée ou plutôt accordée (*attuned*) par l'enfant sur le modèle *MEM*, naturellement ambiant dans sa langue.

Après avoir testé la plausibilité des principes du modèle *MEM* chez les adultes des deux langues (partie II), nous nous sommes donc naturellement tournés, avec la partie IV, vers l'enfant afin d'évaluer cette mise en place ou *attunement* du comportement d'anticipation vocalique. Il s'agissait pour nous d'un enjeu de taille puisque nous ne disposions pratiquement pas de données articulatoires comparatives (si ce n'est celles d'une seule étude, datant de 1980, chez des enfants suédois de 7 à 10 ans). Nos nouvelles données ont été recueillies à dessein pour tester uniquement le *MEM* de *constriction*, celui que les enfants doivent impérativement maîtriser un jour pour se conformer à leur langue. Une première session nous a permis de couvrir la période entre 3 ans et demi et 8 ans, par l'enregistrement de 7 enfants français, ainsi que de deux adultes de référence. Nous avons réenregistré, un an plus tard, les quatre enfants les plus jeunes. Au vu de la littérature du domaine et considérant également les contraintes techniques et méthodologiques d'une telle entreprise, il nous a semblé en effet crucial de concentrer nos efforts sur la période entre 3 et 5 ans, période durant laquelle s'opèrent de grands changements au niveau du contrôle moteur de la parole. Nous verrons que les enfants, s'ils ne sont pas déjà accordés (*attuned*) sur un comportement du type *MEM* après 3 ans, le deviennent au plus tard après 5 ans, avec simplement, comme le permet

Introduction

ce modèle, pour l'adulte comme pour l'enfant, une paramétrisation personnalisée du coefficient d'expansion en anticipation.

**Partie I. Les mouvements articulatoires
à large expansion contrôlée en anticipation
et leurs modèles**

L'objectif de ce chapitre n'est pas de reprendre les excellentes synthèses qui ont été faites sur la coarticulation, par exemple en réunissant sur plusieurs années plusieurs spécialistes dans l'ouvrage finalement paru sous la direction de Hardcastle & Hewlett (1999 ; ce recueil comprenant une présentation par Farnetani & Recasens, pp. 60-61 et Farnetani, pp. 159-161 du modèle *MEM* que nous allons défendre et illustrer). On pourra aussi consulter le n° spécial antérieur de *Language and Speech*, issu du projet européen ACCOR, notamment Fowler & Saltzman (1993) sur la coarticulation et la coordination. Ainsi nous ne mettrons en compétition que les modèles qui ont eu un retentissement sur une assez longue durée (ce qui n'est pas le cas par exemple du *Window Model* de Keating, 1990). Parmi les synthèses françaises, on recommandera celle de Bonnot (1990).

L'objectif que nous avons annoncé dans notre introduction privilégie l'acquisition d'un contrôle à partir d'une coordination de base. C'est un contrôle qui va résulter d'un processus d'*attunement* : on dira qu'une opération de *tuning* s'est produite par *mapping* neural, ou bien encore que le comportement s'est *tuné* ou syntonisé, par un processus d'ajustement ou d'accord avec la langue ambiante. Puisque cet *attunement* se produit par langue, une de nos questions ouvertes sera : l'anglais se comporte-t-il comme le français ? Et si oui, sur quel paramètre contrôlé ? Et encore est-ce seulement par individu ? Cette hypothèse intégrée par le modèle choisi, le *MEM paramétrable par locuteur*, sera à valider sur nos données anglaises, comme elle l'a déjà été sur des données françaises. Plus important encore pour le processus d'acquisition sera une autre question ouverte majeure pour notre travail : quand cet attunement se produit-il au cours de l'histoire de vie de l'enfant avec sa langue ambiante ?

Avec cet objectif de suivre l'ajustement d'un contrôle, il est prioritaire pour nous de tester nos hypothèses sur un phénomène qui soit :

- d'importance majeure pour les phases ou composantes principales du contrôle du comportement (planification, programmation, préparation, exécution, etc.) ;
- d'un intérêt reconnu pour une question de contrôle central et pas un phénomène simple sous-produit périphérique ;
- d'amplitude manifeste (d'un ordre de grandeur important et robuste).

I.1. Quelques conceptions de la coarticulation

La coarticulation est un paradigme de recherche en parole suscitant un vif intérêt depuis plus de cent ans. Notre propos ne sera pas de retracer ici l'origine de la coarticulation, cette tâche ayant déjà été accomplie de manière très détaillée par Kühnert & Nolan (1999). Le terme de *coarticulation* (*Koartikulation*), formulé il y a un siècle par Menzerath & de Lacerda (1933), reflète la conception que la parole est le produit d'interactions entre les articulateurs, de *co-articulations* qui modifient la réalisation canonique d'un phonème tel qu'il serait articulé de manière isolée, interactions dans le temps ainsi que dans l'espace du conduit vocal.

Avant même que le terme de *coarticulation* ne soit proposé par Menzerath & de Lacerda (1933), un certain nombre de travaux avaient déjà pointé l'importance de ce phénomène dans le système de production de la parole, notamment en ce qui concerne le processus d'anticipation labiale. En effet, dès le début des années 1900, les travaux de Rousselot mentionnent ce processus de coarticulation au sein de séquences V_1CV_2 (ou V_2 correspond à la voyelle arrondie [u]) et fait état d'une anticipation du geste d'arrondissement des lèvres pouvant débiter loin dans les consonnes précédant la cible vocalique.

« Entre les deux voyelles, la plus influente est la seconde, comme cela se montre très bien pour a tu dans il a tourné. Dès le milieu de l'a, la ligne des lèvres s'élève à la sollicitation de l'u. » (Rousselot, 1897-1901, p. 947)

Compte tenu des instruments de cette époque, les hypothèses envisagées semblent toujours d'actualité. Les études expérimentales continuées depuis dans ce domaine cherchent en effet à déterminer quelle est l'étendue spatio-temporelle du geste d'arrondissement sur l'entourage phonétique. Le fait que ces premiers travaux soient souvent consacrés à l'observation de l'anticipation d'arrondissement est sans doute dû à la plus grande facilité d'observation et d'analyse des mouvements labiaux (au regard de l'activité linguale ou encore laryngale). Les années 1960 s'ouvrent sur le développement de recherches expérimentales plus systématiques et font de la coarticulation un champ d'investigation dynamique.

Comment définir la coarticulation ? Elle est classiquement replacée dans l'opposition entre unité phonologique et invariante, et unité phonétique variable. C'est ce que font Edda Farnetani & Daniel Recasens (1999), reprenant Fant (1968), dans leur introduction aux modèles de la coarticulation:

« A crucial problem of speech production theories is the dichotomy between the representational and the physical aspects of speech. The former is described as a

system of abstract, invariant, discrete units (the phonemes), the latter as a complex of variable, continuous, overlapping patterns of articulatory movements resulting in a variable acoustic signal, mixture of continuous and discrete events (Fant, 1968). » (Farnetani & Recasens, 1999, p. 31)

Les différents modèles de la coarticulation, élaborés depuis les années 1960, divergent principalement dans la manière d’appréhender cette coexistence entre les aspects phonologiques de la langue et leurs exécutions physiques.

Commentant l’adoption par Kent (1983, p. 57) d’une position toute pragmatique pour définir l’organisation segmentale de la parole (« *I take segmental organization to mean the description of speech patterns in terms of units that are not further divisible within the limits of a particular method of analysis* »), Bonnot (1990) insiste sur l’alliance entre l’approche instrumentale et l’approche théorique :

« Nous estimons cependant que toute recherche sur la coarticulation et sur les contrôles moteurs est doublement articulée : il faut d’une part compter avec l’information délivrée par la ‘technique’ instrumentale et la conception de l’expérience ; il faut d’autre part se placer dans une perspective théorique. C’est alors – et alors seulement – que l’expression ‘d’unités qui ne peuvent plus être divisées dans le cadre d’une méthode d’analyse particulière’ (point de vue strictement physiologique, approche phonétique, perspective phonétique et phonologique) prend tout son sens. » (Bonnot, 1990, p. 37).

Définition pour définition, nous reprendrons celle de Perkell & Matthies (1992), qui couvre tous les effets de la coarticulation:

« Coarticulation is the superposition of multiples influences on the movement of an articulator. Those influences can come from acoustic-phonetic requirements of context and from physical interactions with other articulators. [s.p.n.] » (Perkell & Matthies, 1992, p. 2911)

Les influences mentionnées par les auteurs sont de natures différentes : elles peuvent provenir des besoins acoustiques et phonétiques du contexte phonétique (c’est-à-dire des segments linguistiques adjacents). Le processus de coarticulation consistera à choisir parmi les différents allophones celui qui répond au mieux aux exigences articulatoires du phonème immédiat et de son entourage. Kühnert & Nolan (1999) en donnent un exemple :

« The English phoneme [k] for instance will be articulated further forward on the palate before the front vowel ([ki:] key) and further back before a back vowel ([k :] kaw); and will have a lip position influenced by the following vowel (in particular, with some rounding before the rounded vowel in [kw :]). » (Kühnert & Nolan, 1999, p. 61)

Cet exemple nous permet ainsi d’entrevoir l’un des deux versants de la coarticulation : l’anticipation. Ce phénomène d’anticipation (appelé encore coarticulation régressive) se

traduit par l'influence des mouvements recrutés pour une unité phonétique cible sur le ou les phonèmes avoisinants. L'autre versant de la coarticulation, appelée *persévération* ou *coarticulation progressive* (ou encore *rétentive* comme nous l'avons déjà vu), correspond au phénomène temporel inverse, soit l'influence d'une unité ou d'un segment phonétique observée sur le ou les segments suivants (c'est le cas de la protrusion du [Σ] jusque dans [i] de [Σi], étudiée en français, Abry *et al.*, 1980, Toda *et al.*, 2003).

Il convient de mentionner que le mécanisme de *coarticulation* n'implique pas systématiquement la présence d'autres unités linguistiques, comme l'ont souligné Perkell & Matthies (1992). Celle-ci peut résulter de la simple interaction physique entre différents articulateurs supra-laryngés recrutés pour la production d'un unique phonème. On retiendra parfois l'activité d'un articulateur principal avec des phénomènes de synergie ou encore de compensation de la part d'autres articulateurs.

Le défaut principal de ces conceptions de la coarticulation, essentiellement comme *propagation* (notamment en phonologie), est qu'elles se rattachent à une philosophie de la *contagion* (sans véritable épidémiologie), où tout peut influencer sur tout, ce qui est un degré zéro de la modélisation pour l'observation. Elles manquent tout particulièrement de la notion de contrôle moteur. Ainsi, quand fut évoquée de manière très générale l'inhibition, exemplairement pour l'un des premiers modèles connexionnistes, celui de la dactylographe de Rumelhart & Norman (1982). Différents modèles (celui de Grossberg, depuis Cohen *et al.*, 1988, Plaut & Kello, 1999, etc.) ont surtout des architectures de contrôle (Jordan, 1990, Laboissière, 1992, y compris d'inspiration neurale Guenther *et al.*, 2006), qui pourront en quelque sorte tout « copier » de la phénoménologie de l'anticipation. Certains incorporent des contraintes biomécaniques, mais à notre connaissance, ils ne s'avèrent pas pouvoir rendre compte, pour l'instant (*cf.* plus bas), du phénomène d'anticipation expansible qui est l'objet du modèle *phénoménologique* que nous avons testé dans ce travail.

I.2. L'anticipation comme un effet à courte et longue portée

Dans le numéro spécial hors-série de *Sciences et Vie* (n°2004, septembre 1998) intitulé *Le cerveau et le mouvement. Comment nos gestes construisent notre pensée*, Alain Berthoz, qui a dirigé ce numéro, qu'il introduit sous le titre « Le mouvement au cœur du cerveau », intitule sa propre contribution à la section Mouvement et Perception, sous le titre *Le secret du geste : l'anticipation* (pp. 68-77).

A cette posture de recherche fait écho celle de la linguiste, Joan Bybee, sur le rôle de l'anticipation dans les processus rendant compte de la phonologie mise en forme par l'usage du langage :

« *The temporal directionality of speech is of special importance to phonetic processes. The fluency of motor processing derives from the ability to anticipate the next set of gestures and modify current gestures to accommodate the transition. The greater the predictability of the following configuration, the greater the accommodation or anticipation can be. As Bush (2001) points out, the ability to react to common sequences of events by learning to anticipate the second based on the occurrence of the first is ubiquitous in living organisms. This behaviour is the basis for a range of phenomena from conditioned responses to inferences of cause and effect.* » (Bybee, 2001, p. 164)

En effet, de manière évidente, l'anticipation n'est pas l'apanage de la parole. Keele *et al.* (1990) la considère comme révélatrice des mécanismes de la programmation motrice. Ces auteurs utilisent, comme exemple, une étude sur les mouvements de dactylographe empruntée à Gentner *et al.* (1980).

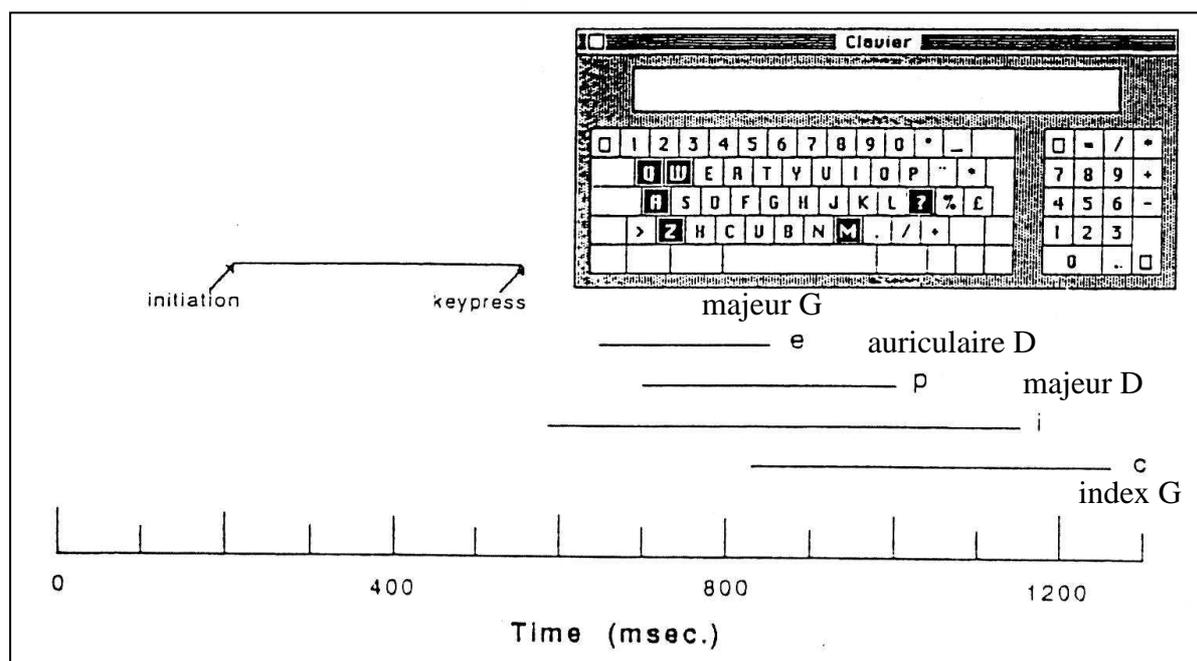


Figure 1 : Dactylographie de la séquence 'an epic'. Le début d'une ligne pleine indique l'instant où un doigt se dirige vers une touche, la fin de cette ligne indique le moment où la touche est pressée. D'après Gentner *et al.*, 1980, in Keele *et al.*, 1980.

Dans cet exemple donné sur la Figure 1 pour la frappe de la séquence « an epic », le mouvement du doigt qui se dirige vers la touche 'i' commence plus de 100 ms avant le mouvement d'un autre doigt qui avance vers 'p'. Notons que ces deux doigts, respectivement le majeur droit pour 'i' et l'auriculaire droit pour 'p', sont portés par la même main.

Suite aux études sur la perception de l'anticipation en parole (Cathiard, 1994), dont nous reparlerons plus loin, les études d'Orliaguet *et al.* (1997 ; Orliaguet, 2004) et Kandel *et al.* (2000a et b) ont permis d'étudier à la fois la production et la perception de l'anticipation dans le geste graphique. Ils ont pu ainsi montrer que, dans une séquence 'ln' opposée à une séquence 'le', l'identification de la seconde lettre était déjà possible dès le *downstroke* (le plein) du 'l'. Dans une étude consécutive d'imagerie cérébrale (PET), Orliaguet, en collaboration avec d'autres chercheurs (Chaminade *et al.*, 2001), ont mis en évidence les circuits neuraux impliqués dans une tâche de pointage et une tâche d'écriture cursive. Les sujets testés, auxquels on présentait seulement les phases de début de mouvement, devaient anticiper le but de la séquence complète. Celle d'écriture était identifiée à 83%. Les activations communes aux deux tâches ont été observées pour le cortex frontal inférieur gauche. Pour l'écriture, l'aire de Broca était en jeu, ainsi que le lobule pariétal supérieur gauche, deux aires qui présentent des lésions lorsque des patients souffrent d'agraphie. Les auteurs concluent que:

« The neural motor system involved in the preparation and execution of action, is also part of a simulation network which is used to interpret the perceived actions performed by others. » (Chaminade et al., 2001, p. 3673)

Partant de cette ouverture générale sur ce phénomène, nous mettrons naturellement en parole l'accent sur les modèles d'anticipation articulatoire. En d'autres termes, seule l'anticipation ou proaction nous intéressera directement. Dans les termes philosophiques de la phénoménologie du temps de Husserl, un classique comme 'rétention' – un terme utilisé de longue date par la phonétique pour la coarticulation qu'on appelle progressive ou rétentive – ne retiendra pas ici notre attention. Beaucoup moins classique en phonétique, mais retenue par les études sur l'intention motrice (pour cet aspect de la cognition motrice, *cf.* Jeannerod, 1994), la protention ou l'aspect protentionnel de l'expérience, ou encore sa structure protentionnelle, a été étudiée jusque dans la schizophrénie (Gallagher, 2004).

Parvenu à ce point, il est nécessaire de rappeler notre objectif, qui est de suivre les différents ajustements ou *attunements* d'un contrôle. Il est, dans cette optique, pour nous prioritaire, comme nous l'avons dit dès l'introduction de cette partie, de tester nos hypothèses sur un phénomène qui soit : (i) d'importance majeure pour les phases ou composantes principales du contrôle du comportement (planification, programmation, préparation, exécution, etc.) ; (ii) d'un intérêt reconnu pour une question de contrôle central et pas un phénomène simple sous-produit périphérique. Nous venons, en retenant l'anticipation, de

satisfaire à ces deux critères. Et nous allons montrer, dans l'examen des travaux suivants, que nous devons aussi satisfaire au troisième critère annoncé : (iii) sélectionner un phénomène d'amplitude manifeste, d'un ordre de grandeur important et robuste.

Ainsi, quelques modélisations réalisées jusqu'à présent en tenant compte des contraintes périphériques (Ostry *et al.*, 1996 ; Perrier *et al.*, 1996) ne se situent clairement pas dans notre rayon d'action (comme le reconnaît en conclusion Perrier, 2004). Ce qui est modélisé par eux relève d'une anticipation par *undershoot* de la mandibule ou de la langue, lorsque ces articulateurs n'ont pas eu le temps d'atteindre leur cible : ce qui donne une anticipation dans le segment adjacent précédent, anticipation d'amplitude plus ou moins faible selon l'amplitude même de l'*undershoot*. Ces phénomènes, non seulement ne relèvent pas de la coarticulation *long range*, sur laquelle nous avons insisté d'entrée, mais ils révèlent d'une vision de la coarticulation comme une influence possible dans toutes les directions – de gauche à droite dans la chaîne parlée ou de haut en bas pour la mandibule –, sans aborder les phénomènes qui relèvent à l'évidence du contrôle, parce que ce sont des phénomènes d'*attunement* ou d'accord, faisant entrer en jeu des adaptations socialement imitées, qui relèvent, eux, de l'apprentissage *cortical* du langage, et pas seulement d'ajustements généraux de contrôles moteurs de bas niveau.

Les phénomènes suivants relèvent, eux, clairement de notre approche. En ce qui concerne le critère de la magnitude du phénomène, il est phonologiquement clairement identifié dans les langues ayant une *harmonie vocalique* courant souvent sur toute la longueur du mot, comme les langues à harmonie vocalique d'Eurasie (exemplairement le turc) et d'Afrique (exemplairement l'akan).

Pour des phénomènes, qui ne sont pas aussi manifestes, mais plus ou moins présents dans l'anglais, le français, etc., de nombreuses études se sont attachées à déterminer la production et la perception de ces comportements articulatoires, y compris pour des objectifs de synthèse et de reconnaissance. Nous citerons les différents mots-clés utilisés par ces études toutes aisément consultables. *Wide-domain effect* en phonologie (Harris, 2006) ; en articulatoire et/ou acoustique *long-distance coarticulation* (West, 2000) ou *long range coarticulation* (Recasens, 1989) ; *long-domain coarticulation* (*long-domain resonance patterns*, *long-domain phonetic dependencies*) en acoustique, perception et/ou synthèse (Local & Kelly, 1986 ; Hawkins *et al.*, 2000, Heid & Hawkins, 2000, Hawkins & Nguyen, 2003, Nguyen *et al.*, 2002, Nguyen *et al.*, 2004, etc.) ; *large temporal span* ou *long span coarticulation effect*

en reconnaissance (Sharma & Hermansky, 1999 ; Dong *et al.*, 2006). Pour ne citer que les plus courants.

Dans cet ensemble de conceptions dénotées par différents termes, on trouve en fait les preuves que des segments relativement loin de l'adjacence immédiate (V_1 dans V_1CV_2 ou C_1 dans C_1VC_2 , etc.) sont influencés par les caractéristiques de segments plus lointains. La phénoménologie que nous visons est classiquement celle de l'empan de l'anticipation dans $V_1C_nV_2$, soit l'*expansibilité* anticipatoire d'un geste articulatoire, et non pas la portée plus ou moins constante d'une influence « long/large... » dans V_1 , C_1 , etc. citée plus haut. Il va de soi que lorsque $n=0$ dans C_n , nous observons des débuts de mouvement vers V_2 dans V_1 , et même toujours dans V_1 lorsque $n=1$. Le modèle *MEM* que nous défendons prévoit que V_2 , lorsque n augmente, et ceci suivant le locuteur et pas la langue, pourra ne plus causer d'influence dans V_1 . Il s'agit donc bien plutôt que d'une longue portée inhérente à un type de trait phonétique (comme [l] ou [r] dans telle langue ; ou d'une harmonie portant sur tout le mot) d'une large potentialité ou *large expansibilité anticipatoire* d'un modèle conçu intrinsèquement comme un modèle d'*expansion* du mouvement.

I.3. La coproduction comme principe de modélisation du contrôle de la parole

Avec cette section, nous entrons dans les modèles articulatoires qui sont davantage interprétables en termes de contrôle du conduit vocal. Nous commencerons par donner un aperçu d'une des conceptions les plus célèbres, celle d'Öhman (1966, 1967), modélisant l'influence des consonnes comme une « perturbation » (au sens de la théorie des perturbations) surimposée sur la transition, qui fait passer le conduit vocal d'une configuration vocalique à une autre (ce modèle aura une longue succession, notamment par Mrayati *et al.*, 1988). Ainsi que nous le verrons plus loin, cette conception phénoménologique simple et originale, où le geste de transition voyelle-voyelle est la base, est toujours en concurrence avec une conception plus « alphabético-phonématique », donnant la priorité au contrôle des segments phonologiques successifs ou « gestes », voyelle-consonne-voyelle (plus ou moins intégrés dans des unités d'ordre supérieur, more, syllabe). Tout à fait révélateur est la résurrection ou *revival* d'un ancêtre de cette dernière conception, abondamment illustrée depuis par les laboratoires Haskins, Martin Joos (1948), grâce à Lindblom *et al.* (2002) pour la confronter dernièrement avec celle d'Öhman. Nous-mêmes bénéficierons d'une reconception du modèle d'Öhman, initialement conçu pour rendre compte de la production chez l'adulte, à la lumière

d'une proposition d'un modèle de développement du contrôle moteur de la parole chez l'enfant (cf. infra, Vilain *et al.*, 1999).

1.3.1. Rappel de quelques données à l'origine du modèle d'Öhman (1966, 1967)

Öhman a réalisé plusieurs études acoustiques portant sur les effets de coarticulation dans différentes langues : le suédois, l'américain et le russe lors de productions de séquences de type V_1CV_2 (où C correspondait aux consonnes occlusives [b], [d] et [g]). Dans une première étude, Öhman (1966) s'intéresse aux productions d'un unique locuteur suédois. Une analyse des transitions formantiques révèle une influence de V_2 sur le segment V_1C précédent. Il remarque ainsi que, lorsque la voyelle finale est un [y] et que la consonne médiane correspond à l'occlusive vélaire [g], il se produit une élévation de la transition du deuxième et troisième formant de la voyelle anté-consonantique [O] ; tandis que lorsque V_2 correspond à la voyelle [A], il observe plutôt une chute de la transition formantique. Ces données acoustiques suggèrent donc que dans une séquence VCV, la voyelle finale exerce une certaine influence sur les segments se situant en amont (phénomène *medium range* d'anticipation), allant jusqu'à modifier la structure formantique de V_1 (du moins au niveau de ses transitions formantiques avec la consonne). Un phénomène similaire est observé lorsque l'on change cette fois la voyelle initiale. Ainsi, dans des séquences de type [ydO] ou [OdO], Öhman (1966) rapporte une influence de V_1 exercée sur V_2 , phénomène que l'on peut décrire comme une propagation des caractéristiques acoustiques sur les segments en aval (persévération). Ces résultats ont été, par la suite, confirmés auprès de locuteurs américains et russes (Öhman, 1967)

Pour le suédois, en examinant l'influence que peut avoir la configuration labiale de V_2 sur les transitions formantiques du segment V_1C dans les séquences [Agy] et [AgA], Öhman remarque que c'est principalement la position de la langue (le corps de la langue), requise pour la production de V_2 , qui exerce une influence sur les segments précédents. Le même patron de résultat est obtenu pour l'américain.

« The cavities may change their shape even during the stop closure. In other words, the tongue is able to make a distorted vowel gesture, while executing the stop consonant. » (Öhman, 1966, p. 166)

Les mouvements de la langue requis pour l'articulation des voyelles et des consonnes sont distincts et autonomes dans le cas de productions isolées mais ces derniers peuvent être initiés simultanément dans le cas d'une suite VCV, permettant des effets coarticulatoires. Par

ailleurs, si ces gestes sont ordonnés séparément au niveau neural, il en résulte tout de même un produit dynamique : la coarticulation.

« The production of the vowels and of the apical and dorsal consonants involves activity in three (probably partly overlapping) sets of muscles. These sets have separate neural representations in the motor networks of the speaker's brain. Articulatory commands may be transmitted over the three neural control channels independently of each other. However, the dynamic response of the tongue to a compound instruction is a complex summation (neural, muscular, and probably mechanical also) of the responses to each of the components of the instruction. »
(Öhman, 1966, p. 166)

Notons que cette théorie sur les canaux de contrôle séparés, neuraux, musculaires, pour les voyelles et les consonnes n'a pas été confirmée par la thèse de Perkell (1974) mais a été reprise par Fowler (1977).

Concernant le russe, Öhman n'observe pas d'influence de V_2 sur les transitions formantiques du segment V_1C précédent (les consonnes sélectionnées correspondaient aux occlusives palatalisées [b], [d], [g] et à leurs homologues vélares). De fait, il semble que les locuteurs des trois langues témoignent d'un comportement de coarticulation différent. Pour l'anglais et le suédois, les transitions formantiques de V_1C sont largement influencées par la voyelle V_2 , témoignant d'une liberté d'anticiper la position de la langue spécifique à la voyelle post-consonantique un peu avant la détente de l'occlusion consonantique (ce qui n'est pas le cas du russe pour lequel l'étendue de l'anticipation semble plutôt restreinte).

Öhman pondère néanmoins cette affirmation en signalant que : (1) la position linguale pour la voyelle ne peut pas être anticipée trop en amont pour des raisons d'intelligibilité de la consonne intervocalique (cas de l'occlusive alvéolaire [d] par exemple) ; (2) il semble peu probable que cette configuration spécifique à la voyelle soit complètement achevée dès les segments précédents, mais plutôt que certaines caractéristiques seulement soient anticipées.

Au vu de ces résultats, Öhman propose un modèle orienté vers la *co-production* des gestes de parole. Selon lui, la production de séquences VCV ne se traduit pas par l'exécution de trois gestes articulatoires successifs, correspondant respectivement aux trois unités phonémiques voyelle-consonne-voyelle, mais plutôt à un geste global de voyelle à voyelle, caractérisé par le passage d'une certaine position de la langue à une autre (en quelque sorte une trajectoire de diphtongue), sur lequel se superpose le geste de la consonne intervocalique. La coproduction des gestes vocaliques et consonantiques dans des séquences VCV est donc largement dominée par les gestes vocaliques qui modèlent la forme globale du conduit vocal, tandis que les gestes

consonantiques sont obtenus par l'ajout d'articulations locales. Typiquement, c'est ce qui se voit clairement avec une occlusive bilabiale dans [ubu], où l'occlusion aux lèvres est réalisée sur une configuration globale du conduit vocal où les lèvres, la langue et le larynx ont la position d'une voyelle [u].

« *Since traces of the final vowel are observable already in the transition from the initial vowel to the consonant, it must be concluded that a motion toward the final vowel starts not much later than, or perhaps even simultaneously with, the onset of the stop consonant gesture. We have clear evidence that the stop-consonant gestures are actually superimposed on a context-dependent vowel substrate that is present during all of the consonantal gesture.* » (Öhman, 1966, p. 165)

Il en résulte un chevauchement temporel de ces gestes, donnant lieu aux phénomènes d'*anticipation* et de *persévération* largement étudiés dans la littérature. La question qui se pose pratiquement est de prédire quelle sera l'étendue de ces chevauchements temporels de gestes dans la production de la parole. Le modèle d'Öhman permet de régler paramétriquement la dominance plus ou moins grande du geste consonantique sur la transition vocalique. On peut considérer que ce réglage est dépendant de la langue.

1.3.2. Le modèle « Task Dynamic »

Le modèle *Task-Dynamic* de Saltzman et collègues (Saltzman 1986, 1991, Saltzman & Munhall, 1989, Saltzman & Kelso, 1987) est l'aboutissement des efforts de modélisation et rend compte des propositions venues des laboratoires Haskins (Carol Fowler puis Louis Goldstein) avec celles de l'approche dynamique héritée via J.A. Scott Kelso de la théorisation de Kugler et Turvey (Kugler *et al.*, 1982). Le modèle *Task-Dynamic* n'est pas – dans sa version originelle – dédié à la description de la coarticulation anticipante, comme c'est le cas d'autres modèles (*Look-ahead, Time-locked, Hybrid*) que nous présenterons plus loin de manière détaillée. Il s'agit donc d'un modèle global de production de la parole incorporant les résultats de plusieurs travaux réalisés au sein du groupe du laboratoire Haskins. Ainsi, certaines fonctions du modèle émanent directement de la *phonologie articulatoire* de Browman & Goldstein (1986, 1992). Celle-ci offre un cadre théorique à l'organisation des unités phonologiques dont les principes sont ensuite implémentés dans le système de production de la parole (synthétiseur articulatoire CASY). Le but de ce modèle est encore et toujours classiquement selon son auteur :

« *To reconcile the linguistic hypothesis that speech involves an underlying sequencing of abstract, context-independent units, with the empirical observation of context-dependant interleaving of articulatory movements.* » (Saltzman, 1995, p. 84)

1.3.2.1. La parole co-produit dynamique des unités phonologiques invariantes de la langue

Cette position de principe a été constamment réaffirmée :

« *This model addresses the manner in which an underlying invariant dynamical system can give rise to lawful, contextually-conditioned surface variability in observed motion patterns of limbs and speech articulators.* » (Byrd & Saltzman, 2003, p. 151)

Le modèle *Task Dynamic* conçoit la parole comme une coordination d'actions spécifiques orientées vers un but (*goal-directed movements*), recrutant certaines structures du conduit vocal, ce que Fowler & Saltzman (1993) appellent plus précisément des *gestes phonétiques dynamiques*.

« *Linguistically, gestures are hypothesized to be the primitives of language-user's phonological system [...]. Physically, they are coordinated movements of the vocal tract that achieve a phonetically significant goal.* » (Fowler & Saltzman, 1993, p. 172)

Se référant à la définition de Saltzman & Munhall (1989), le *geste phonétique* est encore défini comme suit :

« *A member of a family of functionally equivalent articulatory movement patterns that are actively controlled with reference to a given speech-relevant goal (e.g., bilabial closure).* » (Saltzman & Munhall, 1989, p. 334, dans Fowler & Saltzman, 1993, p. 172)

Les coordinations articulatoires sont le produit de contraintes (*coordinative constraints*), auxquelles sont soumis (attelés, *harnessed*, mis sous le harnais ou le joug, *yoked*) les différents articulateurs concernés, créant ainsi entre eux des dépendances. Dans cette perspective, les degrés de liberté des articulateurs sont réduits mais favorisent l'émergence de sous ensembles de gestes phonétiques, requis pour la production du message parlé.

La coarticulation est en soi abordée comme un processus d'influences réciproques entre différentes actions opérées le long de la chaîne parlée. Mais les *gestes phonétiques* sont invariants, indépendants du contexte et conservent leur identité linguistique, quelle que soit la valeur phonologique des unités avoisinantes. A l'inverse du modèle *Look-ahead*, les unités environnantes n'épousent en aucun cas les traits distinctifs associés à une unité cible. En revanche, ce sont les contraintes de coordination qui vont apporter leur caractère dynamique aux gestes phonétiques. Ces contraintes peuvent donner lieu à des situations de compétition entre les différents articulateurs recrutés pour l'exécution de deux gestes successifs. Dans ces cas, on observe alors un chevauchement spatial et/ou temporel de ces gestes (phénomène de *coproduction*).

« *In this view, context-sensitivity in acoustics or articulatory flows arises primarily from the dynamics of gesture interaction during coproduction.* » (Fowler & Saltzman, 1993, p. 173)

Les auteurs insistent sur la nature *transitoire* des contraintes de coordination, qui seraient élaborées en fonction des buts phonétiques successifs (Turvey, 1990) et favoriseraient par exemple la mise en œuvre de stratégies compensatoires en situation de perturbation d'un des articulateurs concernés (Fowler & Saltzman, 1993).

1.3.2.2. Principes d'organisation du modèle Task Dynamic

Leur modèle a été conçu selon deux niveaux fonctionnels distinctifs (Figure 2) : (a) le niveau de la coordination *inter-articulateurs* (ou *niveau de constriction* selon Byrd & Saltzman, 2003) décrivant l'ensemble des articulateurs recrutés pour l'implémentation d'un geste à un moment précis (*e.g.* les lèvres), accompagnés des degrés de constriction réalisées le long du conduit vocal (protrusion des lèvres, point de constriction du dos de la langue etc.) ; (b) le niveau de coordination *inter-gestes* (ou *niveau d'activation*) qui spécifie les degrés d'activation des gestes et les relations temporelles (de phase) entre eux.

Deux gestes peuvent se chevaucher spatialement, c'est-à-dire que leur exécution peut nécessiter un ou plusieurs articulateurs communs. Si ce chevauchement n'implique pas l'ensemble des articulateurs recrutés pour les deux gestes, ces derniers sont coproduits avec peu d'interférences et sans conflit majeur. En revanche, si le même ensemble d'articulateurs est recruté pour chacun des deux gestes, on observera alors une situation d'*interférence mutuelle* caractérisée par des mélanges de gestes à l'intérieur du conduit vocal.

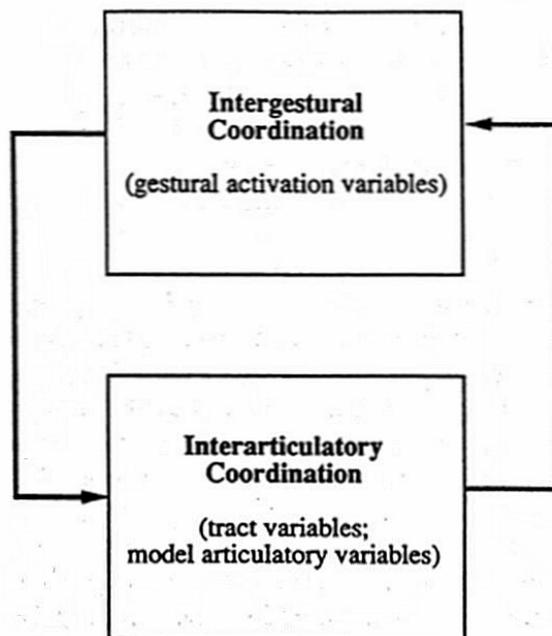


Figure 2 : Représentation schématique des deux niveaux fonctionnels distinctifs élaborés dans le modèle *Task Dynamic*. La flèche de gauche commande l'activation des gestes ; celle de droite matérialise le processus de retour en ligne (*feedback*) des mouvements des articulateurs ou des constrictions réalisées. D'après Saltzman & Munhall, 1989.

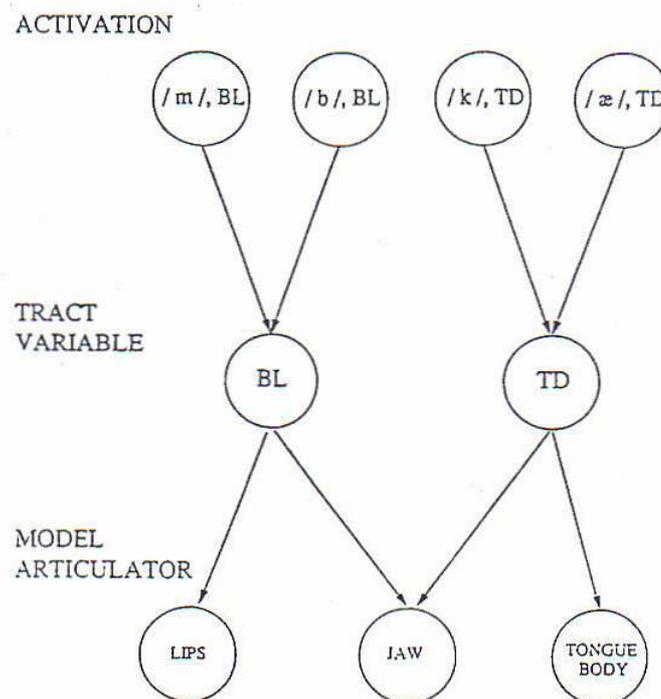


Figure 3 : Organisation hiérarchique du modèle *Task Dynamic*. BL et TD correspondent aux variables du conduit vocal associées aux constrictions bilabiales et linguales (dos de la langue). Au niveau de l'activation, les gestes sont étiquetés en termes d'identité linguistique (e.g. [k]) et d'affiliation en variables du conduit vocal (e.g. TD). D'après Saltzman, 1991.

En ce qui concerne le chevauchement temporel des gestes phonétiques, Saltzman & Munhall (1989), ainsi que Fowler (1990) comparent l'influence d'une unité phonétique sur les

unités avoisinantes à une onde caractérisée par des pentes douces. Durant la phase ascendante, la dominance exercée vis-à-vis des unités adjacentes croît progressivement (phase d'*implémentation graduelle* du geste selon les auteurs). Lors de la phase descendante (*relâchement* du geste), on observe à l'inverse, une influence grandissante des gestes environnants. Dans cette conception de la parole, l'anticipation correspond à la phase d'ascension de l'onde associée à une augmentation progressive de l'influence d'une unité sur l'unité précédente.

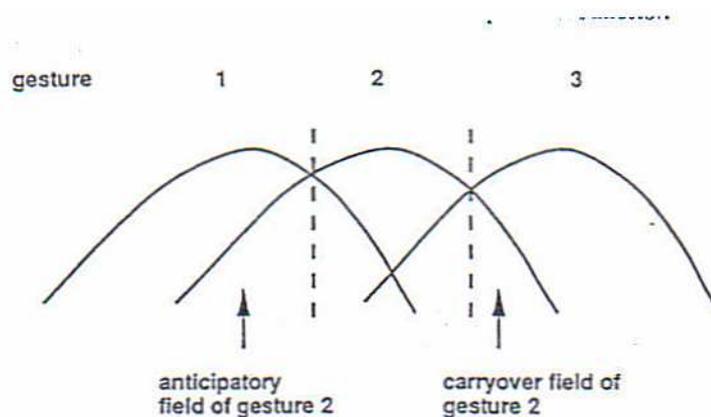


Figure 4 : Représentation schématique du déroulement temporel des gestes de parole tel qu'envisagé par Saltzman & Fowler (1993) dans le modèle *Task Dynamic*. D'après Farnetani & Recasens, 1999, p. 52.

En ce qui concerne la durée de ces phases, Saltzman et Fowler se positionnent en faveur des propositions faites par Bell-Berti & Harris (1974), comme Öhman (1966) et d'autres: chaque geste est activé à l'intérieur d'un cadre temporel limité et son influence ne peut s'étendre loin en amont dans la chaîne parlée.

1.3.2.3. Un modèle Task Dynamic pour l'anglais... et un autre pour le français ?

Ce modèle (qui nous l'avons déjà dit et nous le reverrons) a pour ancêtre la proposition de Joos (1948), a connu, depuis sa popularisation par Löfqvist (1990), un certain succès, notamment dans la synthèse de visages parlants (pour Baldi de Massaro, 1998, et aussi à l'ICP, Le Goff & Benoît, 1997). Il faut dire que la maîtrise de ces fonctions de dominance pour le recrutement des articulateurs qui sont « harnachés » ensemble, afin de produire des mouvements orientés vers un but acoustique, avait été préparée par la technique de décomposition temporelle théorisée par Atal (1983; promue en France par Bimbot *et al.*, 1988).

Ce modèle étant aussi d'inspiration connexionniste, il fait naturellement référence au travail pionnier de Mike Jordan :

« Jordan (1986) showed that sequential networks whose outputs directly represent the distinctive features of speech displayed relatively unconstrained temporal fields of anticipatory behavior. Hence, we tested the hypothesis in our pilot study that 'frame-model-like' behavior requires the addition of side constraints to the sequential net's output units, uniformly applied during both the 'care' and 'don't care' intervals of the network's training phase.» (Rubin et al., 1996, p. 127)

Cependant, en ce qui concerne le but recherché, cette approche peut modéliser des extensions de l'anticipation très différentes. Il ne s'agit plus que d'une question de *programmation impérative* introduisant une contrainte de bord (*side constraint*) *ad hoc* pour éviter que l'anticipation ne se propage au-delà de ce qu'on lui indiquera impérativement de faire. Au lieu d'être réservée aux intervalles sous contraintes spatiales des cibles ou '*care*' *intervals*, cette contrainte de bord (*side constraint*) s'applique aussi dans les intervalles '*don't care*', clairement différente pour l'anglais et le français. La citation suivante n'en fait ainsi qu'une question de degré : ce qui enlève tout pouvoir prédictif au *Frame-model*, pourtant héritier confirmé du modèle *Time-locked* de Haskins (*cf. infra*).

« In a review of the literature, Fowler and Saltzman (1993) concluded that anticipatory coarticulation fields are temporally constrained, and that they do not typically extend very far backward in time from the time of target attainment (*cf. Abry & Lallouache, 1995*). » (*ibid.*)

Plus récemment Saltzman a clairement reconnu que la différence de comportement d'anticipation entre les deux langues n'était pas résolue par une simple commande *ad hoc*, ce qu'il formule par une relation de phase différente pour chaque langue :

« Interestingly, it seems reasonable to hypothesize that the phase angle parameters of the gestural sets of lexical items are language specific. For example, the anticipation intervals of gestures may be defined as the time between gestural onset and target attainment. These intervals have been shown to display different degrees of temporal elasticity across languages (*e.g. Abry & Lallouache, 1995; Saltzman et al., 2000*). By temporal elasticity, we refer to the relationship between a given gesture's activation interval and a preceding interval before target attainment in which there are no conflicting gestures. In English, for example, anticipation intervals are relatively fixed regardless of the length of the preceding no-conflict intervals; in French, however, the anticipation periods stretch in proportion to the length of the no-conflict intervals. » (Byrd & Saltzman, 2003, p. 157)

Sans anticiper sur la partie de notre travail qui montrera que, pourvu qu'on mesure le paramètre pertinent pour le contrôle de l'ouverture du conduit vocal, c'est-à-dire l'aire aux lèvres, les locuteurs anglophones se comporteront comme les francophones, la modélisation ne peut être par principe un simple neuro-mimétisme *ad hoc* de ces comportements.

I.4. Etude d'une des trois dimensions des systèmes vocaliques des langues du monde : l'arrondissement et ses « perturbations » consonantiques

I.4.1. L'arrondissement vocalique

Parmi les trois dimensions articulatoires permettant de caractériser l'ensemble des voyelles des langues du monde, seules deux sont visibles aux lèvres : (i) la dimension de *hauteur* qui concerne l'aperture verticale et oppose par exemple les voyelles [a] à [i] de par leur degré différent d'ouverture du conduit vocal (notons que cette dimension est en outre pertinente pour la distinction des consonnes bilabiales et labiodentales); (ii) la dimension d'*arrondissement* qui permet de différencier la voyelle arrondie [y] de la voyelle non-arrondie ou étirée [i].

Ces deux dimensions sont universellement présentes dans la production des voyelles (*cf.* Abry *et al.*, 1980) et relèvent d'une coordination articulatoire labio-mandibulaire, ou encore parfois labio-linguo-mandibulaire, en ce qui concerne la hauteur. De nombreuses études ont d'ailleurs été dédiées à l'observation de ce type de coordination dans la production de séquences VCV et ses effets vis-à-vis du paramètre de hauteur (notamment les effets de hauteur imposée par la voyelle sur les consonnes adjacentes, Lindblom & Studdert-Kennedy, 1967, et Munhall & Jones, 1998). Nous reviendrons plus en détail sur ce phénomène dans notre partie sur le développement du contrôle de cette coordination articulatoire chez l'enfant.

Dans certaines langues comme le suédois et le français, l'arrondissement constitue une dimension articulatoire cruciale pour la distinction des voyelles. Le français comporte deux séries de voyelles arrondies : les voyelles arrondies antérieures [y, O, ↵] (qui s'opposent aux antérieures étirées [i, e, E] et les voyelles postérieures [u, o, , A] (qui n'ont pas de série étirée correspondantes). A l'intérieur de chaque série, les voyelles diffèrent de par leur aperture buccale.

Les corrélats articulatoires impliqués dans la réalisation de ces voyelles sont désormais bien connus (pour une revue détaillée des gestes labiaux impliqués dans leur réalisation, *cf.* Abry *et al.*, 1980). En ce qui concerne la voyelle postérieure arrondie [u], c'est une voyelle haute qui est réalisée avec une constriction des deux lèvres, assurée principalement par les orbiculaires des lèvres supérieur et inférieur (avec le concours de la houppe du menton et plus ou moins du peaucier du cou). Elle présente un degré maximal de protrusion ou protraction labiale (Abry *et al.*, 1980), et ceci même par rapport à la voyelle haute antérieure arrondie [y], qui est légèrement moins protruse (Benguerel & Cowan, 1974). La voyelle haute [i], qui

contraste de manière minimale avec [y], est produite avec les lèvres étirées, rétractées. Le buccinateur joue principalement ce rôle d'étirement des commissures labiales en arrière, assisté du risorius et du grand zygomatique (Figure 5).

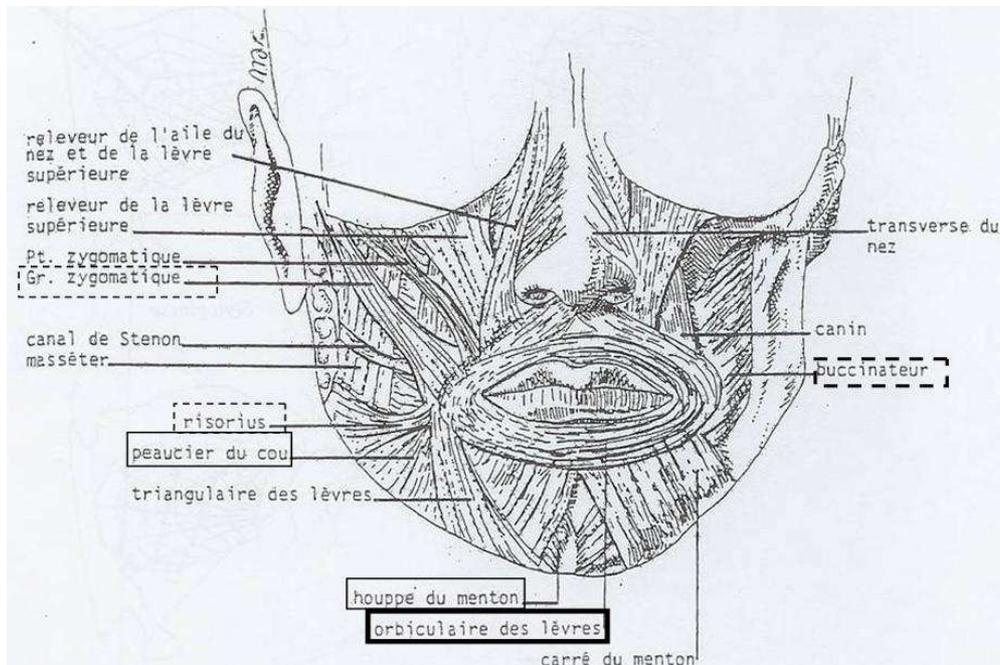


Figure 5 : Les muscles de la face. Ceux qui assurent la réalisation de la voyelle étirée [i] sont encadrés en pointillés tandis que ceux recrutés pour la production des voyelles arrondies [u, y] sont encadrés par des lignes pleines. D'après Bouchet, in Gentil & Boë, 1980, p.38.

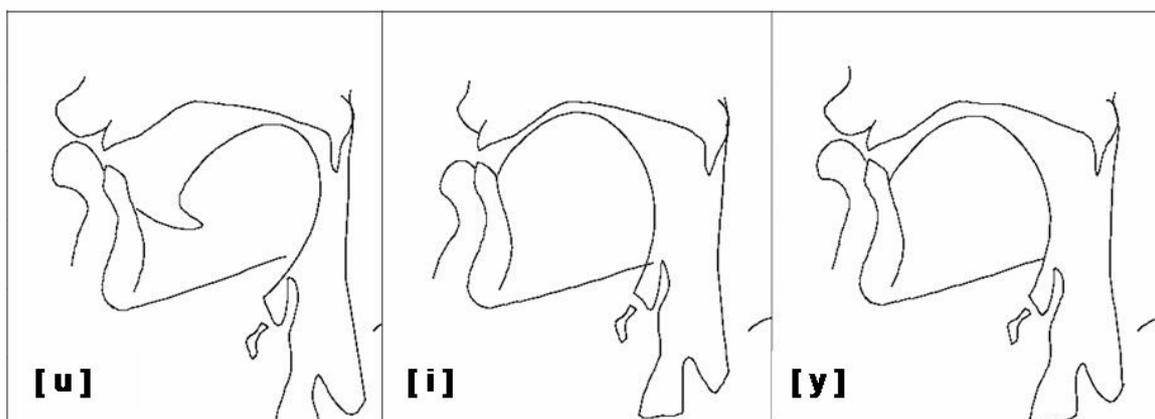


Figure 6 : Coupes sagittales des trois voyelles françaises [u] (à gauche), [i] (au centre) et [y] (à droite). D'après Vilain, 2000.

Certaines langues distinguent trois degrés d'arrondissement. C'est le cas du suédois qui admet la voyelle antérieure non arrondie [i:], dite 'unrounded', la voyelle centrale inrounded' [←:] marquée par une compression verticale des lèvres, comme la voyelle d'arrière [u:], ces dernières ayant toutes deux une protrusion plus faible que l'antérieure arrondie [y:], dite

‘outrounded’, qui est en même temps plus ouverte par l’éversion des lèvres (Ladefoged & Maddieson, 1996).

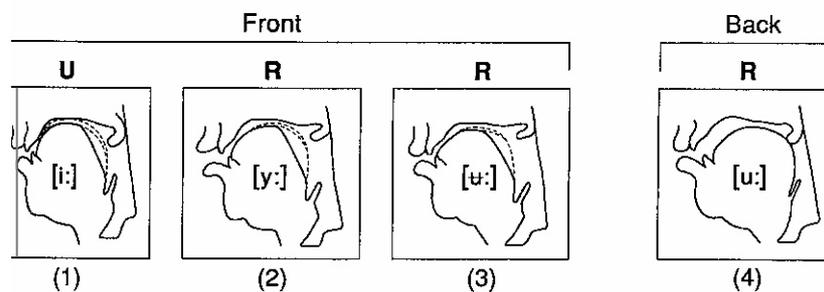


Figure 7 : Coupes sagittales des trois voyelles hautes antérieures (*front*) [i:], [y:], [ɥ:] et de la voyelle haute postérieure (*back*) [u:], appartenant au suédois. Le symbole U caractérise les voyelles non arrondies et R les voyelles arrondies. D’après Fant, 1973, in Farnetani, 1999, p. 145

Il existe, pour la réalisation des voyelles, des corrélations entre l’aperture (verticale) et l’étirement (horizontal) vs. l’arrondissement, ces relations pouvant être différentes selon les langues. Une revue des études comparatives menées dans ce domaine est proposée par Farnetani (1999) qui cite l’article intitulé *Laws for lips* (Abry & Boë, 1986), prolongeant le travail pionnier de Fromkin (1964) et donnant le plus généralement ces relations :

« A parameter used in all languages for distinguishing rounded from unrounded vowels is lip width (i.e. horizontal lip opening), which correlates inversely with all protrusion parameters (Abry & Boë, 1986): rounded vowels have less horizontal lip opening than unrounded vowels. The second dimension is vertical lip opening: rounded vowels tend to have less vertical opening than unrounded vowels. »
(Farnetani, 1999, p. 145)

Dans certaines langues comme le français, le suédois ou encore le cantonnais et le finnois, la dimension d’arrondissement labial est très pertinente du point de vue phonologique (cf. Linker, 1982, qui donne les dimensions comparatives pour ces langues). En revanche, en anglais, l’utilisation de cette dimension semble quelque peu redondante en raison de voyelles arrondies toutes postérieures alors que les voyelles étirées sont toutes antérieures. Farnetani (1999) rapporte plutôt une utilisation de l’étirement horizontal (*horizontal width*) pour la distinction entre voyelles arrondies et non arrondies. Cette tendance est confirmée par les explorations de Perkell & Matthies (1992) ou encore de Fromkin (1964) chez un échantillon de locuteurs américains. Un phénomène intéressant apparemment circonscrit à l’anglais a été remarqué par Fromkin : les voyelles [u] et [ɥ] réalisées avec une protrusion de la lèvre inférieure identique semblent se distinguer par leur étirement ([ɥ] étant nettement plus étirée que [u]) et plus marginalement par leur aperture. Fromkin attribue cet effet à un mouvement

d'arrondissement moindre de l'orbiculaire des lèvres pour [] mais surtout à un avancement de la lèvre inférieure ainsi qu'à un léger abaissement de celle-ci (opéré par le carré du menton). La figure 8 met en relation l'aperture verticale et la protrusion (partie du bas) ainsi que l'étirement labial et la protrusion (partie du haut). On peut apprécier ce comportement d'arrondissement quelque peu atypique. Il semble en effet que, pour ces locuteurs anglais, l'arrondissement – généralement réalisé par le couplage des deux lèvres entraînant une réduction de l'aperture et de l'étirement – puisse être réalisé différemment, c'est-à-dire dans le cas de [] par l'avancement de la lèvre inférieure seule, sans contribution de la lèvre supérieure, expliquant ainsi pourquoi on observe un étirement plus important que pour la voyelle arrondie [u] (Fromkin, 1964, p. 222).

Fromkin conclut donc que, pour l'anglais, la dimension d'étirement apparaît suffisante pour distinguer les voyelles arrondies des voyelles non arrondies:

« It would appear, therefore, that protrusion must be stated as a third parameter in a detailed model accounting for the positions of lips in American English vowels. But as a first approximation we may be content to predict the protrusion from the width of the mouth opening. » (Fromkin, 1964, p. 223).

Par conséquent, il semble que le rôle de l'arrondissement dans la réalisation du contraste vocalique soit fortement lié à la langue.

« [...] two questions arise concerning coarticulation : are there cross-language differences depending on how rounding is implemented ? Are there cross-language differences depending on the phonological function of rounding (distinctive versus redundant)? In other words, how coarticulation is related to the intrinsic articulatory component and to phonology? » (Farnetani, 1999, p. 146)

Notre étude comparative du comportement d'arrondissement chez l'adulte anglophone et francophone tentera d'apporter un éclairage sur ces différences, surtout en ce qui concerne la mesure comparative des mouvements labiaux dans l'organisation temporelle de la coarticulation anticipante.

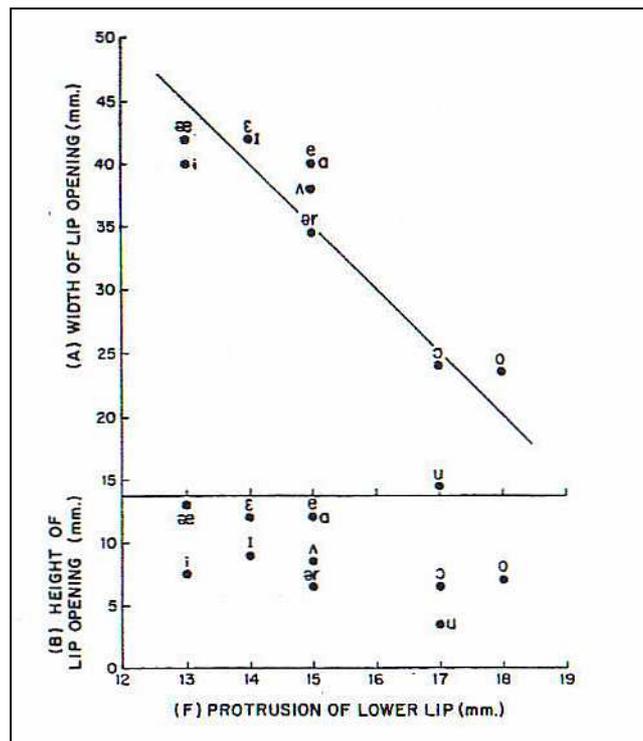


Figure 8 : Relation entre l'étirement horizontal et la protrusion de la lèvre inférieure (partie du dessus) et entre l'aperture verticale et la protrusion (partie du dessous). D'après Fromkin 1964, p. 223.

Nous sommes partis, avant de réussir cette étude, de différences affirmées en ce qui concernait la protrusion, notamment par Lubker & Gay (1982) et les travaux menés à Haskins (que nous rappellerons plus loin). C'est l'idée proposée par Lubker & Gay (1982) d'un comportement moteur pour l'anticipation d'arrondissement acquis et contrôlé selon les exigences du système des oppositions phonologiques ou catégories de la langue :

« We propose that aspects of anticipatory labial coarticulation are language specific and it is therefore a learned behaviour important to some languages, but not to others. » (Lubker & Gay, 1982, p. 438)

Nous soulignerons ici que cette proposition est celle qui a guidé notre démarche expérimentale: tester l'existence d'un tel comportement dans des langues pour lesquelles l'arrondissement ne joue pas *a priori* la même fonction phonologique, et apporter un éclairage sur l'acquisition de ce comportement par l'enfant. Nous verrons que le choix du paramètre mesuré – l'aire aux lèvres et/ou leur protrusion – va se révéler décisif pour généraliser l'existence d'un modèle commun (et son acquisition par l'enfant), tout en préservant la paramétrisation des différences individuelles.

1.4.2. L'arrondissement consonantique

L'arrondissement labial est étroitement lié au contraste vocalique, mais il peut également être observé lors de la production de certaines consonnes. En français, les consonnes constrictives post-alvéolaires [Σ, Z] sont de fait réalisées avec arrondissement labial (Abry *et al.*, 1980 ; Descout *et al.*, 1980). Les autres consonnes pourront être labialisées mais seulement par influence d'un contexte vocalique arrondi. De plus, ainsi que le soulignent Lubker & Gay (1982), le nombre de consonnes labiales (bilabiales ou labiodentales) demeure relativement restreint (que ce soit en français ou dans la plupart des langues du monde), ce qui offre au chercheur un terrain propice à l'observation des possibilités d'extension de l'anticipation d'arrondissement.

« Lip movements reflect the extreme limits of anticipatory coarticulation since many phonemes are unmarked for labiality and thus allow the speaker greater freedom to anticipate labial features. » (Lubker & Gay, 1982, p. 437)

Il existe pourtant un débat quant à l'existence de gestes labiaux spécifiques à la réalisation des consonnes dites non labiales, et tout spécifiquement des fricatives [s, z]. Ce débat a des implications fortes quant à l'interprétation ultérieure des données cinématiques d'arrondissement. Pour certains, ces mouvements auraient une incidence sur le comportement d'anticipation d'arrondissement vocalique et pourraient être à l'origine du manque d'homogénéité des résultats obtenus dans ce domaine ou encore à de mauvaises interprétations des comportements observés (Gelfer *et al.*, 1989 ; Boyce *et al.*, 1990).

1.4.3. Est-il pertinent d'étudier la lèvre inférieure pour comprendre l'arrondissement ?

1.4.3.1. Quand des mouvements mandibulaires sont pris pour des mouvements de la lèvre inférieure

Nous citerons les travaux de Gelfer *et al.* (1989) portant sur l'activité EMG¹ lors de séquences de type [i#tu], [i#ti], [is#tu], [is#ti], [ist#tu], et [ist#ti] chez un locuteur américain. Les auteurs concluent que les pics d'activité musculaire de l'orbiculaire inférieur des lèvres sont dus à la mise en place de l'arrondissement pour la voyelle [u], mais également au mouvement indépendant des consonnes précédentes, puisqu'on les retrouve dans les séquences comportant la voyelle [i] en position finale (Figure 9).

¹ Il s'agit d'enregistrer l'activité électromagnétique des muscles (EMG) en appliquant des diodes infrarouges (LEDS) sur les zones concernées. L'activité musculaire est détectée et stockées sous la forme d'impulsions électriques.

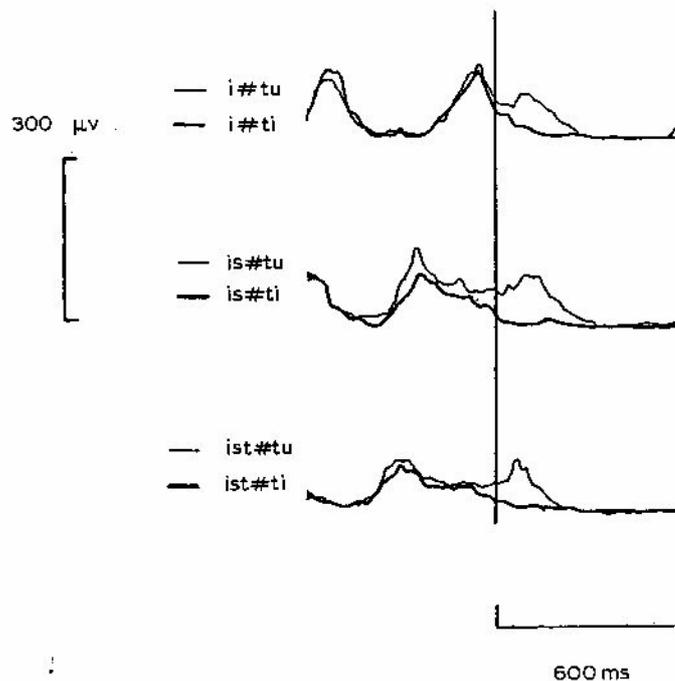


Figure 9 : Tracés moyens de l'activité de l'orbiculaire de la lèvre inférieure d'après 15-20 répétitions de trois paires de séquences. La ligne verticale marque le début acoustique de la voyelle. D'après Gelfer *et al.*, 1989, in Boyce *et al.*, 1990, p. 177.

Ces conclusions sont partagées par Boyce *et al.* (1991). Ces auteurs, comparant les tracés de la lèvre inférieure dans des séquences qui ne comprennent que des [i] (en tireté dans la figure 10) aux tracés qui comprennent une transition [iu], concluent que les petites bosses observées sur ces derniers (signalées par les flèches) correspondent aux bosses de protrusion observées dans les séquences qui n'ont que des [i]. Ces bosses de protrusion reflétant les mouvements propres des lèvres pour les consonnes coronales (ici [t, l]), les profils fléchés dans [iu] seraient la superposition des gestes de protrusion des consonnes et du geste de [u]. Or il s'agit du mouvement de la lèvre inférieure pour lesquels le mouvement d'élévation de la mandibule n'a pas été soustrait. Une grande partie de la protrusion est donc simplement induite par le mouvement d'élévation de la mandibule, qui porte passivement la lèvre inférieure, ladite mandibule coopérant à l'élévation nécessaire de la langue pour la production des consonnes coronales [t, l].

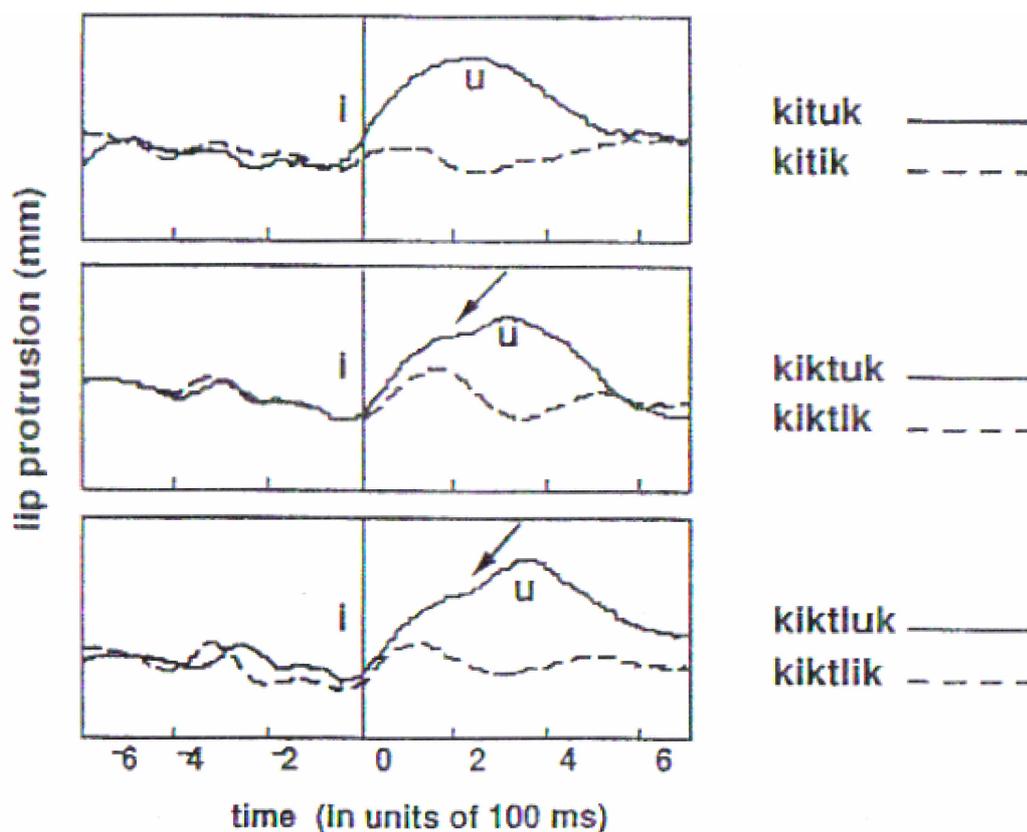


Figure 10 : Tracés de la lèvre inférieure lors de la production des trois paires de séquences [kituk] et [kitik] ; [kiktuk] et [kiktik] ; [kiktluk] et [kiktlik] incorporées à la phrase porteuse « It's a ___ again » par un locuteur américain. D'après Boyce *et al.*, 1990, in Boyce *et al.*, 1991, p. 147.

« Since no rounded vowel is present, the early activity must be attributed to sub-phonemic lip movement for either or both of the intervening consonant(s), and the two peaks of activity seen in, for example, [is#tu], must be assigned separately to the intervocalic consonants and to the rounded vowel. » (Boyce *et al.*, 1990, p. 177)

La confiance attribuée par Boyce *et al.* (1990, 1991), ainsi que par Gelfer *et al.* (1989), à la lèvre inférieure n'est clairement pas justifiée. Insistons sur l'implication du mouvement mandibulaire dans ce qui est considéré ici comme un mouvement contrôlé de protrusion de la lèvre inférieure. Il est désormais bien connu que les déplacements de la lèvre inférieure sont largement tributaires des actions mandibulaires (Fromkin, 1964 ; Munhall & Jones, 1998 ; Tuller & Kelso, 1984 pour n'en citer que quelques-uns). L'étude déjà ancienne de Fromkin (1964)², réalisée auprès de locuteurs anglophones, a révélé une absence de corrélation entre l'aperture verticale et la protrusion de la lèvre inférieure pour l'ensemble des voyelles de l'anglais (Figure 7, p. 223), phénomène que Fromkin attribue aux mouvements d'élévation et d'abaissement de la mâchoire (indépendants des muscles labiaux recrutés pour la protrusion).

² Dans cette étude, Fromkin a réalisé simultanément des mesures par rayons X (photos prises de face), des photos classiques et des moulages en plâtre des lèvres des locuteurs

Tuller *et al.* (1981) ont également noté une grande élévation de la mâchoire pour la production de la fricative [s] par rapport à d'autres consonnes telles que l'occlusive vélaire [k].

Ce questionnement sur un mouvement spécifique des lèvres pour la fricative [s] sera repris dans l'étude de Perkell & Matthies (1992), qui sera décrite plus loin dans le cadre du test des modèles de l'anticipation d'arrondissement vocalique. Les petits mouvements de la lèvre inférieure observés lors des séquences [i#ski] (*cf.* Figure 11) ont été clairement attribués, par les auteurs, aux déplacements de la mandibule pour accompagner ceux de la langue vers la position du [s].

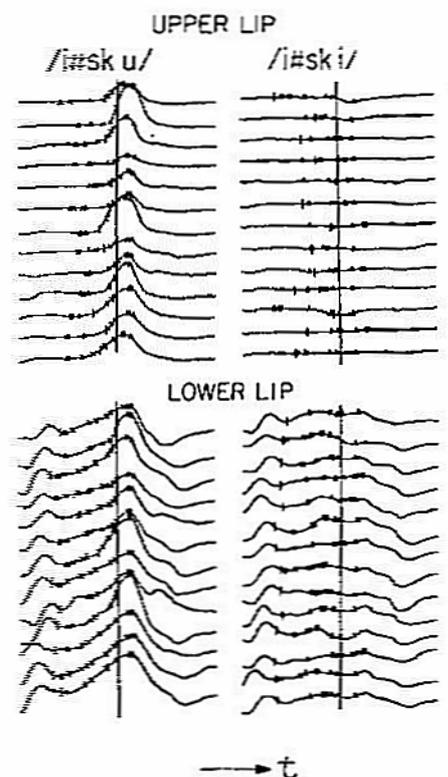


Figure 11 : Tracés des déplacements antéropostérieurs (protrusion) de la lèvre supérieure (en haut) et de la lèvre inférieure (en bas). Les lignes verticales marquent le début acoustique de la voyelle sur lequel les signaux ont été alignés. D'après Perkell & Matthies, 1992, p. 2913.

On retrouve sur leurs séquences [i#sku] (même Figure 11) le résultat d'un mouvement spécifique d'arrondissement pour la voyelle [u], ainsi que des fluctuations similaires à celles qui ont été observées lors des séquences [i#ski], et qui seraient, toujours selon les auteurs, le résultat du mouvement coopératif de la mandibule pour produire les fricatives. Ces fluctuations feront l'objet d'une analyse quantitative chez Perkell & Matthies (1992) que nous rapporterons plus loin. Retenons pour l'instant que ces derniers conclueront qu'il faut

absolument travailler sur la protrusion de la lèvre supérieure, les mouvements de protrusion sur la lèvre inférieure étant trop perturbés par ceux de la mandibule qui aident à produire les consonnes intervocaliques.

1.4.3.2. Existe-il un mouvement de la lèvre supérieure spécifique pour la réalisation des consonnes coronales non-labialisées ?

La plupart des études sur l'arrondissement se réfèrent principalement au déplacement de la lèvre supérieure qui semble témoigner plus nettement du mouvement d'arrondissement vocalique et de son décours temporel (Lubker & Gay, 1982 ; Perkell & Matthies, 1992 ; Farnetani, 1999).

Il existe cependant des études suggérant des mouvements spécifiques de la lèvre supérieure lors de la production des consonnes dites non labiales. Boyce *et al.* (1990) ont relevé des petits mouvements de protrusion de la lèvre supérieure via un système optoélectronique³ chez un locuteur anglais américain dans des séquences de type [kiktluk] (déjà vues) intégrées dans une phrase porteuse : « It's a __ again ». Des phrases contrôles comportant la voyelle [i] en position finale ont été ajoutées afin d'isoler l'activité labiale induite par la commande d'arrondissement vocalique de celle due à l'exécution des consonnes (Figure 12). Les auteurs remarquent la présence d'un mouvement d'arrondissement dans les séquences [kiktlik], qui semble aligné temporellement avec le second pic d'accélération (carrés noirs) dans les séquences [kiktluk]. Par ailleurs, le début du geste de protrusion pour les séquences [kiktluk] (identifié par le premier carré blanc) semble coïncider, dans les séquences [kiktlik], avec une rétraction maximale pour le [i], suivie d'une inflexion attribuée selon les auteurs à l'exécution des consonnes adjacentes.

Nous avons plusieurs remarques sur l'interprétation formulée à partir de ces données cinématiques. Tout d'abord, les paires de signaux ont été choisies pour l'illustration : il n'y a pas de correspondance statistiquement significative.

On peut également remarquer que lorsqu'il existe un mouvement de protrusion de la lèvre supérieure dans [kitlik] – et ce n'est pas le cas pour tous les signaux –, l'amplitude de ce mouvement apparaît très minime par rapport à celui pour atteindre la cible arrondie [u] dans [kitluk].

³ Il s'agit d'un système composé de petites diodes infrarouges (LEDS) collées sur les parties du corps que l'on souhaite tester tel le système Optotrak et permettant de recueillir le déplacement bidimensionnel des diodes.

Nous pourrions d'ailleurs imaginer que cette *inflexion*, induite selon les auteurs par les consonnes, correspondrait plutôt au relâchement du mouvement d'étirement des commissures labiales pour la voyelle [i] vers une position plus neutre (effet de l'élasticité musculaire), ce qui peut très bien se manifester sur les tracés par un petit effet de bosse.

Nous avons volontairement réalisé un agrandissement de cette inflexion sur l'un des signaux pour souligner la très faible amplitude de protrusion de la lèvre supérieure attribuée selon les auteurs aux consonnes coronales [tɫ].

Quoi qu'il en soit, même si les consonnes intervocaliques exercent une faible interférence sur le comportement d'arrondissement mis en œuvre pour la réalisation d'une cible vocalique arrondie, nous n'avons pas la preuve que celles-ci contraignent pour autant le déroulement temporel du mouvement d'arrondissement. Rappelons que les deux lèvres sont couplées dans l'arrondissement et que le mouvement de la mandibule qui monte pour aider à l'articulation des coronales peut pousser la lèvre supérieure.

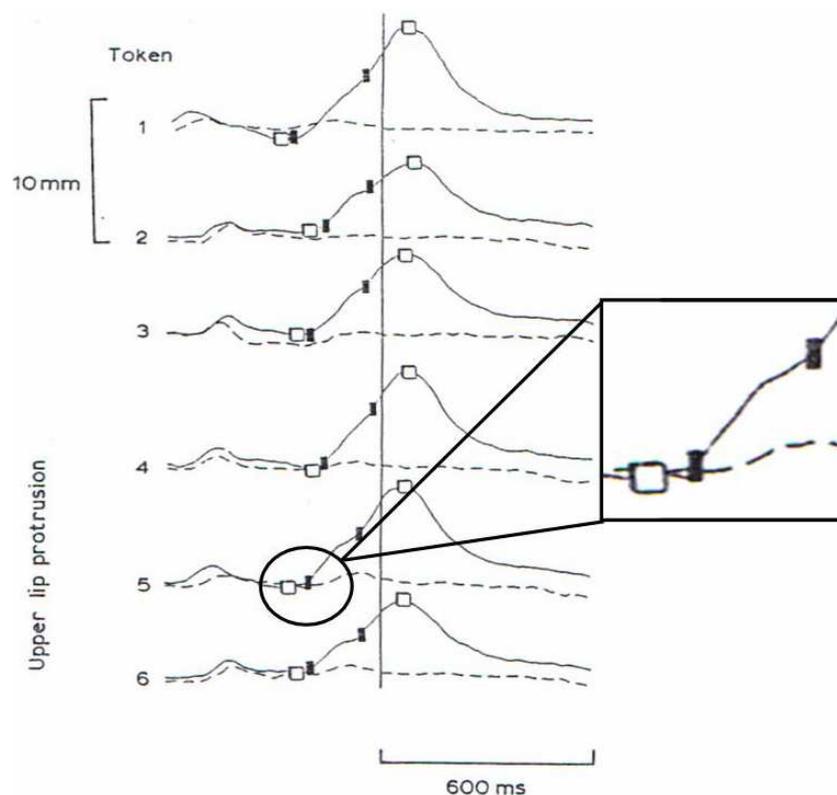


Figure 12 : Tracés de la lèvre supérieure lors de la production de séquences [kiktɫuk] (ligne pleine) et [kiktɫik] (ligne pointillée). La ligne verticale correspond au début acoustique de la voyelle [u], les carrés blancs au début et fin de protrusion (passages par zéro de la vitesse), les carrés noirs aux pics d'accélération. Le zoom fait apparaître les bosses (d'un signal « bumpy ») que Boyce voit se correspondre entre elles : la soi-disant protrusion de la consonne (en tireté) se retrouvant en bosse dans la protrusion de la voyelle (ligne continue). D'après Boyce *et al.*, 1990, p. 183.

Rappelons pour conclure sur les perturbations possibles de la lèvre supérieure que dans une étude explorant l'activité EMG chez quatre locuteurs Suédois, Lubker & Gay (1982) avaient mis en évidence une forte variabilité interlocuteur, se manifestant parfois par le phénomène de double pic d'activité observé par Gelfer *et al.* (1989 ; mais ne se manifestant cependant pas chez tous les locuteurs). Par ailleurs, Lubker et Gay ont démontré que l'activité EMG générée par un muscle n'impliquait pas pour autant le mouvement de la structure qui lui est associée (Figure 13).

« A great deal of published research in speech production, including some of our own, operate on the assumption that EMG activity generated by a given muscle implies movement of its associated structure. This assumption is, of course, open to question. For example, electrical activity generated by a muscle may be agonistic to the movement of any given structure, antagonistic to the movement of that or another structure, or simply stabilizing for either of them. » (Lubker & Gay, 1982, p. 442)

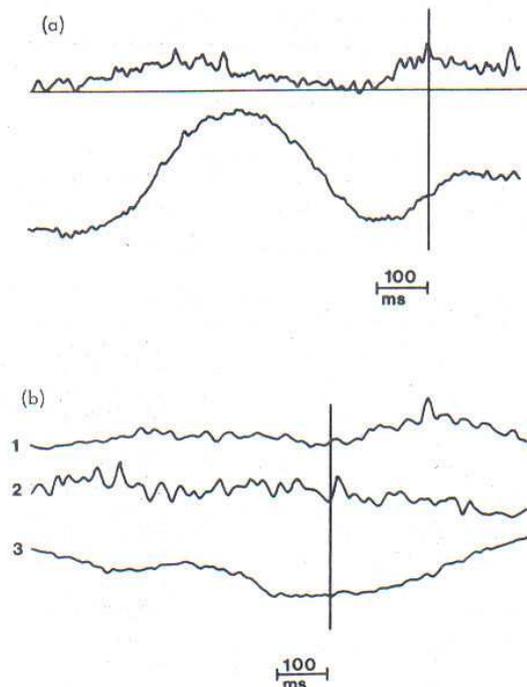


Figure 13 : (a) Illustration d'un phénomène de double pic d'activité de l'orbiculaire des lèvres supérieur (signal du haut) et du mouvement antéropostérieur (signal du bas). Moyenne de 15 occurrences de la séquence [ukstu:] par un locuteur suédois. (b) Exemple d'absence de double pic d'activité pour (1) l'orbiculaire supérieur des lèvres, (2) le modiole; (3) correspond au tracé du mouvement antéropostérieur. Moyenne de 15 occurrences de la séquence « you too » par un locuteur américain. D'après Lubker & Gay, 1982, p. 442.

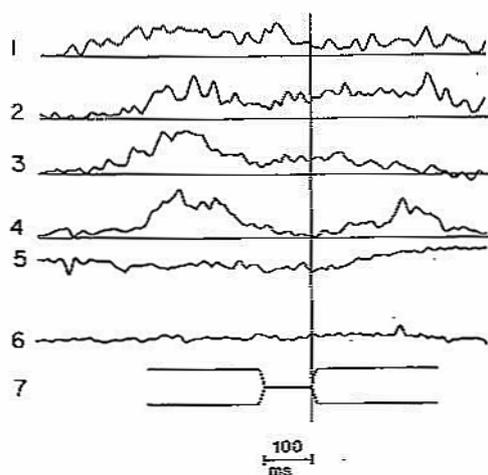


Figure 14 : Exemple de signaux retraçant l'activité EMG de différents muscles labiaux avec les mouvements correspondant. (1) orbiculaire inférieur des lèvres, (2) orbiculaire supérieur des lèvres, (3) modiole, (4) carré du menton, (5) mouvement labial vertical, (6) mouvement labial antéropostérieur ; (7) correspond au signal acoustique (stylisé). Ces signaux constituent une moyenne de 15 productions de la séquence « you too » par un locuteur anglais américain. Les signaux sont alignés sur le début acoustique de la voyelle arrondie. D'après Lubker & Gay, 1982, p. 442.

Ces différents résultats nous invitent donc à rester prudent sur l'interprétation des données cinématiques recueillies à partir de la seule EMG, que ce soit pour la lèvre supérieure, mais également pour la lèvre inférieure dans les études cités précédemment (Gelfer *et al.*, 1989 ; Bell-Berti & Harris, 1974 ; Gay, 1977). Il est possible d'observer une forte activité électromyographique de l'orbiculaire des lèvres (comme c'est le cas dans la figure 14, ainsi que dans Boyce *et al.*, 1990) sans pour autant constater un mouvement distinct, contrôlé, d'arrondissement labial vers la cible vocalique arrondie.

1.4.3.3. Pour ajouter à l'instabilité des gestes labiaux : note sur le phénomène du trough (effet d'auge)

Le phénomène de « trough effect » a été observé dans bon nombre d'études portant sur l'activité linguale ou labiale dans la production dynamique de la parole, plus précisément lors de la production de séquences VCV comportant la même voyelle en position anté-consonantique et post-consonantique. Celui-ci se manifeste, par exemple, dans une séquence [uCu] par un creux de protrusion dans la phase consonantique. Ce creux ou auge (*trough*) proviendrait d'une désactivation momentanée de l'activité musculaire liée à la voyelle lors de la réalisation de la consonne médiane. Dans un geste [iu], ce phénomène correspondrait à ce que Perkell (1986) décrit comme « *an apparent discontinuity in anticipatory coarticulation* » (p. 48).

Cet événement peut paraître très surprenant au regard des assertions largement acceptées dans la littérature sur la dominance de la voyelle en terme de configuration globale du conduit vocal. De fait, il nous semble intéressant de clore cette partie en mentionnant brièvement quelques-unes des études traitant de ce phénomène, car ces dernières peuvent certainement nous apporter des éléments de réflexion supplémentaires par rapport au débat sur l'arrondissement consonantique. Gardons cependant à l'esprit que notre étude ne testera pas spécifiquement ce phénomène puisque nous étudierons des séquences [iC_nu] ou [iC_ny], c'est-à-dire le geste de transition d'une voyelle non-arrondie à une voyelle arrondie, avec classiquement (comme dans les études comparables) [iC_ni] comme contrôle, sans avoir besoin de faire appel à [uC_nu] ou [yC_ny].

Nous citerons en particulier les résultats obtenus en anglais, notamment ceux de Bell-Berti & Harris (1974) et Gay (1977). Ces derniers ont examiné l'évolution de l'activité EMG pour l'orbiculaire inférieur des lèvres lors de la production de séquences [uCu] et ont constaté une brève diminution de cette activité à la fin de la voyelle initiale [u]. Le même phénomène a été rapporté pour le suédois par McAllister (1978) et Engstrand (1980). Gay (1978) avait déjà conclu de ces résultats que la parole est articulée dans le cadre restreint d'unités syllabiques (*syllable-sized articulatory unit*) contrairement au postulat d'Öhman qui, rappelons-le, défend un geste de transition uni de voyelle à voyelle. C'est cette position opposée à Öhman que Lindblom *et al.* (2002) ont tenté plus récemment de faire revivre, par un rappel de la proposition ancienne de Joos (1948) sur la parole contrôlée par des ondes d'activation chevauchantes. C'est une proposition que nous examinerons plus loin.

Deux études menées par ailleurs par Perkell (1986) et McAllister & Engstrand (1992) ont montré que les communautés linguistiques réagissaient différemment face à ce phénomène. Perkell (1986) a en effet démontré l'existence du *trough effect* au niveau de l'activité de la lèvre inférieure chez quatre Américains et un Espagnol, mais pas chez les deux locuteurs français testés, pour lesquels le geste d'arrondissement semble stable durant toute la séquence. Les résultats de McAllister & Engstrand (1992) obtenus pour les séquences [ipi]⁴ corroborent ceux de Perkell (1986) : le phénomène de *trough effect* est présent chez les locuteurs anglais, allemands, irlandais et suédois, mais pas chez les français.

A partir de ces résultats, l'hypothèse d'une coarticulation des gestes spécifique à la langue serait séduisante, mais les études consacrées au *trough effect* font état d'une variabilité interlocuteurs et même intra-locuteur trop importante pour aboutir à de telles généralisations.

⁴ Leurs données cinématiques ont été obtenues à partir de mesures par électropalatographie (EPG).

I.5. Les modèles concurrents d'anticipation d'arrondissement vocalique

De nombreux modèles de la coarticulation ont été proposés de manière à réconcilier les propriétés invariantes, phonologiques, avec leurs réalisations phonétiques, dynamiques, provenant du flux de la parole. La plupart des modèles de la coarticulation anticipante ont été conçus à partir de l'observation des mouvements labiaux notamment d'arrondissement. Parmi les travaux portant sur d'autres articulateurs, ceux sur le velum en particulier ont alimenté directement le débat sur les modèles de la coarticulation labiale (Bell-Berti, 1993 ; Bonnot *et al.*, 1986). Les modèles que nous décrirons dans cette partie ont ainsi émergé essentiellement à partir de données cinématiques. C'est le cas des résultats de Benguerel & Cowan (1974), faisant suite à la thèse de Henke (1966), qui ont été interprétés en faveur du modèle de propagation des traits de ce dernier, le modèle dit *Look-ahead*. Ainsi que des expériences de Perkell & Chiang (1986) qui ont abouti, devant la complexité des patrons de mouvements, à la formulation d'un modèle dit *Hybrid*, à deux phases, pour tenter de concilier avec le *Look-ahead* le modèle *Time-locked* de Bell-Berti & Harris (1981).

En ce qui concerne l'anticipation d'arrondissement, le point qui nous intéresse ici se situe au niveau de l'organisation temporelle des mouvements articulatoires assurant l'arrondissement vocalique. Peut-on prédire le moment précis de son initialisation ? Le moment de son aboutissement ? Quels sont les facteurs modifiant éventuellement son déroulement temporel dans le flux de la parole ?

Nous présenterons dans cette partie les modèles d'anticipation de l'arrondissement considérés comme références dans la littérature. Notre exposé ne sera donc pas exhaustif, ayant seulement pour but de proposer un aperçu des études cinématiques, et parfois acoustiques, qui ont contribué à étayer les théories sur l'organisation temporelle de la parole.

1.5.1. Le modèle Time-locked (Bell-Berti & Harris, 1981)

Dans le débat portant sur l'organisation motrice de la parole, Bell-Berti & Harris (1981) se positionnent en faveur des prédictions formulées par les modèles de *coproduction*. La coarticulation résulte de la coproduction des gestes articulatoires nécessaires à la production des unités linguistiques. Ceci étant, ces derniers se distinguent quelque peu des modèles de coproduction par l'accent mis sur *l'organisation temporelle* de la parole (*timing* résultant des unités gestuelles). Ils considèrent ces unités de parole comme des *gestes articulatoires* dynamiques (et non comme des configurations articulatoires statiques chevauchantes). Cette

considération importante pour les caractéristiques temporelles de la coarticulation se répercute jusque dans le nom donné à leur modèle : *Time-locked Model*. Le déroulement temporel des gestes articulatoires (*timing*) pour la production d'un phonème est en effet considéré comme une propriété fondamentale, intrinsèque à la dynamique de la parole et non comme une caractéristique assujettie aux besoins acoustiques :

« *The model offered here considers time and timing relationships to be intrinsic to speech motor organization and the units of speech to be dynamic gestures rather than static vocal tract configurations.* » (Bell-Berti & Harris, 1981, p. 13)

Ce modèle est élaboré selon trois principes :

- I. considérant tout articulateur, la période articulatoire inhérente à la production d'une unité est plus longue que sa période acoustique (Figure 15) ;
- II. cette période articulatoire demeure spécifique à chaque articulateur. L'activité des différents articulateurs est initialisée de manière asynchrone. Les auteurs ajoutent toutefois que la nature d'une telle organisation reste à préciser : ce décalage résulte-t-il de contraintes biomécaniques du système moteur de la parole ou s'agit-il simplement de caractéristiques de planification motrice des segments de parole ?
- III. le déroulement temporel de la période articulatoire est enfin totalement indépendant de la longueur de la chaîne consonantique précédant le phonème en question.

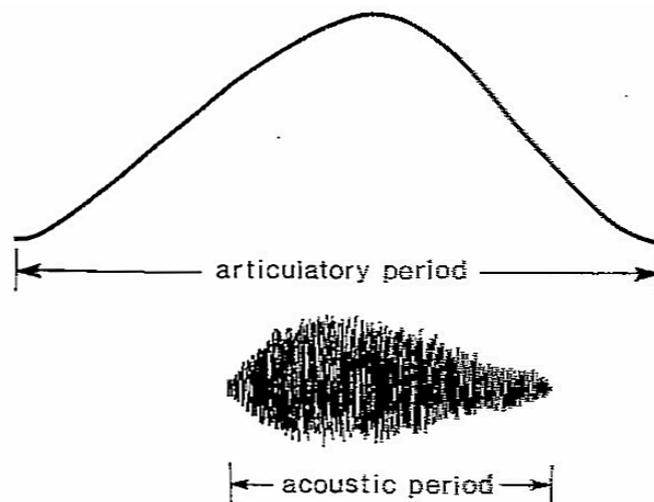


Figure 15 : Représentation schématique de la 'période articulatoire' par rapport à la 'période acoustique' selon Bell-Berti & Harris (1981). Cette dernière est plus courte que la période articulatoire durant laquelle le geste d'arrondissement est exécuté. D'après Bell-Berti & Harris, 1981, p. 13.

Précisons que ce dernier point, concernant l'invariance du geste, est valable pour un style de réalisation de la parole donné, c'est-à-dire sous les mêmes conditions de vitesse d'articulation ou de *focus*.

Dans le cas spécifique de *l'anticipation d'arrondissement vocalique*, le geste débute à date fixe par rapport au début acoustique de la voyelle arrondie suivant la chaîne consonantique (les auteurs précisent que ce principe s'applique lorsque la chaîne consonantique est constituée de consonnes non labiales, qui n'induisent pas de geste d'arrondissement pour leur propre exécution). Ainsi il n'y a pas d'ajustement temporel du mouvement d'arrondissement par rapport à la durée de l'intervalle consonantique. Il s'agit plutôt d'une caractéristique motrice intrinsèque, invariable quelle que soit la durée de la chaîne intervocalique (Figure 16). Ces *gestes* apparaissent donc comme des entités indépendantes et stables vis-à-vis du contexte segmental.

« The nonlabial consonants do not acquire a rounding feature; rather, there is cooccurrence of vowel-associated lip rounding with tongue activity associated with the consonant string. » (Bell-Berti & Harris, 1981, p. 15)

Toutefois, comme nous l'avons déjà signalé, les auteurs admettent que les caractéristiques de débit et de style sont à prendre en considération car ils peuvent avoir des incidences majeures sur le processus d'anticipation d'arrondissement, en particulier sur l'initialisation du geste à date fixe par rapport au début acoustique de la cible vocalique.

Les graphiques présentés à la Figure 16 montrent la relation entre le début de l'activité EMG pour l'arrondissement vocalique et la durée de la chaîne consonantique entre les deux voyelles. On remarque une certaine variabilité des comportements entre les différents locuteurs, une variabilité que l'on peut retrouver par ailleurs dans les productions d'un même locuteur. Ces résultats ne peuvent corroborer l'hypothèse d'un comportement d'anticipation d'arrondissement invariant, puisque l'écart entre les débuts d'activité EMG chez un locuteur varie jusqu'à 100 ms selon la longueur de l'intervalle consonantique.

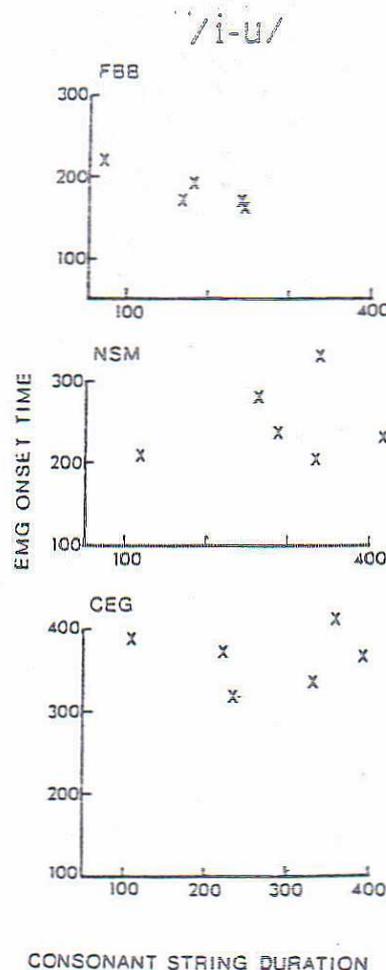


Figure 16 : Relation entre le début de l'activité EMG et la durée de l'intervalle consonantique chez trois locuteurs anglais. D'après Bell-Berti & Harris, 1982, p. 451.

1.5.2. Le modèle *Look-ahead* (Henke, 1966)

1.5.2.1. Présentation du modèle

Le modèle *Look-ahead*, formulé par Henke (1966), se trouve dans la lignée des modèles de propagation des traits (*feature spreading theory*), proposition qui figure notamment chez Chomsky & Halle (1968).

D'un point de vue général, le mécanisme de coarticulation serait régi par des règles phonologiques universelles et non par les contraintes purement physiologiques du système de production de la parole, soient les contraintes des mouvements réalisés dans le conduit vocal. En effet, pour les défenseurs des modèles de traits (Daniloff & Moll, 1968 ; Daniloff & Hammarberg, 1973 ; Hammarberg, 1976 ; Benguerel & Cowan, 1974), il s'agit de rendre compte d'un conflit apparent entre les *intentions* du locuteur, qui porteraient sur le trait catégorisant un segment, et la réalisation temporelle de ce trait, son *exécution* physique.

« *The assumption is that the intention underlying a speaker's pronunciation is invariant. This intention or abstract goal is only changed discretely in time, thus allowing the desired transformation from a discrete phonemic input to a continuous articulatory output.* » (Benguerel & Cowan, 1974, p. 42)

D'après les auteurs, il serait possible de résoudre ce conflit par un processus de propagation de traits phonologiques distinctifs au-delà du phonème clef, si bien que les articulateurs se conformeraient à l'implémentation physique de ces traits. Ces derniers sont déterminés par des règles phonologiques invariantes et universelles, permettant de catégoriser les phonèmes et de les distinguer les uns des autres (Chomsky & Halle, 1968). Chaque trait est exprimé de manière binaire et reçoit les valeurs 'positive' ou 'négative'. Par exemple, dans une langue comme le français où l'arrondissement est une dimension phonologique pertinente, la voyelle [u] sera caractérisée par la valeur [+rond] pour ce trait.

Selon le modèle *Look-ahead*, les traits phonologiques distinctifs attribués à une unité se propageraient donc aux segments précédents, pour lesquels ce trait n'a pas été encore 'spécifié'. Au niveau de leur exécution articulatoire, les mouvements inhérents à la réalisation de cette unité cible (appelés *buts articulatoires*) se propageraient donc à travers la configuration du conduit vocal requise pour la production des consonnes précédentes. Ces effets anticipatoires pourraient en outre s'étendre loin en amont dans la chaîne consonantique, tant que les buts articulatoires du phonème cible n'entrent pas en conflit avec des buts articulatoires immédiats.

Pour une séquence $V_{\text{non arrondie}} C_n V_{\text{arrondie}}$ comportant un nombre important de consonnes non labiales (avec le trait d'arrondissement 'non spécifié'), le mouvement d'arrondissement labial débiterait dès la fin acoustique de la voyelle non arrondie et s'étendrait, avec l'allongement de l'intervalle consonantique, entre les deux voyelles. En d'autres termes, la durée du mouvement d'anticipation serait proportionnelle à la durée de l'intervalle acoustique consonantique. La Figure 17, tirée de Perkell & Matthies (1992), propose une illustration du déroulement temporel du geste d'arrondissement en fonction de l'intervalle consonantique précédant la cible arrondie (d'après le modèle *Look-ahead*).

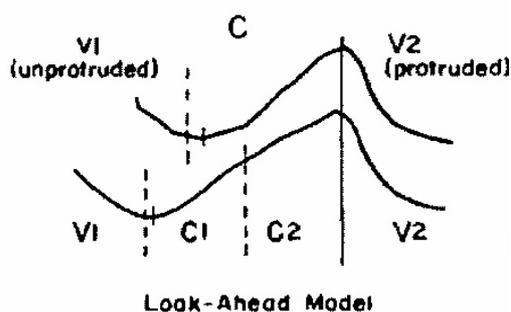


Figure 17 : Représentation schématique du tracé de la lèvre supérieure pour une séquence [iC_nu] selon les prédictions du modèle *Look-ahead*. La ligne pleine marque le début acoustique de la voyelle arrondie [u] et les lignes pointillées correspondent aux frontières acoustiques des autres segments de la séquence : fin acoustique de la voyelle non arrondie [i] et de la ou des consonnes intervocaliques suivantes. Les deux petites marques caractérisent le début du mouvement d'arrondissement. D'après Perkell & Matthies, 1992, p. 2912.

Au niveau expérimental, plusieurs travaux pour soutenir les principes du modèle *Look-ahead* ont été réalisés dans plusieurs communautés linguistiques. Nous en exposerons les principaux résultats.

1.5.2.2. L'étude de Benguérel & Cowan (1974) sur le français

Benguerel & Cowan (1974) ont réalisé une étude sur l'anticipation d'arrondissement en mesurant le mouvement de la lèvre supérieure chez 6 adultes de langue maternelle française. Ce choix a été motivé par les résultats d'une étude cinématique (Roberts, 1972) montrant une certaine instabilité des mouvements de la lèvre inférieure lors de la production de consonnes constrictives telles que [s] et [z] par rapport à ceux de la lèvre supérieure.

L'activité de la lèvre supérieure a été suivie, par une cellule photo-électrique. Il s'agissait de séquences du type V₁CV₂, où V₁ et V₂ pouvaient être les voyelles [i], [y] ou [u], et où C correspondait à un nombre variable de consonnes, de manière à manipuler la durée de l'intervalle acoustique entre les deux cibles vocaliques. Les consonnes produites dans cet intervalle étaient [p, b, m, t, d, n, k, g, s, z, r, l].

Un autre corpus était constitué des chaînes consonantiques [rskr], [rstr] et [kstr] suivies d'une voyelle arrondie ou non arrondie. Une frontière de mots pouvait être placée à différents endroits de l'intervalle consonantique, afin de tester si elle influençait l'étendue temporelle du geste d'arrondissement. Enfin, les sujets devaient répéter deux séquences comportant un intervalle consonantique maximal, correspondant aux 6 consonnes [strstr] placées entre les deux cibles vocaliques : « une **sinistre structure** » et « une **sinistre stricture** ».

La Figure 18 présente la méthode employée par Benguerel & Cowan (1974) pour le calcul de différents événements et intervalles proposés pour la description des comportements d'anticipation d'arrondissement. On notera l'intervalle BE' caractérisant la durée du mouvement de protrusion (*transition time*) de la voyelle étirée vers la voyelle arrondie ; le segment E'E correspondant à l'amplitude de protrusion (*extent of protrusion*), le calcul de la vitesse se faisant enfin par le rapport entre E'E et BE'.

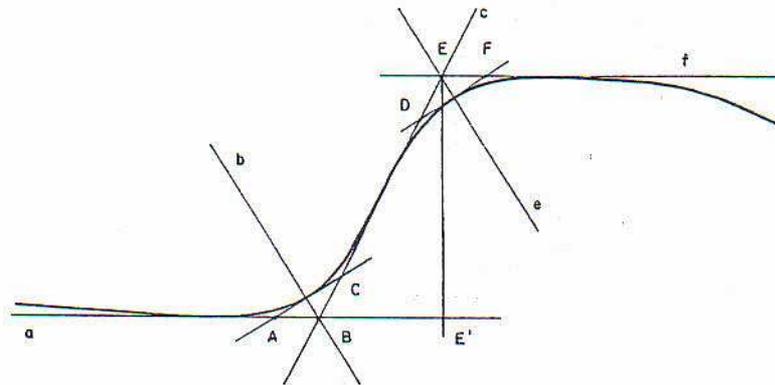


Figure 18 : illustration de la méthode employée pour le calcul du début de protrusion. L'événement A marque le début du geste de protrusion. La phase déterminée par les lettres B et E' délimitent la durée du mouvement (*transition time*) ; celle comprise entre les bornes E' et E caractérise l'étendue ou amplitude de protrusion. La vitesse du geste de protrusion est calculée par le rapport $E'E/BE'$. D'après Benguerel & Cowan, 1974, p. 47.

Pour l'ensemble des locuteurs, Benguerel et Cowan observent une certaine amplitude de protrusion pour les consonnes bilabiales et les constrictives [f, v, Σ , Z], mais qui demeure toujours en deçà du mouvement observé pour les voyelles arrondies [u] et [y]. Les séquences comportant ces consonnes ont donc été écartées de leurs analyses.

En ce qui concerne le mouvement d'arrondissement initié pour les voyelles arrondies, les auteurs ont observé une amplitude de protrusion significativement plus importante pour le [u] que pour le [y] ($p < 0.001$). Par ailleurs, ils notent que l'ajout d'une consonne entre les voyelles [i] et [u] provoque une diminution significative de l'amplitude de protrusion (segment E'E sur la Figure 18) et de la vitesse du mouvement (rapport $E'E/BE'$), mais pas de sa durée (segment BE'). Autre phénomène intéressant : dans les séquences comportant une pause, telles que [iC_n#u] ou [i#C_nu], des tests statistiques (t de student) ont révélé que la pause avait un effet d'allongement de la durée du mouvement de protrusion (*transition time*) dû, selon les auteurs, à une plus grande amplitude de protrusion. Cet effet disparaît pour les séquences comportant simplement une pause, mais pas de consonnes (cas de [i#u]).

Dans la plupart des séquences composées des segments consonantiques [rskr], [rstr] ou [kstr], Benguerel et Cowan ont observé une initialisation très précoce du geste d'arrondissement, c'est-à-dire dès le début de la première consonne de la chaîne intervocalique, en particulier lorsque celle-ci était composée de peu de consonnes (4 selon les auteurs).

Selon les auteurs, les résultats obtenus sont congruents avec les prédictions formulées par Henke (1966) dans son modèle *Look-ahead*. En effet, le trait phonologique [+rond] attribué à la voyelle [u] semble se propager par anticipation à travers les consonnes neutres du point de vue de la labialité. Le mouvement démarre d'autant plus tôt que l'intervalle consonantique est long (graphique 8, Figure 19). Une transposition phonologique de ce phénomène en fonction des règles de propagation des traits est proposée dans la Figure 20.

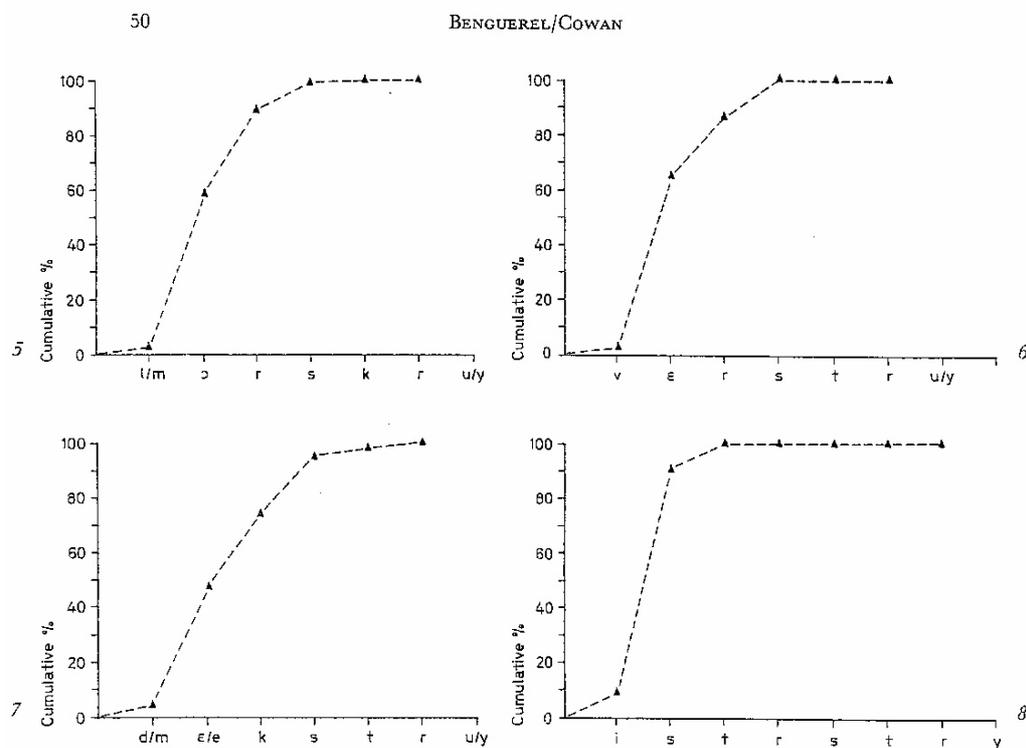


Figure 19 : Distribution des débits de mouvements de protrusion dans les séquences contenant la chaîne consonantique [rskr] (graphique 5), [rstr] (graphique 6), [kstr] (graphique 7), [strstr] (graphique 8). Benguerel & Cowan, 1974, p. 50.

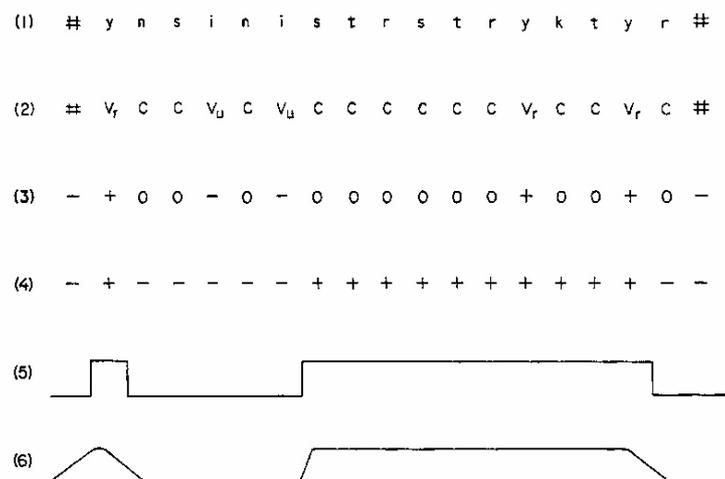


Figure 20 : Dérivation du mouvement de protrusion pour la séquence « Une sinistr(e) structure » en utilisant le modèle de Henke (1966). La ligne (1) correspond à la transcription phonétique avec les frontières de syntagme ; (2) aux catégories consonantiques et vocaliques (Vu = voyelle non-arrondie ; Vr = voyelle arrondie), auxquelles appartiennent les segments ; (3) aux valeurs du trait d'arrondissement au niveau phonémique ([o] pour non-marqué) ; (4) à la spécification du trait au niveau du planning articulatoire ; (5) à ces dernières spécifications sous forme de présence/absence ; et (6) est une approximation grossière d'une sortie du modèle avec une fonction non discontinue (les pentes dépendant des contraintes biomécaniques et linguistiques propres au geste de protrusion. D'après Benguerel & Cowan, 1974, p. 53.

1.5.2.3. L'étude de Sussman & Westbury (1981) sur l'anglais

Sussman & Westbury (1981) ont également réalisé des mesures cinématiques chez trois adultes anglais afin de tester l'étendue temporelle du geste d'arrondissement pour le contraste vocalique [i] vs. [u]. L'activité de l'orbiculaire supérieur des lèvres a été enregistrée par électromyographie (EMG) lors de la production de séquences dépourvues de sens ([tiku], [taku], [tuki], [kikstu], [kakstu]). Pour un des locuteurs, le mouvement de protrusion de la lèvre supérieure (déplacement sur l'axe antéropostérieur) a été mesuré.

Leurs conclusions s'orientent en faveur d'un mécanisme d'anticipation d'arrondissement *Look-ahead*. En effet, le début du mouvement de l'orbiculaire des lèvres ne débute pas à date fixe par rapport au début de la cible vocalique [u] ; celui-ci semble plutôt débiter très tôt lorsque la chaîne consonantique n'est formée que d'une seule consonne, et un peu plus tard lorsque celle-ci est de trois consonnes.

En ce qui concerne le décours temporel de la protrusion pour un des locuteurs testés, les auteurs remarquent que ce mouvement est initié de manière précoce par rapport au début de l'activité EMG associée à la voyelle [u]. Selon les auteurs, ce phénomène déjà mentionné

dans une étude antérieure (Abbs & Eilenberg, 1976) doit être relié à des effets d'élasticité des tissus labiaux qui n'ont aucun lien avec l'activité musculaire induite pour la réalisation du [u]. Au contraire, cette initiation précoce du mouvement de protrusion labiale serait plutôt liée à la fin du geste d'étirement des commissures labiales par le *risorius* pour la voyelle [i].

« Considerable upper lip displacement towards protrusion was accomplished merely by cessation of risorius spreading action for the /i/ and allowance of lip tissue to return to a mechanically neural position. » (Sussman & Westbury, 1981, p. 21)

Même si les résultats obtenus chez ces trois locuteurs infirment le modèle *Time-locked*, Sussman & Westbury (1981) émettent néanmoins des réserves quant à l'acceptation d'une version forte du modèle *Look-ahead* de Henke. En effet, ces derniers ont observé que le mouvement d'arrondissement labial était initié plus tôt dans les séquences comportant la voyelle [i] en position initiale plutôt que la voyelle [a].

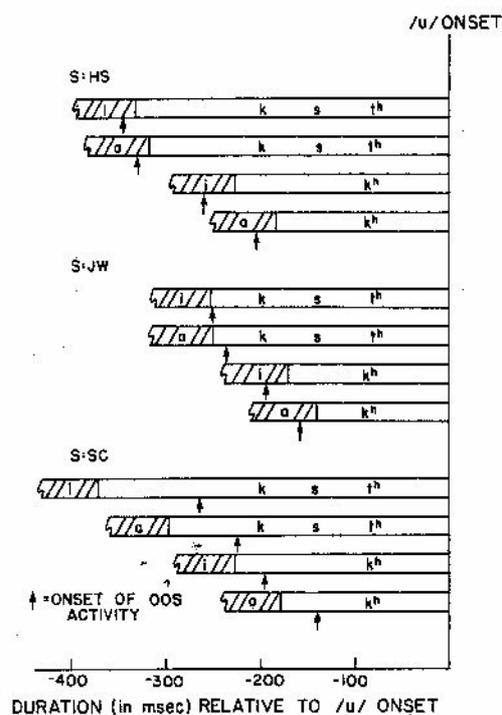


Figure 21 : Représentation schématique du début de l'activité EMG de l'orbiculaire inférieur des lèvres lors du mouvement d'arrondissement attendu pour la voyelle [u] selon que la voyelle anté-consonantique est [a] ou [i]. Les séquences sont représentées selon la nature de l'intervalle consonantique et leur durée par rapport au début acoustique de la cible arrondie (axe des abscisses). D'après Sussman & Westbury, 1981, p. 19.

La figure 21 montre que ce mouvement peut même débiter à l'intérieur de la voyelle [i]. Or, selon les prédictions du modèle *Look-ahead*, nous devrions observer le phénomène inverse en raison de la neutralité de la voyelle [a] par rapport au trait d'arrondissement. Selon le principe du *Look-ahead scanning*, l'instruction d'arrondissement devrait se propager de

manière régressive jusqu'à ce point où deux buts articulatoires entrent en conflit, c'est-à-dire à travers les segments neutres précédant la voyelle arrondie jusque dans la voyelle [a].

Dans les séquences de type [kikstu], pour lesquelles le trait phonologique [-rond] est attribué à la voyelle [i], le locuteur se trouve à l'inverse dans une situation de conflit entre deux buts articulatoires et ne devrait pas avoir la possibilité d'exécuter la commande de propagation du trait d'arrondissement tant que le geste contradictoire n'est pas terminé. La Figure 22 reprend les différentes séquences testées en spécifiant la valeur pour le trait phonologique d'arrondissement pour chacun des phonèmes composant ces séquences. Sussman et Westbury montrent qu'à l'inverse des prédictions du modèle *Look-ahead*, le geste d'arrondissement peut se propager loin en amont jusque dans la voyelle étirée en dépit du trait antagoniste.

$V_1 = /i/$						$V_1 = /a/$					
\bar{k}	\bar{i}	\bar{k}	\bar{s}	\bar{t}	\bar{u}	\bar{k}	\bar{a}	\bar{k}	\bar{s}	\bar{t}	\bar{u}
-	0	0	0	+	-	0	0	0	0	+	-
	t	i	k	u		t	a	k	u		
	-	0	+			0	0	+			

Figure 22 : Spécification physiologique du trait phonologique d'arrondissement dans les séquences [kikstu] et [kakstu], [tiku] et [taku]. D'après Sussman & Westbury, 1981, p. 22.

Les auteurs proposent donc l'explication suivante :

« A perceptually-based explanation for the results might rely on a contrast principle. Since English does not have a high front rounded vowel to contrast with [i], rounding onset can occur more easily without creating any perceptual confusion with any other English vowels. However, if rounding intruded into the [a] vowel space a perceptual confusion could possibly arise due to similarity to [], a low rounded vowel. However, the onset of lip rounding was relatively late in the vowel environment and any acoustic effects of protrusion would be minimal at best. »
(Sussman & Westbury, 1982, p. 22)

Ainsi, selon les auteurs, l'étendue du geste d'anticipation pourrait être associée à des contraintes perceptives. En effet, comme il n'existe en anglais aucune voyelle arrondie équivalente à la voyelle [i] en terme d'aperture et d'antériorité, le locuteur a toute la liberté d'initier son geste d'arrondissement *dès que possible*, y compris durant la voyelle [i] sans créer de confusion perceptive. En revanche, si l'on remplace celle-ci par la voyelle [a], le démarrage précoce du mouvement d'arrondissement aurait pour conséquence de rendre la voyelle initiale ambiguë du point de vue perceptif.

1.5.2.4. L'étude de Lubker & Gay (1982) sur le suédois

Il nous paraît très à propos de présenter les travaux de Lubker & Gay (1982) car ces derniers ont mené une étude comparative de l'anticipation d'arrondissement chez des locuteurs issus de deux communautés linguistiques différentes, le suédois et l'américain. Ce choix retient particulièrement notre attention puisqu'il permet de tester la valeur phonologique accordée à la dimension d'arrondissement dans ces deux langues grâce à des mesures cinématiques.

L'activité labiale a été enregistrée par électromyographie (EMG) et magnétoétrie au cours de la production de séquences du type $V_1C_nV_2$ par 6 adultes suédois et 5 américains. Les deux voyelles étaient séparées par un nombre variable de consonnes non labiales pour chacune des deux langues étudiées.

Les auteurs ont mis en évidence une variabilité de comportement d'anticipation d'arrondissement entre les locuteurs des deux langues testées. En effet, il semblerait d'après le calcul du début du geste de protrusion par rapport à la longueur de la chaîne consonantique (Figure 23) que ce geste soit amorcé plus tôt chez les Suédois que chez les Américains. Néanmoins, une forte variabilité apparaît également à l'intérieur des groupes de langue. Deux des locuteurs suédois (AG et LK) initient leur mouvement de protrusion dans une fenêtre temporelle peu variable, quelle que soit la longueur de l'intervalle consonantique tandis que les trois autres (ER, BK et EH) ajustent leur comportement d'arrondissement en fonction de la longueur de l'intervalle consonantique précédant le début acoustique de la voyelle arrondie [u].

A partir des résultats des locuteurs suédois, Lubker & Gay (1982) concluent que le comportement des deux premiers locuteurs relève d'un patron temporel invariant de type *Time-locked* (caractérisé par une pente nulle, Figure 23) tandis que les trois autres se conforment aux prédictions formulées dans le modèle *Look-ahead* (pente de 1).

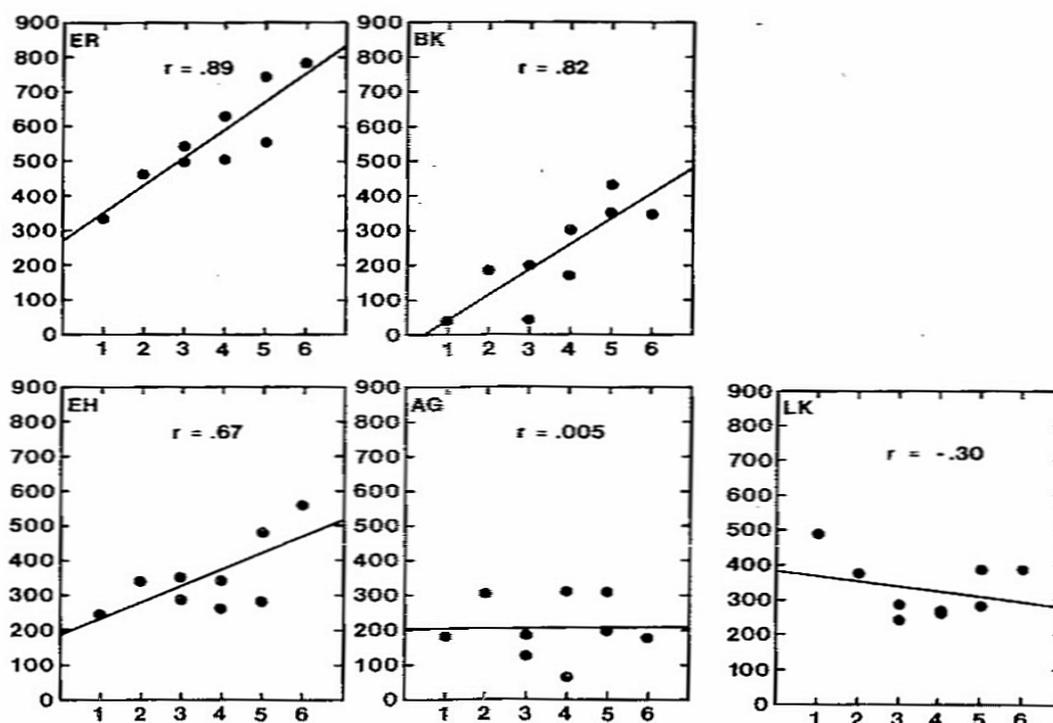


Figure 23 : Relation entre le nombre de consonnes intervocaliques (axe des abscisses) et le début du mouvement d'arrondissement labial (axe des ordonnées) par rapport au début acoustique de la voyelle arrondie chez 5 locuteurs suédois. D'après Lubker & Gay, 1982, p. 445.

Nous souhaitons ici faire une remarque quant à la validité de l'interprétation formulée par les auteurs. En effet, les coefficients de corrélations obtenus à partir de régressions linéaires pour les deux locuteurs associés au profil *Time-locked* sont statistiquement non significatifs (voir les deux graphiques en bas à droite, Figure 23). Il nous semblerait donc plus exact d'avancer que les données recueillies chez ces deux locuteurs sont trop dispersées pour que l'un des deux modèles puisse rendre compte de leur évolution temporelle. En revanche, les coefficients de corrélation significatifs obtenus pour les trois autres locuteurs (respectivement 0,89 ; 0,82 et 0,65) corroborent l'hypothèse d'un comportement *Look-ahead*.

Des résultats similaires ont été observés dans une étude préalable explorant le décours temporel de l'arrondissement chez trois locuteurs suédois (Lubker, 1981). Lors de cette étude, Lubker avait mis en relation le début de l'arrondissement labial (mesuré par l'activité électromyographique) et la durée de l'intervalle consonantique entre les deux cibles vocaliques (Figure 24)

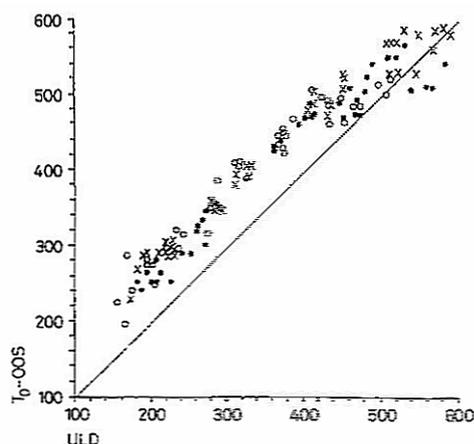


Figure 24 : Relation entre le début de l'activité EMG pour la lèvre supérieure (en ordonnée) et la durée de l'intervalle consonantique (en abscisse) pour l'ensemble des séquences chez les trois locuteurs. D'après Lubker (1981), p. 57.

La droite tracée correspond au cas où $y=x$, c'est-à-dire où l'on aurait une relation proportionnelle entre le début de l'activité EMG pour la lèvre supérieure et la durée de l'intervalle consonantique. Or, on peut voir que la plupart des données sont au-dessus de cette droite, le début du geste étant globalement initié 100 ms avant le début de l'intervalle acoustique entre les deux voyelles.

Lubker (1981) conclut ainsi :

« [...] *at least the strong version of the articulatory syllable model* [le modèle Time-locked] *fails when one notes that, as shown in legend 4* [correspond au graphique ci-dessus] *by the y intercepts of the regression lines for the several subjects, muscle activity for the rounding gesture is initiated approximately 100 ms before the onset of the consonant string. That is to say, muscle activity for rounding begins in the nonround vowel preceding the consonant string well in advance of the boundary of the articulatory syllable of Kozhevnikov and Chistovich [1965].* » (Lubker, 1981, p. 59)

En ce qui concerne les deux sujets américains (Figure 25), les auteurs affirment que le mouvement d'arrondissement est initié plus tard que chez les sujets suédois. Mais les données ne sont pas suffisantes pour apporter une lecture décisive : en dépit de coefficients de corrélations élevés, ces derniers ne sont pas pour autant statistiquement significatifs ($p < 0,2$ selon la table de Bravais-Pearson).

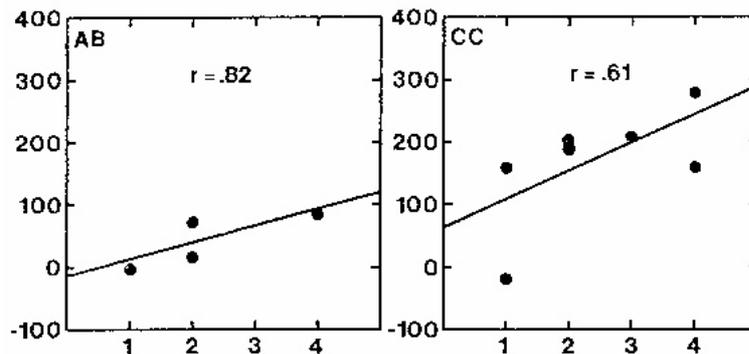


Figure 25 : Relation entre le nombre de consonnes intervocaliques (axe des abscisses) et le début du mouvement d'arrondissement labial (axe des ordonnées) par rapport au début acoustique de la voyelle arrondie chez les 2 locuteurs américains. D'après Lubker & Gay, 1982, p. 445.

1.5.3. Le modèle Hybrid (Perkell & Chiang, 1986)

La première conclusion qui peut être formulée à partir des études que nous venons de citer est que la variété des modèles d'anticipation semble incontestablement liée aux résultats empiriques contrastés. En effet, comme nous le rappellent Lubker & Gay (1982), les données cinématiques et électromyographiques recueillies dépendent fortement de facteurs à la fois biologiques, expérimentaux et linguistiques. De fait, on s'aperçoit très vite que les modèles classiques envisagés pour décrire l'anticipation d'arrondissement tels que le modèle *Look-ahead* et le modèle *Time-locked* ne peuvent pas, par exemple, rendre compte des comportements d'anticipation de locuteurs appartenant à des communautés linguistiques différentes. Ainsi Boyce (1988) distingue l'américain (selon un comportement du type *Time-locked*) du turc (patron de type *Look-ahead*). Lubker & Gay (1982) notent que la protrusion est plus contrastive chez les Suédois que chez les Américains, et que, par conséquent, les comportements d'anticipation, très différents pour ces deux langues, ne peuvent être modélisés de la même manière.

1.5.3.1. Présentation du modèle

Devant les contradictions soulevées par ces études expérimentales, un troisième modèle – le modèle *hybrid* – a été développé par Perkell & Chiang (1986) avec pour objectif de s'adapter à la variabilité des comportements d'anticipation rencontrés. Ce modèle envisage le mouvement d'arrondissement selon deux phases : une première phase caractérisée par un démarrage assez lent du geste d'arrondissement qui débiterait dès la fin acoustique de la

voyelle non arrondie. Cette première phase se rapproche d'un comportement de type *Look-ahead* caractérisé par une pente de 1 (cf. la Figure 27 récapitulant les caractéristiques des trois modèles). Celle-ci serait suivie d'une phase plus rapide marquée par un pic d'accélération (Figure 26). Notons que ce point d'inflexion du mouvement débiterait à date fixe par rapport au début acoustique de la voyelle arrondie (donc de type *Time-locked*). Dans cette perspective, le modèle *Hybrid* apparaît comme un compromis entre les deux modélisations précédentes *Time-Locked* et *Look-Ahead*, puisque la durée de la première phase est ajustée en fonction de la longueur de l'intervalle consonantique, alors que la durée de la seconde phase reste constante.

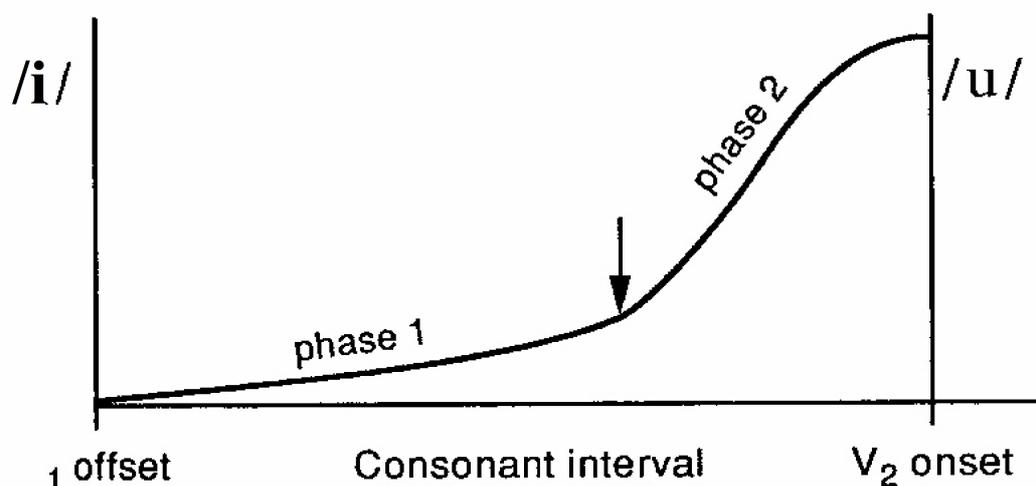


Figure 26 : Illustration des prédictions du modèle Hybrid caractérisé par une première phase lente (phase 1) et une phase rapide (phase 2) initiée par un pic d'accélération (identifié par la flèche). D'après Farnetani, 1999, p. 157.

1.5.3.2. Test des trois modèles classiques de l'anticipation d'arrondissement par Perkell & Matthies (1992)

L'étude de Perkell & Matthies (1992) fait référence dans la littérature sur l'anticipation d'arrondissement. Ces auteurs ont en effet testé les prédictions des trois modèles classiques (*Time-locked*, *Look-ahead* et *Hybrid*) quant à l'organisation temporelle du geste d'arrondissement (Figure 28).

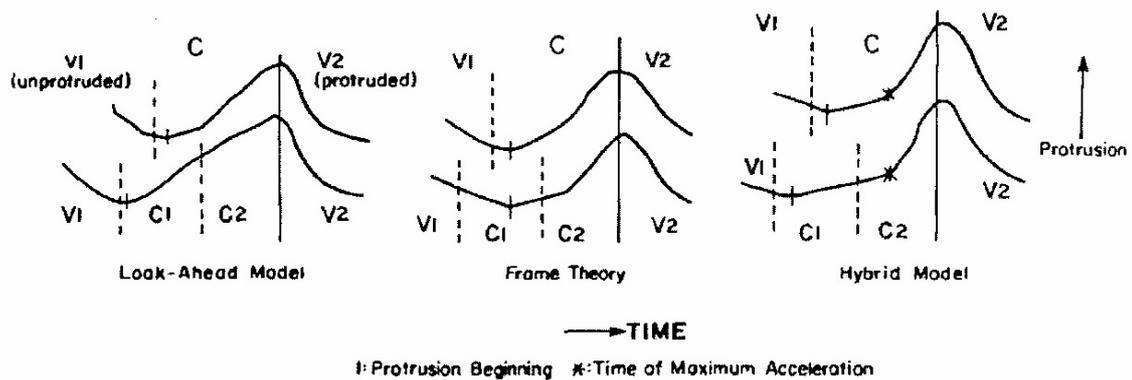


Figure 27 : Illustration du tracé de la lèvre supérieure pour une séquence [iC_nu] selon les prédictions du modèle *Look-ahead*, *Time-locked*, et *Hybrid*. Les lignes pleines définissent le début acoustique de la voyelle arrondie [u] et les lignes pointillées correspondent aux frontières acoustiques des autres segments de la séquence : fin acoustique de la voyelle non arrondie [i] et de la ou des consonnes intervocaliques suivantes. Les petites marques noires caractérisent le début du mouvement d'arrondissement d'après les prédictions de chacun des trois modèles. D'après Perkell & Matthies, 1992, p. 2912.

Pour cette étude, les mouvements verticaux et horizontaux (axe antéropostérieur) des lèvres supérieures et inférieures ont été enregistrés chez 4 locuteurs américains avec un système optoélectrique. Le corpus comprenait, en ordre aléatoire, des énoncés du type [iC_nu] (avec [iC_ni] pour contrôle), où C_n correspondait à un groupe consonantique variable composé des quatre consonnes suivantes : [t], [s], [k], [h]. Une frontière de mot (notée #) a parfois été ajoutée à différents endroits de la chaîne consonantique. Ces énoncés étaient intégrés dans des phrases porteuses du type « It's a lee coot again ». Une attention particulière a été accordée aux éventuels mouvements cinématiques propres aux consonnes intervocaliques.

La Figure 28 propose un échantillon des signaux de protrusion de la lèvre supérieure dans des séquences [iC_ni] vs. [iC_nu]. La différence de comportement de la lèvre supérieure, selon ces deux situations, est très claire : aucun mouvement de protrusion n'est exécuté vers la cible de la seconde voyelle lorsqu'elle est étirée, et ce quel que soit l'environnement consonantique.

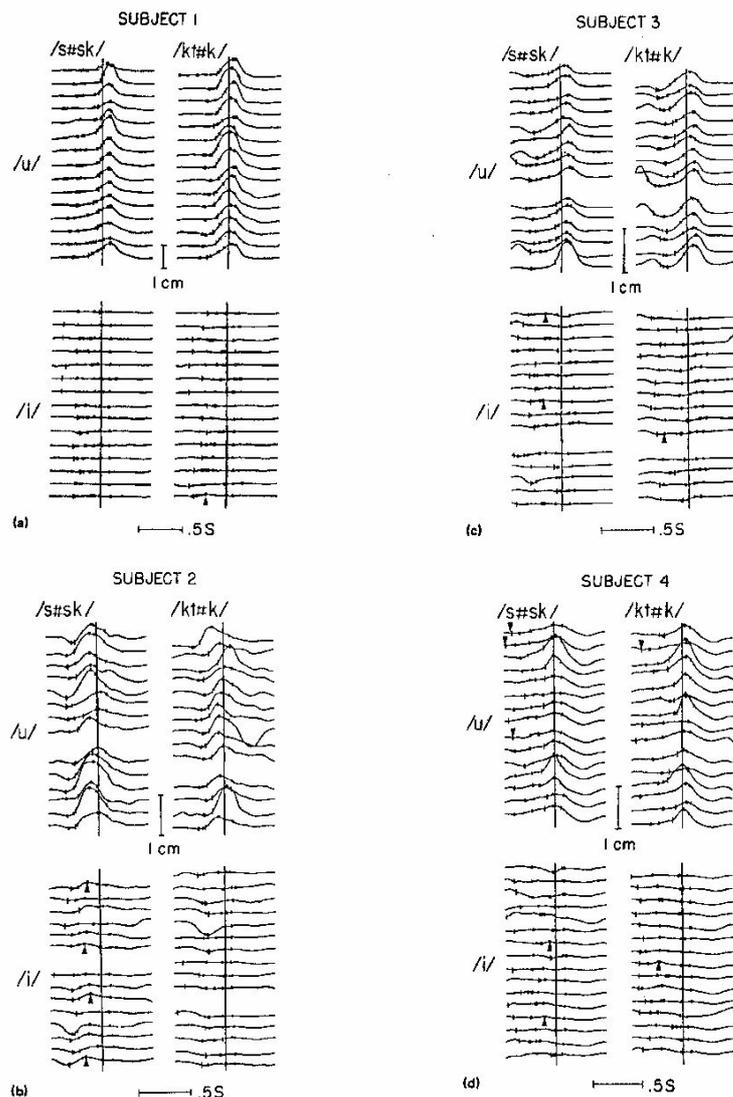


Figure 28 : Tracés des déplacements antéro-postérieurs (protrusion) de la lèvre supérieure pour des séquences [iCnu] (dans le panneau supérieur de chaque sujet) et [iCni] (dans le panneau inférieur). Les lignes verticales marquent le début acoustique de la voyelle sur lequel les signaux ont été alignés. D'après Perkell & Matthies, 1992, p. 2916.

Afin de tester les trois modèles, les auteurs ont repéré différents événements sur les signaux acoustiques et cinématiques de protrusion (Figure 29) :

Sur la fonction de protrusion :

- le temps correspondant à l'initialisation du geste de protrusion (MBeg), déterminé par une valeur minimale de protrusion accompagnée de la dernière valeur positive pour la vitesse ;
- le temps correspondant au maximum d'accélération (ACC MAX)
- le temps marquant la fin du geste de protrusion (MEnd), délimité par le premier maximum de protrusion, correspondant à la première valeur négative pour la vitesse.

Sur le signal acoustique :

- les temps délimitant les bornes de début et fin acoustique des voyelles [i] (i End) et [u] (V Beg).

Par ailleurs, Perkell et Matthies ont considéré quatre phases, décrites dans la figure XX :

- le 'Movement interval' est déterminé par l'écart entre le début du mouvement (MBeg) et le début acoustique de la voyelle arrondie (VBeg) ;
- l' 'Onset interval' est délimité par le pic d'accélération (ACC MAX) et le début acoustique de la voyelle arrondie ;
- l' 'Offset interval' caractérise la durée entre le début acoustique de la voyelle [u] et la fin du mouvement de protrusion (MEnd) ;
- la 'Protrusion duration' est définie par le pic d'accélération (ACC MAX) et la fin du mouvement de protrusion (MEnd).

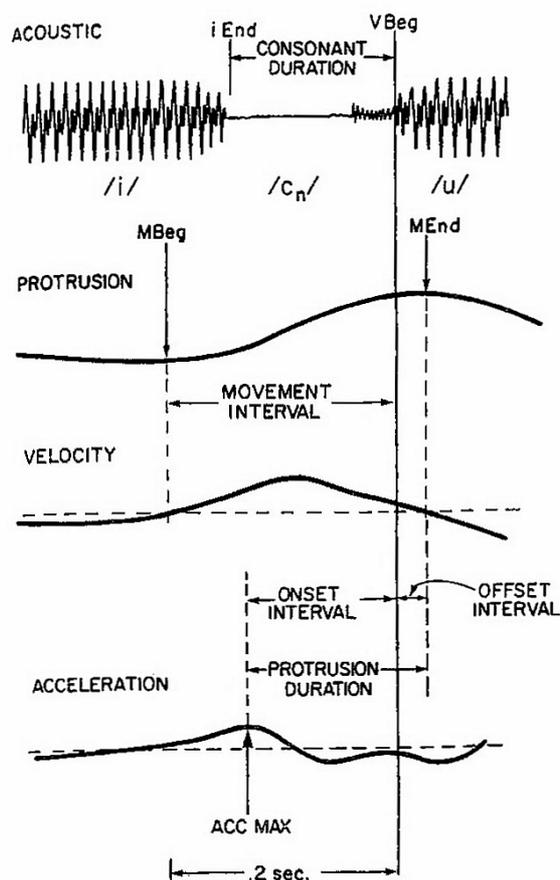


Figure 29 : Représentation schématique des événements repérés sur les signaux de protrusion de la lèvre supérieure. Présentation des signaux de haut en bas : le signal acoustique, le décours temporel du mouvement de la lèvre supérieure, le tracé de sa vitesse, le tracé de l'accélération du mouvement. Les deux événements acoustiques correspondent à la fin acoustique de la voyelle [i] (iEnd) et le début acoustique de la voyelle [u] (VBeg). Les trois événements articulatoires retenus sont : le début du mouvement (MBeg), la fin du mouvement (MEnd) et le pic d'accélération (Accmax). D'après Perkell & Matthies, 1992.

Les auteurs ont observé une forte variabilité du mouvement de protrusion de la voyelle arrondie [u] entre locuteurs et selon la longueur des séquences. D'autre part, ces derniers ont remarqué, pour l'ensemble des sujets, que la fin du geste d'arrondissement ne dépendait pas du début de la voyelle arrondie. Celui-ci se produit en effet à différents moments par rapport au début du voisement pour la voyelle arrondie (parfois jusqu'à 100 ms après le début du voisement ou au contraire jusqu'à 80 ms avant pour l'un des locuteurs dont le débit était plus rapide). Ces premières constatations semblent plutôt infirmer les versions fortes des deux modèles *Look-ahead* (proportionnel) et *Time-locked* (invariance absolue).

En ce qui concerne le décours temporel de l'arrondissement, la figure 30 présente les résultats obtenus pour trois des quatre intervalles cités plus haut en relation avec la durée de l'intervalle consonantique. Au vu de ces données, les comportements d'arrondissement sont très variables d'un locuteur à l'autre.

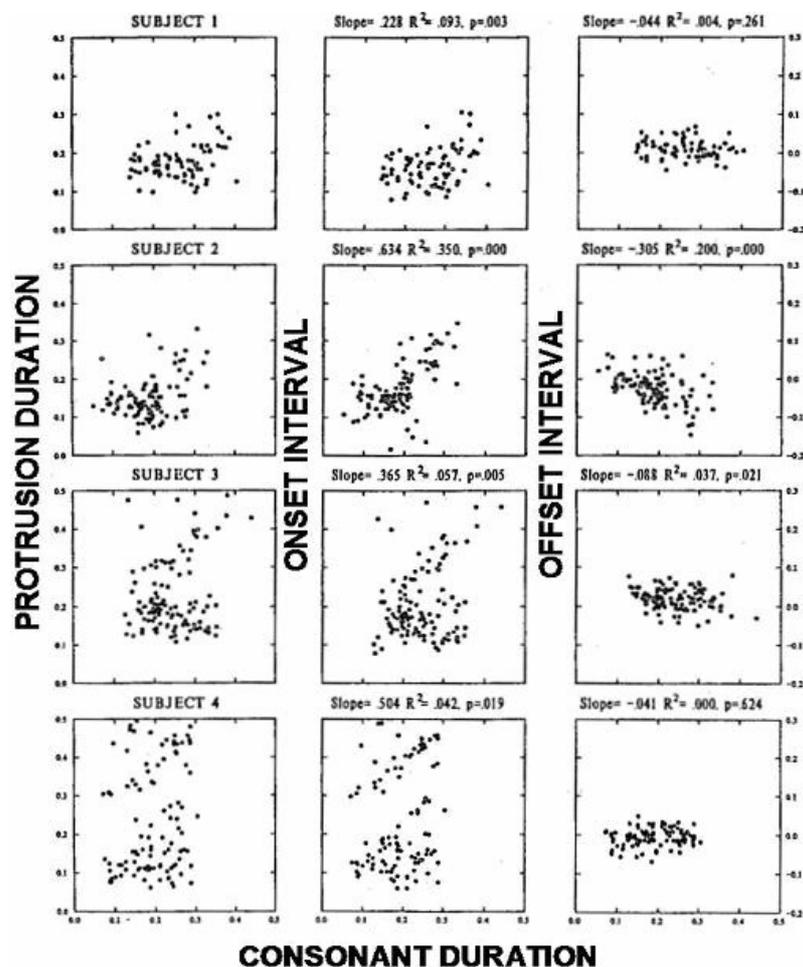


Figure 30 : Illustration des trois intervalles de mouvements (en secondes) mis en relation avec la durée de l'intervalle consonantique (en secondes) pour toutes les séquences (excepté celles comportant un [s]) chez les 4 locuteurs américains (de haut en bas). Les résultats de l'analyse par régression linéaire sont indiqués en haut du graphique. D'après Perkell & Matthies, 1992, p. 2921.

A partir de ces graphiques, Perkell & Matthies (1992) ont émis les conclusions suivantes. Seules les régressions pour l'*onset interval* (colonne du milieu) sont significatives pour les quatre locuteurs. En ce qui concerne l'*offset interval* (colonne de droite), seuls les locuteurs 2 et 3 ont un coefficient de corrélation significatif. Néanmoins, les auteurs soulignent que les coefficients de corrélation (R^2) obtenus rendent seulement compte de 4% à 35% de la variance des données. Le Tableau 1 récapitule les coefficients r obtenus pour ces deux intervalles.

	Onset interval	Offset interval
Sujet 1	r= 0,30	r= 0,06 ns
Sujet 2	r=0,59	r=0,45
Sujet 3	r=0,24	r=0,20
Sujet 4	r=0,20	r=0 ns

Tableau 1 : Distribution des coefficients de corrélation r obtenus pour l'*onset interval* et l'*offset interval* chez les quatre sujets américains ainsi que leurs seuil de significativité d'après la table de Bravais-Pearson. Calculs réalisés d'après la Figure 30 (Perkell & Matthies, 1992)

On a noté que seul le deuxième locuteur obtient un r élevé de 0,59 pour l'*onset interval* (expliquant 35% de la variance). Et pourtant ce dernier résultat ne peut cependant être pris en compte car il souffre, ainsi que les autres d'ailleurs, d'un artefact statistique, ce que nous allons démontrer. En effet, l'*onset interval* (comme l'*offset interval*) partage une borne commune avec la durée de l'intervalle consonantique ou *consonant duration*, en l'occurrence la borne de début acoustique de la voyelle arrondie. Or, il a été démontré par Benoît (1986) qu'il existe une corrélation « tout-partie » supérieure à 0,70, même si les deux parties ne sont pas corrélées, en raison de cette borne commune (Figure 31). Une corrélation de 0,59 est donc un simple résultat artéfactuel.

Ce phénomène avait déjà été discuté auparavant par Barry (1983), sans être validé statistiquement. Benoît (1986) a donc démontré mathématiquement l'existence de cet artefact statistique, présent dans de nombreuses études sur l'organisation temporelle de la parole (entre autres, dans l'étude EMG de Lubker 1979, Almé & McAllister, 1984 pour le français, le danois et le suédois).

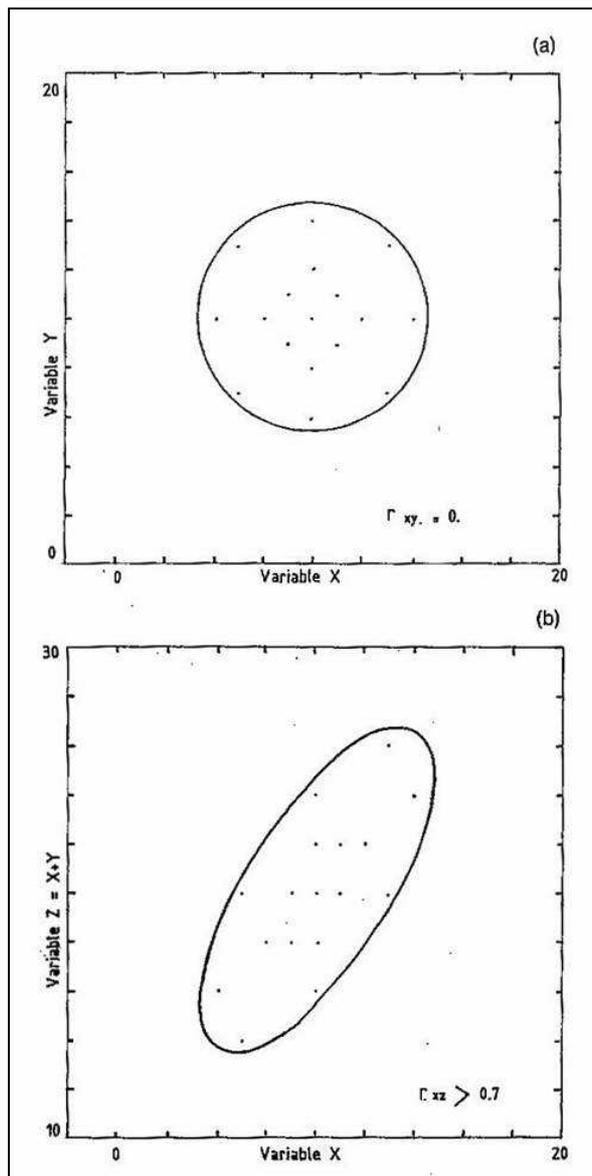


Figure 31 : Ellipses de dispersion pour des données de démonstration, illustrant ce qui se passe lorsque 2 variables x et y non corrélées en (a), donnent lieu à un coefficient de corrélation r plus grand que 0,7, lorsque l'une des variables x est corrélée à la somme $x+y$. D'après Benoît, 1986, p. 1847.

D'après les données cinématiques obtenues pour les quatre locuteurs (en éliminant les séquences comportant un [s]), Perkell & Matthies (1992) émettent l'hypothèse de trois contraintes agissant sur l'organisation de la coarticulation :

- la fin de la protrusion interviendrait pendant la partie voisée de la voyelle arrondie [u] ;
- le locuteur choisirait la durée de mouvement d'arrondissement la plus adéquate ;
- le début du mouvement pour la cible vocalique [u] serait également déterminé en lien avec des contraintes perceptives. De fait, il s'agirait de ne pas interférer avec la

configuration articulatoire requise pour la voyelle non arrondie [i] afin de garantir l'efficacité perceptive.

Les auteurs soulignent la nécessité d'obtenir des données plus quantitatives, afin de mieux apprécier la compétition de ces trois contraintes sur l'organisation temporelle de l'anticipation d'arrondissement.

Dans un article de 1990, Perkell concluait déjà qu'en l'état actuel des données, aucun des trois modèles ne peut rendre compte de la grande variabilité des comportements recueillis (variabilité intra et interlocuteur).

« In spite of the large amount of observed variability, there are relationships in the data which allow us to reject strong versions of all three models as means of accounting for the results of this particular experiment. » (Perkell, 1990, p. 280)

Perkell & Matthies (1992) concluent également que le modèle *Hybrid* n'est pas capable de rendre compte des comportements d'anticipation observés et que l'hypothèse d'une segmentation du mouvement d'arrondissement en deux phases peut correspondre en réalité aux mouvements isolés et très variables des consonnes, en sus du mouvement d'arrondissement attendu pour la voyelle arrondie (p. 2923). Plus précisément, ils suggèrent que l'événement MBeg (début du mouvement) traduit un début d'arrondissement lié à la consonne [s] (tout en soulignant la grande variabilité des données observées) tandis que le maximum d'accélération témoignerait davantage du geste d'arrondissement associé à la voyelle [u].

« Evidence which was originally interpreted as support for the hybrid model (Perkell & Chiang, 1986) appears to stem, at least in parts, from consonant effects such as those discussed above for /s/. » (Perkell & Matthies, 1992, p. 2923)

En effet, Perkell & Matthies ont testé minutieusement cet effet supposé du [s] sur la lèvre supérieure. Après avoir essayé d'analyser l'influence du [s] sur l'onset intervalle (le seul qui expliquerait, artéfact à part, un pourcentage appréciable de la variance), ils concluent :

« However, for each /s/-containing utterance type, there was no statistically significant correlation of onset interval with consonant duration, so the comparisons could not be made. » (Perkell & Matthies, 1992, p. 2919)

Cette question, comme celle des modèles, ne semblant pas avoir été davantage résolue par Perkell et Matthies, du fait de la petite amplitude des gestes et de la variabilité de ces effets supposés des consonnes intervocaliques, nous proposerons, pour éviter ces effets qui peuvent se manifester aussi dans le paramètre d'aire aux lèvres – qui nous servira à généraliser notre

modèle d'anticipation du français à l'anglais –, de comparer les performances des sujets en mâchoire libre et en mâchoire bloquée.

I.6. L'anticipation vue comme un phénomène 'longue distance' : Le *Modèle d'Expansion du Mouvement (MEM)* (Abry & Lallouache, 1991)

De manière générale, les trois modèles décrits dans l'étude de Perkell & Matthies (1992) tentent de déterminer si la durée du mouvement d'arrondissement labial est liée à la fin acoustique de la voyelle non arrondie (profil *Look-ahead*), ou au début acoustique de la voyelle arrondie [u] (comportement du type *Time-locked*), ou contrainte par les deux événements acoustiques (modèle *Hybrid*).

Pour le français, un modèle alternatif à ceux décrits précédemment a été élaboré par Abry & Lallouache (1991). Ce modèle, appelé *Modèle d'Expansion du Mouvement* propose, au vu des résultats controversés des travaux expérimentaux menés dans le domaine, une nouvelle approche de l'anticipation d'arrondissement. En effet, il s'agit de rendre compte de l'anticipation comme d'un phénomène pouvant se développer sur une longue distance temporelle (*long range anticipation*), sur laquelle le locuteur ajuste son comportement d'arrondissement selon sa propre loi d'expansion du mouvement en fonction de la longueur, de la durée de la chaîne consonantique. Il faut insister sur l'optique de ce modèle qui est de rechercher un comportement du locuteur personnel mais *lawful*, c'est-à-dire comportant une expansion régulière et non erratique, en l'occurrence un coefficient personnel d'expansion du mouvement, dans une fonction linéaire statistiquement significative.

Abry et Lallouache se sont dans un premier temps intéressés à *la protrusion de la lèvre supérieure* (paramètre le plus largement étudié dans la littérature) puis ont testé leur modèle sur le paramètre de *l'aire aux lèvres* (constriction).

Nous présenterons dans cette partie chacune de ces deux études.

1.6.1. Le MEM de protrusion (Abry & Lallouache, 1991, 1995a)

1.6.1.1. Présentation du paradigme expérimental

Abry & Lallouache (1991, 1995a) ont tenu à adopter, pour comparaison avec les résultats obtenus par Perkell & Matthies (1992) qui ont testé le comportement d'arrondissement de leurs 4 locuteurs anglophones, le même paradigme expérimental, pour 4 locuteurs francophones. Ainsi ces derniers ont enregistré des séquences [iC_ny], dans lesquelles C_n était formée d'une chaîne consonantique variable de 0 à 5 consonnes, de manière à observer les

effets de l'allongement de l'intervalle acoustique entre les deux voyelles sur le déroulement temporel de l'arrondissement. Les auteurs ont tiré profit du contraste vocalique [i] vs. [y]. En effet, même si cette dernière voyelle est un peu moins protruse que la voyelle [u], elle bénéficie d'une position linguale très voisine de [i] ce qui permet de tester essentiellement le contraste produit par le geste d'arrondissement des lèvres.

L'intervalle acoustique le plus long était composé de la suite miroir [kstsk] et inséré dans la phrase porteuse « Ces deux Sixte sculptèrent », formant ainsi la chaîne consonantique maximale. Notons que ce type d'intervalle consonantique existe en français dans les cas de jointures de mots (jusqu'à 6 consonnes : « un directeur de société mixte scrupuleux » où, en français standard, l'élision du *e* permet de prononcer la suite [kstsk[®]]; contrairement à l'exemple « sinistre structure » de Benguerel & Cowan, 1974, qui est soit prononcé [st[®]↔st[®]] ou [stst[®]], jamais [st[®]st[®]]). Les items suivants ont été formés en enlevant une à une les consonnes du centre du groupe, jusqu'à donner la simple transition vocalique sans consonne [iy] (Tableau 2).

Séquences porteuses Groupe nominal-Verbe	Transitions étudiées V1...C _n ...V2	Nombre de consonnes (de 0 à 5)
Ces deux Scies ultèrent	[iy]	0
Ces deux Scies cultèrent	[iky]	1
Ces deux Sikhs cultèrent	[ikky]	2
Ces deux Sikhs sculptèrent	[iksky]	3
Ces deux Sixes sculptèrent	[ikssky]	4
Ces deux Sixtes sculptèrent	[ikstsky]	5

Tableau 2 : Présentation des 6 séquences élaborées pour le corpus d'arrondissement [iC_ny] chez le locuteur français.

Les sujets ont été enregistrés de face et de profil par le poste « Visage-Parole » ou *ICP-Lip-Shape-Tracker* élaboré par Lallouache (1991) (une description détaillée de ce système sera fournie dans notre partie III).

1.6.1.2. Événements et phases proposés dans le Modèle MEM pour la description des comportements de protrusion de locuteurs Français

Différents événements acoustiques et cinématiques ont été relevés. La Figure 32 propose une illustration de ces événements pour la suite miroir [kstsk]. Sur le signal acoustique, les auteurs ont défini l'intervalle d'obstruence (IO) correspondant à l'intervalle acoustique entre les deux voyelles. Celui-ci est délimité par la fin acoustique de la voyelle non arrondie [i]

(*Vocalic Voice Termination*) et le début de la voyelle arrondie [y] (*Vocalic Voice Onset*), identifiés par deux flèches sur la figure. Ces deux étiquettes, apposées manuellement sur le signal acoustique correspondant à la fin de la structure formantique pour le [i] et son début pour le [y].

Sur le décours temporel du signal de protrusion, quatre événements cinématiques ont été repérés : le temps correspondant au minimum de protrusion [i] (cercle noir), au pic d'accélération (triangle blanc), celui marquant un pic de vitesse (triangle noir) et enfin le temps correspondant à la valeur maximale pour la protrusion (cercle blanc).

A partir des événements décrits ci-dessus, les auteurs ont calculé la durée du mouvement de protrusion MT (maximum de protrusion – minimum de protrusion) afin de décrire l'évolution temporelle du mouvement d'arrondissement (exprimée en ms) par rapport à l'évolution de la durée de l'intervalle consonantique entre les deux voyelles, mesuré sur le signal acoustique (nommé *Intervalle d'Obstruence IO*).

1.6.1.3. Modélisation du comportement de protrusion

La Figure 33 présente les résultats obtenus pour le locuteur Jean-Luc. L'analyse par régression linéaire révèle qu'il existe une augmentation linéaire du mouvement d'arrondissement par rapport à IO, avec une pente de 0,42 validée par un coefficient de corrélation significatif ($p < 0,00001$ sans les séquences [iy]). Ce premier constat réfute donc l'hypothèse d'un comportement de type *Time-locked* (qui serait caractérisé par une pente nulle) pour ce locuteur.

Par ailleurs, les auteurs relèvent plusieurs phénomènes importants :

- Le maximum de protrusion se produit autour du début acoustique de la voyelle arrondie [y] (cercles blancs, Figure 35).
- Ce mouvement débute de plus en plus tôt par rapport au début de la cible en fonction de l'augmentation de IO, et s'accompagne des événements de pics d'accélération et de vitesse observés en Figure 34.
- Plus précisément, le geste d'arrondissement peut débiter dès la fin acoustique du [i] lorsque IO est petit (40-80% de la consonne [k]) ou plus tardivement lorsque IO est grand (à 20 –40% de la durée de la chaîne de consonnes [kstsk]).
- Il existe une durée du mouvement (MT) incompressible pour la transition simple de [i] à [y] (appelé *geste de base*) ; il s'agit de la *constante d'exécution* (autour de 140ms pour ce locuteur). Cette durée est similaire pour une séquence où IO est

composé d'une seule consonne (cf. la droite rejoignant la moyenne des transitions [iy] à celles des [iky] dans la Figure 33).

- La durée du mouvement de protrusion des lèvres augmente ainsi de façon linéaire par rapport à l'intervalle d'obstruence selon un certain coefficient d'expansion. Les événements de pic d'accélération et de vitesse suivent également une relation linéaire par rapport à IO (Figure 34). Ces résultats vont à l'encontre des prédictions du modèle *Hybrid* en ce qui concerne l'initialisation à date fixe du pic d'accélération par rapport au début de la voyelle arrondie.

Par ailleurs, on remarque sur la Figure 34 que les départs des minimums de protrusion suivent une hyperbole (dont la corrélation de 0,93 est très significative). Ces événements se produisent majoritairement tôt dans la voyelle non arrondie [i], voire même avant le début de la voyelle dans les séquences où IO correspond à une seule consonne. Le comportement de ce locuteur ne se conforme donc pas aux prédictions du modèle *Look-ahead* quant à un démarrage du mouvement d'arrondissement fixé à la fin de la voyelle [i].

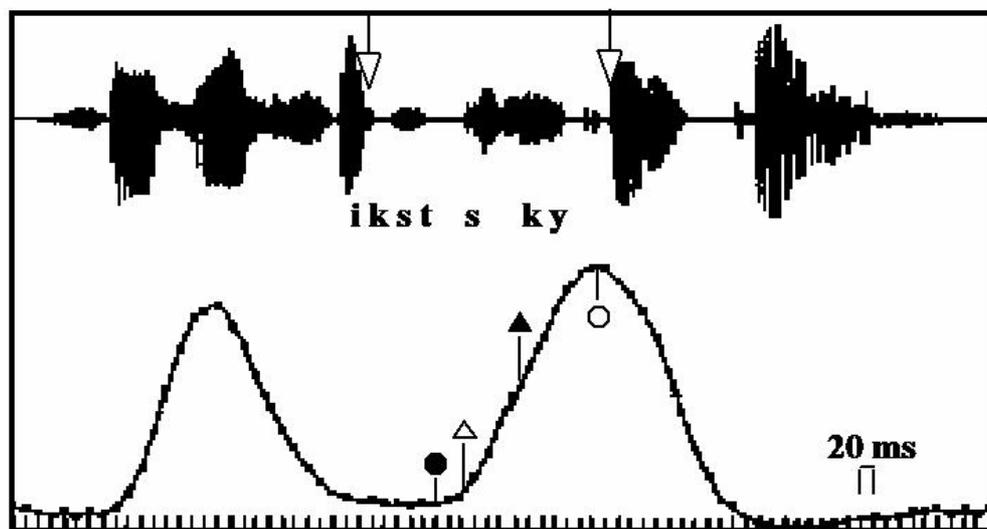


Figure 32 : Tracé du signal acoustique (signal du haut) et du déroulement temporel du mouvement de protrusion de la lèvre supérieure pour la séquence [ikstsky]. D'après Abry & Lallouache, 1991, p. 223.

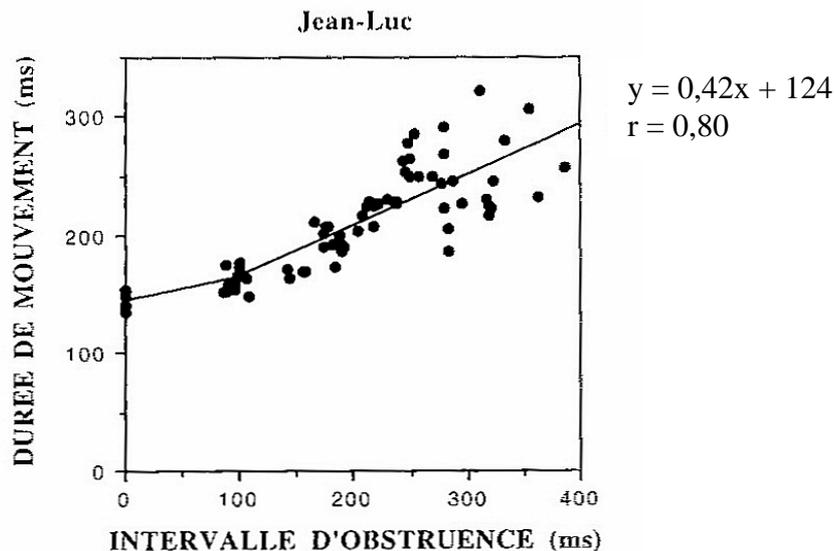


Figure 33 : Relation entre la durée du mouvement de protrusion (en ms) et l'intervalle d'obstruence (en ms). La régression et le coefficient de corrélation r sont indiqués à droite. Ils n'incluent pas les valeurs obtenues pour les transitions vocaliques minimales [iy]. La droite rejoignant la moyenne de ces transitions à celles de [iky] est tracée par commodité. D'après Abry & Lallouache, 1995a, p. 94.

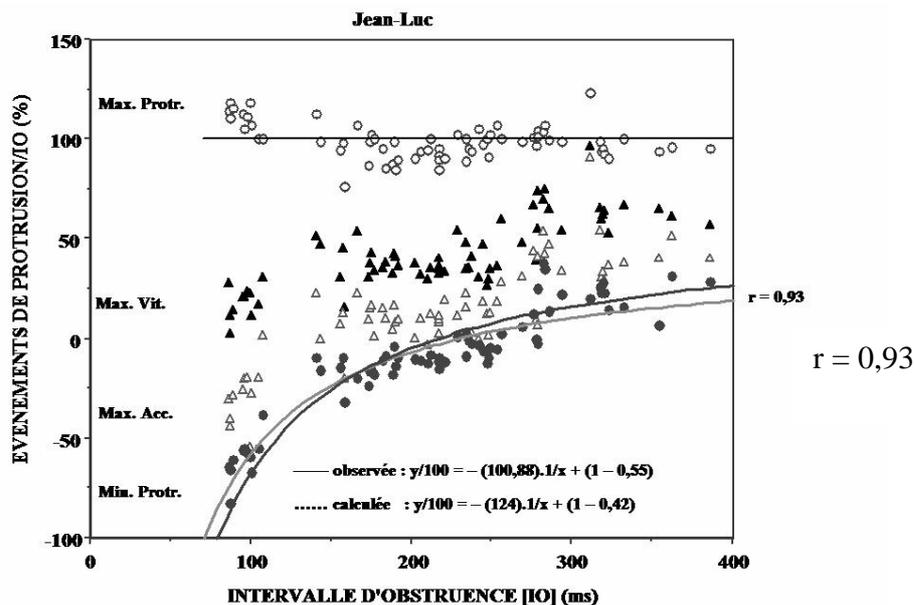


Figure 35 : Organisation temporelle des événements repérés sur le signal de protrusion (%) par rapport à l'Intervalle d'Obstruence (ms). Ces événements sont exprimés en pourcentage de la durée de IO. La droite horizontale au niveau de la valeur 100% marque le début acoustique de la cible vocalique arrondie [y]. Les deux hyperboles observée et calculées sont ajustées à partir du départ du geste de protrusion (Min. Protr.). D'après Abry & Lallouache 1995a, p.92

La Figure 36 présente seulement les résultats obtenus pour deux des trois autres locuteurs testés. En effet, les signaux de protrusion du troisième locuteur 'Christophe' avaient une trop faible amplitude pour être exploitables (nous reviendrons en détail sur le cas de ce locuteur

pour le modèle de constriction). Les analyses par régression linéaire révèlent des comportements d'anticipation d'arrondissement similaires à celui observé chez le premier locuteur ('Jean-Luc'). Pour les trois locuteurs, Abry et Lallouache observent une durée du mouvement (MT) d'environ 140 ms pour la séquence minimale [iy] (*geste de base*) et cette valeur reste stable pour la séquence [iky]. A partir des séquences comportant une consonne intervocalique, la durée du mouvement d'arrondissement croît de manière linéaire, selon un coefficient d'expansion spécifique à chaque locuteur. Ce phénomène se traduit par une pente différente selon le locuteur dont la validité est soutenue par des coefficients de corrélations forts. Ces corrélations nous éloignent des données fortement bruitées comme celles obtenues dans Perkell & Matthies (1992) ou de certaines données de Lubker & Gay (1982). Nous avons donc affaire à un comportement individuel mais *lawful*, soit régulier.

Le mouvement d'anticipation d'arrondissement étant ainsi paramétré par locuteur, chacun ajuste son geste d'arrondissement avec son propre « pas » d'expansion. Dans cette perspective, un profil d'arrondissement de type *Look-ahead* est tout à fait envisageable. Il s'agit d'ailleurs du cas d'un des locuteurs testés : 'Annie' pour lequel on peut observer une pente proche de 1 (c'est-à-dire un mouvement proportionnel à la longueur de l'intervalle d'obstruence). En revanche, les auteurs soulignent que, par définition du *MEM* qui est paramétrable par locuteur, ce modèle *Look-ahead* ne peut être généralisable à l'ensemble des locuteurs d'une communauté linguistique (ici les deux autres sujets ont respectivement des pentes de 0,42 et 0,54).

D'autre part, les auteurs ne relèvent aucun comportement de type *Time-locked* (pente de 0) avec un mouvement invariant, débutant à date fixe par rapport au début de la cible vocalique [y].

Les pentes des trois locuteurs se situent donc entre les deux pentes théoriques extrêmes suggérées par les modèles *Time-locked* et *Look-ahead*, avec cependant une tendance pour le sujet 'Annie' à se comporter selon les prédictions de ce dernier modèle (Figure 37).

Il existe ainsi une variabilité inter-locuteur qui s'explique par le fait que chacun adopte son propre comportement d'anticipation du mouvement. Cette variabilité n'induit pas forcément de variabilité intra-locuteur puisque chacun *contrôle* l'extension du mouvement d'arrondissement selon un coefficient d'expansion dont la stabilité est confortée par les coefficients de corrélations significatifs. Contrairement aux trois autres modèles, le *Modèle d'expansion du Mouvement (MEM)* semble parvenir à rendre compte de la *régularité* des comportements personnels observés. Ce modèle intègre en effet la variabilité individuelle

comme composante intrinsèque, paramétrée, de l'organisation de l'anticipation d'arrondissement.

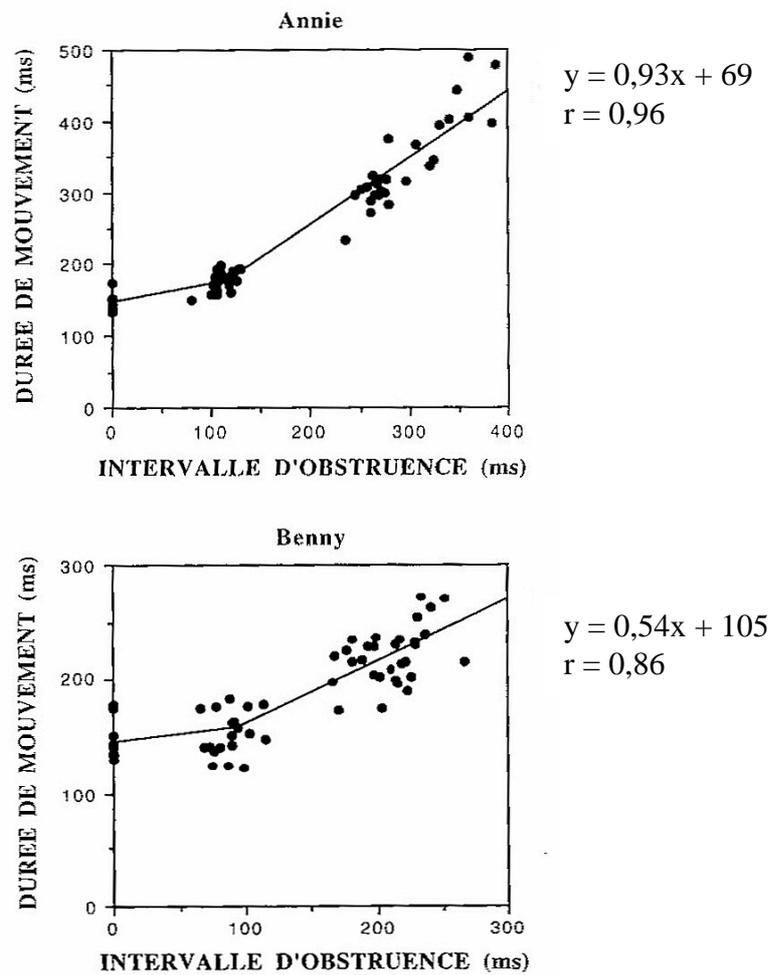


Figure 36 : Relation entre la durée du mouvement de protrusion de la lèvre supérieure (ms) et l'intervalle d'obstruence (ms) pour les trois locuteurs. La régression et le coefficient de corrélation r sont indiqués à droite. Une petite droite a été tracée entre la moyenne des transitions [iy] et celle des transitions [iky]. D'après Abry & Lallouache, 1995a, p. 94.

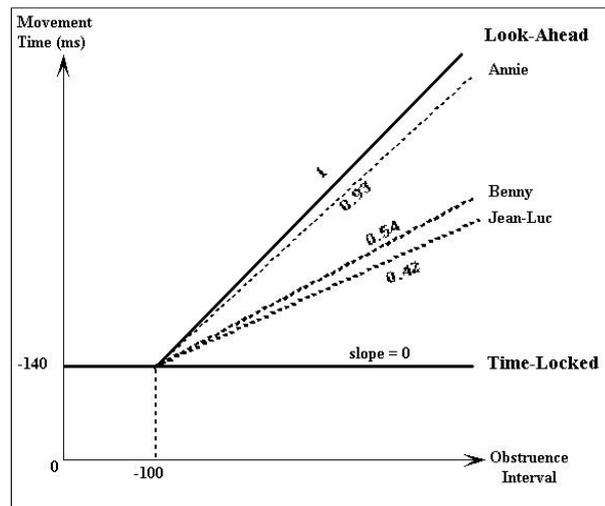


Figure 37 : Illustration des pentes obtenues en fonction de l'*Intervalle d'Obstruence* pour les trois locuteurs. Celles-ci se situent entre le profil *Look-ahead* (pente 1) et le profil *Time-locked* (pente 0). D'après Abry & Lallouache, 1995a, p. 96.

1.6.2. Adaptation du modèle à la description d'un paramètre plus robuste, l'aire intérolabiale : le MEM de constriction

Comme nous l'avons mentionné plus haut, un des locuteurs montrait une trop faible protrusion pour être analysable. Le *Modèle d'Expansion du Mouvement* a été adapté à l'étude de la constriction intérolabiale (Abry & Lallouache, 1995b ; Abry *et al.*, 1996a et b) afin de prédire l'évolution temporelle du mouvement d'arrondissement pour les locuteurs ne montrant pas de protrusion. La constriction est en effet un paramètre articulatoire réputé crucial pour ses conséquences acoustiques en sortie du conduit vocal (Badin *et al.*, 1994 ; Perkell & Matthies, 1992), jouant par conséquent un rôle majeur dans la perception du message articulé par l'interlocuteur. Soulignons que l'étude que nous décrivons ici s'avère la première tentative d'exploitation de ce paramètre dans le cadre de la coarticulation anticipante. Selon Abry *et al.* (1996a), ce phénomène serait lié au fait qu'aucun système suffisamment robuste n'avait jamais été conçu pour ce genre d'analyse cinématique, de même qu'aucune procédure de détection d'événements stables et pertinents pour la description des comportements d'anticipation.

1.6.2.1. Description des événements et phases proposés dans le Modèle MEM de constriction

Rappelons que les quatre locuteurs avaient été enregistrés de profil (pour la protrusion) et simultanément de face, pour obtenir l'aire aux lèvres. Sur les fonctions temporelles de la constriction (aire aux lèvres), plusieurs références robustes ont pu être détectées :

- l'aire maximale pour la cible [i] (événement 1 sur la Figure 38) ;
- l'aire minimale pour la voyelle [y] (événement 4) ;
- le temps correspondant à une chute de 10% de l'aire entre [i] et [y] (événement 2). Les auteurs ont choisi cet événement plutôt que le traditionnel maximum d'aire pour garantir l'amorce du geste d'arrondissement labial et éviter les éventuels effets de plateau autour du maximum et minimum d'aire aux lèvres.
- le temps correspondant à une chute de 90% de l'amplitude d'aire intérolabiale vers la voyelle arrondie (événement 3) ;
- le temps correspondant à 10% d'augmentation de l'aire aux lèvres suivant l'aire minimale de la voyelle [y] (événement 5).

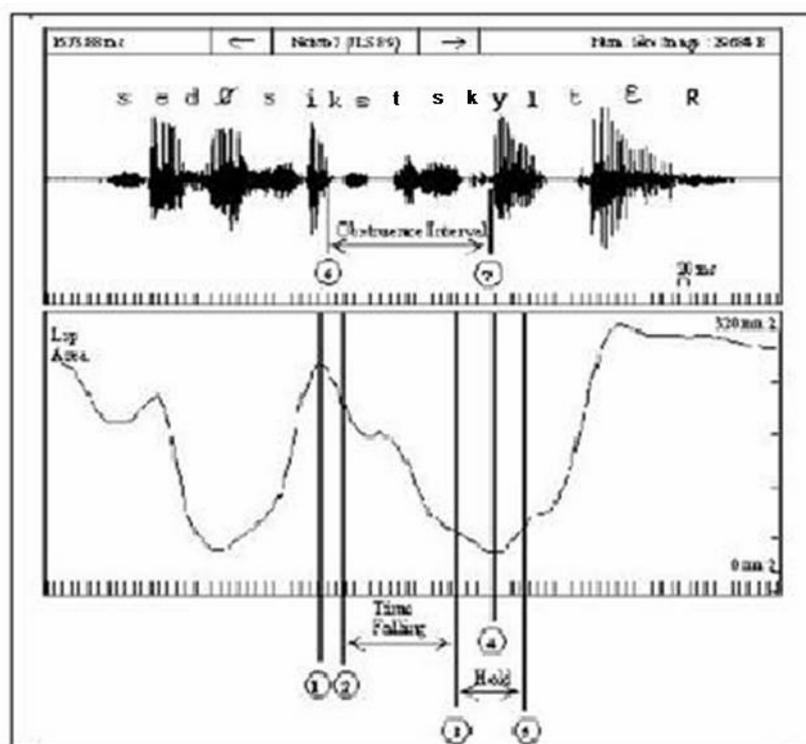


Figure 38 : Signal acoustique de la séquence « Ces deux Sixtes sculptèrent » (signal du haut), décours temporel de l'aire aux lèvres (signal du bas). L'événement (1) correspond au maximum d'aire pour la voyelle [i], (2) à une chute de 10% de l'aire entre [i] et [y], (3) à une chute de 90% de l'aire entre [i] et [y], (4) au minimum d'aire pour la voyelle [y], (5) à 10% d'augmentation de l'aire aux lèvres suivant l'aire minimale de la voyelle [y]. La phase de *Time Falling* est délimitée par les événements 2 et 3 et La phase de *Hold* est délimitée par les événements 3 et 4. L'*Intervalle d'Obstruence* est défini par les bornes 6 et 7 correspondant à la fin acoustique de la voyelle [i] et au début de [y]. Figure de Abry et Lallouache, publiée dans Cathiard *et al.*, 2003.

Ces différents événements ont été utilisés pour constituer deux phases articulatoires décrivant le mouvement d'arrondissement. La phase de *Time Falling* est déterminée entre l'événement 2 et 3 et constitue la chute du geste d'arrondissement vers la cible vocalique

(Figure 38). La phase de *Hold* est délimitée par les événements 3 et 5 et représente une phase de tenue du geste de constriction minimale pour la voyelle arrondie, durant laquelle l'efficacité acoustique de la constriction est maximale. Ces deux phases sont traditionnellement utilisées en électromagnétisme pour décrire les courbes d'hystérésis.

Sur le signal acoustique, l'intervalle d'obstruence (IO) a été déterminé sur le spectrogramme entre la fin acoustique de la voyelle [i] (caractérisée par la disparition de la structure formantique) et le début de la voyelle arrondie [y] (apparition de la structure formantique).

1.6.2.2. Modélisation du comportement de constriction

Afin de rendre compte du comportement d'anticipation d'arrondissement des locuteurs, Abry & Lallouache (1995b) ont mis en relation la durée totale de la constriction (phase de *Time Falling* + *Hold*) avec la durée de l'*Intervalle d'Obstruence* (IO). Une analyse de régression linéaire a été réalisée pour chacun des quatre sujets (sans les séquences [iy]). Ces derniers ont pris soin de corrélérer des phases articulatoires et acoustique ne partageant pas de borne commune afin d'éviter la génération de corrélations tout-parties.

De manière générale, les auteurs ont mis en évidence une relation linéaire entre le mouvement d'arrondissement labial et la durée de IO, tout comme lors de leur étude précédente sur le décours temporel de la protrusion (Figures 39 et 40).

La pente relative à cette relation est de nouveau paramétrée par locuteur selon un coefficient d'expansion spécifique. Les coefficients de corrélations sont significatifs (de 0,97 à 0,99 selon le locuteur). De nouveau, le sujet 'Annie' présente un comportement d'anticipation d'arrondissement proche du profil *Look-ahead* (avec une pente proche de 1) mais il s'agit d'un seul cas, puisque les trois autres locuteurs ont des pentes de 0,79 ; 0,69 et 0,74 (Figure 39).

Un comportement du type *Look-ahead* peut donc tout à fait être observé et être décrit dans le cadre des principes du modèle *MEM*. En revanche, l'hypothèse d'un comportement de type *Time-locked* associé à une pente nulle et à un mouvement invariant temporellement est rejetée.

Il semblerait donc que tous les locuteurs présentent un comportement d'expansion linéaire selon un coefficient d'expansion et une pente spécifique à chaque sujet. Plus précisément, les auteurs remarquent que lorsque l'Intervalle d'Obstruence (IO) est petit, le mouvement d'arrondissement débute dès la voyelle [i]. Lorsque IO est plus long (par exemple constitué de

5 consonnes), ce mouvement est initié plus tard et s'étend de manière linéaire. Par ailleurs, on remarque une durée du mouvement pour le geste de base [iy] d'environ 140 ms, très stable pour l'ensemble des locuteurs (Figure 39). Notons que le même phénomène avait été observé pour la protrusion.

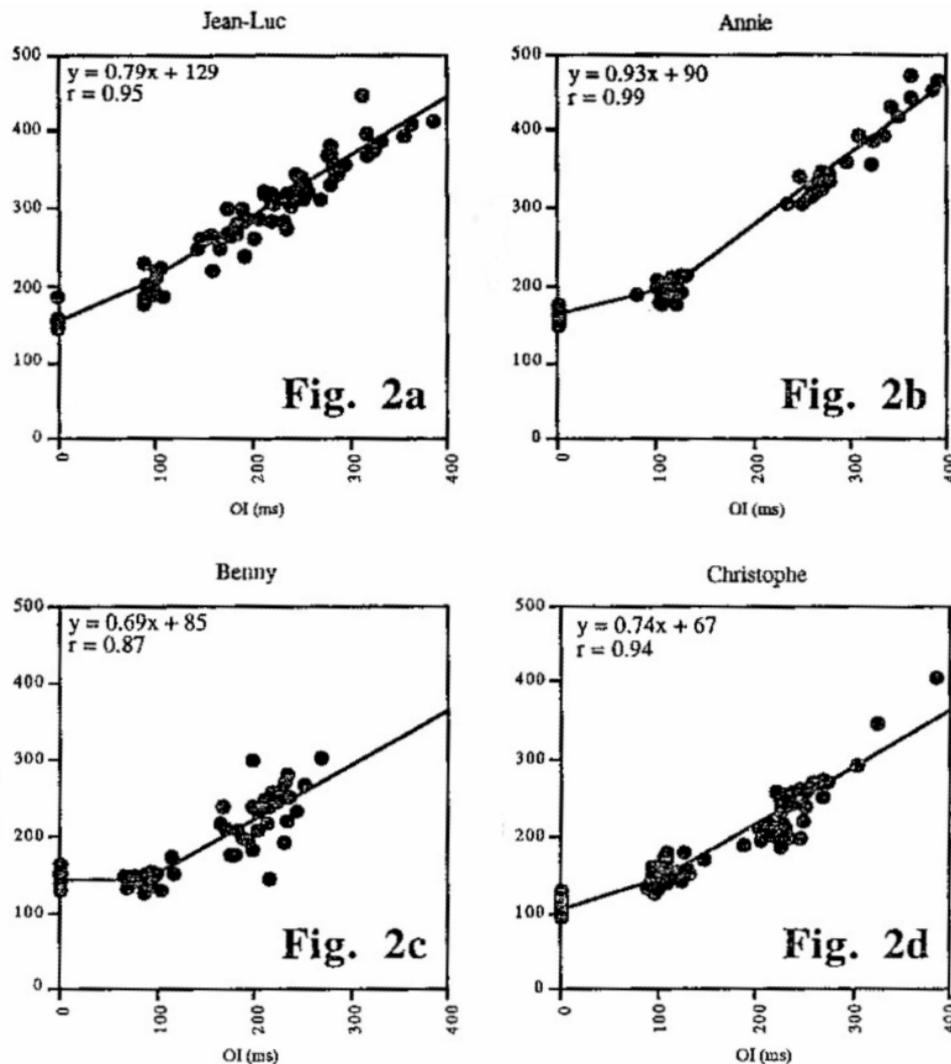


Figure 39 : Relation entre la durée du mouvement d'arrondissement (TF+H) (ms) et l'intervalle d'obstruction (ms) pour les quatre locuteurs. La régression et le coefficient de corrélation r sont indiqués en haut à gauche. Une petite droite a été tracée entre la moyenne des transitions [iy] et celle de transitions [iky]. D'après Abry & Lallouache, 1995b, p. 155

En revanche, les auteurs notent qu'il n'existe pas de continuum entre les séquences comportant une seule consonne intervocalique et les autres séquences plus longues (surtout en ce qui concerne les locuteurs 'Annie' et 'Benny'). Ils font l'hypothèse que certains paramètres prosodiques tels que le débit, peut-être insuffisamment contrôlés durant l'enregistrement, pourraient être la cause de ces irrégularités temporelles.

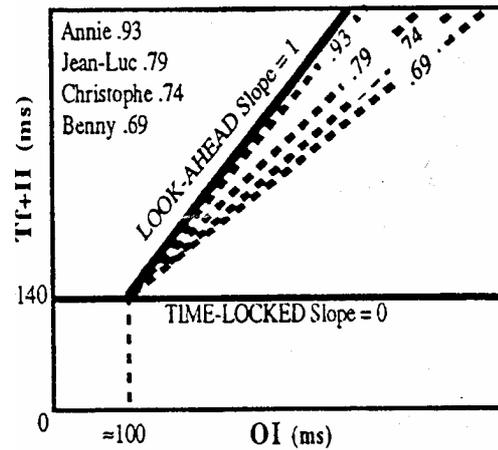


Figure 40 : Représentation schématique des pentes obtenues pour les quatre locuteurs après régression linéaire. D'après Abry & Lallouache, 1995b, p. 152.

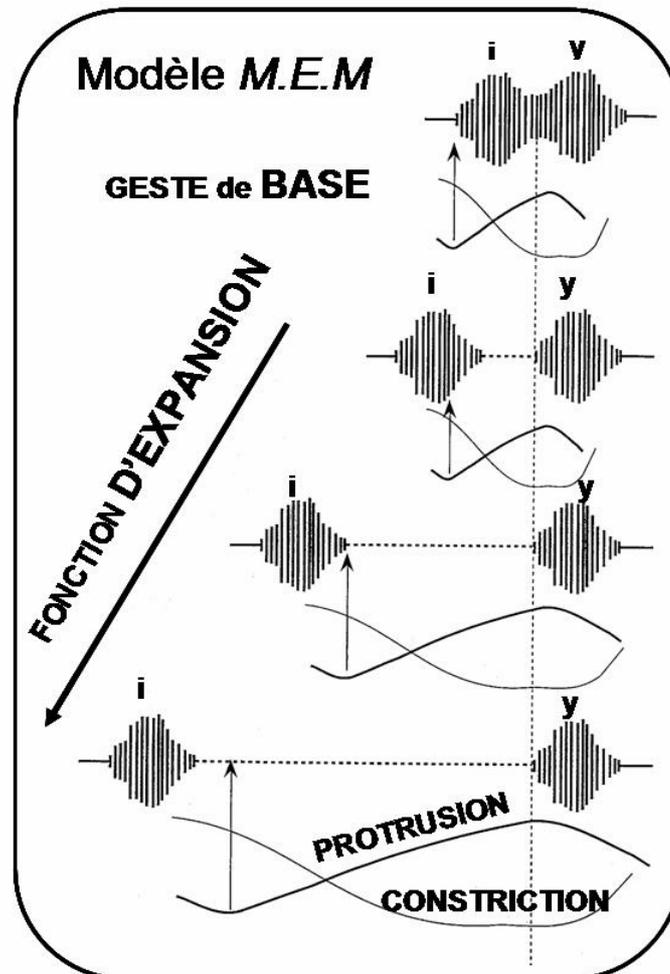


Figure 41 : Illustration de a fonction d'expansion des profils d'anticipation d'arrondissement (protrusion et constriction), partant du geste de transition de base sans consonne, puis avec une consonne, jusqu'à un intervalle de plusieurs consonnes. Notez que les gestes ne débutent pas sur la fin du [i] (même si cela peut arriver) : ils commencent avant cette fin pour les deux premiers cas et après dans le dernier. L'avant-dernier cas étant une simple coïncidence due au coefficient d'expansion propre à ce sujet. (Abry, communication personnelle).

La Figure 41 propose une illustration des profils attendus pour le décours temporel de la protrusion et constriction selon le modèle *MEM*.

Abry *et al.* (1996) ont montré que cette variabilité ne peut être expliquée exclusivement par aucun des trois modèles *Look-ahead*, *Time-locked* ou *Hybrid*. La différence essentielle avec Perkell étant que ce dernier n'explique aucun comportement individuel par un modèle régulier. Le *MEM*, pour prendre une approche connue ailleurs en linguistique chomskienne, applique le même *principe*, celui d'une fonction linéaire d'expansion à partir de son geste de base dans [iy] et [iCy], seul le *paramètre* de pente variant selon le sujet.

Rappelons que la plupart des modèles que nous avons présentés dans ce chapitre sont des modèles de contrôle. Le *MEM* reste jusqu'à présent un modèle phénoménologique. Mais rien n'empêche de le considérer comme un modèle de contrôle où le paramètre d'expansion est ajusté par le sujet au cours de ses interactions dans son *attunement* personnel à sa langue ambiante.

Le comportement d'anticipation tel que modélisé par Abry et Lallouache est le fruit d'une coordination articulatoire (ou co-articulation) spécifique au locuteur, se déployant le long d'un intervalle temporel contraint pour une articulation fluide des segments successifs de parole. Il s'agit là d'un phénomène dit de 'longue distance' (*long range anticipation*), pouvant se propager en français jusqu'à six consonnes en amont de la cible vocalique arrondie. Il faut souligner encore une fois que la variabilité observée suit la propre loi du locuteur, variabilité appelée *lawful variability* par les auteurs (notion reprise par Noiray *et al.* 2006).

Plusieurs points émergent des études décrites précédemment. Tout d'abord en ce qui concerne le ou les dimensions articulatoires sélectionnées. La protrusion ne semble pas adaptée pour la description du comportement d'anticipation d'arrondissement chez tous les locuteurs. Les travaux de Abry & Lallouache (1995a) ont montré que, même chez un locuteur français, la protrusion peut ne pas être utilisée de manière distinctive, bien que cette dimension soit réputée pertinente dans cette langue. La revue proposée par Farnetani (1999) sur la coarticulation labiale fait état d'une différence importante entre protrusion et arrondissement dans de nombreuses langues. Rappelons tout bonnement que dans leur étude sur l'anglais, Perkell & Matthies (1992) avaient déjà souligné la difficulté de recruter des locuteurs ayant une protrusion suffisante pour la réalisation de l'arrondissement. Nous rencontrerons les mêmes comportements de protrusion de très faible amplitude, selon les sujets, dans l'étude qui va suivre.

En se concentrant sur la constriction intérolabiale, qui est sans aucun doute la plus efficace acoustiquement, on constate que le *Modèle d'Expansion du Mouvement* parvient à rendre compte des comportements d'anticipation de tous les sujets francophones. De fait, la constriction semble, ainsi que nous le verrons par la suite, un descripteur articulatoire utilisable dans tous les cas.

Avec ce descripteur robuste, sera-t-on mieux à même de comparer l'organisation temporelle du processus d'anticipation d'arrondissement dans différentes langues et d'y trouver des caractéristiques communes? Cela nous permettra-t-il de ne pas envisager la coarticulation labiale comme un contrôle spécifique à la langue, comme cela a été encore resuggéré en ce début du XXIe siècle par Byrd & Saltzman (2003) ? En quelque sorte un compromis, que nous avons déjà signalé, entre nos résultats pour le français et ceux de Haskins pour l'anglais...

« [...] *the anticipation intervals of gestures may be defined as the time between gestural onset and target attainment. These intervals have been shown to display different degrees of temporal elasticity across languages (e.g. Abry & Lallouache, 1995; Saltzman et al., 2000). By temporal elasticity, we refer to the relationship between a given gesture's activation interval and a preceding interval before target attainment in which there are no conflicting gestures. In English, for example, anticipation intervals are relatively fixed regardless of the length of the preceding no-conflict intervals; in French, however, the anticipation periods stretch in proportion to the length of the no-conflict intervals.* » (Byrd & Saltzman, 2003, p. 157)

Et si ce compromis théorique, *ad hoc* par langue, n'avait rien d'empirique et n'était dû qu'à des comportements sur un descripteur aussi variable que la protrusion, ces auteurs n'ayant pas considéré la constriction d'arrondissement, qui a un rôle crucial dans la génération du but acoustique du geste ?

**Partie II. Extension du *MEM*
de protrusion et de constriction
à des adultes franco-québécois et anglo-
américains**

Dans la partie I, nous avons présenté l'adéquation descriptive d'un modèle, le *MEM*, aux comportements des locuteurs français. La question posée par le compromis « un modèle pour l'anglais, un autre modèle pour le français », proposé en dernier lieu par Byrd & Saltzman (2003), est maintenant: dans quelle mesure ceci est-il acceptable dans le contexte des explications linguistiques qui oscillent toujours entre « language-specific » et UG (grammaire universelle) ? Une approche chomskienne *principe et paramètre ajustable par langue* a pu faire croire que ce dilemme était enfin soluble. La question reste posée en dépit de l'arrivée du programme minimaliste de Chomsky depuis plus de 12 ans. Pour rester au sein de notre modèle, le simple fait qu'il soit ajustable par locuteur, ne garantit en rien qu'il suive une loi générale, une fonction d'expansion régulière, et non pas une dispersion quasiment aléatoire comme dans les comportements anglo-américains mesurés jusqu'à présent, y compris quand les gestes d'anticipation étaient proposés comme « invariants » : en fait ils ne permettaient pas de rejeter l'*Hypothèse Nulle*.

Le pari que nous avons voulu relever, dans ce premier volet expérimental de notre travail, concerne l'exploration du comportement d'arrondissement de locuteurs adultes francophones québécois et anglophones, et cela par l'étude des deux paramètres labiaux articulatoires cruciaux pour cette dimension d'arrondissement : la protrusion et la constriction des lèvres. Les comportements de protrusion et de constriction ont été analysés, pour chacun des locuteurs anglo-américains, dans le cadre du modèle *MEM*. Nous avons bénéficié, pour notre comparaison programmée avec le comportement réputé protrus du français, de l'occasion de tester des membres de la communauté francophone québécoise.

Nous avons pu constater, dans la revue de la littérature, que c'est la protrusion qui a été jusqu'à ce jour le paramètre le plus largement pris en compte, que ce soit par les études EMG (Lubker, 1981 ; Bonnot *et al.*, 1986) ou en suivi par système optoélectrique (Perkell, 1990) ou par traitement d'images (Abry & Lallouache, 1995a). Bien que la constriction des lèvres soit le paramètre reconnu comme le plus robuste du point de vue de l'acoustique (Badin *et al.*, 1994) pour la dimension d'arrondissement, ce paramètre a été relativement peu souvent pris en compte dans les modèles d'anticipation vocalique, au moins ceux proposés pour l'anglais, en raison, semble-t-il, de sa difficulté de mesure. Nous citerons à ce propos les deux remarques de Perkell & Matthies (1992):

« *The labial kinematic parameter that has the most direct relation to the vowel acoustics is the time-varying vocal-tract area function in the labial region; however, such a kinematic measure is currently unobtainable. One feasible approximation requires an elaborate video transduction system (Lallouache, 1991). We did not have such a system, so as a compromise we recorded antero-posterior and vertical displacements of midsagittal points on the vermilion borders of the upper and lower lips.* » (Perkell & Matthies, 1992, p. 2913)

Ce qui les a empêché, à notre avis, d'utiliser un paramètre robuste et pertinent, comme ils le reconnaissent :

« *Another major source of variation may be related to the fact that we have not been examining the only or the best combination of articulatory contributions to production of the vowel /u/. As we have acknowledged, protrusion of the upper lip gives only a fair indication of the lip aperture cross-sectional area and length (cf. Lallouache, 1991).* » (Perkell & Matthies, 1992, p. 2923).

Ainsi le choix du dispositif expérimental apparaît crucial. Nous avons pu bénéficier du poste « Visage-Parole » de l'ICP mis au point par Lallouache (1991), avec son système de détection automatique des contours labiaux grâce au maquillage des lèvres en bleu (*Lip-Shape-Tracker*, réactualisé au niveau de l'implémentation par Audouy, 2000), permettant la détection de paramètres labiaux de face comme de profil. C'est avec ce poste qu'Abry & Lallouache ont réalisé les enregistrements de locuteurs adultes français ayant conduit à l'élaboration du MEM de protrusion (Abry & Lallouache, 1991, 1995a) et du MEM de constriction (Abry & Lallouache, 1995b ; Abry *et al.*, 1996a, 1996b).

Le recueil de nos données d'adultes a été réalisé lors de notre séjour de 3 mois au *Laboratoire de Phonétique* de l'Université du Québec à Montréal (UQAM), dirigé par Lucie Ménard, avec comme objectif de tester la validité du MEM sur deux communautés linguistiques particulières : le franco-québécois et l'anglais. Nous avons donc « transporté » cette technique de maquillage des lèvres en bleu au Québec avec un enregistrement vidéo de face de nos locuteurs pour la détection de la constriction (le laboratoire de phonétique ne disposant que d'une seule caméra vidéo), que nous avons couplé à un système Optotrak (celui du *Motor Control Lab* de l'Université McGill à Montréal), largement utilisé en parole pour la détection de la protrusion. Soulignons que cette technique d'enregistrement simultané par vidéo et Optotrak était en elle-même un beau défi à relever, à notre connaissance jamais tenté.

II.1. Protocole expérimental

II.1.1. Sujets

Nous avons enregistré cinq adultes franco-québécois (deux hommes et trois femmes), ayant une moyenne d'âge de 28 ans. Ainsi que cinq adultes anglo-américains (deux hommes et trois femmes), ayant une moyenne d'âge de 22 ans. Dans cette étude, nous les appellerons constamment anglo-américains (américains au sens d'habitants d'Amérique du Nord, sans aucune connotation de nationalité, du Canada ou des Etats-Unis), n'ayant pas relevé, étant donné leur niveau d'éducation, d'accent particulier les identifiant comme membre d'un dialecte marqué de l'anglais d'Amérique du Nord. Tous résidaient dans la ville de Montréal depuis plus ou moins longtemps. Les sujets ont été recrutés par voie d'affichage au sein de l'Université McGill et de l'UQAM. Les enregistrements ont eu lieu individuellement au *Motor Control Lab*.

Sujets franco-québécois			Sujets anglo-américains		
Nom	Age	Sexe	Nom	Age	Sexe
ANN	35	F	LAU	20	F
JOH	28	F	SAR	24	F
LUC	31	F	SHU	24	F
ANT	22	H	DAN	20	H
JER	24	H	KAL	20	H

Tableau 3 : Présentation des sujets franco-québécois et anglo-américains.

II.1.2. Corpus

Nous avons choisi d'étudier le comportement d'anticipation d'arrondissement pour le contraste vocalique [i] vs. [u]. Nous avons retenu la voyelle arrondie [u], et non la voyelle [y] (comme cela a été le cas lors des précédentes études sur le français), pour des besoins de comparaison inter-langue, la voyelle [u] étant phonémiquement présente aussi bien en français qu'en anglais.

La tâche du locuteur consistait en la production aléatoire d'une série de courts énoncés ayant une construction de type [iC_nu] avec un nombre n de consonnes C variant de 0 et 5. Notons que, si nous avons choisi des groupes consonantiques pouvant atteindre cinq consonnes, c'est parce que ces groupes peuvent réellement exister dans les deux langues testées, avec ou sans jointure de mots. Dans des séquences du type « ces deux Sixte sculptèrent » en français (Abry & Lallouache, 1995a) ou en anglais (avec un maximum de 4

consonnes chez Perkell & Matthies, 1992 : « It's a leaked scoot »), ici aussi jusqu'à 5 dans des logatomes comme « Two keakst skookeek ».

Au vu des travaux expérimentaux menés sur les mouvements labiaux dans la production des voyelles et des consonnes (cf. partie I sur les modèles), il nous est apparu important d'éliminer tout geste éventuel de fermeture des lèvres, en excluant de notre corpus les consonnes occlusives bilabiales ou labiodentales. Pour la formation de l'intervalle consonantique, nous avons choisi les trois consonnes suivantes [k, s, t]. Nous avons pu ainsi obtenir la suite miroir [kstk] formant la chaîne consonantique maximale; puis en enlevant, une à une, les consonnes centrales, les suites [ksk], [kk] et [k], auxquelles nous avons ajouté la suite [kstk]. Ce type de suites miroirs a été précédemment testée par Abry & Lallouache (1995a) avec la voyelle arrondie [y] en position vocalique finale.

Bien que d'un point de vue phonologique, les consonnes sélectionnées soient généralement considérées comme relativement neutres au regard de la dimension d'arrondissement, certains chercheurs, notamment Boyce *et al.* (1990), Gelfer *et al.* (1989) ont pu observer un mouvement des lèvres lors de tâches de productions similaires. La conclusion que nous avons tirée précédemment de leurs données est que l'existence d'un tel effet consonantique sur ces deux effecteurs reste à confirmer (surtout en ce qui concerne la lèvre supérieure, la lèvre inférieure relevant plutôt des mouvements mandibulaires).

Les consonnes sélectionnées pour notre corpus sont réalisées essentiellement par un mouvement de la langue (apex ou dos de la langue), soit dans la région alvéolaire (l'occlusive orale apico-alvéolaire [t]), soit dans la région vélaire (l'occlusive dorso-vélaire [k]).

En ce qui concerne la production de la consonne constrictive pré-dorso alvéolaire [s], il a été montré que les muscles orofaciaux recrutés pour sa réalisation sont les mêmes que pour la voyelle étirée [i] (Abry *et al.*, 1980) : une activité importante du buccinateur et du risorius pour l'étirement des commissures labiales vers l'arrière, assistés du grand zygomatique.

Toutefois, même si la constrictive [s] n'est pas marquée du point de vue de l'arrondissement, on ne peut négliger une éventuelle modification de l'aire intérolabiale de par l'élévation de la mandibule. La remarque formulée par Abry & Perrier (1996) nous invite donc à rester prudent :

« Rappelons que l'aire intérolabiale peut être sujette aux variations produites par les recrutements consonantiques sur la mandibule et pas seulement par la commande vocalique sur les lèvres. De fait pour élaborer un modèle complet d'anticipation sur tout le geste d'arrondissement, il faudrait tenir compte de toute la coordination bilabio-mandibulaire, l'aire aux lèvres n'étant que la résultante de

cette coordination. Il reste qu'il n'est pas toujours facile de "soustraire" les composantes du mouvement qui sont dues en propre aux commandes consonantiques. » (Abry & Perrier, 1996, p. 70)

Considérant ces interrogations, nous avons élaboré un paradigme expérimental, comprenant l'addition d'une condition à mâchoire bloquée (avec un *bite-block*) afin d'éliminer l'influence d'éventuels gestes consonantiques mandibulaires sur le mouvement labial dans son déroulement pour la production de la voyelle arrondie [u]. Il nous apparaît en effet possible de tester uniquement l'action des lèvres en limitant les mouvements de la mandibule par cette contrainte expérimentale de *bite-block*.

Les sept items suivants : [iku] [ikku], [iksku], [ikstku], [ikstsku] et la transition minimale [iu] ont été intégrés à des séquences porteuses (dépourvues de sens) de type « Two keak cookeec » pour l'anglais (Tableau 5) ou « Deux kixes cookiquent » pour le franco-québécois (Tableau 4).

Séquences porteuses Déterminant-Nom-Verbe	Transitions étudiées V1 C _n V2	Nombre de consonnes (de 0 à 5)
Deux ki oukiquent	iu	0
Deux ki coukiquent	iku	1
Deux kikes coukiquent	ikku	2
Deux kixes coukiquent	iksku	3
Deux kixtes coukiquent	ikstku	4
Deux kixtes scoukiquent	ikstsku	5
Deux kixtes skikiquent	ikstski	5

Tableau 4 : Description des séquences sélectionnées pour le corpus d'arrondissement [iC_nu] chez les locuteurs franco-québécois.

Séquences porteuses Déterminant-Nom-Verbe	Transitions étudiées V1 C _n V2	Nombre de consonnes (de 0 à 5)
Two kea ookeek	iu	0
Two kea cookeek	iku	1
Two keak cookeek	ikku	2
Two keaks cookeek	iksku	3
Two keakst cookeek	ikstku	4
Two keakst skookeek	ikstsku	5
Two keakst skeekeek	ikstski	5

Tableau 5 : Description des 7 séquences élaborées pour le corpus d'arrondissement [iC_nu] chez les locuteurs anglo-américains.

Nous avons choisi de ne pas doubler systématiquement l'ensemble des séquences avec la voyelle [i] en position vocalique finale, afin de ne pas trop alourdir le corpus. En revanche, nous avons ajouté, comme contrôle, l'énoncé [ikstski] comportant la chaîne consonantique maximale (5 consonnes). Lors d'études précédentes explorant le geste d'arrondissement en français (Abry & Lallouache, 1991 ; 1995a ; Abry *et al.*, 1996a et b), cette précaution n'avait pas été prise pour des raisons de taille du corpus. Or, la possibilité de gestes labiaux inhérents aux consonnes a été souvent débattue dans la littérature, un point que nous avons longuement présenté dans notre partie I. Il nous est donc apparu indispensable de tester la transition d'arrondissement [iC_nu] avec au moins une séquence contrôle comportant la transition vocalique de [i] à [i] afin de vérifier de possibles effets des consonnes.

Chaque séquence cible a été produite par l'ensemble de nos locuteurs entre 10 et 12 fois, ce nombre ayant varié en fonction de contraintes pratiques lors de l'enregistrement.

Bien que les séquences testées soient dénuées de sens, nous avons pris soin de préserver une structure syntaxique régulière, de type [Déterminant-Nom-Verbe] pour l'ensemble du corpus. Chaque locuteur a reçu la consigne de produire les séquences aussi naturellement que possible en respectant la syntaxe proposée. Nous avons demandé aux participants de ne pas réaliser de pause entre le nom et le verbe, afin d'éviter de maximiser les effets de jointure de mots sur le mouvement d'arrondissement. Nous leur avons également demandé d'adopter un rythme de parole régulier, auquel les sujets ont été entraînés avant enregistrement et qu'ils devaient garder autant que possible constant tout au long de la séance d'enregistrement.

II.1.3. Dispositif d'enregistrement des données audiovisuelles

Nous avons procédé à une seule session d'enregistrement pour chaque sujet. Celle-ci était divisée en deux parties, correspondant respectivement aux deux conditions expérimentales : condition *bite-block* vs. sans *bite-block*. La durée totale de la session, incluant le temps de préparation ainsi que l'enregistrement des deux corpus, n'excédait pas une heure.

Quelques jours avant l'enregistrement, chaque locuteur a reçu un document dans lequel nous lui indiquions les objectifs principaux de l'expérience et le descriptif du corpus.

Chaque enregistrement s'est déroulé selon trois phases. En premier lieu, les locuteurs prenaient connaissance des objectifs de l'étude et remplissaient un formulaire de consentement (un exemple est proposé en Annexe VII). Après notification des consignes, une phase d'entraînement leur était proposée afin que chaque sujet se familiarise avec l'ensemble

du corpus et le rythme de parole à adopter. La dernière phase consistait en l'enregistrement du corpus. A la fin de l'étude, chaque sujet recevait une compensation financière⁵ en remerciement de sa participation.

Lors de l'enregistrement, les séquences étaient projetées en ordre aléatoire pour chaque sujet sur un écran d'ordinateur portable placé devant le locuteur⁶ (Figure 42). La projection successive des séquences était contrôlée manuellement par l'un des expérimentateurs et non par le sujet lui-même pour éviter tout mouvement de sa part. En outre, le réglage manuel de la projection permettait au locuteur de produire successivement chacune des séquences affichées sur l'écran lorsqu'il le souhaitait. Enfin, il est à noter que nous avons opté pour un affichage graphique des séquences plutôt qu'une lecture préalable à haute voix (comme cela sera le cas pour le recueil des productions enfantines) – en particulier dans le cas des locuteurs anglophones – afin de ne pas perturber, de par notre accent français, leurs productions. Nous souhaitons en effet que leurs productions soient le reflet de phrases produites avec une intonation la plus naturelle possible et propre à chaque locuteur.

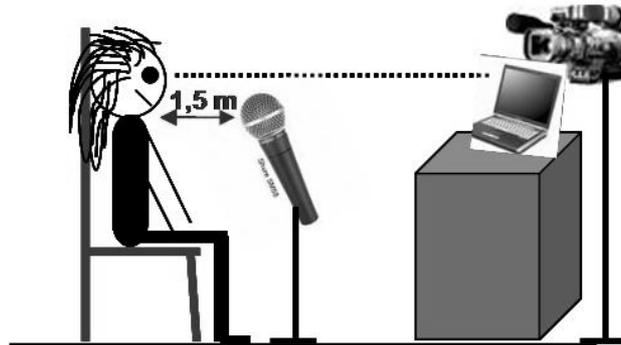


Figure 42 : Dispositif expérimental général pour l'enregistrement des adultes.

Comme indiqué plus haut, le corpus a été dupliqué de façon à obtenir deux conditions expérimentales : une condition d'enregistrement sans *bite-block* et une condition avec *bite-block*. De petits *bite-blocks* en plexiglas, ayant une épaisseur de 4 mm, ont ainsi été confectionnés au sein de notre laboratoire. Le *bite-block* était placé entre les molaires gauches du locuteur, au niveau de la deuxième ou troisième molaire en fonction du degré de confort

⁵ Cette compensation financière, ainsi que les coûts relatifs aux enregistrements, ont été financés dans le cadre des projets FQRSC (Fond Québécois de Recherche sur la Société et la Culture) et SSHRC (Conseil de Recherche en Sciences Humaines du Canada).

⁶ Les séquences ont été générées, via le logiciel *Random Slideshow*, qui permet de créer des présentations au format Power Point en ordre aléatoire.

ressenti. Avant l'enregistrement du corpus en condition *bite-block*, le locuteur bénéficiait d'une période de familiarisation, afin de ne pas ressentir de gêne lors de la production des phrases.

Chaque locuteur se tenait assis dans un fauteuil fixe et avait pour consigne de garder les pieds au sol et maintenir la tête immobile, contre le dossier du fauteuil. Pour la détection ultérieure de la constriction intérolabiale, nous avons maquillé les lèvres des locuteurs avec un bleu à lèvres résistant à l'eau. La pigmentation bleue étant naturellement absente du visage, son application uniforme sur les lèvres permet, après traitement, une détection homogène des contours labiaux.

Les préréglages réalisés au niveau de la caméra permettent de garantir la qualité de l'image pour les analyses ultérieures :

- réalisation de la balance des blancs avec un plan blanc placé devant le visage du locuteur ;
- création d'un quadrillage de référence pour la conversion des pixels en centimètres. Pour cela, un plan blanc quadrillé (carreaux de 1cm x 1cm) était placé devant le visage du locuteur et filmé durant un court instant par notre caméra. Ce quadrillage devait servir de référence pour la détection ultérieure des contours labiaux.

Les sujets étaient filmés avec une caméra⁷ positionnée face au visage du locuteur, en mode zoom pour une meilleure visibilité des lèvres. Les données vidéo collectées avec une fréquence de 30Hz (norme américaine, au format NTSC⁸), étaient stockées sur des minis cassettes DV (Digital Video cassettes). Le signal sonore était enregistré par un microphone⁹ placé à environ 1,5 mètres du locuteur, dont le signal était tout d'abord pré-amplifié avant d'être enregistré par la carte d'acquisition¹⁰ de la caméra. Ce son, synchrone à la vidéo, avait une fréquence d'échantillonnage de 48 kHz.

⁷ Les sujets ont été enregistrés avec le caméscope Panasonic AG-DVC30 disponible au laboratoire de Phonétique, UQAM, Montréal.

⁸ Le format d'acquisition et de lecture NTSC (National Television System Comitee) est une norme de codage de données vidéo. Elle correspond à un format de 29,97 images par secondes soit un total de 525 lignes avec une résolution d'images 640X480.

⁹ Nous avons utilisé un microphone à condensateur unidirectionnel, modèle SHURE SM-86.

¹⁰ Il s'agit d'une carte Canopus ADVC-100.

II.1.4. Acquisition des données de la protrusion de la lèvre supérieure

En ce qui concerne l'enregistrement de la protrusion de la lèvre supérieure, nous avons utilisé le système Optotrak (version Optotrak 3020) du *Motor Control Lab*. Il s'agit d'un système de détection optique, qui permet de suivre en temps réel le déplacement tridimensionnel de petites diodes infrarouges (IREDS) avec une résolution spatiale nominale de 0.1 mm.

Pour notre étude, nous avons choisi de positionner trois diodes sur la ligne médiane du vermillon de la lèvre supérieure et inférieure ainsi que sur la ligne médiane du menton (Figure 43). Chaque diode était en outre collée avec de l'adhésif double face. Lors de l'installation du dispositif, nous avons pris soin de ne pas traverser le contour interne des lèvres avec les fils reliant chaque diode au système central de l'Optotrak.



Figure 43 : Positionnement respectif des trois diodes (IREDS) pour la détection du déplacement en temps réel de la lèvre supérieure, inférieure et de la position du menton.

Outre ces trois senseurs, nous avons également utilisé un cadre de référence en plexiglas que nous fixions aux lunettes du locuteur (Figure 44). Ce cadre servait au positionnement de quatre diodes de références, utilisées par la suite pour la correction des mouvements de tête. D'autre part, nous avons mesuré l'orientation du plan occlusal, en demandant au locuteur de mordre, durant un court instant, un cadre en plexiglass sur lequel étaient fixées trois diodes (Figure 45). Cette procédure nous permettait par la suite de représenter les déplacements des trois *sensors* enregistrés par l'Optotrak, selon un axe en trois dimensions (x, y, z) correspondant au déplacement vertical, horizontal (axe antérieur/postérieur) et latéral. Avant le début de l'enregistrement, une phase de test était réalisée afin de s'assurer du fonctionnement des *sensors*.

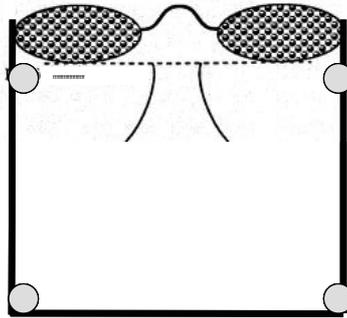


Figure 44 : Positionnement des quatre sensors (cercles gris) sur le cadre de référence.

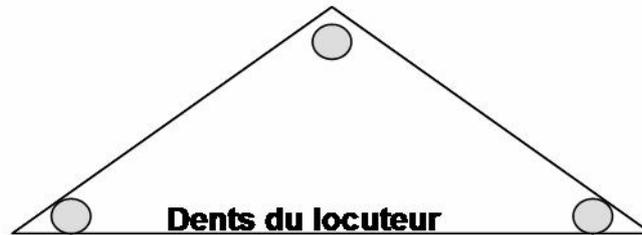


Figure 45 : Positionnement des trois sensors (cercles gris) sur le plan occlusal, servant à la correspondance des données par rapport aux trois dimensions verticale, horizontale et latéral x, y, z.

II.1.5. Prétraitements des signaux audiovisuels

Les données audiovisuelles recueillies ont été traitées à l'Institut de la Communication Parlée de Grenoble. Elles ont nécessité plusieurs traitements préliminaires avant de pouvoir être exploitables par le système de détection automatique des contours labiaux. Le corpus audiovisuel brut, stocké sur les minis cassettes DV, a tout d'abord été numérisé sous la forme de films au format AVI¹¹ (Audio Video Interleaved), puis converti en séries d'images bitmap à l'aide du logiciel Adobe Première Pro (version 6.5).

En ce qui concerne le traitement du signal acoustique (synchrone avec la vidéo), un étiquetage phonétique manuel a été réalisé à l'aide du logiciel de traitement de signaux de

¹¹ Le format AVI est un format de fichier conçu pour stocker des données audio et vidéo et les lire simultanément.

parole PRAAT¹². Cet étiquetage se présente sous la forme de TextGrids et a été élaboré selon trois niveaux hiérarchiques de segmentation.

- I. Un premier niveau nous permet d'étiqueter les deux bips repérés au début et à la fin de l'enregistrement du corpus. Ces derniers sont par la suite utilisés pour la resynchronisation des différents signaux provenant de l'Optotrak et de la caméra.
- II. Le second niveau hiérarchique correspond à la délimitation de chaque séquence produite par le locuteur sur l'ensemble du signal acoustique. Ces séquences sont délimitées par des « bornes » (événements) apposées manuellement sur le signal sonore et numérotées afin de faciliter leur regroupement ultérieur avec les séries d'images correspondantes.
- III. Enfin, un dernier niveau de segmentation est consacré à l'étiquetage précis des transitions cibles. Ce dernier niveau est essentiel pour notre étude puisqu'il permet de définir la nature de l'intervalle d'obstruence (IO) entre les deux voyelles ainsi que sa durée. Notons que nous n'avons pas segmenté précisément les consonnes à l'intérieur de l'intervalle d'obstruence car ces informations n'étaient pas nécessaires pour nos analyses ultérieures. Concernant la segmentation des voyelles [i] et [u], nous avons pris comme référence le début et la fin de la structure formantique sur le spectrogramme pour délimiter manuellement les bornes de début et fin de voyelle.

Nous avons utilisé plusieurs scripts PRAAT afin de diviser le signal sonore global en petits extraits sonores correspondant à chaque séquence, accompagné de son TextGrid et des images qui lui sont associées. Celles-ci sont ensuite classées manuellement dans des dossiers portant le nom et le numéro de la séquence en question.

II.1.6. Détection et extraction des trajectoires de l'aire aux lèvres

A la suite de ces manipulations, nous procédons à l'extraction automatique du paramètre d'aire aux lèvres avec le logiciel TACLE (Traitement Automatique des Contours des Lèvres, Audouy, 2000) développé à l'ICP.

Ce système permet de détecter automatiquement différents paramètres du contour interne des lèvres sur le visage vu de face : entre autres, la séparation verticale des lèvres (B), leur écartement horizontal (A) et enfin l'aire intérolabiale (S) (Figure 46).

¹² PRAAT est un logiciel d'analyse et de transcription phonétique élaboré par Paul Boersma et David Weenink.

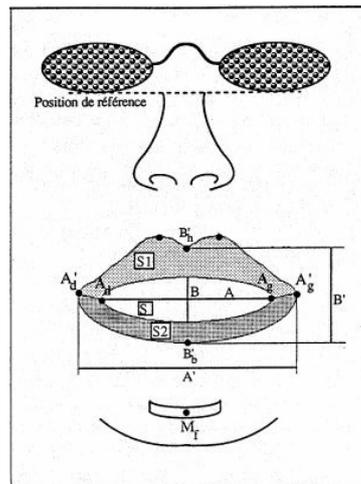


Figure 46 : Représentation des différents paramètres labiaux calculés par le logiciel TACLE à partir d'une trame de face. D'après Audouy, 2000.

Pour chaque locuteur, il s'agit tout d'abord de fixer différentes valeurs de paramètres de traitement des images nécessaires au suivi efficace des contours labiaux. La procédure de traitement est la suivante.

- I. Réalisation de la calibration des images. Cette procédure nous permet d'obtenir la conversion des pixels en centimètres (sur l'axe horizontal et vertical) à l'aide du plan quadrillé filmé par la caméra avant le début de l'enregistrement. Ces paramètres de calibration sont sauvegardés dans un fichier de calibration (.cal) et appliqués à l'ensemble du corpus audiovisuel pour un locuteur.
- II. Définition des paramètres de traitement pour l'application du seuillage par Chroma-key numérique (technique décrite ci-après). Ceux-ci correspondent au réglage des paramètres de teinte, luminance et saturation (norme TSL) et sont sauvegardés dans un fichier de paramètres (.par)
- III. Définition de la fenêtre de traitement dans laquelle s'effectue le seuillage par Chroma-key. Pour la recherche des contours, il est également nécessaire de définir un point de départ en cliquant avec la souris.
- IV. Créer un fichier Report (fichier écrit au format texte) pour chaque séquence vidéo dans lequel sont indiqués les *Time Codes* de chaque image (exprimée en heures : minutes : secondes : images) ainsi que le répertoire et le nom de chaque image constituant la séquence en question. Cette étape peut être réalisée manuellement ou avec une fonction Matlab qui crée automatiquement les fichiers report correspondant à la séquence.

Exemple de ligne écrite dans le fichier Report pour la séquence [ikku] :

00:00:00:01 D:\Aude\Ann_Ufra_bb_f15\Ann_Ufra_bb_f15_ikku1\Ann_Ufra_bb_f15_0657.bmp

00:00:00:02 D:\Aude\Ann_Ufra_bb_f15\Ann_Ufra_bb_f15_ikku1\Ann_Ufra_bb_f15_0658.bmp

V. Réalisation d'un fichier Batch (fichier texte) récapitulant les informations suivantes :

- Le répertoire et le nom des fichiers report pour chaque séquence
Exemple : *D:\Aude\Luc_Ufra_bb_f17\Luc_Ufra_bb_f17_ikku6\ikku6.txt*
- Le répertoire et le nom des fichiers paramètres pour chaque séquence
Exemple : *D:\Aude\Luc_Ufra_bb_f17\Luc_Ufra_bb_f17_ikku6\ikku6.par*
- Le répertoire et le nom du fichier de calibration
Exemple : *D:\Aude\Luc_Ufra_bb_f17\Luc.cal*
- Le répertoire et le nom du fichier de résultats généré par le logiciel TACLE pour chaque séquence une fois le traitement par Chroma-key et la détection automatique des contours labiaux effectués.
Exemple : *D:\Aude\Luc_Ufra_bb_f17\res_tac\ikku6.tac*

Lors de la détection automatique des paramètres labiaux, un processus de désentrelacement des images est tout d'abord réalisé pour chaque séquence (fréquence initiale de 30Hz). Cette étape consiste à extraire les trames paires et impaires d'une image, pour ensuite reconstituer deux nouvelles images séparées de 16 millisecondes. Les lignes manquantes sont alors recalculées par interpolation. Nous obtenons ainsi, à partir d'une image (durant 40 ms), deux trames A et B (de 20 ms chacune).

Le logiciel TACLE applique ensuite un traitement par Chroma-key numérique sur l'ensemble des images sélectionnées. Plus précisément, il s'agit d'appliquer un seuillage en noir sur les zones de pixels bleus repérées sur chaque image, dans notre cas, les lèvres des sujets que nous avons au préalable maquillées en bleu (Figure 47). Le réglage du seuillage est défini manuellement pour chaque locuteur de manière à obtenir une détection optimale des contours labiaux.

Les images sont ensuite lissées à l'aide d'un filtre médian qui supprime les éventuels pixels noirs isolés pouvant interférer avec le contour des lèvres. Le logiciel détecte les contours des zones noires résultantes (Figure 47), calcule les axes d'inertie et les points de mesure intermédiaires, réalise des approximations paraboliques pour l'extraction finale des paramètres labiométriques.

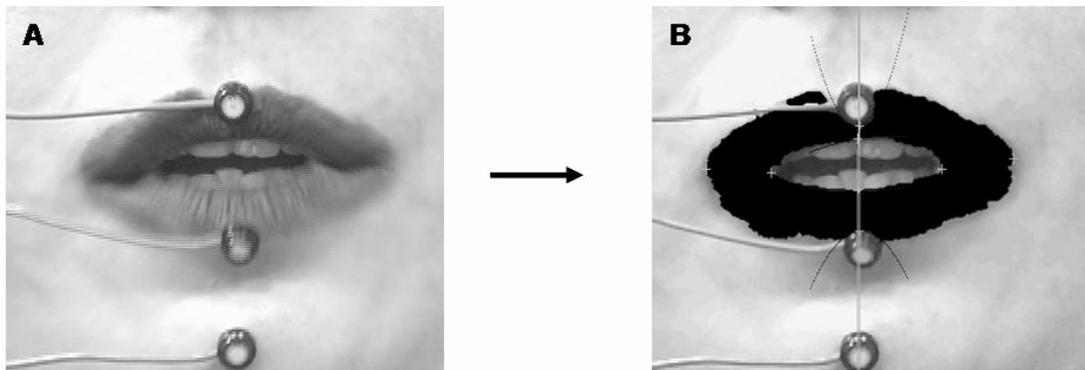


Figure 47 : image A : Exemple d'une image brute acquise lors de l'enregistrement vidéo. Image B : image obtenue après traitements : seuillage des zones bleues par traitement Chroma-key. L'aire intérolabiale sera donc calculée à partir du contour interne des lèvres.

La détection automatique de ces paramètres est effectuée pour chaque trame. Les valeurs ainsi obtenues sont exprimées en centimètres carrés dans le cas de l'aire intérolabiale. Ces valeurs sont sauvegardées dans un fichier (.tac) lisible avec le logiciel Excel et peuvent être visualisables, soit sous la forme d'une courbe avec l'interface TRAP (*TRAI*tement des signaux de Parole développé à l'ICP par H. Loevenbrück et Ch. Savariaux), soit avec l'application ATAC (*Affichage des fichiers .tac* également développé par Ch. Savariaux) permettant d'afficher simultanément et en synchronie les signaux acoustique et articulatoires.

Lors de la visualisation du décours temporel des données labiales, nous avons parfois repéré des sauts de détection, se manifestant par une valeur nulle d'aire intérolabiale par défaut. Ces erreurs de détection sont généralement dues à de petites variations de la position de la tête du locuteur qui entraîne un décalage de la zone à traiter hors de la fenêtre de traitement. Lorsque le logiciel Tacle décelait de telles erreurs, il était possible d'ajuster les valeurs de seuillage choisies pour le traitement Chroma-key et de réinitialiser la détection. Dans notre étude, certaines erreurs étaient dues à des contraintes expérimentales plus complexes et qui feront l'objet ci-après d'une description détaillée.

Afin de nous assurer qu'il n'y ait pas de petites erreurs de détection – parfois non visibles sur l'ensemble du signal d'aire aux lèvres mais néanmoins importantes pour nos interprétations futures –, nous avons préféré réaliser une vérification *manuelle* de toutes les séquences. Cet examen minutieux de tous les signaux, bien que fastidieux, nous a garanti une qualité optimale de l'ensemble des signaux analysés.

II.1.7. Extraction des trajectoires de protrusion de la lèvre supérieure

L'enregistrement Optotrak nous a permis d'obtenir deux types de signaux : cinématiques et sonores. Etant de nature différente, ces signaux ne pouvaient être analysés dans leur forme brute. Un certain nombre de prétraitements ont donc été nécessaires, en particulier concernant la synchronisation des signaux.

D'une part, nous avons recueilli les déplacements des lèvres inférieure, supérieure et de la position du menton selon un axe tridimensionnel x , y , z correspondant au déplacement horizontal, vertical et latéral des trois senseurs. Ces trajectoires ont été enregistrées par l'interface Ndi¹³ de l'Optotrak et converties au format ASCII de manière à ce que les données brutes de l'Optotrak soient analysable avec les programmes Matlab. Les signaux de protrusion ont été lissés en utilisant un filtre passe-bas¹⁴ de 7 Hz de fréquence de coupure.

Par ailleurs, un enregistrement sonore du corpus a été obtenu via la carte son de l'Optotrak (SoundMax Digital Audio). Bien que synchrone avec les trajectoires des *sensors*, ce signal sonore avait une fréquence d'échantillonnage relativement basse (10 kHz). Par conséquent, nous nous sommes référés au son enregistré par la caméra – de meilleure qualité – pour l'étiquetage phonétique du corpus. Nous avons néanmoins conservé le signal sonore de l'Optotrak pour la synchronisation des données audiovisuelles et cinématiques de protrusion.

En effet, le signal sonore enregistré par la caméra n'étant pas synchrone avec celui de l'Optotrak et des trajectoires correspondantes, nous avons utilisé des algorithmes disponibles avec Matlab pour la resynchronisation des différents signaux. A l'aide d'une fonction automatique PRAAT, nous avons tout d'abord converti le signal acoustique de la caméra pour obtenir une fréquence d'échantillonnage identique à celle du son issu de l'Optotrak (10 kHz). Puis, nous avons utilisé une fonction de corrélation croisée (*cross-correlation*) nous permettant de mesurer la similarité entre les deux signaux acoustiques. Cet algorithme compare les échantillons dans une fenêtre temporelle prédéfinie par l'utilisateur et détermine la différence d'échantillons entre les deux signaux. Cette différence est ensuite comblée manuellement en ajoutant ou enlevant une portion de silence au début du signal sonore de l'Optotrak.

¹³ Ndi est le sigle de la compagnie distribuant le système Optotrak et son interface de travail. Il s'agit de la Northern Digital Incorporate.

¹⁴ Il s'agit d'un filtre appelé « Tchebychev » qui permet de supprimer sur le signal acoustique les fréquences supérieures à 7Hz.

Une fois les signaux acoustiques et cinématiques synchrones, nous avons procédé au découpage du signal global de protrusion de l'Optotrak en petites trajectoires correspondant aux séquences étudiées. Cette procédure a été réalisée de manière automatique, avec la même fonction PRAAT utilisée auparavant lors du découpage du fichier sonore global en petites séquences.

A la suite de ces diverses manipulations, il nous était possible, pour une séquence, d'associer le signal acoustique et sa description phonétique (Textgrid) ainsi que le mouvement de protrusion de la lèvre supérieure correspondant.

II.1.8. Analyse des données articulatoires

L'analyse des signaux de protrusion de la lèvre supérieure et de l'aire intérolabiale de chaque séquence a été réalisée avec l'interface TRAP. Celle-ci permet, entre autres, de visualiser et analyser les paramètres labiaux générés par le logiciel TACLE ainsi que tout autres mesures au format ASCII. En outre, il est possible de calculer d'autres valeurs relatives aux paramètres labiaux testés, tels que la vitesse des mouvements enregistrés (première dérivée) ou encore l'accélération (seconde dérivée). Nous avons volontairement écarté les mesures d'accélération en raison de la grande instabilité de ces événements révélée par Perkell & Matthies (1992) et Abry & Lallouache (1995a). Nous utiliserons pour finir d'autres événements plus robustes que les passages par zéro de la vitesse (*cf.* plus loin).

II.1.8.1. Fluctuation des mouvements d'arrondissement : comparaison des séquences [ikstski] vs. [ikstsku] pour la protrusion

Comme nous l'avons mentionné au début de ce chapitre, nous avons souhaité recueillir des séquences contrôles [iC_ni] afin de nous assurer de l'absence de mouvements de protrusion caractérisés au cours de la réalisation des consonnes intervocaliques. Bien que nous ayons choisi des consonnes dites non labiales, il était important, par rapport à la littérature que nous avons discuté sur l'effet de la labialité consonantique (partie I), de vérifier que les mouvements d'arrondissement ne soient pas maximalement influencés par les perturbations dues à la mandibule par la production des consonnes.

Nous proposons, dans la Figure 48, un exemple de séquence [ikstski] produite par un locuteur franco-québécois, pour lequel nous avons observé une dynamique de protrusion assez importante. Notons que l'évolution de la protrusion est affichée en valeurs inversées pour faciliter leur comparaison ultérieure avec les signaux de constriction (aire aux lèvres).

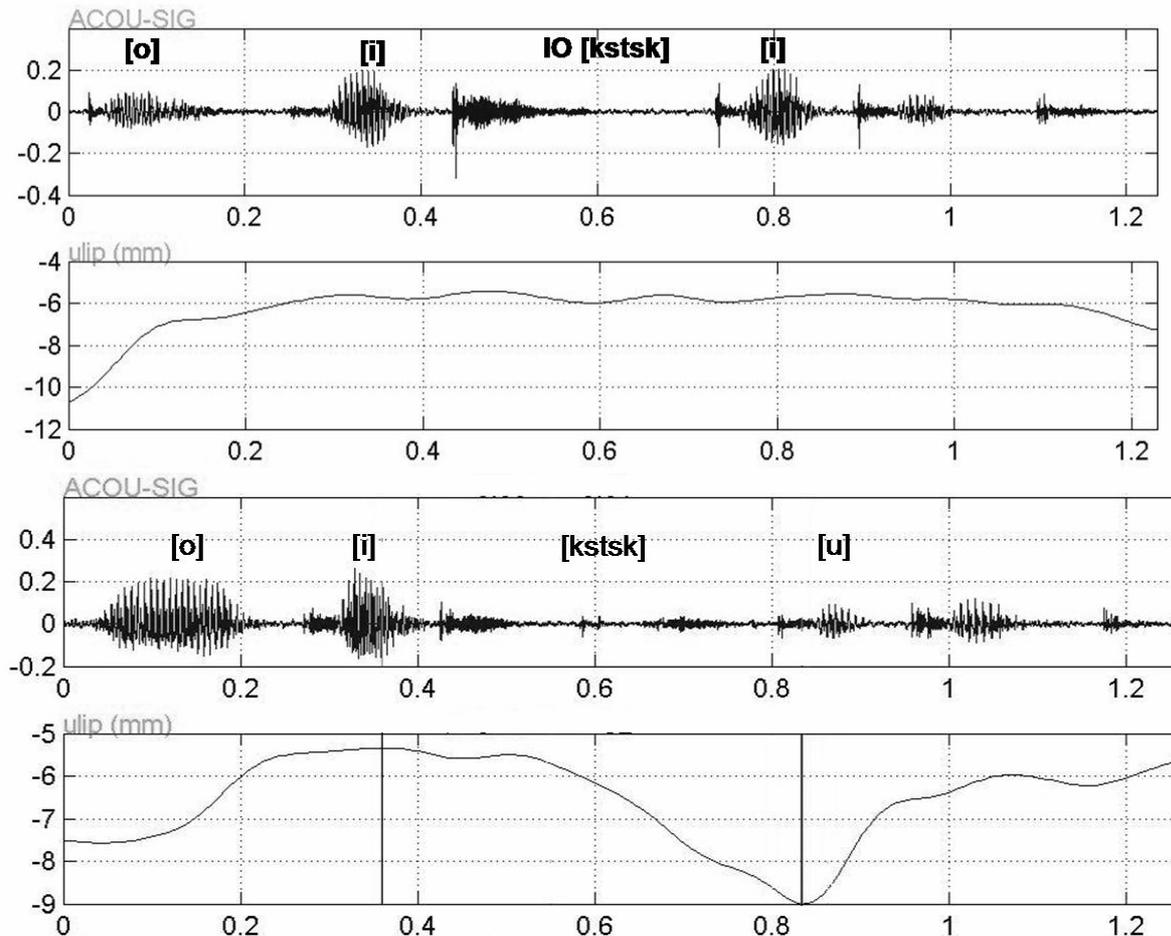


Figure 48 : Signal acoustique et tracé de la protrusion de la lèvre supérieure (en mm) pour les énoncés [ikstski] (figure du haut) et [ikstsku] (figure du bas). Les repères de minimum et maximum de protrusion sont indiqués par une ligne pleine pour l'énoncé [ikstsku]. Les signaux de protrusion sont inversés.

Pour comparaison, nous présentons une séquence [iC_nu] comportant le même groupe de consonnes intervocaliques, commençant toutes deux, comme c'est prévu dans le corpus, par une voyelle arrondie avec un maximum de protrusion (visible à gauche au début du signal avec la convention d'inversion). Pour la séquence [ikstski], on peut remarquer l'absence évidente de dynamique labiale. Les valeurs de protrusion fluctuent sur moins d'1 mm depuis le début du premier [i] jusqu'à la fin de la séquence. Notons que le seuil d'amplitude de 2 mm, que nous nous sommes fixés en début d'étude pour prendre en considération nos mouvements d'arrondissement, est d'autant plus pertinent lorsque l'on observe ces petites variations de protrusion tout le long des séquences. Pour la séquence [ikstsku], on observe clairement une augmentation de la protrusion en se dirigeant vers le [u] sur une dynamique d'environ 4,5 mm.

Un seuil que nous n'utiliserons pas pour l'arrondissement, les petites variations d'aire pouvant être très significatives étant donné la non-linéarité de la relation articulatoire-acoustique entre l'aire et les formants. Dans les deux cas, nous choisirons des événements robustes n'utilisant les passages par zéro que pour caractériser les gammes d'amplitude.

II.1.8.2. Événements cinématiques et phases déterminés sur le signal d'aire aux lèvres

Différents événements et phases ont été repérés et étiquetés sur les signaux articulatoires pour tester le *Modèle d'Expansion du Mouvement (MEM)*. Certains ont été sélectionnés manuellement par le biais de l'interface TRAP, d'autres, selon une procédure que l'on peut qualifier de semi-automatique.

La Figure 49 fournit un aperçu de ces différents événements et phases repérés sur le signal acoustique et sur le signal articulatoire d'aire intérolabiale pour la séquence « Deux kixes coukquent », produite par un locuteur franco-qubécois en condition de *bite-block*.

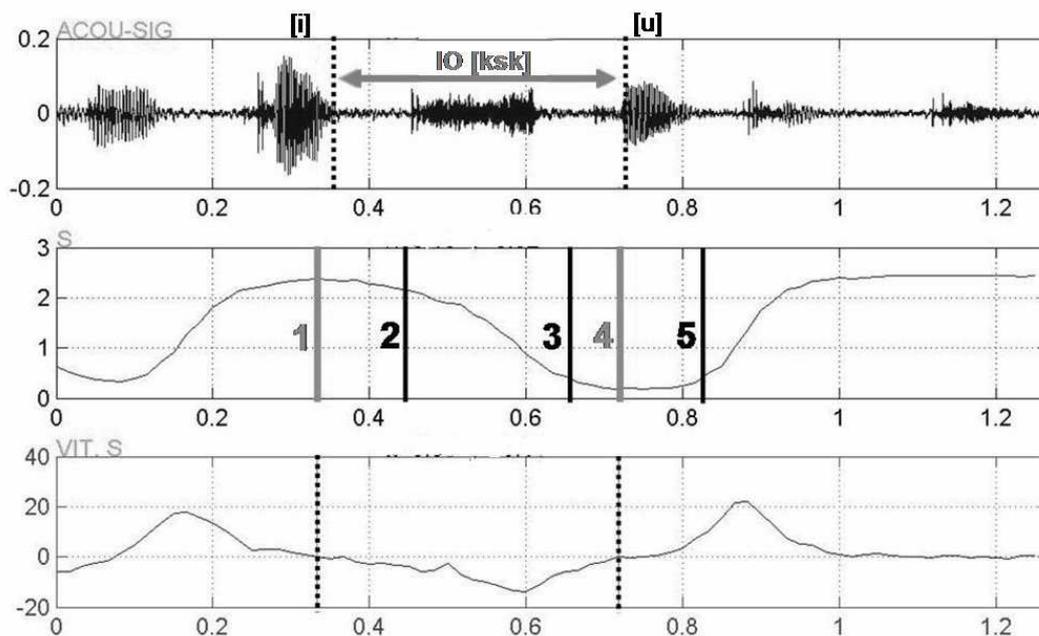


Figure 49 : Visualisation du signal acoustique (en haut), tracés de l'aire intérolabiale en cm^2 (au milieu) et de sa vitesse en cm/s (en bas). L'intervalle d'obstruence (IO) est délimité sur le signal acoustique par les lignes pointillées. Les étiquettes correspondent à 1 : l'aire maximale de [i] ; 2 : 10% de la différence d'aire entre [i] et [u] ; 3 : 90% de la différence d'aire entre [i] et [u] ; 4 : aire minimale de [u] ; 5 : 10% de la différence d'aire entre [i] et [u] suivant la valeur minimale de [u]. L'intervalle entre les événements 2 et 3 détermine la phase de *Time Falling*, celui entre les événements 3 et 5, la phase de *Hold*.

Sur le signal acoustique, nous avons défini deux événements importants pour la délimitation de l'intervalle d'obstruence (IO) : la fin de la voyelle [i] (caractérisée par la fin de la structure formantique vocalique) et le début de la voyelle [u] (caractérisé par le début de la structure formantique). Cet intervalle temporel (exprimé en secondes) varie en fonction du nombre et de la nature des consonnes insérées entre les deux voyelles. Il est extrait automatiquement à partir de l'étiquetage phonétique (TextGrid) réalisé au préalable avec le logiciel PRAAT.

Sur la fonction d'aire intérolabiale, les étiquettes correspondant respectivement à la valeur maximale d'aire pour la voyelle [i] (événement 1) ainsi qu'à la valeur minimale du [u] (événement 4) ont été apposées manuellement pour chaque séquence. Ces événements étaient systématiquement déterminés par un passage par zéro de la vitesse (lignes pointillées sur la dérivée vitesse servant à établir les événements 1 et 4 sur le signal d'aire aux lèvres).

A partir de ces deux événements repérés sur la fonction d'aire, il a été possible de calculer les deux phases proposées dans le modèle *MEM* pour rendre compte du comportement d'anticipation des locuteurs, c'est-à-dire le *Time Falling* (chute des valeurs de l'aire aux lèvres dans l'établissement de la constriction) et le *Hold* (phase de tenue du geste pour la voyelle).

Pour le calcul de ces deux phases – suivant en cela la procédure adoptée par Abry & Lallouache (1995b) pour l'étiquetage de la fonction temporelle de l'aire –, nous avons considéré trois événements robustes sur la fonction temporelle d'aire aux lèvres. Une fois déterminée l'amplitude du geste d'arrondissement par les valeurs maximale et minimale d'aire (aux passages par zéro de sa dérivée vitesse), ces événements (1 et 4), étant trop dépendants des fluctuations des max et min, ne sont pas retenus pour l'étude du *timing* d'anticipation. Par contre, la date correspondant à 10% de l'écart entre la valeur maximale et minimale d'aire (événement 2) est un repère robuste qui indique que le geste d'arrondissement labial vers la voyelle [u] a démarré dans une proportion sensible.

Nous avons également considéré le temps correspondant à 90% de cet écart (événement 3), date où la cible de petite aire est suffisamment approchée. Enfin, le temps correspondant à une remontée de 10% de la valeur de l'aire, qui suit l'aire minimale (événement 5), est le signe que la phase de petite aire aux lèvres du [u] est déjà en bonne partie abandonnée pour aller vers le [i]. La phase de *Time Falling* (exprimée en ms) représente donc l'intervalle temporel délimité entre les événements 2 et 3. Elle caractérise en effet une phase où l'aire intérolabiale diminue clairement pour établir la constriction d'arrondissement. Puis la phase

de maintien d'une aire suffisamment petite pour cet arrondissement de la voyelle, *Hold*, est délimitée entre les événements 3 et 5.

Le calcul de ces deux phases a été réalisé de manière semi-automatique avec le logiciel TRAP. En effet, pour chaque séquence, nous avons délimité manuellement la zone dans laquelle les trois repères décrits ci-dessus devaient être détectés. Le calcul final de ces valeurs ainsi que leurs temps respectifs se faisait alors de manière automatique.

II.1.8.3. Événements cinématiques et phases déterminés sur le signal de protrusion

En ce qui concerne la détection des événements et phases sur le signal de protrusion, la même procédure que pour la constriction a été employée. On a déjà noté que, sur l'interface de visualisation (Figure 50), l'évolution de la lèvre supérieure est présentée en échelle inversée, dans le simple but de permettre une meilleure comparaison avec les fonctions de constriction : ainsi une valeur négative élevée correspond à une forte protrusion de la lèvre.

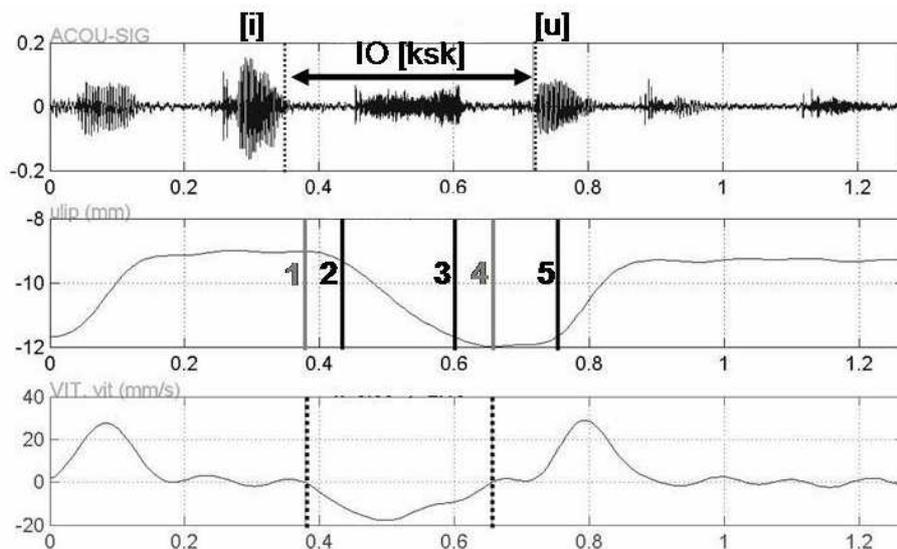


Figure 50 : Visualisation du signal acoustique (en haut), tracés de la protrusion de la lèvre supérieure en mm (au milieu ; en valeurs inversées) et de sa vitesse en mm/s (en bas). L'intervalle d'obstruence (IO) est délimité sur le signal acoustique par les lignes pointillées. Les étiquettes correspondent à 1 : la protrusion minimale pour [i] ; 2: 10% de la différence de protrusion entre [i] et [u] ; 3: 90% de la différence de protrusion entre [i] et [u] ; 4: la protrusion maximale pour [u] ; 5: 10% de la différence de protrusion entre [i] et [u] suivant la valeur maximale de [u]. L'intervalle entre les événements 2 et 3 détermine la phase de *Time Falling*, celui entre les événements 3 et 5 la phase de *Hold*.

Nous avons détecté, pour chaque séquence, l'instant de protrusion minimale pour [i] (événement 1) et l'instant de protrusion maximale pour [u] (événement 4). Comme pour la constriction, ces événements ne sont pas retenus dans l'analyse du *timing*, mais ils servent simplement à déterminer l'amplitude du geste, laquelle sert par la suite à repérer les

événements robustes. C'est en l'occurrence l'instant correspondant à 10% de la différence de protrusion entre le minimum de protrusion de [i] et le maximum pour [u] (événement 2), et celui correspondant à 90% de la différence de protrusion entre le minimum de protrusion de [i] et le maximum pour [u] (événement 3). Enfin l'instant du 10% de la différence de protrusion entre [u] et [i] qui suit la valeur maximale de [u] (événement 5). Le calcul de l'intervalle entre les événements 2 et 3 a permis de déterminer la phase de *Time Falling* (soit l'établissement de la protrusion, *Falling* convenant ici simplement à la convention d'inversion) et celui entre les événements 3 et 5, la phase de *Hold* (ou tenue).

II.1.9. Aléas techniques rencontrés au cours de l'étude

Toute étude expérimentale, quelle qu'elle soit, rencontre un certain nombre de difficultés techniques, méthodologiques ou encore humaines avec lesquelles l'expérimentateur doit composer. Aussi, lors de notre étude, nous avons rencontré une de ces difficultés avec l'un de nos locuteurs anglais (SHU) dont le débit de parole s'est avéré beaucoup plus rapide que l'ensemble de nos autres locuteurs. Nous avons évidemment tenté de rectifier ce problème en lui demandant de ralentir sa vitesse d'élocution. Mais il n'est pas parvenu à maintenir un débit constant tout au long du corpus. Nous avons donc obtenu un enregistrement très hétérogène (composé alternativement de séquences en débit normal et débit rapide). Les données labiométriques ainsi obtenues pour ce locuteur étaient donc difficilement comparables avec les autres locuteurs. En outre, nous avons remarqué que pour ses séquences comportant 4 à 5 consonnes, nous avons des difficultés à distinguer perceptivement les consonnes en question (ce qui s'est confirmé *a posteriori* lors de l'étiquetage phonétique du corpus). Nous avons donc décidé d'écarter l'ensemble des données de ce locuteur de nos analyses.

Au cours de l'acquisition des données de notre corpus, nous avons rencontré un certain nombre de difficultés liées au couplage des deux systèmes d'extraction de données articulatoires que sont l'Optotrak et le système de détection des contours labiaux (*Lip-Shape-Tracker*). En effet, ce dernier nécessite un dispositif expérimental particulier dont certaines contraintes ne sont pas toujours faciles à mettre en œuvre. En premier lieu, une détection optimale des contours labiaux requiert une lumière vive, de type halogène ou lumière du jour, afin de minimiser les zones d'ombres sur le visage. Le système Optotrak fonctionnant à l'aide de diodes infrarouges (IREDS), il n'est donc pas recommandé d'utiliser une lumière artificielle trop vive car celle-ci pourrait altérer le suivi du mouvement des diodes.

Lorsque nous avons réalisé un pré-enregistrement pour nous assurer du bon fonctionnement simultané des deux systèmes, nous n'avons rencontré aucun problème lié à la lumière : le suivi du mouvement des *sensors* ainsi que celui de l'aire intérolabiale s'est fait correctement. Or, lors de l'analyse des données finales, nous avons parfois rencontré des erreurs de détection des contours labiaux. Nous suggérons que ces erreurs soient dues à des conditions de luminosité différentes entre les pré-tests et la session d'enregistrement finale (la lumière du jour entrante devait être moins vive que lors des pré-enregistrements).

Ces difficultés ont néanmoins pu être résolues en appliquant un filtre¹⁵ aux séquences concernées avec le logiciel Adobe Première et en réinitialisant la détection automatique des contours labiaux. Lorsque les erreurs ne concernaient que quelques images, il était possible de réaliser des corrections manuelles avec un logiciel de retouche d'images tel que Paint Shop Pro. Malheureusement, pour l'un de nos locuteurs franco-québécois (JOH), les images obtenues après application d'un filtre n'étaient toujours pas satisfaisantes. Nous avons dû également écarter les données de ce locuteur.

Enfin, nous avons été confronté à des erreurs de détection lorsque le bleu du maquillage « bavait » sur les dents de certains locuteurs. Dans ce dernier cas, nous avons éliminé les séquences de nos analyses.

Les divers aléas rencontrés dans cette étude, que ce soit au cours de l'enregistrement ou du traitement ultérieur des données cinématiques, ont aussi conduit à la sélection d'un nombre variable de séquences chez chaque locuteur (nous en tiendrons évidemment compte dans les analyses statistiques).

II.2. L'extension du MEM aux adultes franco-québécois et anglo-américains

Dans une première section, nous rendrons compte des comportements d'anticipation du geste de protrusion de la lèvre supérieure pour les deux communautés linguistiques. Dans une seconde section, nous nous intéresserons au geste de constriction en suivant l'évolution de l'aire aux lèvres.

¹⁵ L'application d'un tel filtre correspond à un réglage manuel des propriétés de lumière, de saturation et de proportion de bleu contenues dans une image. Une fois que l'on a spécifié ces différents paramètres en fonction du but recherché, ces derniers sont appliqués automatiquement sous la forme d'un filtre sur l'ensemble du corpus par le logiciel Adobe Première Pro.

II.2.1. L'anticipation du mouvement de protrusion

Lors de l'analyse des signaux de protrusion, nous avons observé chez certains de nos locuteurs, en particulier les sujets anglo-américains, une amplitude de protrusion très faible. Comme l'avaient déjà souligné Perkell & Matthies (1992) dans leur étude sur l'anticipation d'arrondissement, les anglophones semblent en effet utiliser préférentiellement la dimension d'aperture plutôt que la protrusion pour le contraste d'arrondissement vocalique.

« *We observed informally that about half of our potential subjects pronounced /u/-containing utterances with minimal lip protrusion, instead using lip closure to accomplish the same acoustic-phonetic result.* » (Perkell & Matthies, 1992, p. 2914)

Bien que la précision nominale de l'Optotrak soit de 0.1 mm lorsque les mesures sont réalisées sur un support rigide et lisse, nous avons choisi de ne considérer que les signaux de protrusion montrant une amplitude d'au moins 2 mm sur l'ensemble de la séquence [iC_nu], en tenant compte du fait que nos mesures sont appliquées sur la structure molle des lèvres. Ce seuil a été adopté pour l'ensemble de nos locuteurs, afin d'écarter de nos analyses des signaux éventuellement bruités, ne correspondant pas clairement au geste d'anticipation de l'arrondissement vocalique.

II.2.1.1. Locuteurs franco-québécois

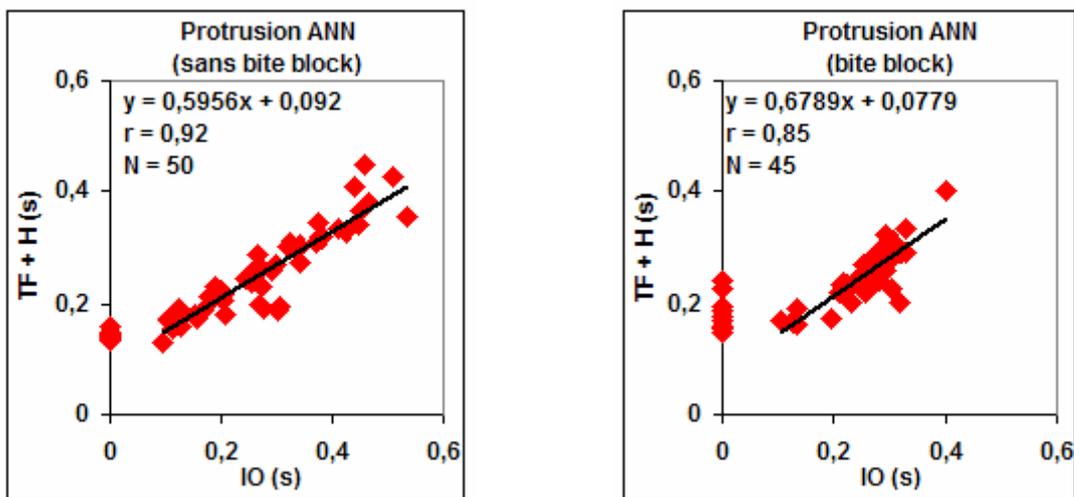
II.2.1.1.1. Présentation des données

Pour rendre compte du déroulé temporel de l'anticipation de protrusion vocalique de nos locuteurs, nous avons sélectionné les deux phases *Time Falling* (TF) et *Hold* (H) proposées par le modèle *MEM*. D'abord seulement pour le *MEM* de constriction (Abry *et al.*, 1996) : l'appliquer au *MEM* de protrusion comme au *MEM* de constriction est, ici, une première. Ces phases ont été mises en relation avec l'*Intervalle d'Obstruence* (IO). Le choix de ces deux phases, TF et H, sans aucune borne commune avec IO, nous permet de calculer l'expansion de l'anticipation chez nos locuteurs, et ceci sans nous confronter à l'artefact statistique d'une corrélation « tout-partie » avec IO, qui génère de fait des coefficients de corrélation artificiellement élevés (comme cela a été démontré par Benoît, 1986, un artefact qui se rencontre dans l'analyse des données de Perkell & Matthies, 1992).

Nous avons choisi de montrer l'évolution de ces deux phases par rapport à IO, et non par rapport au nombre de consonnes, comme c'est parfois le cas dans certaines études (Lubker & Gay, 1982, pour les adultes; Abelin *et al.*, 1980 pour les enfants). Ce type de lecture ne nous

paraît pas réellement pertinent pour notre analyse du décours temporel de l'arrondissement puisque l'augmentation du nombre de consonnes intervocaliques n'implique pas forcément un allongement de la durée de l'intervalle entre les deux voyelles ; celui-ci peut être d'une durée plus longue avec deux consonnes qu'il ne le sera avec trois. Nous avons donc préféré nous référer à l'*Intervalle d'Obstruence* qui témoigne plus fidèlement de l'intervalle de temps disponible entre les deux cibles vocaliques.

Les résultats seront présentés pour chacun des 4 sujets (Figure 52) dans chaque condition expérimentale (avec et sans *bite-block*). Une analyse des données individuelles par régression linéaire a été réalisée à chaque fois afin d'obtenir la pente et le coefficient de corrélation r de Bravais-Pearson. Les séquences [iu] étant caractérisées par un *Intervalle d'Obstruence* nul, les droites de régressions linéaires ont été calculées uniquement sur les séquences de [iC_nu] (n variant de 1 à 5 consonnes). Rappelons que les aléas techniques rencontrés au cours du traitement des données nous ont contraints à éliminer certaines séquences de nos analyses. Il en résulte donc une variabilité dans le nombre total des séquences analysées pour chaque locuteur, dont l'analyse statistique tient compte.



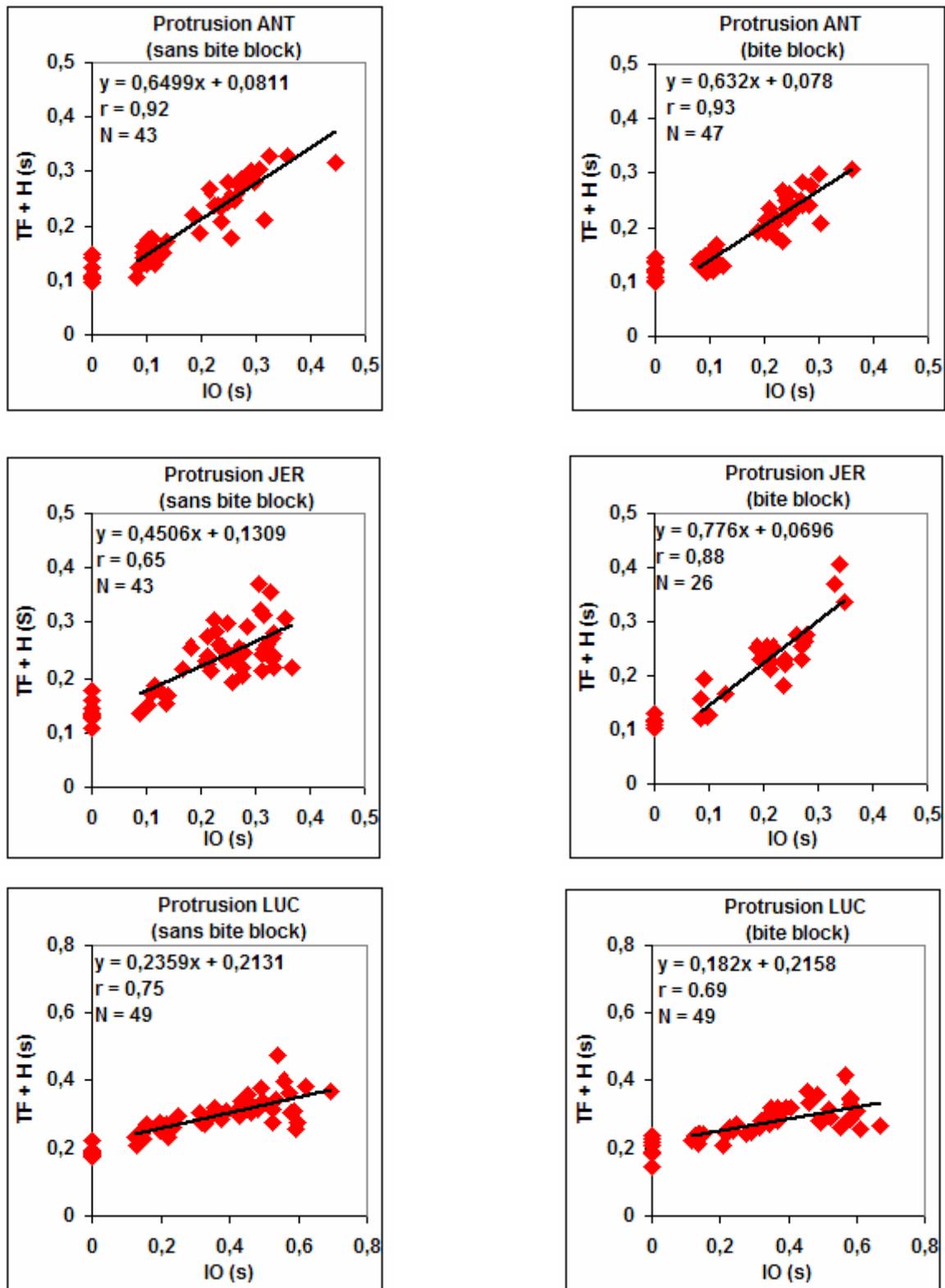


Figure 52 : Relation entre la durée du mouvement de protrusion (TF+H) et la durée de IO pour les 4 locuteurs franco-qubécois en condition *bite-block* (figures de gauche) et sans *bite-block* (figures de droite). Coefficients de corrélation r significatifs à $p < 0,00001$. N correspond au nombre de séquences analysées dans chaque condition.

Considérons tout d’abord les séquences du geste de base [iu]. Le Tableau 6 donne, pour chaque locuteur, les durées moyennes du mouvement d’arrondissement (TF+H) pour ces transitions sans consonne intervocalique, soit les *constantes d’exécution* ou durées minimales du geste. Ces durées sont variables d’un locuteur à l’autre (de 120 à 180 ms) avec un locuteur (LUC) montrant une durée de mouvement pour la transition [iu] bien plus élevée que les autres locuteurs (190-210 ms). Pour les séquences avec une consonne [iku], la durée du mouvement varie selon le locuteur de 140 à 180 pour 3 locuteurs et de 230 à 250 ms pour LUC. Malgré la variabilité interindividuelle, on peut remarquer que les durées moyennes de mouvement pour [iku] sont égales ou supérieures (jusqu’à 50 ms) à celles des séquences [iu].

Comparaison des constantes d’exécutions calculées pour les séquences [iu] avec les durées moyennes de mouvement (TF + H en ms) pour les séquences [iku] Locuteurs franco-québécois – Protrusion –																
Locuteurs	ANN				ANT				JER				LUC			
Type de séquences	[iu]		[iku]													
Condition expérimentale	ssbb	bb	ssbb	bb												
Durée de TF + H (ms)	140	180	170	180	120	120	140	140	140	120	170	150	190	210	240	230

Tableau 6 : Constantes d’exécutions calculées pour les séquences minimales [iu]. Comparaison avec les durées moyennes de mouvement (TF+H en ms) pour les séquences [iku] chez les 4 locuteurs franco-québécois en condition sans *bite-block* (ssbb) et *bite-block* (bb).

Ces données nous montrent qu’il est important de considérer les séquences [iu] à part vis-à-vis des autres séquences lors du calcul de la droite de régression. En effet, si l’on trace, sur les Figures 52, la fin de la droite de régression jusqu’à l’intercept avec l’axe des ordonnées y (correspondant à un *Intervalle d’Obstruence* nul), on obtient la plupart du temps des durées de mouvements plus petites que les durées réelles. Les séquences [iu] nous permettent donc de faire émerger une *constante d’exécution* spécifique à chaque locuteur correspondant à la durée minimale du mouvement de protrusion pour une séquence sans consonne intervocalique. Selon Abry & Lallouache (1995a), le calcul de cette durée minimale du mouvement est crucial pour une bonne prédiction de l’ensemble de l’anticipation d’arrondissement, depuis le cas d’une séquence n’offrant aucun laps de temps pour réaliser le geste d’arrondissement entre les deux cibles antagoniques, jusqu’aux séquences comportant un intervalle acoustique suffisamment large pour permettre à chacun de déployer son mouvement d’anticipation de manière préférentielle (p. 95). Les auteurs avaient mis en évidence pour la transition [iy] des

durées moyennes de 140 ms stables chez l'ensemble de leurs sujets, durées très comparables à celles que nous trouvons pour 3 de nos sujets (en condition sans *bite block*).

A partir de plus d'une consonne dans [iC_nu], le mouvement d'arrondissement s'étend selon un coefficient d'expansion spécifique à chaque sujet. Pour l'ensemble de nos sujets, nous avons obtenu des coefficients de corrélations statistiquement significatifs à $p < 0,00001$, quelle que soit la condition expérimentale. Nous commenterons, dans une prochaine section, les différences éventuellement observées entre les deux conditions expérimentales avec et sans *bite-block*. Trois des quatre locuteurs (ANN, ANT et JER) présentent des coefficients d'expansion du mouvement qui sont loin d'être faibles (avec une pente entre 0,45 et 0,78), traduisant le fait que lorsque l'*Intervalle d'Obstruence* consonantique augmente, ces locuteurs ajustent leur temps de mouvement, en l'augmentant régulièrement (*lawfully*), ce que montrent les *r* élevés (de 0,65 à 0,93).

Nous pouvons visualiser, sur la Figure 53, cette expansion du mouvement en fonction de la longueur de l'intervalle consonantique en comparant une réalisation de la séquence « Deux kikes coukiquent » et une réalisation de la séquence « Deux kixtes scoukiquent » par le locuteur ANN (condition sans *bite block*). Dans le premier cas, le démarrage du geste s'effectue au cours de la voyelle [i] et le maximum de protrusion est atteint bien avant le début acoustique de la voyelle arrondie [u]. Dans la seconde séquence avec un intervalle consonantique long, le geste, qui a démarré aussi durant la voyelle non-arrondie, s'étire davantage tout en se terminant aussi avant la voyelle [u].

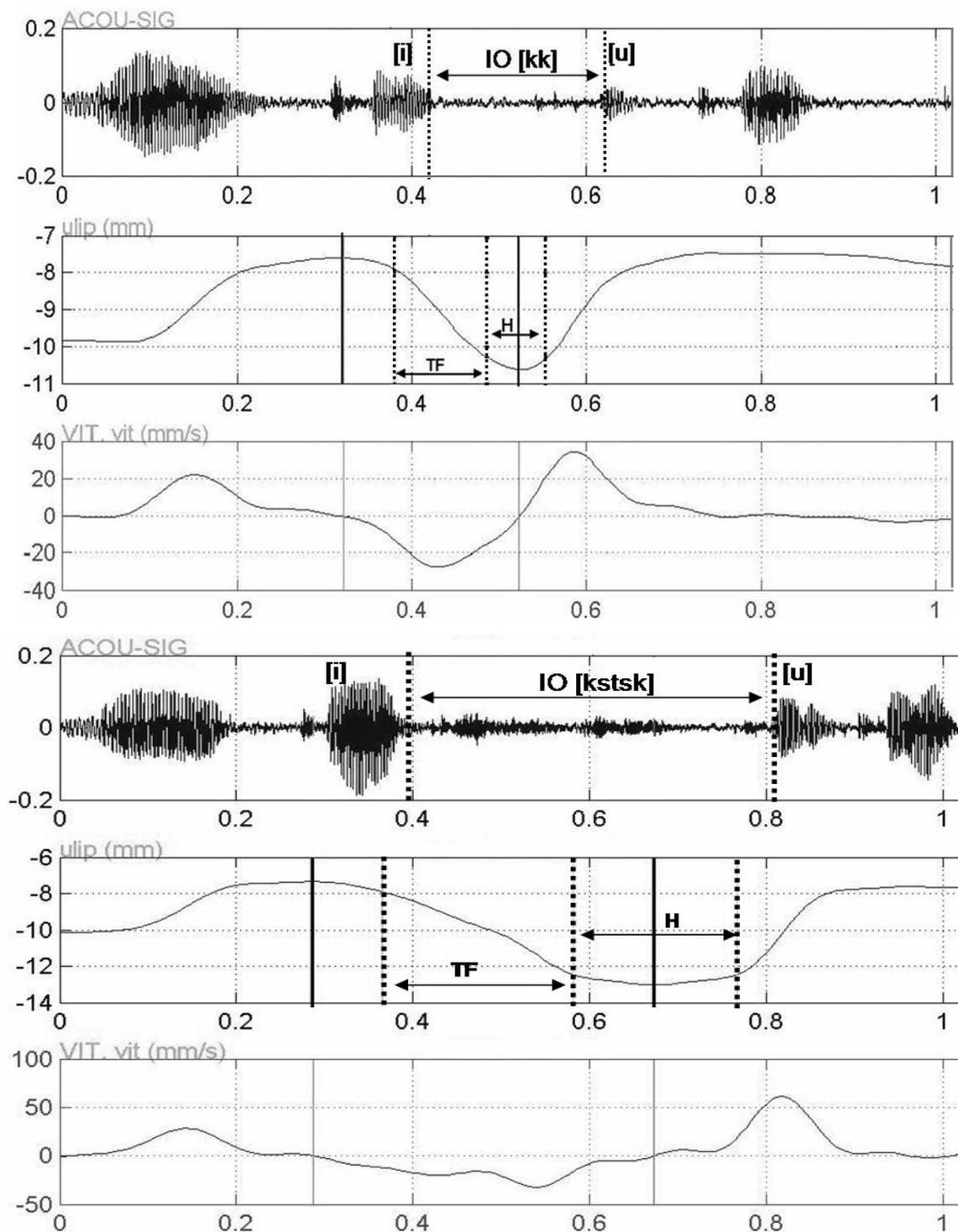


Figure 53 : Signal acoustique et évolution temporelle de la protrusion et de sa dérivée vitesse, en condition sans *bite-block*, pour une séquence « Deux kikes coukiquent » (figure du haut) et une séquence « Deux kixtes scoukiquent » (figure du bas) produites par le locuteur franco-québécois ANN. Les signaux de protrusion sont inversés. Sur le signal acoustique, la fin de la voyelle [i] et le début de [u] sont indiqués par des lignes verticales pointillées. Sur le signal de protrusion, le minimum et le maximum de protrusion sont indiqués par des lignes verticales continues et les phases *Time Falling* (TF) et *Hold* (H) sont situées entre les lignes pointillées.

II.2.1.1.2. Un comportement *Time-locked* chez un des locuteurs ?

Un de nos sujets (LUC) présente un profil d'anticipation d'arrondissement clairement différent de ceux des trois autres locuteurs avec une expansion faible du mouvement d'arrondissement qui pourrait sembler proche d'un profil *Time-locked*, avec une pente de 0,24 en condition sans *bite-block* et de 0,18 en condition avec *bite-block*. C'est celui pour lequel nous avons déjà remarqué des durées particulièrement longues pour le geste de base.

Peut-on davantage caractériser son profil de geste ? Nous l'illustrerons par la visualisation de deux séquences particulières, l'une avec une transition [iku] dans « Deux kikes coukiquent » (Figure 54, en haut) et l'autre avec une transition [ikstsku] dans « Deux kixtes scoukiquent » produites par ce locuteur (Figure 54, en bas).

La séquence [iku] comportant une seule consonne est caractérisée par un minimum de protrusion se produisant à l'intérieur de la voyelle [i], avec une phase de *Time Falling* débutant vers la fin acoustique de la voyelle. Le maximum de protrusion est atteint aux alentours du début de la voyelle [u]. C'est aussi ce que nous avons pu observer, pour ce même locuteur, pour les séquences [ikku] à deux consonnes. En revanche, lorsque IO grandit (à partir de 3 et jusqu'à 5 consonnes), le minimum de protrusion est détecté bien après la fin acoustique de la voyelle [i], comme c'est le cas dans la séquence « Deux kixtes scoukiquent » visualisée dans la Figure 54. La phase de *Time Falling* va, par conséquent, démarrer beaucoup plus tard, c'est-à-dire bien à l'intérieur de l'intervalle consonantique, au cours de l'occlusion du [t]. Ceci nous montre que, lorsque ce locuteur dispose d'un laps de temps relativement long pour initier son mouvement d'arrondissement, il maintient sa position d'étirement des commissures labiales au-delà de la voyelle [i] et amorce son geste tardivement. Au final, le fait de retarder l'établissement de la protrusion dans les séquences à plusieurs consonnes aboutit à un coefficient d'expansion très faible. Ce comportement est, rappelons-le, tout à fait un comportement *MEM* avec une paramétrisation du coefficient d'expansion propre au locuteur. Mais si l'on tenait absolument à le caractériser comme un exemple de *Time-locked*, on aurait, contrairement à ce qui se passe la plupart du temps pour le *Time-locked*, le moyen statistique de rejeter l'*Hypothèse Nulle*. Nous insisterons sur le fait que la question ne se pose pas davantage que pour les exemples *Look-ahead*, ce qui compte pour le *MEM* est un comportement *lawful*, validé par la statistique, la pente étant une affaire individuelle.

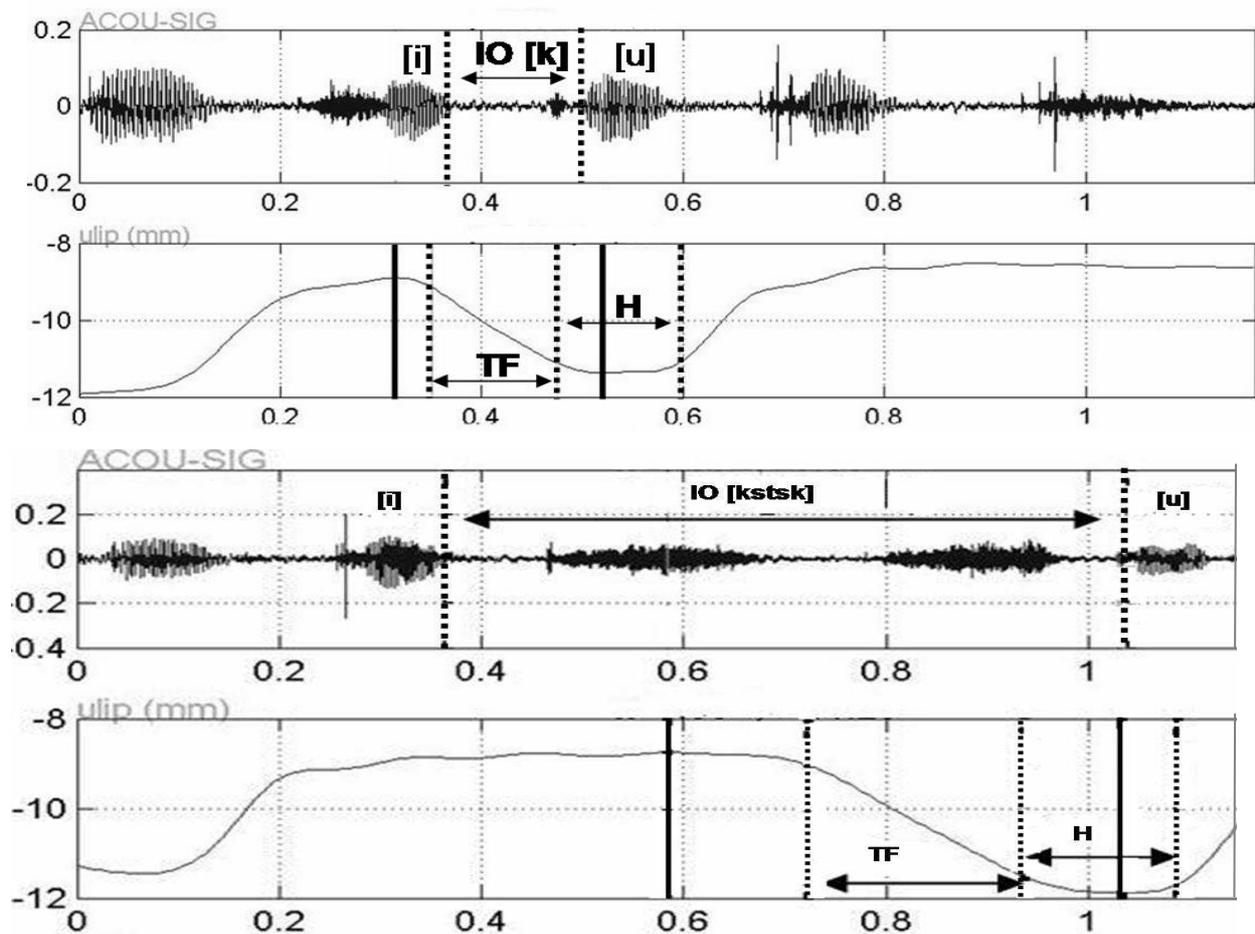


Figure 54 : Signal acoustique et évolution temporelle de la protrusion, en condition sans *bite-block*, pour une séquence « Deux ki coukiquent » (figure du haut) et une séquence « Deux kixtes scoukiquent » (figure du bas) produites par le locuteur franco-québécois LUC. Les signaux de protrusion sont inversés. Sur le signal acoustique, la fin de la voyelle [i] et le début de [u] sont indiqués par des lignes verticales pointillées. Sur le signal de protrusion, le minimum et le maximum de protrusion sont indiqués par des lignes verticales continues et les phases *Time Falling* (TF) et *Hold* (H) sont situées entre les lignes pointillées.

II.2.1.1.3. Comparaison des productions avec et sans *bite-block*

Les pentes obtenues pour la condition avec *bite-block* et sans *bite-block* sont, pour chaque sujet, relativement stables à l'exception du locuteur JER qui a une pente de 0,45 en condition sans *bite-block* alors qu'elle est de 0,78 en condition avec *bite-block*. Nous avons testé la possibilité d'une différence entre les pentes des deux conditions par un test t de Student. Ces analyses ont révélé, pour chaque locuteur, une différence non significative (ANN : $t=1,14$ $ddl=93$; ANT : $t=0,31$ $ddl=88$; LUC : $t=-1,32$ $ddl=96$) sauf pour le locuteur JER ($t=2,76$ $ddl=67$ $p<0,01$). On peut donc en déduire que les profils d'anticipation sont proches d'une condition à l'autre. Rappelons que l'usage du *bite-block* visait à éliminer l'influence du

mouvement de la mandibule lors de la réalisation des consonnes sur le geste d'arrondissement.

Comment expliquer la différence de pentes obtenue chez notre locuteur JER ? Nous suggérons que cette variabilité soit liée à une plus grande dispersion des actions d'arrondissement dans la condition moins contrainte, soit sans *bite-block*, dispersion possiblement due à l'excursion mandibulaire. Mais ce n'est sans doute pas la seule explication. Ce locuteur a en effet globalement montré une faible dynamique de protrusion pour de nombreux signaux, que nous avons en conséquence éliminés. Ce comportement est d'autant plus remarquable que cette dynamique faible est bien plus présente en condition *bite-block*. L'analyse par régression linéaire n'a donc porté que sur 26 signaux en *bite-block* contre 43 sans *bite-block*. Cette réduction d'amplitude a simplement réduit la dispersion ($r = 0,88$ contre $0,65$ en *bite-block*).

II.2.1.1.4. Le MEM de protrusion a sauté l'Atlantique vers la Belle Province

D'après les résultats obtenus en protrusion, nous pouvons apprécier la validité des prédictions du modèle MEM pour les quatre locuteurs franco-qubécois (cf. la Figure 55 résumant schématiquement les pentes observées). Rappelons que, selon ce modèle, chaque locuteur peut adopter son propre comportement d'expansion de l'anticipation. Le mouvement d'arrondissement peut être amorcé à l'intérieur de la voyelle [i] ou plus tard si l'*Intervalle d'Obstruence* entre les deux voyelles est grand. Ce mouvement n'est pas lié à la fin acoustique de la voyelle non arrondie, comme le proposait le modèle *Look-ahead*, pas plus qu'il n'est fixé par rapport au début de la cible arrondie (modèle *Time-locked*). Un des principes forts du modèle MEM est d'être paramétrable par locuteur, c'est-à-dire que l'on cherche à rendre compte de la variabilité intrinsèque à la parole, et plus particulièrement du phénomène de coarticulation anticipante en faisant émerger la propre régularité ou loi du locuteur (*speaker-specific lawful variability*). Dans cette perspective, la seule question qu'on peut se poser pour des locuteurs francophones de la Métropole ou de la Belle Province relèverait d'une possible répartition sociolinguistique de l'arrondissement. Or, à en juger par le sondage que nous permet le déploiement de moyens d'analyse articulatoire exceptionnels, les locuteurs qubécois s'insèrent parfaitement dans la variabilité rencontrée pour les Français, avec juste un locuteur bien différent au niveau de la durée de son geste d'arrondissement de base et de son comportement peu expansif en anticipation. On verra

d’ailleurs que ce comportement individuel rentrera dans la ligne des autres expansions pour son MEM de constriction.

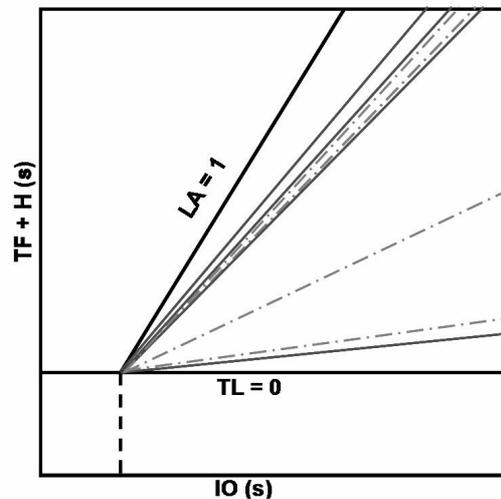


Figure 55 : Représentation schématique des pentes de protrusion des 4 locuteurs franco-québécois par rapport aux pentes suggérées par le modèle *Look-ahead* (pente de 1) et modèle *Time-locked* (pente de 0). Les lignes pleines correspondent aux pentes obtenues en condition *bite-block*, les lignes pointillées en condition sans *bite-block*.

II.2.1.2. Un locuteur anglo-américain protrudeur

En ce qui concerne nos 4 locuteurs anglo-américains, 3 d’entre eux (KAL, LAU, SAR) ont montré une amplitude de protrusion globalement inférieure à 2 mm. Ayant pu récupérer à peine une vingtaine de signaux atteignant 2 mm entre les deux conditions avec et sans *bite-block*, nous avons préféré ne pas réaliser d’analyse à partir d’un nombre aussi faible de données labiales valables. Pour tous les locuteurs francophones dont nous avons étudié le geste de protrusion, nous pouvons constater (Tableau 7) qu’ils ont une amplitude maximale de protrusion comprise entre 3,7 et 6,9 mm (rappelons que le plus grand protrudeur des trois locuteurs à protrusion d’Abry & Lallouache (1995a), Jean-Luc, avait une protrusion pouvant aller jusqu’à 8 mm). Le seul locuteur anglo-américain protrudeur que nous avons rencontré – contrairement à ses co-locuteurs qui confirment la réputation de faibles protrudeurs que leur avaient déjà fait Perkell & Matthies (1992) –, est DAN qui a une amplitude de mouvement maximale tout à fait honorable de 6,9 mm.

Amplitude maximale de protrusion (mm)									
Locuteurs franco-québécois								Locuteur Anglo-américain	
ANN		ANT		JER		LUC		DAN	
ssbb	bb	ssbb	bb	ssbb	bb	ssbb	bb	ssbb	bb
4,6	6,3	3,7	6,7	3,7	4,5	4,9	4,5	4,5	6,9

Tableau 7 : Amplitude de protrusion maximale (mm) pour les 4 locuteurs franco-québécois et le seul locuteur anglo-américain ayant montré une dynamique de protrusion supérieure à 2 mm. Condition sans *bite-block* (ssbb), condition *bite-block* (bb).

Nous présentons, dans la Figure 56, les données obtenues pour ce locuteur anglo-américain.

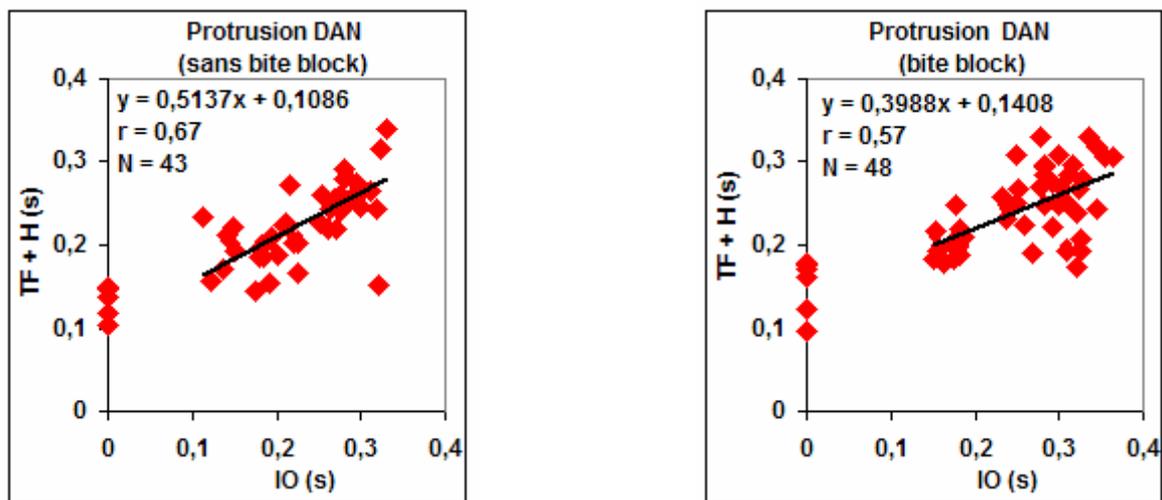


Figure 56 : Relation entre la durée totale du mouvement de protrusion (TF+H) et la durée de l'*Intervalle d'Obstruence* (IO) du seul locuteur anglo-américain DAN montrant une amplitude de protrusion d'au moins 2 mm. Condition sans *bite-block* (à gauche) et *bite-block* (à droite). Coefficients de corrélation r significatifs à $p < 0,0001$. N correspond au nombre de séquences considérées pour l'analyse par régression linéaire pour chaque session.

Comme pour les 4 adultes franco-québécois, son comportement d'arrondissement évolue de manière linéaire selon un coefficient d'expansion spécifique (pentes de 0,51 en condition sans *bite-block* et de 0,40 en condition *bite-block*), autour desquelles ses données sont bien regroupées, soit *lawful* ($r = 0,67$ et $0,57$, $p < 0,0001$; différence non significative $t = 0,94$ $ddl = 89$). Rappelons que si son coefficient de corrélation est proche du meilleur obtenu par Perkell & Matthies (1992; soit 0,59), ce dernier valait pour l'*Onset Interval* en corrélation tout-partie avec l'intervalle consonantique, c'est-à-dire qu'il n'était qu'artéfactuel. La constante d'exécution moyenne de notre locuteur (durée de mouvement pour les séquences

[iu]) est de 135 ms en condition sans *bite-block* et 150 ms en condition avec *bite-block* (pour les séquences [iku], la durée moyenne de mouvement étant de 200 ms en condition sans *bite-block* et 224 ms en condition avec *bite-block*).

Au vu des résultats recueillis auprès des 3 autres locuteurs anglo-américains, il semble donc bien que le paramètre de protrusion de la lèvre supérieure ne constitue pas un corrélat articulatoire révélateur en anglais pour la distinction voyelle arrondie vs. voyelle étirée. Les données de 3 de ces locuteurs corroborent cette hypothèse et viennent par ailleurs conforter les conclusions de Perkell et Matthies, mentionnées en introduction, ainsi que celle de Lubker & Gay (1982) :

« *Failure on the part of speakers of American English to protrude the upper lip during the production of rounded vowels must imply different production goals for them than for speakers of Swedish.* » (Lubker & Gay, 1982, p. 444)

Rappelons que nous avons également été confrontés à ce phénomène, pour l'un de nos locuteurs franco-qubécois (JER), pour lequel les signaux de protrusion avaient bien souvent une amplitude inférieure à 2 mm. Tout comme l'avaient déjà signalé Abry *et al.* (1996a), pour leur locuteur français 'Christophe', qui ne se servait pas non plus de la protrusion pour marquer la distinction entre [i] et [y]. Quoi qu'il en soit notre locuteur anglophone protrudeur DAN (Figure 57), se situe honorablement par rapport à la pente des locuteurs français les moins expansifs (Benny 0,54 et Jean-Luc 0,42, cf. Abry & Lallouache, 1995a). Les trois sont battus en minimalisme d'expansion par le comportement de notre locuteur franco-qubécois LUC (0,18).

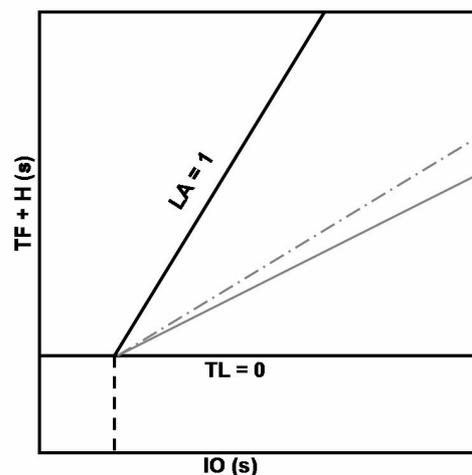


Figure 57 : Représentation schématique des pentes de protrusion du seul locuteur anglo-américain DAN ayant une dynamique labiale supérieure à 2 mm. Ses pentes sont exprimées par rapport aux pentes prédites par le modèle *Look-ahead* (pente de 1) et le modèle *Time-locked* (pente de 0). La ligne pleine correspond à la pente obtenue en condition *bite-block*, celle en pointillées en condition sans *bite-block*.

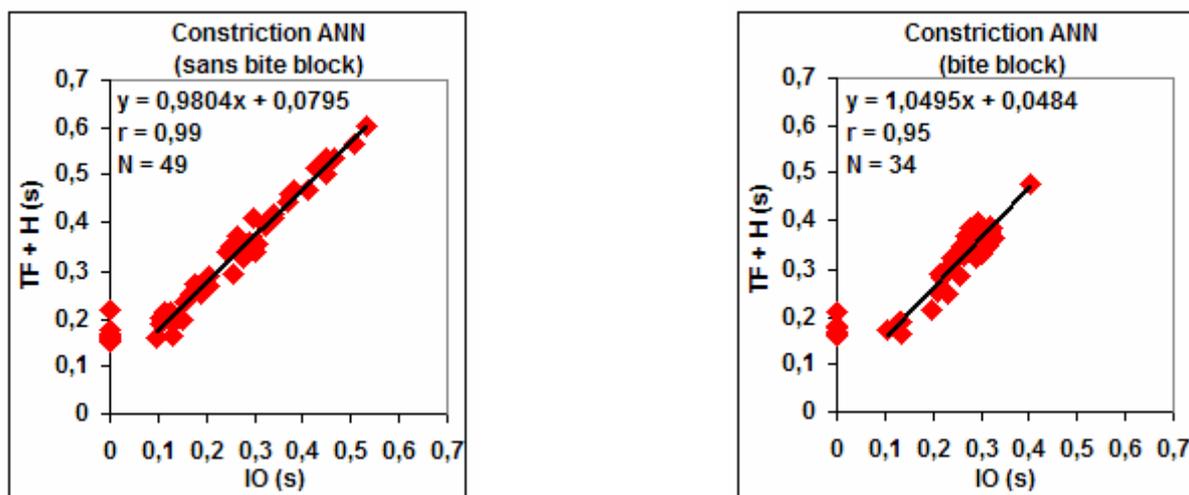
II.2.2. L'anticipation du mouvement de constriction labiale

Afin d'établir les profils généraux d'anticipation de constriction de nos locuteurs, nous avons suivi la même procédure que celle employée précédemment pour l'analyse de la protrusion. Nous avons analysé l'évolution de la durée totale du mouvement de constriction (représentée par les deux phases TF+H) par rapport à la durée de IO. De même l'analyse par régression linéaire a été réalisée sur toutes les séquences comportant des séquences [iC_nu], excluant les séquences minimales [iu] qui correspondent à une valeur nulle pour IO.

Nous présenterons dans l'ordre les patrons d'anticipation des locuteurs franco-québécois lors des deux conditions expérimentales (*bite-block* et *sans-bite-block*), puis ceux des anglo-américains.

II.2.2.1. Locuteurs franco-québécois

Nous présentons en Figure 58 les résultats de nos 4 locuteurs franco-québécois. L'analyse par régression linéaire a de nouveau révélé des comportements d'anticipation d'arrondissement idiosyncratiques. Le mouvement d'arrondissement suit une pente spécifique à chaque locuteur, allant de 0,85 à 0,99 en condition sans *bite-block* et de 0,85 à 1,05 en condition *bite-block*. Les données sont nettement regroupées autour des droites de régression puisque les coefficients de corrélation r sont supérieurs à 0,90 et significatifs à $p < 0,00001$ pour l'ensemble des sujets.



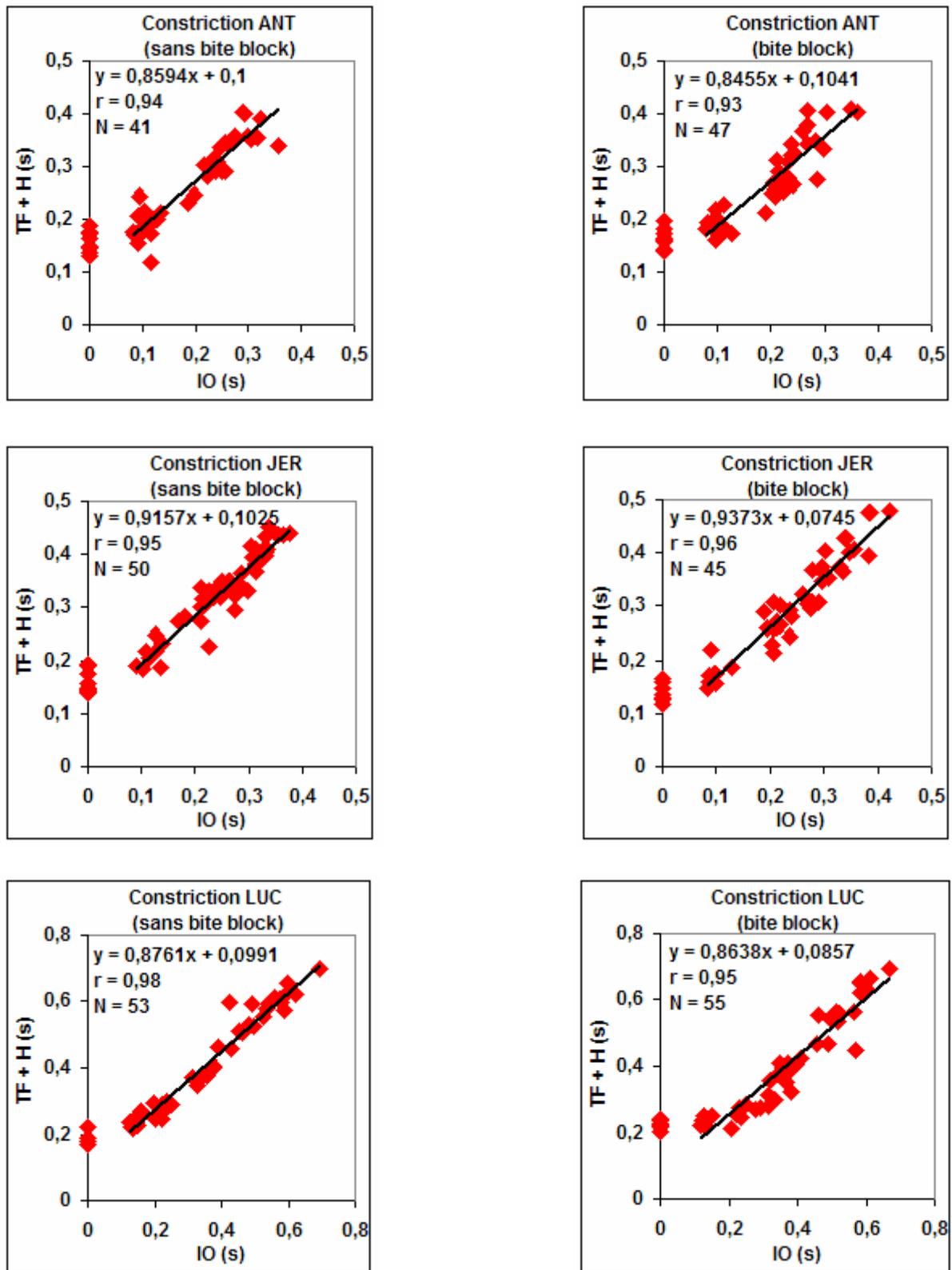


Figure 58 : Relation entre la durée totale du mouvement de constriction (TF+H) et IO chez 4 locuteurs franco-québécois en condition sans *bite-block* (figures de gauche) et *bite-block* (figures de droite). Coefficients de corrélation r significatifs à $p < 0,00001$. N correspond au nombre de séquences considérées pour l'analyse par régression linéaire pour chaque session.

Pour chaque locuteur, le test de comparaison de pentes n'indique pas de différence significative entre les deux conditions avec et sans *bite-block* (ANN : $t=1,20$ ddl=81 ; ANT : $t=-0,19$ ddl=86 ; JER : $t=0,36$ ddl=93 ; LUC : $t=-0,25$ ddl=106).

L'ensemble de nos sujets ayant prononcé les séquences [iu] sans réaliser de pause silencieuse, nous avons pu calculer la durée moyenne du mouvement d'arrondissement (TF+H) ou *constante d'exécution*, pour toutes ces transitions minimales. Le Tableau 8 rend compte des valeurs obtenues pour chaque locuteur. On note que celles-ci sont relativement proches pour 3 des 4 locuteurs testés (de 140 à 180 ms) et restent assez stables entre les deux conditions expérimentales. Le quatrième locuteur (LUC) présente, pour sa constriction comme il le montrait déjà en protrusion, les durées de mouvement pour les séquences [iu] les plus élevées : de 190 ms en condition sans *bite-block* à 230 ms en condition avec *bite-block*. En comparant avec les séquences [iku], on remarque que la durée moyenne de TF+H est, pour tous les locuteurs, toujours légèrement supérieure à celle obtenue pour la transition minimale [iu].

Comparaison des constantes d'exécutions calculées pour les séquences [iu] avec les durées moyennes de mouvement (TF + H en ms) pour les séquences [iku] Locuteurs franco-qubécois – Constriction –																
Locuteurs	ANN				ANT				JER				LUC			
Type de séquences	[iu]		[iku]													
Condition expérimentale	ssbb	bb	ssbb	bb												
Durée de TF + H (ms)	170	180	190	200	160	160	180	190	160	140	200	170	190	230	240	240

Tableau 8 : Constantes d'exécutions calculées pour les séquences minimales [iu]. Comparaison avec les durées moyennes de mouvement (TF+H en ms) pour les séquences [iku] chez les 4 locuteurs franco-qubécois en condition sans *bite-block* (ssbb) et *bite-block* (bb).

En ce qui concerne l'expansion du mouvement avec l'augmentation de l'*Intervalle d'Obstruence*, nous pouvons remarquer que tous nos locuteurs ont une pente assez élevée (de 0,85 à 1,05).

Nous détaillerons le comportement du locuteur ANN (Figure 59). En observant chacun des signaux de constriction recueillis pour ce locuteur, nous avons remarqué que le mouvement d'arrondissement était, dans la majorité des cas, amorcé près de la fin de la voyelle [i], quelle que soit la condition expérimentale (l'ajout du *bite-block* ne change pas le profil observé). Deux illustrations de son profil d'arrondissement sont proposées dans la Figure 59 ci-après, lors des séquences « Deux ki coukiquent » et « Deux kixes coukiquent ». Les deux lignes verticales continues sur le décours temporel de la constriction (signal du milieu) indiquent le maximum d'aire intérolabiale pour la voyelle [i] et le minimum d'aire

pour la voyelle arrondie [u] (déterminés par les passages par zéro de la vitesse). Les lignes verticales pointillées sur le même signal permettent de délimiter les phases de *Time Falling* et de *Hold*.

On remarque clairement que le maximum d'aire aux lèvres est localisé au tout début de la voyelle étirée et que le mouvement d'arrondissement (phase de *Time Falling*) débute dans les deux cas avant la fin acoustique de la voyelle. Dans les deux cas, le minimum d'aire aux lèvres est atteint au début de la voyelle [u] et le geste global (TF+H) est terminé à la fin de celle-ci. On remarque donc que lorsque l'*Intervalle d'Obstruence* augmente, la durée du geste global augmente, avec en particulier une expansion la plus importante sur la phase de *Time Falling*.

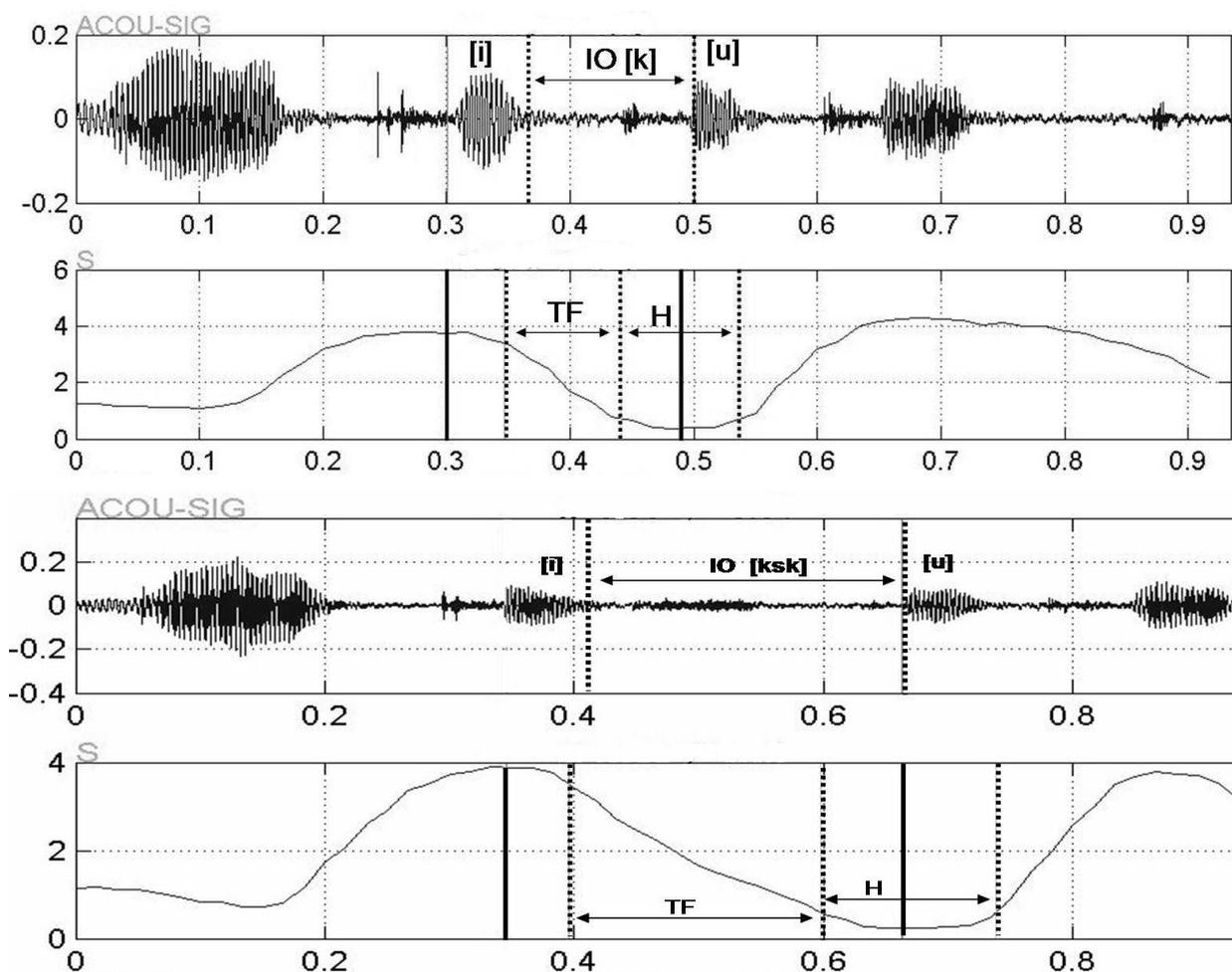


Figure 59 : Signal acoustique et évolution temporelle de l'aire intérolabiale (en cm^2), lors de la production, en condition *bite-block*, de la séquence « Deux ki coukiquent » (graphique du haut) et de la séquence « Deux kixes coukiquent » (graphique du bas), par le locuteur franco-québécois ANN. Sur le signal acoustique, la fin de la voyelle [i] et le début de [u] sont indiqués par des lignes pointillées. Sur le signal de constriction, le maximum et le minimum de constriction sont indiqués par des lignes verticales continues et les phases *Time Falling* (TF) et *Hold* (H) sont situées entre les lignes pointillées.

Afin d'avoir une vue d'ensemble de ce qui se passe chez ce locuteur, nous présenterons l'évolution de deux événements temporels pour toutes les séquences. La Figure 60 présente l'organisation temporelle de l'événement correspondant à 10% de la différence d'aire entre les valeurs maximale pour [i] et minimale pour [u] – en d'autres termes cet événement marque l'amorce du geste d'arrondissement ou début du *Time Falling* – en fonction de la durée de l'*Intervalle d'Obstruence* pour toutes les séquences produites par ANN en condition sans *bite-block* et condition *bite-block*. Nous avons également indiqué les événements de maximum de constriction (temps max. constr. [i] sur la figure) par rapport à IO pour comparaison. Ces événements sont exprimés en pourcentages de la durée de IO. La droite horizontale à 0% correspond à la fin acoustique de la voyelle [i], soit le début de IO.

Pour la majorité des séquences, le geste d'arrondissement est amorcé avant la fin de la voyelle [i]. Plus précisément, il existe une relation hyperbolique entre l'amorce du mouvement d'arrondissement et l'évolution de IO : les débuts de gestes ne sont donc pas alignés sur la fin de [i], car les points se trouvent majoritairement en-dessous de la ligne horizontale à 0% (Figure 60). Une relation hyperbolique similaire avait été mise en évidence par Abry & Lallouache (1995a) concernant leurs données de protrusion. Les auteurs avaient aussi observé, pour deux locuteurs, un démarrage du mouvement d'arrondissement très précoce, pouvant se produire dès 50% de la voyelle [i].

Ces résultats confortent une fois de plus les prédictions du modèle *MEM* quant à la possibilité d'extension du mouvement d'arrondissement. Le locuteur adapte son geste d'arrondissement par rapport à l'intervalle de temps disponible. Lorsque IO est petit, le mouvement est initié de manière très précoce, c'est-à-dire bien à l'intérieur de la voyelle [i], alors que, lorsque IO est grand (4 à 5 consonnes), le mouvement démarre à la fin de la voyelle, voire même un peu après. Ce phénomène est d'autant plus visible en condition sans *bite-block* où l'étendue temporelle des *Intervalles d'Obstruence* est plus importante. L'évolution des événements Maximum de constriction suit également une courbe en hyperbole. Lorsque IO est petit, le maximum de constriction peut se produire jusqu'à 60 % avant la fin de la voyelle étirée [i], alors que lorsque IO s'allonge, le maximum d'aire intérolabiale intervient plus tard (mais jamais après la fin acoustique de la voyelle [i]).

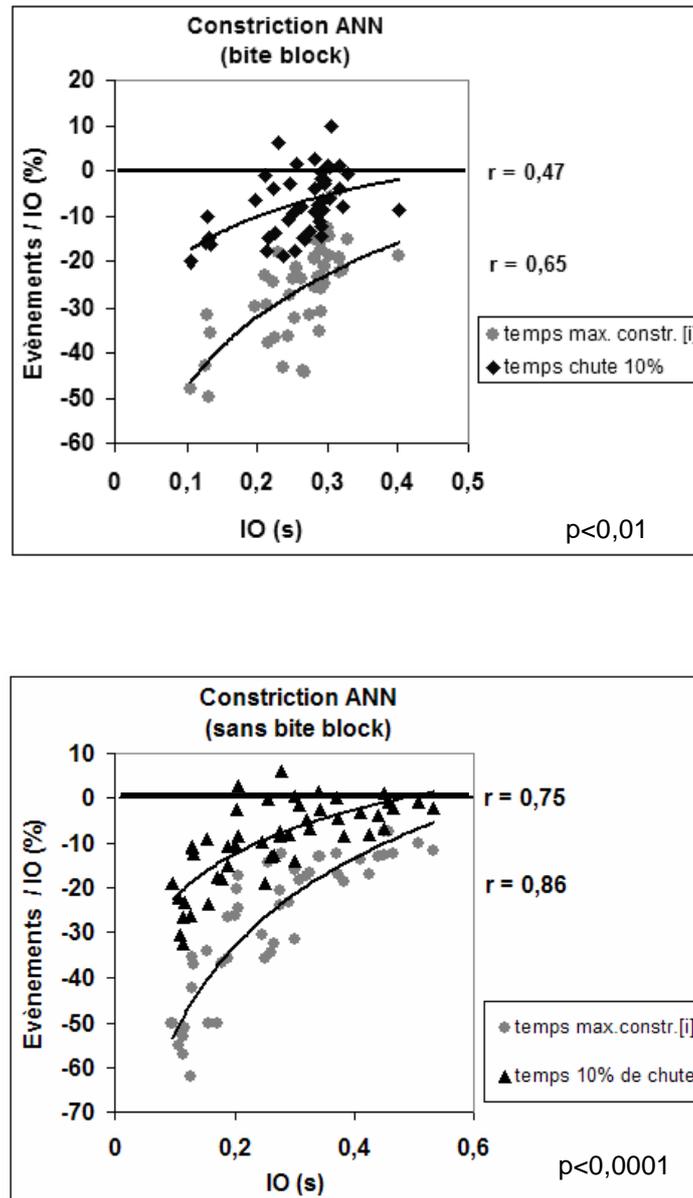


Figure 60 : Organisation temporelle de l'événement correspondant à 10% de la différence d'aire entre les valeurs maximale d'aire pour [i] et minimale pour [u] (triangles noirs) et le temps correspondant au maximum d'aire pour [i] (cercles gris) en condition *bite-block* (figure du haut) et sans *bite-block* (figure du bas). Ces événements sont exprimés en pourcentages de la durée de l'*Intervalle d'Obstruence* (s). La droite horizontale à 0% correspond à la fin acoustique de la voyelle [i] ou encore au début de IO.

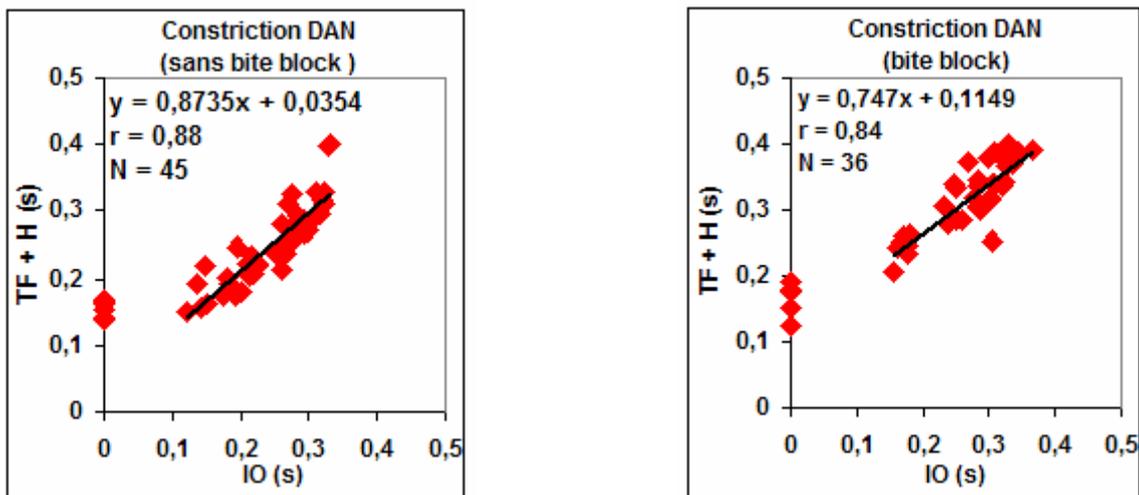
Quelle lecture pouvons-nous faire de ces signaux vis-à-vis du modèle de propagation des traits (modèle *Look-ahead*) ?

Au vu des résultats pour ce locuteur, il semble que le mouvement d'arrondissement ne soit pas contraint par les traits distinctifs attribués aux cibles précédentes. D'après le modèle *Look-ahead*, celui-ci devrait débuter à la rencontre de segments neutres au regard du trait d'arrondissement, sans interférer avec des buts immédiats (dans notre cas, la voyelle [i])

associée au trait phonologique [-rond]). Or, nous avons observé une majorité de signaux labiaux pour lesquels le mouvement d'arrondissement débutait dans la voyelle [i]. Chaque locuteur déploie son geste d'arrondissement de manière préférentielle sans que celui-ci soit déterminé par des règles phonologiques de propagation des traits. Il semblerait que cette similarité de la durée du mouvement avec la durée de IO ne soit pas systématique : le mouvement peut être initié avant la fin de la voyelle [i] suggérant que les articulateurs supra-laryngés sont très tôt en position pour amorcer le geste d'arrondissement. De même, le minimum de constriction peut également être atteint avant le début acoustique de la voyelle.

II.2.2.2. Locuteurs anglo-américains

A l'inverse de la protrusion, le paramètre de constriction s'est avéré tout à fait robuste pour rendre compte de l'anticipation d'arrondissement vocalique chez nos quatre sujets anglo-américains. Nous avons obtenu des patrons d'anticipation spécifiques à chaque locuteur, validés par de forts coefficients de corrélation (Figure 61).



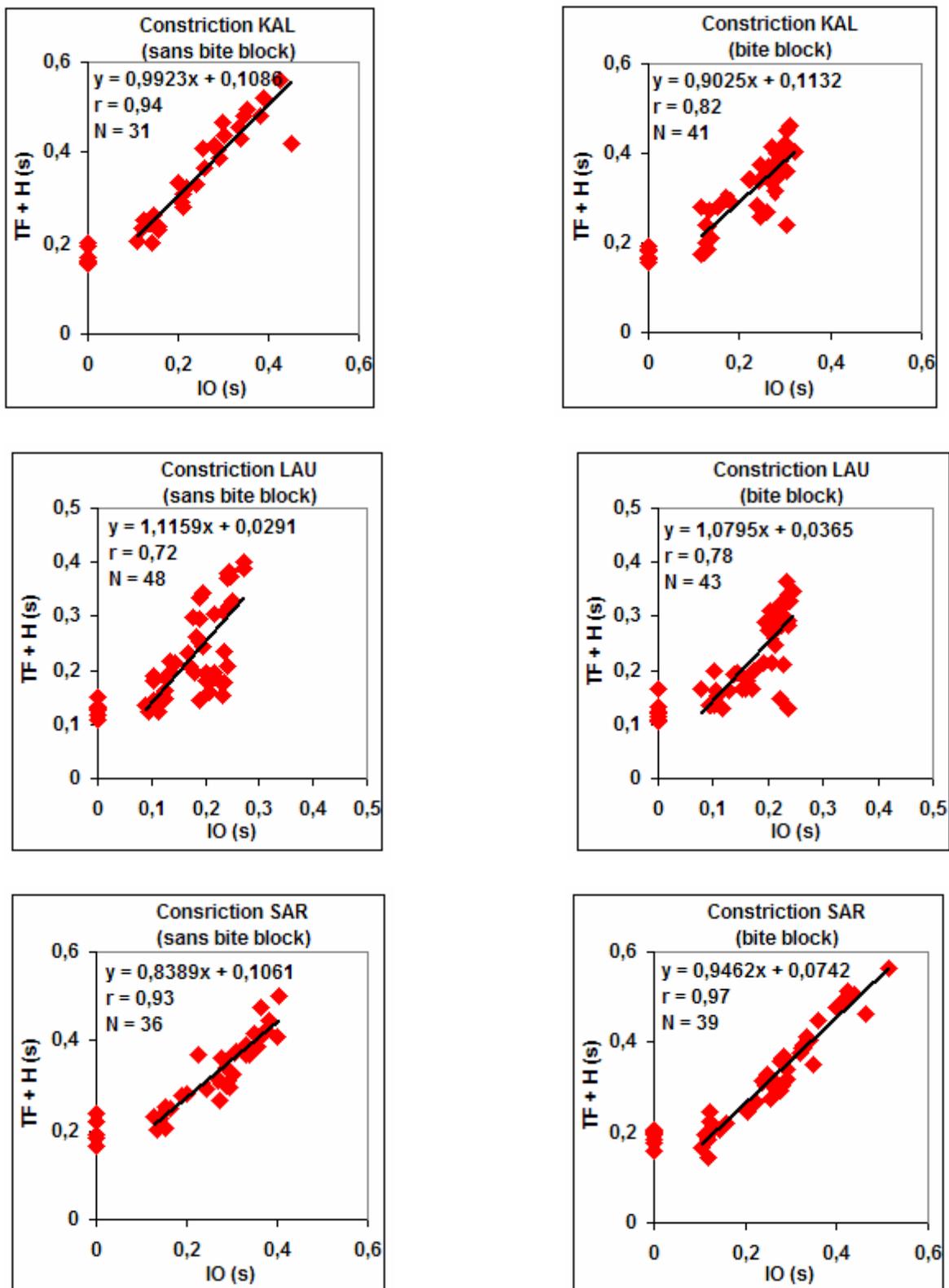


Figure 61 : Relation entre la durée totale du mouvement de constriction (TF+H) et la durée de IO pour les 4 locuteurs anglo-américains en condition sans *bite-block* (figures de gauche) et *bite-block* (figures de droite). Coefficients de corrélation r significatifs à $p < 0,0001$. N correspond au nombre total de séquences considérées pour l'analyse par régression linéaire.

En ce qui concerne les *constantes d'exécution* calculées sur les séquences [iu], nous avons observé des valeurs évidemment variables d'un locuteur à l'autre, de 130 à 210 ms (Tableau 9). On remarque que ces constantes d'exécution des séquences [iu] sont généralement toujours inférieures aux durées de mouvement des séquences [iku] (variant de 140 à 270 ms).

Comparaison des constantes d'exécutions calculées pour les séquences [iu] avec les durées moyennes de mouvement (TF + H en ms) pour les séquences [iku] Locuteurs anglo-américains – Constriction –																
Locuteurs	DAN				KAL				LAU				SAR			
Type de séquences	[iu]		[iku]													
Condition expérimentale	ssbb	bb	ssbb	bb												
Durée de TF + H (ms)	150	160	190	270	170	180	240	240	130	130	160	150	210	190	230	200

Tableau 9 : Constantes d'exécutions calculées pour les séquences minimales [iu] pour chaque locuteur anglo-américain en condition sans *bite-block* (ssbb) et *bite-block* (bb). Comparaison avec les durées moyennes de mouvement (TF+H en ms) pour les séquences [iku].

En ce qui concerne l'expansion du mouvement au cours des séquences [iC_nu], les pentes sont globalement assez fortes, comprises entre 0,72 et 1,12. De plus, la différence de pente entre les deux conditions avec et sans *bite-block* ne s'est avérée significative pour aucun des locuteurs (DAN : $t=-1,15$ ddl=79 ; KAL : $t=0,74$ ddl=70 ; LAU : $t=-0,18$ ddl=89 ; SAR : $t=-1,55$ ddl=73). Nous trouvons donc des coefficients d'expansion du mouvement de constriction aussi régulièrement élevés pour nos locuteurs anglo-américains, que ceux que nous avons trouvés précédemment pour nos locuteurs franco-québécois (pente de 0,85 à 1,05).

Rappelons que les quatre locuteurs français d'Abry & Lallouache (1995b) avaient également un coefficient d'expansion relativement élevé, variant de 0.69 à 0,93 selon le locuteur.

II.2.2.3. Y a-t-il une contribution égale, d'un locuteur à l'autre, des phases de *Time Falling* et *Hold* ?

Avant d'entrer dans la discussion générale des résultats obtenus en protrusion et en constriction pour tous nos locuteurs, nous souhaitons nous poser la question de la contribution respective de chaque phase - celle de *Time Falling* et celle de *Hold*-, à la réalisation du geste global d'arrondissement. Y aurait-il *systématiquement* une phase plus spécifiquement responsable de l'expansion du mouvement observée pour chacun de nos locuteurs ? En d'autres termes, y aurait-il vraiment nécessité de spécifier, dans le modèle, la phase

d'établissement du geste d'arrondissement et sa phase de tenue, que ce soit généralement ou par locuteur ? De manière à discuter globalement de cette question, nous avons décidé de la traiter aussi bien du point de vue des données obtenues en constriction qu'en protrusion.

Nous avons donc mis en relation séparément chacune des deux phases avec l'*Intervalle d'Obstruence* pour observer leur évolution séparée et calculer la droite de régression pour chaque cas. Nous présenterons les valeurs des coefficients d'expansion obtenues pour chaque locuteur franco-qubécois et anglo-américain en constriction (Tableau 10) et en protrusion (Tableau 11). Ces tableaux indiquent pour chaque locuteur, dans chaque condition avec et sans *bite-block*, la pente obtenue, quand le geste global (TF+H) est étudié, ainsi que pour la phase de *Time-Falling* (TF) et de *Hold* (H) séparément. Les figures présentant l'ensemble des observations, qui ont permis de déterminer les coefficients d'extension, sont données dans les Annexes I, II, III et IV.

En ce qui concerne la constriction (Tableau 10), pour tous les locuteurs francophones, c'est la phase de *Time Falling* qui semble maximalement contribuer à l'expansion du geste : le coefficient obtenu en phase Tf est relativement plus proche de celui obtenu sur le geste global TF+H, que ne l'est celui correspondant à H - et ceci pour les deux conditions avec et sans *bite-block*. Il en va de même pour les locuteurs anglophones KAL et SAR. Mais pas pour le locuteur LAU, pour lequel c'est la phase de *Hold* qui contribue très clairement le plus à l'expansion du geste global. Ni pour le locuteur DAN : contribution plus importante de la phase de *Hold* en condition sans *bite-block*, et de celle de *Time Falling* (mais à un moindre degré) en condition *bite-block*.

En ce qui concerne la protrusion (Tableau 11), nous pouvons constater que c'est la phase de *Time Falling* qui contribue le plus à l'expansion du geste, pour le locuteur anglo-américain protrudeur et trois des quatre locuteurs franco-qubécois. C'est encore vrai pour le locuteur franco-qubécois ANN en condition sans *bite-block*, mais pas en condition avec *bite-block*, condition pour laquelle les deux phases contribuent à égalité.

Pentes obtenues pour les phases <i>Time Falling + Hold (TF+H), Time Falling (TF), Hold (H)</i> en constriction - Locuteurs franco-québécois -																								
Locuteurs	ANN						ANT						JER						LUC					
Condition expérimentale	Sans bite block			Bite block			Sans bite block			Bite block			Sans bite block			Bite block			Sans bite block			Bite block		
Phases	TF+H	TF	H	TF+H	TF	H	TF+H	TF	H	TF+H	TF	H	TF+H	TF	H	TF+H	TF	H	TF+H	TF	H	TF+H	TF	H
pentés	0,98	0,88	0,1 ns	1,05	1,02	0,03 ns	0,86	0,67	0,19	0,85	0,57	0,28	0,92	0,66	0,25	0,94	0,67	0,27	0,88	0,80	0,07	0,86	0,83	0,03 ns

Pentes obtenues pour les phases <i>Time Falling + Hold (TF+H), Time Falling (TF), Hold (H)</i> en constriction - Locuteurs anglo-américains -																								
Locuteurs	DAN						KAL						LAU						SAR					
Condition expérimentale	Sans bite block			Bite block			Sans bite block			Bite block			Sans bite block			Bite block			Sans bite block			Bite block		
Phases	TF+H	TF	H	TF+H	TF	H	TF+H	TF	H	TF+H	TF	H	TF+H	TF	H	TF+H	TF	H	TF+H	TF	H	TF+H	TF	H
pentés	0,87	0,21	0,66	0,75	0,45	0,30 ns	0,99	0,98	0,01 ns	0,90	0,73	0,16 ns	1,12	0,15 ns	0,96	1,08	0,21 ns	0,86	0,84	0,70	0,14	0,95	0,83	0,11

Tableau 10 : Comparaison des pentes obtenues pour le geste global de constriction (TF+H) et séparément pour les phases de *Time Falling* (TF) et *Hold* (H) par rapport à IO, pour les 4 locuteurs franco-québécois (en haut) et les 4 locuteurs anglo-américains (en bas) dans chaque condition expérimentale. Les pentes sont validées par des coefficients de corrélations à $p < 0,01$, à l'exception de celles accompagnées du symbole ns.

Pentes obtenues pour les phases <i>Time Falling + Hold (TF+H), Time Falling (TF), Hold (H) en protrusion</i> - Locuteurs franco-québécois -																								
Locuteurs	ANN						ANT						JER						LUC					
Condition expérimentale	Sans bite-block			Bite-block			Sans bite-block			Bite-block			Sans bite-block			Bite-block			Sans bite-block			Bite-block		
Phases	TF+H	TF	H	TF+H	TF	H	TF+H	TF	H	TF+H	TF	H	TF+H	TF	H	TF+H	TF	H	TF+H	TF	H	TF+H	TF	H
pentes	0,60	0,48	0,12	0,68	0,35	0,32	0,65	0,60	0,04	0,62	0,54	0,09	0,45	0,37	0,07	0,78	0,63	0,14	0,24	0,17	0,06	0,18	0,12	0,06

Pentes obtenues pour les phases <i>Time Falling + Hold (TF+H), Time Falling (TF), Hold (H) en protrusion</i> - Locuteurs anglo-américains -																								
Locuteurs	DAN						KAL						LAU						SAR					
Condition expérimentale	Sans bite-block			Bite-block			Sans bite-block			Bite-block			Sans bite-block			Bite-block			Sans bite-block			Bite-block		
Phases	TF+H	TF	H	TF+H	TF	H	TF+H	TF	H	TF+H	TF	H	TF+H	TF	H	TF+H	TF	H	TF+H	TF	H	TF+H	TF	H
pentes	0,51	0,66	-0,14	0,40	0,46	-0,05 ns																		

Tableau 11 : Comparaison des pentes obtenues pour le geste global de protrusion (TF+H) et séparément pour les phases de *Time Falling* (TF) et *Hold* (H) par rapport à IO, pour les 4 locuteurs franco-québécois (en haut) et le locuteur anglo-américain DAN (en bas ; les 3 autres locuteurs n'ayant pas de protrusion suffisante) dans chaque condition expérimentale. Les pentes sont validées par des coefficients de corrélations à $p < 0,01$, à l'exception de celles accompagnées du symbole ns.

Quelle conclusion peut-on tirer de ces observations ? Certes, nous pourrions tenter de défendre que, pour la plupart des locuteurs, c'est la phase de *Time Falling* qui rend compte de l'expansion du geste d'arrondissement. Et chercher à défendre aussi des styles plus personnels pour quelques-uns de nos sujets, comme le locuteur anglo-américain LAU, qui varie essentiellement sa phase de tenue du geste. Mais nous défendrons qu'adopter cette ligne de lecture de nos données reviendrait surtout à perdre en partie la cohérence du comportement *MEM* qui se vérifie sur l'ensemble des locuteurs, telle que nous avons réussi à l'obtenir sur le geste global, en particulier sur le geste de constriction (cf. la Figure 62 récapitulative de nos 8 locuteurs), une cohérence dont nous allons discuter dans la section suivante.

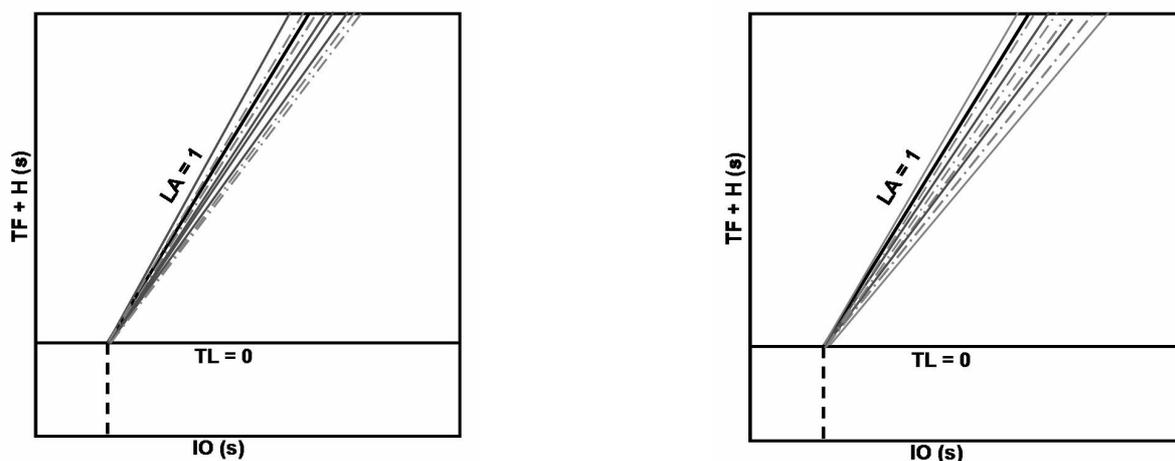


Figure 62 : Représentation schématique des pentes de constriction des 4 locuteurs anglo-américains (à gauche) et des 4 locuteurs franco-québécois (à droite) par rapport aux pentes suggérées par le modèle *Look-ahead* (pente de 1) et modèle *Time-locked* (pente de 0). Les lignes pleines correspondent aux pentes obtenues en condition *bite-block*, les lignes pointillées en condition sans *bite-block*.

II.3. Discussion générale

L'un des objectifs de cette étude cinématique était d'étendre un paradigme expérimental jusqu'ici testé seulement pour la description de l'anticipation d'arrondissement en français à des locuteurs issus de différentes communautés linguistiques, c'est-à-dire le franco-québécois et l'anglo-américain. Nous avons choisi de recueillir deux paramètres labiaux dynamiques susceptibles de rendre compte de l'évolution de l'arrondissement, la protrusion de la lèvre supérieure – figurant comme paramètre privilégié dans la littérature sur la coarticulation anticipante – et la constriction intérolabiale. Pour le recueil simultané de ces deux paramètres, nous avons couplé deux techniques d'extraction de données labiales, l'Optotrak ainsi que le système de capture en ligne des contours labiaux (*Lip-Shape-Tracker*).

II.3.1. De l'influence des consonnes dans les comportements observés

Dans notre première partie sur les modèles d'anticipation, nous avons évoqué le débat animé qui subsiste sur l'arrondissement consonantique. Nous avons pu dégager, des différentes études présentées, que les mouvements d'arrondissement attribués aux consonnes relevaient :

- I. principalement de l'action d'élévation de la mandibule ;
- II. sont observés essentiellement au niveau de la lèvre inférieure, la lèvre supérieure étant jugée indépendante vis-à-vis de la porteuse mandibulaire.

Bien que les consonnes [k, s, t] présentes dans notre corpus soient considérées comme phonologiquement *neutres* au regard de l'arrondissement, certaines études ont fait état d'un mouvement labial, dû par exemple au couplage avec la mâchoire lors de la production de consonnes coronales (Perkell & Matthies, 1992). Pour ces raisons, notre corpus a été dupliqué avec une condition *bite-block* afin de bloquer les éventuels mouvements mandibulaires d'arrondissement consonantique et ne conserver que le mouvement d'arrondissement vocalique. Il s'agit là d'une simple précaution méthodologique, notre objectif n'étant pas d'analyser précisément les effets consonantiques, mais de les neutraliser. Nous pouvons dégager de ces résultats que même si l'on a parfois rencontré des signaux dont l'apparence suggérait une légère influence des consonnes, celle-ci n'a pas d'incidence sur le déroulement temporel de l'anticipation d'arrondissement puisque l'on retrouve pour tous les locuteurs des profils idiosyncratiques stables entre les deux conditions expérimentales (sans différence significative entre les pentes obtenues en condition *bite block* et *sans bite-block*, à l'exception du locuteur JER dont nous avons discuté le cas).

II.3.2. Un paramètre universel: la constriction

Au travers de cette étude sur l'arrondissement vocalique, nous avons souhaité tester deux paramètres labiaux contribuant à la réalisation distinctive des voyelles arrondies *vs.* non arrondies (ou étirées). Le premier de ces paramètres, la protrusion de la lèvre supérieure, a été le paramètre privilégié des modèles d'anticipation élaborés pour l'anglais (*Look-Ahead*, *Time-Locked* et *Hybrid*). Nous avons longuement exposé, dans notre partie I, l'échec de ces modèles à rendre compte de manière cohérente du comportement d'anticipation d'arrondissement vocalique, au moins dans leurs versions fortes (selon Perkell, 1990, et Perkell & Matthies, 1992). En ce qui concerne le dernier modèle élaboré pour le français, le *MEM* (Abry & Lallouache, 1995a; Abry *et al.*, 1996a), leurs auteurs avaient rencontré des

difficultés avec un de leurs sujets qui ne présentait tout simplement pas de protrusion labiale. Nous avons d'ailleurs rappelé que d'autres auteurs (Perkell & Matthies, 1992, et Lubker & Gay, 1982) avaient aussi rencontré ce problème. Et nous-mêmes n'avons pas toujours pu mesurer la protrusion, et ce sur plus d'un de nos locuteurs. Remarquons que c'est bien un problème d'amplitude insuffisante du geste de protrusion chez certains locuteurs, anglophones principalement mais aussi francophones, et non pas un problème technique de mesure. Lorsque nous pouvons mesurer la protrusion, nous pouvons le faire aussi bien avec une technique de détection de contours labiaux de profil (avec la technique du « bleu à lèvres » de Lallouache, 1991) ou avec l'Optotrak (comme dans notre étude menée au Québec).

Ainsi nous défendrons l'idée qu'un modèle de l'anticipation basé sur la protrusion ne pourra être proposé que pour des sujets présentant un minimum d'au moins 2 mm d'amplitude labiale en protrusion (nous avons eu 3 locuteurs sur 8 sans protrusion suffisante) ! Mais, pour ces sujets, le modèle *MEM* de protrusion d'Abry & Lallouache (1995a) offre une paramétrisation du comportement propre à chaque locuteur, quelle que soit sa langue.

Dans tous les cas, si l'usage distinctif de la protrusion relève des normes phonologiques d'une langue et peut suffire à décrire les comportements d'anticipation de l'arrondissement vocalique dans un univers linguistique – et nous avons vu que ce n'est pas systématiquement le cas pour tous les locuteurs, même dans une langue comme le français, où l'arrondissement a un rôle phonologique distinctif – ce paramètre devient quel que peu *ad hoc* lorsque l'on s'attache à comparer les stratégies de planification de l'arrondissement inter-langues. En revanche, la constriction s'est révélée être un corrélat articulatoire particulièrement efficace de la dimension d'arrondissement, en français, en franco-qubécois et en anglais.

En effet, il semble que la variation de l'aire aux lèvres ou constriction permet à tous les coups de mettre en évidence ce contraste. Selon nous, il s'agit même du seul paramètre articulatoire suffisamment robuste du point de vue articulatoire et acoustique pour rendre compte du contrôle de l'anticipation d'arrondissement vocalique chez l'adulte, sur une dimension quasi universelle des voyelles (Schwartz *et al.*, 1997). Ce point avait déjà été soulevé par Perkell & Matthies (1992, p. 2923) lors de leur étude sur le décours temporel de la protrusion.

Pourquoi ce paramètre d'aire aux lèvres a-t-il alors été aussi peu exploré jusque-là, par les modélisateurs de l'anticipation labiale ? Sans doute en raison de contraintes techniques, le

suivi par points (*flesh-points*) obtenu par exemple avec un système Optotrak ne délivrant pas ce paramètre de constriction. Le *Lip-Shape-Tracker* de Lallouache (1991), qui comme son nom l'indique, est dédié au suivi de formes et non pas simplement de points cutanés, s'avère une technique tout à fait adaptée, précise et robuste pour la détection de la constriction labiale, dont les petites aires sont si cruciales acoustiquement. Et nous avons pu la mettre en œuvre, pour la première fois, avec une détection simultanée de la protrusion par Optotrak.

II.3.3. Le succès d'un modèle

Au vu des résultats très variables fournis pour l'anglais par les 3 modélisations classiques que sont les modèles *Look-ahead*, *Time-locked* et *Hybrid* (Perkell & Matthies, 1992) et ceux *a contrario* cohérents obtenus chez 4 locuteurs français dans le cadre du *Modèle d'Expansion du Mouvement* (Abry & Lallouache, 1995a), nous avons fait le choix, au début de notre travail de thèse, de prendre ce modèle pour référence pour la description du comportement d'anticipation d'arrondissement vocalique. Nous avons pu mettre en évidence des profils d'anticipation variables d'un locuteur à l'autre, mais pouvant être décrits dans le cadre d'une variabilité régie par la propre loi du locuteur (*lawful variability*). Celui-ci étend son mouvement d'arrondissement selon un coefficient d'expansion qui lui est propre, donnant lieu à des pentes ajustées ou attunées (validées par des coefficients de corrélations significatifs). Lorsque l'intervalle consonantique disponible est court, l'empan de l'anticipation est lui-même réduit et celui-ci ne se réduit pas en-deçà d'une *constante d'exécution* identifiable, pour chaque locuteur, pour le geste de base [iu]. Celle-ci témoigne de la pénétration maximale du geste d'arrondissement attendu pour V_2 dans V_1 . Au-delà d'une consonne, le geste d'arrondissement s'étend de manière linéaire.

La modélisation par le modèle *MEM* du geste global d'arrondissement, avec détection d'événements et de phases permettant de décrire au mieux ce geste global, offre une prédiction adéquate de l'évolution idiosyncratique du comportement d'anticipation, quelle que soit la communauté linguistique des sujets. Les phases retenues nous paraissent plus appropriées pour rendre compte d'une telle variabilité, que la sélection d'événements temporellement toujours fragiles, tels que les maximums de constriction ou minimums de protrusion, car ces derniers peuvent être associés à des plateaux et n'impliquent pas forcément le démarrage imminent du geste d'arrondissement (c'était le cas du locuteur LUC en protrusion).

Dans leur article consacré à la description des mouvements labiaux, Ramsay *et al.*, (1996) insistaient sur le fait qu'il était en effet relativement complexe de capturer leurs caractéristiques essentielles, lorsqu'on prélève sur des fonctions temporelles des mesures ponctuelles, pour résumer, par des moyennes, différentes mesures spatio-temporelles (durée, amplitude, etc.) de ces points – le danger étant de récupérer des informations approximatives sur le déroulement temporel des coordinations articulatoires.

« [...] *there is the problem of studying the dynamic or functional character of the process. One of the most common simplifying assumptions is to restrict the analysis to scalar summaries of the movement trajectories such as the average duration, amplitude and peak velocity of markers attached to the individual tissues points. [...] But using point summary measures of continuous functions presumes that these are sufficient for understanding the underlying process and moreover, that the time-varying detail of the movement trajectories is relatively unimportant.* » (Ramsay *et al.*, 1996, p. 3718)

En ce qui nous concerne, le choix d'événements robustes, comme le 10% et 90% de chute d'aire intérolabiale, nous permettent de définir un intervalle décrivant bien l'amorce du geste d'arrondissement (*Time Falling*), tandis que la phase de *Hold* nous garantit d'atteindre la cible acoustique associée à la constriction minimale pour le [u]. Ensemble, ces deux phases illustrent l'expansion du mouvement par rapport à l'*Intervalle d'Obstruence* sans tomber dans l'artefact statistique de la corrélation « *tout-partie* ». Les événements de pics d'accélération ont, quant à eux, été écartés de nos analyses car ils ne représentent pas, de notre point de vue, des critères suffisamment stables, ces derniers pouvant être associés à des effets consonantiques ou refléter des données bruitées.

Le modèle *MEM*, en tant que modèle paramétrable par locuteur, peut rendre compte d'un comportement d'arrondissement, dont l'évolution est étroitement liée à l'intervalle consonantique disponible entre les deux voyelles, avec notamment une pente proche de 1, sans qu'il s'apparente par principe à un modèle *Look-ahead* qui, lui, dans son principe même, borne le déploiement du geste d'arrondissement par rapport à la fin acoustique de la voyelle [i] et le début acoustique de la voyelle arrondie [u], ce qui n'est pas du tout le cas du *MEM*.

Nous avons défendu qu'un comportement de type *Time-locked* (pente 0 et $r = 0$) est irrecevable dans le cadre d'un test statistique, puisque ce serait, en termes statistiques, accepter l'*Hypothèse Nulle*. Evidemment, les études ayant défendu le principe d'invariance temporelle du mouvement, quelle que soit la durée de l'intervalle consonantique, n'ont pu démontrer expérimentalement cette proposition. Si l'on reprend les données d'électromyographie de Bell-Berti & Harris (1982) (Figure 16, partie I.5.1), on remarque

qu'aucun des 3 locuteurs anglais testés n'initie son mouvement à date fixe par rapport au début de la voyelle arrondie. Suivant la durée de l'intervalle consonantique, le départ du geste varie de 100 ms. C'est également le cas des 2 locuteurs dans l'étude de Lubker & Gay (1982) dont les comportements sont décrits en termes de modélisation « *Time-locked* » alors que leurs départs d'activité EMG sont clairement dispersés dans une fenêtre temporelle de plus de 200 ms. Par conséquent, il semble que les hypothèses du modèle *Time-locked* résistent tant que les données restent artificiellement compressées par la représentation graphique adoptée (*sic* !). Mais lorsque les données sont, comme il se doit, normalisées sur les deux axes x et y, celles-ci apparaissent à l'évidence très dispersées. Cette dispersion est alors attestée statistiquement par des coefficients de corrélation à chaque fois non significatifs.

Le Tableau 12 reprend l'ensemble des pentes – et leurs coefficients de corrélations associés – pour tous nos locuteurs, à partir de l'analyse de la protrusion et de la constriction intérolabiale.

Droites de régression et coefficients de corrélation (p<0,0001)				
	Protrusion de la lèvre supérieure		Constriction (aire aux lèvres)	
Locuteurs anglo-américains	condition bite block	condition sans bite block	condition bite block	condition sans bite block
DAN	0,40 (0,57)	0,51 (0,67)	0,75 (0,84)	0,87 (0,88)
KAL	Locuteurs n'ayant pas d'amplitude de protrusion suffisante		0,90 (0,82)	0,99 (0,94)
LAU			1,08 (0,78)	1,12 (0,72)
SAR			0,95 (0,97)	0,84 (0,93)
Locuteurs franco-québécois				
ANN	0,68 (0,85)	0,60 (0,92)	1,05 (0,95)	0,98 (0,99)
ANT	0,62 (0,93)	0,65 (0,92)	0,85 (0,93)	0,86 (0,94)
JER	0,78 (0,88)	0,45 (0,65)	0,94 (0,96)	0,92 (0,95)
LUC	0,18 (0,69)	0,24 (0,75)	0,86 (0,95)	0,88 (0,98)

Tableau 12 : Pentés et coefficients de corrélation (entre parenthèses) calculées pour les 4 locuteurs anglo-américains et franco-québécois suite à l'analyse de la protrusion et de la constriction intérolabiale en condition *bite-block* et sans *bite-block*.

Au vu de ces résultats, on voit clairement que le comportement de protrusion pour nos locuteurs ne suit pas celui de la constriction intérolabiale : les pentés décrivant l'évolution de la constriction sont toutes plus élevées, quelle que soit la condition expérimentale. C'est un phénomène qui est déjà présent, même s'il est moins marqué, dans les données d'Abry &

Lallouache (1995) : les coefficients d'expansion de leurs 3 locuteurs protrudeurs (Jean-Luc : 0,42 ; Benny : 0,54 et Annie : 0,93) sont inférieurs ou égaux à ceux trouvés en constriction (Jean-Luc : 0,79 ; Benny 0,69 et Annie : 0,93).

Nous avons précédemment discuté (en section II.2.2.3.) de la contribution respective des phases de *Time Falling* et de *Hold* et défendu le point de vue selon lequel la description du comportement d'anticipation d'arrondissement donnant la loi la plus générale est obtenue, pour chacun de nos locuteurs, en considérant le geste global, et non simplement l'une de ses phases.

Néanmoins, l'intérêt de considérer chaque phase peut résider dans le fait d'éclairer au mieux la relation qui peut exister entre la mise en place différenciée, pour certains locuteurs, du geste de protrusion et du geste de constriction des lèvres, qui permettent au final de réaliser la configuration articulaire de la voyelle arrondie pour assurer sa qualité acoustique.

Nous illustrerons ainsi le cas du sujet franco-québécois LUC qui présente peu d'expansion en protrusion (pente de 0,24 sans *bite-block* et 0,18 avec *bite-block*) mais qui, en constriction, à un coefficient d'expansion très élevé (0,88 sans *bite-block* et 0,86 avec *bite-block*). Nous présentons, en Figure 63, deux séries de signaux de constriction et de protrusion comme illustration de ce patron.

Les 3 premiers signaux de la figure correspondent à la réalisation acoustique d'une séquence « Deux ki coukiquent » produite par le locuteur LUC en condition *bite-block*, accompagnée du décours temporel de la protrusion (ulip, en mm) et de celui de la constriction (S, en cm²). Les 3 signaux du bas proposent les mêmes informations pour la séquence « deux kixtes scoukiquent » comportant donc un *Intervalle d'Obstruence* plus long. Lorsque IO est petit, les signaux d'aire aux lèvres et de protrusion de la lèvre supérieure (ce dernier signal étant inversé pour la commodité de la comparaison) évoluent de manière relativement semblable avec des phases TF et H de durée comparable. Lorsque IO augmente, les deux profils de geste deviennent clairement différents. On constate que la phase de Hold, en protrusion et constriction, varie peu en durée par rapport à la réalisation avec un petit *Intervalle d'Obstruence*. La variation importante concerne la phase de *Time Falling* : celle de protrusion n'augmente que très peu avec l'augmentation de IO, tandis que celle de l'aire aux lèvres croît énormément.

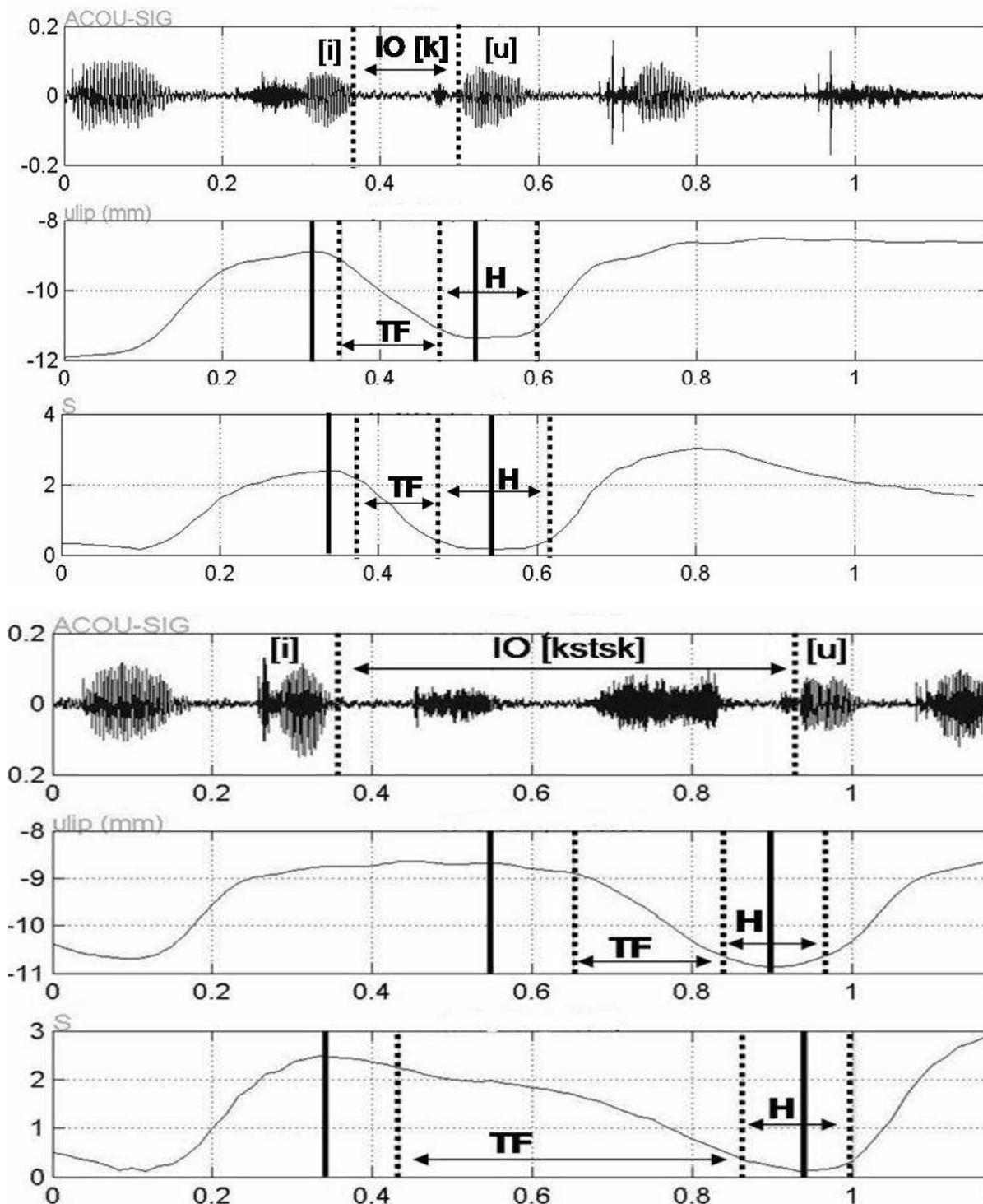


Figure 63 : Signal acoustique et décours temporel de la protrusion de la lèvre supérieure (ulip) et de l'aire aux lèvres (S), pour les séquences « Deux ki coukiquent » (3 premières figures en haut) et « Deux kixtes scoukiquent » (3 dernières figures en bas) produites par le locuteur franco-québécois LUC, en condition sans *bite-block*. Le signal acoustique est présenté avec les repères en pointillés de fin acoustique de la voyelle [i] et de début de [u]. Les minimums et maximums de constriction et protrusion sont indiqués par des lignes verticales continues et les phases *Time Falling* (TF) et *Hold* (H) sont situées entre les lignes pointillées.

On remarque donc qu'il y a désynchronisation, chez ce locuteur, entre les deux composantes qui permettent de mener à bien l'arrondissement : alors que ce locuteur garde longtemps les lèvres étirées après la voyelle [i] et initie un geste de protrusion seulement deux consonnes avant la consonne arrondie [u], il anticipe précocement, soit dès la première consonne de la chaîne de 5 consonnes, son geste de constriction, augmentant ainsi fortement sa phase de *Time Falling*. Ce *phasage* différent de son geste de constriction permet ainsi en quelque sorte à ce locuteur de rejoindre les autres locuteurs dans l'expansion temporelle du geste de constriction. Nous retrouvons là encore un argument supplémentaire de robustesse de ce paramètre.

Nous reprendrons les différents points de succès de ce modèle en conclusion. Mais d'ores et déjà, nous pouvons annoncer – ce qui ne dépend pas de l'étude sur les enfants qui va suivre – que le compromis qui avait été proposé par Byrd & Saltzman (2003), soit un modèle pour l'anglais et un modèle pour le français, avec une contrainte de bord *ad hoc*, a vécu.

« *By temporal elasticity, we refer to the relationship between a given gesture's activation interval and a preceding interval before target attainment in which there are no conflicting gestures. In English, for example, anticipation intervals are relatively fixed regardless of the length of the preceding no-conflict intervals; in French, however, the anticipation periods stretch in proportion to the length of the no-conflict intervals.* » (Byrd & Saltzman, 2003, p. 157)

La citation que nous venons de lire traitant de l'*élasticité* – qui est une propriété fondamentale dans l'adaptabilité du contrôle du mouvement, et donc pour la parole –, se termine sur une constatation (*statement*) pour l'anglais qui devient caduque. En effet, celui-ci se range, par nos expériences, définitivement sous la règle du MEM de constriction, comme le français, pour prédire l'expansibilité d'un mouvement-orienté-vers-un-but-acoustique – ceci à partir d'une *constante d'exécution*, tout aussi spécifique au locuteur que son coefficient d'expansion selon sa propre loi linéaire.

Partie III. L'acquisition du contrôle vocalique

Dans cette partie nous traiterons essentiellement de l'acquisition du contrôle vocalique en vue de comprendre par la suite le réglage (*attunement*) du geste d'anticipation de la voyelle. Avant d'exposer les connaissances sur les coordinations du geste vocalique dans la parole, nous nous interrogerons sur les précurseurs de ces gestes oro-laryngés chez le primate non-humain en utilisant les découvertes les plus récentes dans ce domaine. La question de l'imitation notamment est revenue au premier plan des études sur l'enfant et en primatologie. Nous privilégierons les données sur le contrôle labio-laryngé qui est l'objet principal de notre thèse.

III.1. Imitation, émulation dans la production et la perception d'un contrôle labio-laryngé chez le primate non-humain

L'existence d'un comportement d'imitation chez les singes est une question qui reste empiriquement ouverte, que ce soient chez les singes (ou *monkeys*) autres qu'anthropoïdes, ou même chez ceux-ci ou grands singes (en anglais *apes*, comprenant l'orang-outan, le gorille, le chimpanzé, le bonobo). La question qui est posée est celle de la continuité de cette capacité chez les hominidés (*cf.* la Figure 64 pour une idée des distances de spéciation en millions d'années).

Objet de discussions passionnées dans des domaines aussi divers que la philosophie, la psychologie, la biologie ou encore la sociologie, le concept d'*imitation* a longtemps souffert de définitions relativement floues, si bien qu'il reste délicat – en ce qui concerne plus particulièrement le domaine de la parole – de dégager ses implications pour l'apprentissage du langage.

Concernant spécifiquement la question des précurseurs de la production de la parole, l'avis nuancé de Hauser, Chomsky *et al.* (2002) est rapporté de manière catégorique par Pinker & Jackendoff (2005):

« Non-human primates are also notoriously resistant to training of their vocalizations (Hauser, 1996), and as HCF themselves note, they show no ability to learn vocalizations through imitation. HCF try to downplay the difference between humans and primates by pointing out that vocal imitation is not uniquely human. » (Pinker & Jackendoff, 2005, pp. 208-209).

Ces derniers finissent par faire appel à Darwin en citant le babillage canonique :

« And as Darwin noted, the innate vocal babbling of human infants is one of the clearest signs that "man has an instinctive tendency to speak. » (ibid., p. 209)

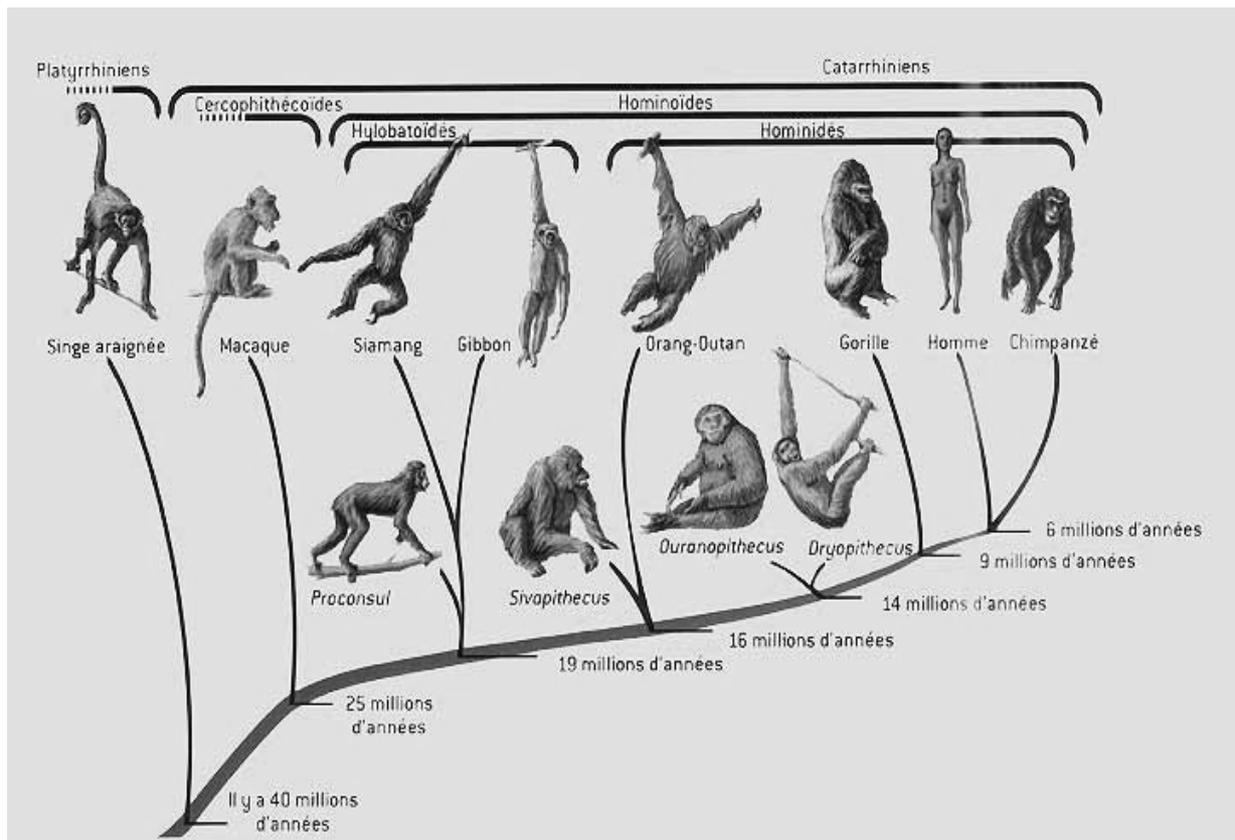


Figure 64 : Arbre phylogénétique (cladistique) des espèces de singes comprenant l'homme, avec leurs distances génétiques en millions d'années. D'après Picq, 2003.

Pour aborder plus en profondeur notre sujet du point de vue de l'ontogénèse, une revue substantielle a été réalisée récemment par Serkhane (2005) dans le cadre de ses travaux de thèse sur « *L'étude et la modélisation des mécanismes d'exploration vocale et d'imitation orofaciale dans le développement de la parole* ». C'est une revue grâce à laquelle on peut apprécier les tenants philosophiques, psychologiques ou éthologiques du processus d'imitation.

Serkhane (2005) a attiré notre attention sur une proposition du grand psychologue de la Gestalt et de l'animal (directeur de la station de recherche sur les anthropoïdes de l'Académie prussienne des sciences à Ténérife aux Canaries, de 1913 à 1920), Wolfgang Köhler, auteur du livre paru en allemand (1917), en anglais (1921) puis en français *L'intelligence des singes supérieurs* (1927). Celui-ci distinguait déjà entre deux formes d'imitation. Dans ses propres termes, l'imitation de la *substance*, c'est-à-dire selon la reformulation de Bard & Russell (1999), l'*émulation*, soit l'obtention par différents moyens du résultat ou de l'objectif final. Et l'imitation de la *forme*, c'est-à-dire *l'imitation à l'identique* du comportement observé, y compris du chemin (ou *path*) pour l'atteinte du but. Ce paradigme distinguant émulation *versus* imitation nous paraît également puissant en termes d'apprentissage du langage. Nous

l'aborderons aussi bien à travers l'exploration de quelques travaux récents sur l'imitation chez les primates non-humains qu'en relation avec ce que nous connaissons de l'imitation chez le jeune enfant. Nous essaierons de montrer que ce mécanisme est aussi porteur d'enjeux capitaux en termes de communication des intentions et d'intégration sociale. Nous verrons que l'imitation de comportements moteurs et de gestes corporels peut mener à l'imitation plus fine des gestes orofaciaux pour la communication interpersonnelle et, plus loin encore, au réglage (*attunement*) du contrôle du conduit vocal pour l'acquisition de la norme phonologique de la langue ambiante.

III.1.1. L'imitation comme transmission de la norme sociale chez les chimpanzés

Whiten *et al.* (2005) ont réalisé une étude suggérant que certaines communautés de primates tels que les chimpanzés utilisent l'imitation comme vecteur de *transmission culturelle* (selon les termes mêmes des auteurs).

Pour tester l'existence d'un tel processus, Horner et ses collègues ont élaboré un protocole expérimental basé sur l'extraction de nourriture placée dans des boîtes. La nourriture pouvait être obtenue via deux techniques (chemins ou *paths*): soit en poussant une petite baguette le long du tube, libérant ainsi la boulette de nourriture ; soit en levant la baguette et en la tirant vers soi ce qui entraînait le glissement de la nourriture à l'extérieur du tube. Chacune de ces stratégies est illustrée dans la Figure 65.

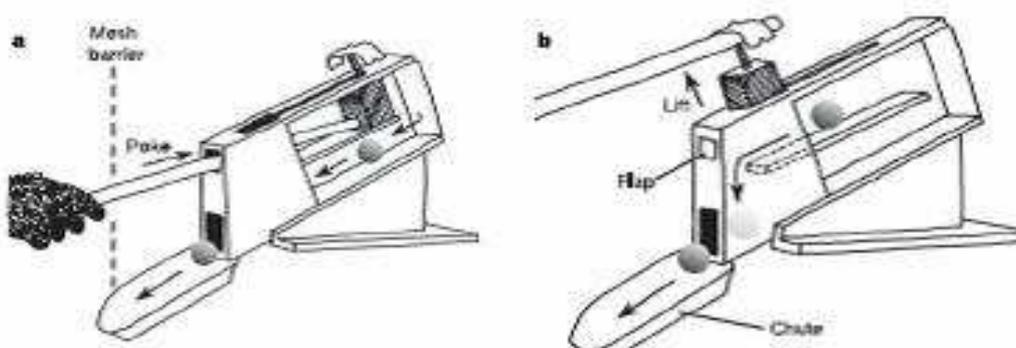


Figure 65 : Illustration des deux techniques d'extraction de la nourriture enseignées à un des membres de deux communautés de chimpanzés. L'image A représente la technique du poussoir ou *poke technique*. L'image B est la technique de rapprochement de la baguette ou *lift technique*. D'après Whiten *et al.*, 2005, p. 737.

Deux groupes de seize chimpanzés chacun n'étant pas habitués aux interactions avec l'homme ont été sélectionnés pour cette étude. L'une des deux techniques a été enseignée à un membre de chaque groupe, momentanément isolé du reste de son groupe. Lorsque le modèle

s'est familiarisé avec la stratégie d'extraction de la nourriture, les expérimentateurs mettent en place une phase d'observation de sept jours durant laquelle le reste du groupe peut observer le modèle exécuter la tâche sans y avoir eux-mêmes accès. Puis les expérimentateurs disposent l'une des deux boîtes dans chaque groupe. Le résultat observé est que lorsque l'on soumet la tâche d'extraction de la nourriture à l'ensemble du groupe, la majorité d'entre eux (quinze membres) copie la méthode employée par le chimpanzé modèle.

En référence à Köhler, ce comportement relève d'un processus d'imitation, se traduisant par la copie du *but* ainsi que par la copie du *chemin* (*path*), c'est-à-dire la coordination gestuelle appropriée pour l'atteinte du but. C'est une coordination que l'on retrouvera chez l'homme pour la parole, lorsque pour le but de production d'une voyelle, sera recrutée une coordination particulière entre la mandibule et la langue et les lèvres. Nous soulignons que cette coordination comprend implicitement le contrôle de son déroulement temporel.

Avant le début des tests, un troisième groupe, appartenant à l'une des deux communautés a été isolé afin de contrôler la validité des résultats obtenus. Les expérimentateurs ont proposé l'une des deux boîtes à l'ensemble des membres de ce troisième groupe en leur laissant explorer les possibilités de résolution de la tâche, sans leur enseigner aucune des deux techniques précédentes. Chaque membre tente d'extraire la nourriture, certains découvrent l'une ou les deux méthodes. Pour ce groupe, le comportement initié par les chimpanzés correspond à une réponse d'*émulation*, c'est-à-dire qu'ils copient uniquement le but quel que soit le moyen d'extraire la nourriture de la boîte.

Ce qui est frappant c'est que lorsque ce groupe rejoint sa communauté, celui-ci se conforme à la technique adoptée au sein de la communauté. On retrouve ce patron même chez les chimpanzés qui avaient découvert les deux stratégies d'extraction de la nourriture ; ces derniers abandonnent en effet leur méthode au profit de celle acceptée au sein de la colonie. Cette étude place le processus d'imitation comme un outil d'apprentissage en conformité à la norme sociale du groupe.

Quelques années auparavant, dans un article rapportant les observations de sept études longitudinales menées chez diverses communautés de chimpanzés d'Afrique (*Pan Troglodytes*), Whiten *et al.* (1999) avaient mentionné l'existence d'une forte diversité culturelle à travers ces communautés. Cette variabilité se traduisait par des patrons ou modèles de comportements ayant une valeur hautement distinctive à l'intérieur de chaque groupe.

« [...] *the combined repertoire of these behaviour patterns in each chimpanzee community is itself highly distinctive, a phenomenon characteristic of human cultures but previously unrecognised amongst non-human species.* ». (Whiten et al., 1999, p. 683)

Bien que le terme de *culture* soit polysémique et manipulé et interprété différemment selon les disciplines, les auteurs se réfèrent à son acception en biologie qui semble tout à fait convenir aux comportements observés chez les chimpanzés.

« *In the biological sciences, a more inclusive definition is accepted in which the significance of cultural transmission is recognised as one of the only two important processes that can generate evolutionary change; inter-generation transmission of behaviour may occur either genetically or through social learning, with processes of variation and selection shaping biological evolution in the first case and cultural evolution in the second. From this perspective, a cultural behaviour is one that is transmitted repeatedly through social or observational learning, to become a population-level characteristic.* » (Whiten et al., 1999, p. 683)

De nombreuses études ont relevé la valeur de ces traditions culturelles dans diverses colonies de chimpanzés. Ce qui est particulièrement intéressant dans l'étude de Whiten *et al.* (2005), c'est qu'il est également possible d'assister à cette *transmission culturelle* selon un paradigme contrôlé, à laquelle Whiten et ses collègues font référence. En effet, un comportement clé (dans notre cas, l'une des deux méthodes d'extraction de la nourriture) est transmis via un modèle (un des chimpanzés appartenant au groupe) à l'ensemble des membres de ce groupe. Ce comportement est aussitôt adopté et réutilisé à chaque fois qu'un membre du groupe est confronté à la tâche. Tout autre comportement non normé, c'est-à-dire non conforme à celui en vigueur dans la communauté, est abandonné. Ce processus de transmission de la norme semble perdurer puisque, lorsque l'on teste ces communautés deux mois plus tard, tous les membres du groupe (à l'exception d'un chimpanzé non-conformiste dans chacun des groupes testés) utilisent la méthode acquise deux mois auparavant. Dans cette perspective, l'imitation permet de synchroniser les actions, celles d'un comportement avec celles de l'environnement social.

Depuis plus d'une décennie, le développement de méthodes expérimentales et de protocoles rigoureux a permis de répliquer ce type d'expériences (Matsuzawa, 1994). Parmi l'ensemble des travaux réalisés, certains sont dédiés au contrôle moteur chez le singe avec un intérêt croissant pour l'exploration de l'activité neurale fonctionnelle du cerveau du singe lors de tâches motrices. Un grand nombre d'études réalisées dans ce domaine sont le fruit du groupe de Giacomo Rizzolatti à Parme (Italie) et de celui de Dave Perrett à St-Andrews (Ecosse).

III.1.2. De l'imitation des gestes oro-faciaux orientés vers la communication chez le macaque rhésus

Il existe une population de neurones visuo-moteurs localisés dans l'aire fonctionnelle F5 (cf. Figure 66) du macaque qui se trouve être activée spécifiquement en condition d'observation chez autrui d'une action dirigée vers un objet et de reproduction de cette action, ainsi que lors de la reproduction de cette action. En raison de leurs propriétés mimétiques, ces neurones découverts par Rizzolatti *et al.* (1996 ; Gallese *et al.*, 1996), ont été appelés *neurones miroirs*. Ces derniers se distinguent d'une autre catégorie de neurones appelés *neurones canoniques* qui ne répondent qu'en condition de présentation par un tiers de certains objets qui demandent une saisie bien particulière (on dit aussi une *affordance*).

Dans l'étude princeps de Rizzolatti *et al.* (1996), depuis donnée en exemple, l'expérimentateur place une boulette de viande sur un plan horizontal devant le singe. On entend (via leur amplification par haut-parleur) certains neurones de la région F5 décharger (crachoter). Ensuite, lorsque le macaque saisit à son tour la boulette de viande, la même population de neurones est activée. En revanche, si l'expérimentateur réitère l'expérience en plaçant cette fois la nourriture avec un instrument, ces neurones ne s'activent pas. De même, lorsque l'expérimentateur mime l'action sans objet (sans transitivité), aucune activité de ces neurones n'est observée. Par conséquent, ces neurones situés dans la région F5 déchargent uniquement lorsqu'une action est réalisée par un effecteur tel que la main ou la bouche ; et ils sont, en outre, plus ou moins spécialisés selon la forme ou même la phase de l'action.

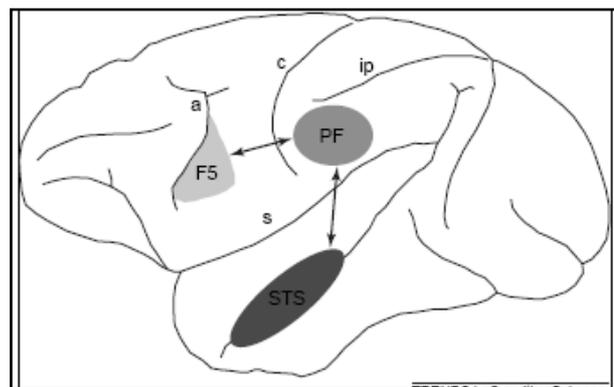


Figure 66 : Localisation chez le macaque des trois zones de neurones spécialisés dans la cognition sociale : neurones dits « miroirs » et neurones dits « canoniques », dans le cortex pré-moteur F5 (analogue de Broca) et la région PF du lobule pariétal inférieur (7b analogue de BA40 ou gyrus supramarginal) ; et neurones seulement « canoniques » dans le sulcus temporal supérieur STS. (a, c, ip et s désignent respectivement les sillons ou sulci arqués, central, intrapariétal et de Sylvius). D'après Keysers & Perrett, 2004, p. 501.

D'autres études ont permis de préciser ces premiers résultats en montrant que l'activation de certains groupes de neurones dans la zone F5 était associée spécifiquement aux mouvements buccaux. Ferrari *et al.* (2003) ont identifié différents groupes fonctionnels de neurones. Si la plupart d'entre eux répondent aux actions d'ingestion (ingestion de nourriture ou liquide, comme illustré dans les deux premières paires d'images, Figure 67), un tiers de ces neurones répondent également aux actions de communication (imitation de mouvements buccaux tels que la protrusion des lèvres ou de la langue, *lipsmack* ou claquement des lèvres, Figure 67).



Figure 67 : Illustration des mouvements buccaux reproduit par le singe macaque à partir des mouvements réalisés par l'expérimentateur. De bas en haut : saisie d'un morceau de nourriture, succion du jus d'une seringue et protrusion des lèvres. D'après Ferrari *et al.*, 2003.

La découverte des neurones miroirs chez le singe est importante car elle nous permet d'entrevoir l'imitation comme un outil de compréhension des actions d'autrui. Ces neurones de l'imitation coordonnent en effet la *perception de l'action* chez autrui (action parfois complexe) et le *contrôle de l'action* (la reproduction de cette même action). En ce sens, ils assurent une fonction de *représentation de l'action* par le recours à une modalité commune aux deux individus : la modalité motrice.

Ainsi, pour reprendre la terminologie employée par Köhler, l'*action* consiste pour le singe à reproduire à l'identique celle du modèle grâce à la copie du *chemin*, c'est-à-dire la coordination gestuelle la plus appropriée pour l'atteinte du *but*.

D'autres neurones miroirs ont été observés chez le singe macaque dans le lobule inférieur du cortex pariétal (Gallese *et al.*, 2002). Et d'autres neurones canoniques (mais pas miroirs)

dans le Sulcus Temporal Supérieur (STS) en ce qui concerne les actions manuelles (Perret 1989) et la cognition sociale (Keysers & Perrett, 2004).

III.1.3. Le défi du contrôle volontaire oro-laryngé dans la production de vocalisations oro-faciales chez Charlie et Patty de Parme, deux macaques à queue de cochon

Le contrôle volontaire de la vocalisation a été très longtemps dénié aux singes inférieurs. Pourtant dès les années 1970, Sutton *et al.* (1974) obtinrent des résultats sur le contrôle de la durée de « coos » dans le cortex cingulaire antérieur. Très récemment, Coudé *et al.* (2006) ont identifié dans la région F5 une population de neurones répondant uniquement en condition de production de vocalisations volontaires. Deux macaques à queue de cochon (*macaca nemestrina*) Charlie et Patty ont été entraînés à produire une vocalisation d'affiliation, un *coo*, via une procédure de rétribution par la nourriture. Cette vocalisation est généralement d'une durée d'environ 200 ms, voisine de la taille d'une syllabe (*cf.* Figure 68, « sonogram of a coo call »). De manière intéressante pour notre étude sur l'arrondissement, cette vocalisation comporte ce type de contrôle, et même, elle est caractérisée par une anticipation de protrusion. Grâce à l'obligeance de Gino Coudé, Francesca Rodà et Leonardo Fogassi, Christian Abry a eu l'occasion, dans l'automne 2006, d'assister à une séance de test de Charlie. Tout d'abord Charlie réalise une légère protrusion et une constriction (rapprochement) des lèvres. La figure 68 donne à voir la différence de configuration orofaciale du singe en condition de vocalisation (ou de tentative de vocalisation) et en condition de repos : on note très clairement la différence de configuration labiale entre ces deux situations. Puis Charlie fait vibrer ses cordes vocales par une expulsion de l'air de ses poumons (« *chest pulse* »). Nous n'avons pas d'information sur la configuration de sa langue et il semble que l'attaque de ses « coos » soit plus une attaque glottique qu'une détente de type [k] : cela ressemble davantage au son produit en soufflant sur une chandelle (pour l'estimation de la longueur du conduit vocal et sa fonction avec ce même type de vocalisation, chez les rhésus, *cf.* Ghazanfar *et al.*, 2007).

En fait ce processus de vocalisation implique la maîtrise par Charlie du contrôle coordonné de trois systèmes de l'appareil vocal lui permettant de générer un tel « *coo* » :

- I. la mise en place de la configuration du conduit vocal supra-glottique, dont nous observons surtout les lèvres ;
- II. la mise en place de la configuration glottique (au larynx) ;
- III. pour que le passage de l'air des poumons permette aux cordes vocales de vibrer.

On retrouve la *coordination orolaryngée* essentielle chez l'homme pour produire la parole, ne serait-ce qu'une voyelle après pause, comme nos [y].

Chose importante : il semble que cette coordination connaisse son déroulement naturel une fois déclenchée. En bref, il n'est pas sûr qu'en dehors de la nécessité d'avoir obtenu une configuration supra-glottique appropriée pour un « *coo* », Charlie contrôle le timing de l'anticipation de ses lèvres par rapport à son larynx. Mais quoi qu'il en soit, le timing naturel de cette production sera essentiel pour obtenir l'intégration bimodale audio-visuelle, ainsi que nous le verrons plus loin.

On notera que Charlie, tout particulièrement dans les conditionnements hors cage, ne produit pas systématiquement des vocalisations volontaires voisées. De même pour Patty. Par contre, ce que l'équipe de Parme a étiqueté comme tentative de vocalisation (*attempt to vocalize*) est en réalité une réussite du positionnement anticipant de la configuration du conduit vocal, mais sans effet Bernoulli (voisement), les cordes vocales n'étant sans doute pas suffisamment rapprochées : c'est là la seule partie de « l'échec » dans cette coordination. En relisant les données de cette manière, on s'aperçoit que Patty anticipe la protrusion dans toutes les conditions, en cage comme avec la tête libre ou fixée (Figure 69), soit dans plus de 80% des cas. Charlie a, lui, autant de réussites que de tentatives.

L'activité neurale des aires F1, F4, F5 du cortex a été enregistrée de même que les sons émis. « *The preliminary data on vocalization suggest that lateral F5 contains a population of neurons that can control voluntary vocal production.* » (cf. Figure 67) L'activité neurale varie en fonction des différentes conditions d'expérimentation. On remarque une forte activité, traduite par des *spikes* précédant le signal voisé, en condition de production de vocalisation. Mais ce n'est pas le cas en condition *play-back*, lorsqu'on rejoue sa vocalisation au singe (Figure 68). L'aire F5 est donc sollicitée lors de la production de mouvements buccaux (Ferrari *et al.*, 2003 ; Rizzolatti *et al.*, 2001) mais également lors de la production de vocalisations volontaires.

Configuration labiale

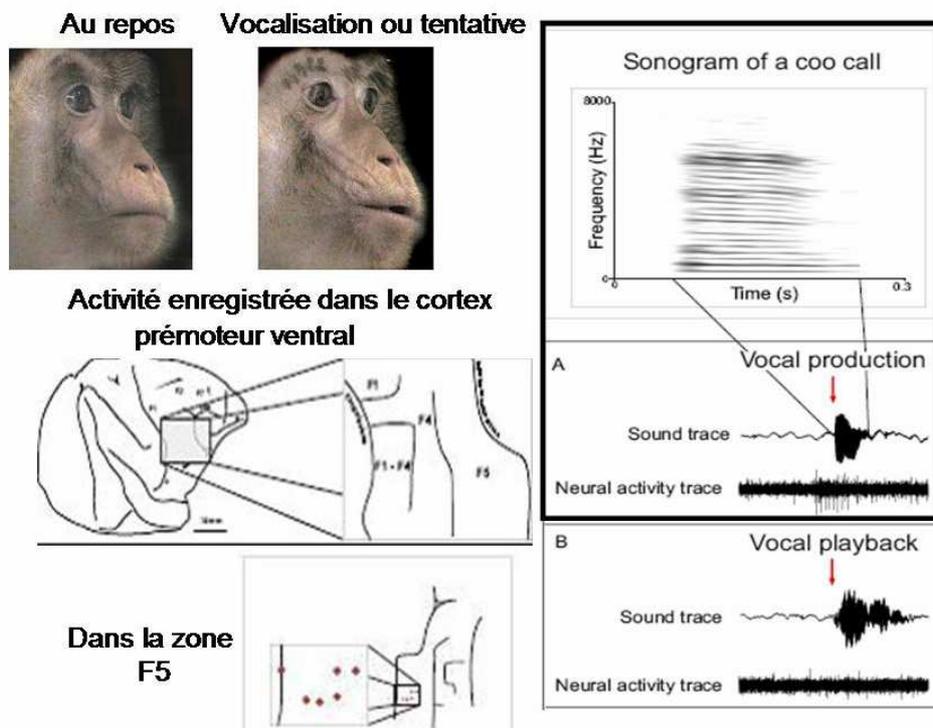


Figure 68 : Configuration labiale du singe en condition de repos et en condition de vocalisation ou tentative de vocalisation (en haut), présentation de la région où l'activité neurale a été enregistrée (régions F1, F4, F5 du cortex pré-moteur ventral (en bas), tracé de l'activité neurale en condition de production volontaire ou de play-back (à droite). D'après Coudé *et al.*, 2006.

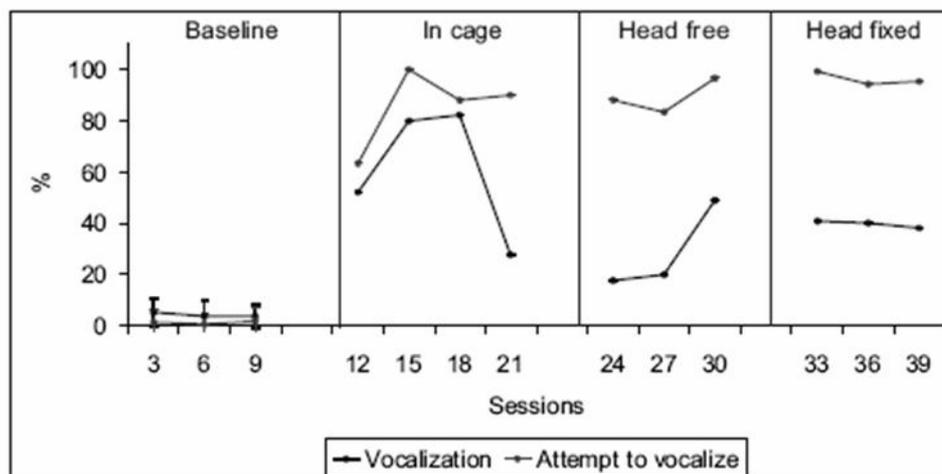


Figure 69 : Pourcentage de vocalisations (lignes du bas) ou de tentatives de vocalisations (lignes du haut) réalisées en fonction du nombre de sessions d'enregistrement dans quatre conditions : avant le test en condition de base, en cage, en ayant la tête libre ou la tête fixée (en bas). D'après Coudé *et al.*, 2006.

Il y a donc une réelle volonté de la part de ces macaques *nemestrina* de produire la vocalisation à laquelle ils ont été conditionnés, avec bien sûr, selon les conditions et le nombre de sessions d'entraînement, une variabilité des résultats. Lorsqu'un macaque parvient

à ce but avec les moyens de sa coordination poumons-larynx-conduit vocal, les auteurs observent une activation spécifique des neurones de l'aire F5. D'autres travaux ont trouvé des éléments d'une représentation des mouvements buccaux et du larynx dans la partie ventrale de l'aire F5 (Hast *et al.*, 1974 ; Gentilucci *et al.*, 1988). Ce travail en cours pourrait donc par la suite venir conforter les données cumulatives de Ferrari *et al.* (2003) sur un site où se situent les *neurones miroirs* de l'imitation volontaire.

III.1.4. Vers un ancrage phylogénétique de la parole ?

Ce nouveau résultat nous paraît des plus fascinants car il ouvre la porte à de nouvelles interprétations quant à l'ancrage phylogénétique de la parole. En effet, il est désormais presque unanimement admis que la région F5 d'activation de ces neurones (et ses subdivisions) correspond anatomiquement chez l'homme à l'aire de Broca (Rizzolatti *et al.*, 1996). Cette assertion suscite encore quelques interrogations, soulevées notamment par Ghazanfar & Miller (2007). Ces derniers suggèrent que le moyen le plus sûr d'accréditer le lien entre F5 et Broca serait de démontrer une activité de F5 en termes de production de vocalisation.

« The best evidence that monkey PMv [cortex prémoteur ventral] is related to Broca's area would be a demonstration that it is involved in the production of vocalization. » (Ghazanfar & Miller, 2007, p. 880)

Il semblerait donc que les résultats pionniers de Coudé *et al.* (2006) fournissent une première réponse à cette interrogation de Ghazanfar & Miller (2007). Est-ce donc une coïncidence que l'on retrouve chez l'homologue de Broca une activité neurale liée à la communication volontaire chez le macaque? A partir de ces résultats, peut-on inférer de nouvelles hypothèses quant à la possibilité d'un langage articulé chez le singe ? Nous n'avons pas la prétention de défendre à partir de ces premières données l'hypothèse d'un langage émergent à partir de ces vocalisations primitives. Cette étude nous donne simplement l'opportunité de bâtir un pont entre la capacité initiale du primate à reproduire des gestes manuels et buccaux contrôlés pour des actions dites intransitives ou de communication et l'extension du contrôle de ces gestes – avec les propriétés des neurones miroirs responsables de cette relation perception-action – pour la production de vocalisations plus « ouvertes » non liées à des appels¹⁶ en situation d'urgence. Bien entendu, les vocalisations de Charlie ou Patty

¹⁶ Il existe de nombreuses vocalisations de type 'appels' à travers les différentes espèces de primates (entre 12 et 15 chez les macaques Rhesus selon Ghazanfar & Miller, 2007). Un répertoire d'appels peut être constitué par exemple pour avertir de la venue de différents prédateurs.

ne correspondent pas aux coordinations articulatoires complexes que l'on observe chez l'homme dans la production de syllabes CV. De fait, si les « *coos* » semblent être des gestes *discrets*, les *nemestrina* connaissent aussi des gestes répétitifs, cycliques comme les *lipsmacks* (Maestriperi, 2005). La liaison entre gestes cycliques et gestes discrets n'est certes pas évidente pour passer des vocalisations des macaques à la parole. Coudé *et al.* (2006) demeurent d'ailleurs prudents car, selon eux, ces premières données indiquent clairement qu'il s'agit bien d'un apprentissage d'une utilisation volontaire et non d'un processus de production courant chez Charlie et Patty, qui connaissent comme tous leurs congénères, ces vocalisations sous leurs formes affectives. Ainsi en dépit du fait que ces neurones de F5 soient liés à la vocalisation volontaire :

« *However, both behavioural and neuronal data seem to indicate that this control is an emerging property, probably reflecting a first step in the evolution of speech production.* » (Coudé *et al.*, 2006)

Pourrait-on alors imaginer que ces neurones miroirs, via l'imitation, puissent conduire à l'acquisition de vocalisations plus complexes chez le macaque et par extension aient favorisé l'évolution du langage chez l'homme? Pour tenter d'apporter un éclairage sur cette question, il est important d'indiquer que, parallèlement à ces travaux, d'autres études ont démontré l'existence d'un système de neurones miroirs similaires chez l'homme, liant également l'*observation* et l'*action* (Fadiga *et al.*, 1995). Ceux de Buccino *et al.* (2001) ont notamment signalé une activation de ces neurones dans les aires prémotrices et pariétales en situation d'observation et d'exécution d'actions motrices transitives ou intransitives, telles que des gestes orofaciaux (buccaux), manuels (préhension) ou encore des actions reliées au pied (par exemple frapper un ballon). Les résultats obtenus suggèrent que le système de neurones miroirs dédié aux actions buccales et manuelles chez l'homme est principalement localisé dans l'aire de Broca.

Dans un article intitulé « *Language within our grasp* », Rizzolatti & Arbib (1998) mettent en relation cette activité motrice – repérée dans l'aire F5 chez certains primates non-humains et son homologue chez l'homme – avec l'émergence de la parole d'un point de vue phylogénétique et également ontogénétique. A partir de données d'imagerie cérébrale, ils s'interrogent donc sur le rôle du système de neurones miroirs dans le développement de la parole :

« *Humans and monkey data indicate that, in primates, there is a fundamental mechanism for action recognition. We argue that individuals recognize actions made by others because the neural pattern elicited in their premotor areas during action observation is similar to that internally generated to produce that action. This mechanism in humans is circumscribed to the left hemisphere. [...] Our proposal is*

that the development of the human lateral speech circuit is a consequence of the fact that the precursor of Broca's area was endowed, before speech appearance, with a mechanism for recognizing actions made by others. » (Rizzolatti & Arbib, 1998, p. 190)

D'une *transmission* volontaire par le geste, comme nous le montre l'étude déjà citée de Whiten *et al.* (2005), reposant sur un processus triangulaire primitif orienté vers l'action – reconnaissance-interprétation-reproduction de l'action (mouvements manuels ou orofaciaux) –, l'homme aurait développé une forme plus élaborée de communication grâce au contrôle progressif du système de neurones miroirs orienté vers la transmission vocale intentionnelle. L'évolution d'un tel système serait expliquée par les auteurs de la manière suivante. Lorsqu'un individu se trouve en situation d'exécution d'une action ou bien d'observation de cette action, les aires prémotrices sont sollicitées. Dans ce dernier cas, même si l'individu est doté d'un système inhibiteur qui lui permet de ne pas reproduire systématiquement l'action observée, il lui arrive parfois d'en reproduire un fragment lorsque cette action revêt une importance particulière pour lui. Ce comportement a alors des implications fortes pour chacun des deux individus:

« The actor will recognize an intention in the observer, and the observer will notice that its involuntary response affects the behavior of the actor. The development of the capacity of the observer to control his or her mirror system is crucial in order to emit (voluntarily) a signal. When this occurs, a primitive dialogue between observer and actor is established. This dialogue forms the core of language. [...] This new association should have yielded enormous benefits of adaptive value for the group of individuals that have started to make use of it, providing the selective pressure for the extension of communicative capacities to larger groups. » (Rizzolatti & Arbib, 1998, p. 191)

Nous proposons dans la Figure 70 une décomposition du processus de communication intentionnelle, telle que formulé par Rizzolatti & Arbib (1998).

Selon les auteurs, l'individu aurait donc bénéficié d'un système spécialisé dans certaines actions motrices (une activation de l'aire de Broca a plusieurs fois été rapportée pour des actions manuelles, *cf.* Bonda *et al.*, 1994) et il aurait ensuite progressivement adapté le mécanisme de perception-action à la production de parole.

Chez les primates non-humains, ces premiers actes de communication correspondraient aux gestes orofaciaux dont certains chercheurs ont fait état, en particulier les mouvements de claquement des lèvres et de la langue (*lipsmacking, tonguesmacking*) parfois comparés à l'oscillation mandibulaire comme cadre primitif de la parole chez l'enfant (MacNeilage, 1998).

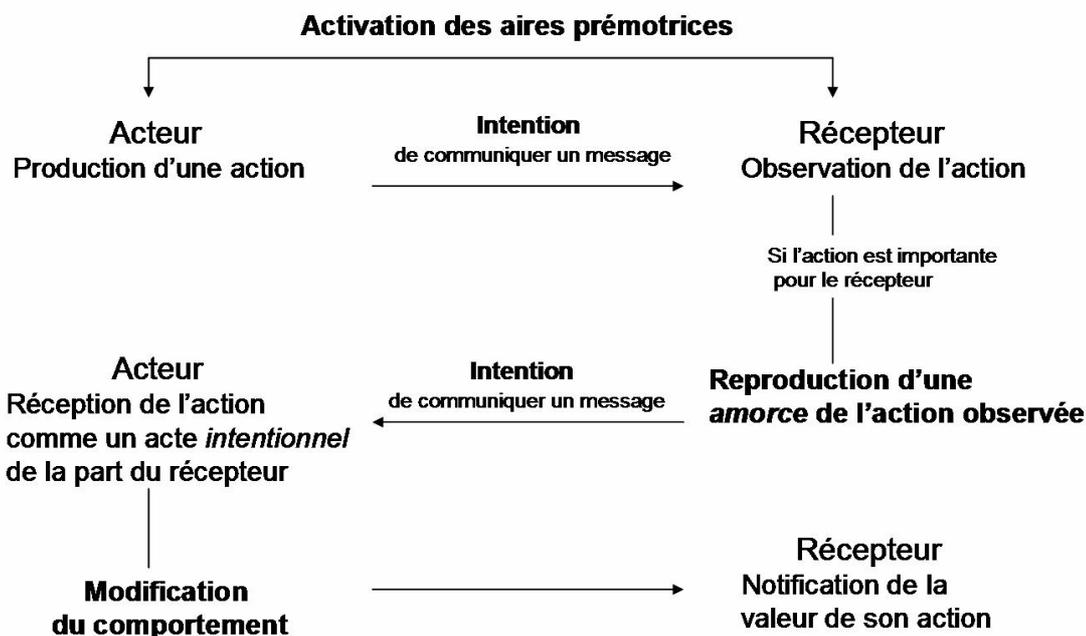


Figure 70 : L'émergence de la communication intentionnelle. D'après Rizzolatti & Arbib, 1998.

Nous retrouvons à travers les hypothèses formulées par Rizzolatti & Arbib (1998) l'idée d'un contrôle *émergent* de ces neurones – tel qu'observé par Coudé *et al.* (2006), chez les macaques *nemestrina* – qui aurait contribué au développement de nouvelles formes de communications intentionnelles. Notons également l'importance de la *diffusion* du comportement, s'établissant progressivement comme la norme, ainsi qu'Whiten *et al.* (1999, p. 683) l'ont signalé à une plus petite échelle comme une *transmission culturelle* dans une communauté de chimpanzés.

Est-ce que l'émergence du langage chez l'homme reposerait à la fois sur un mécanisme de contrôle progressif du système de neurones miroirs, orientant l'imitation gestuelle vers la communication vocale intentionnelle – ce qu'Arbib définit comme le 'chaînon manquant' (*the missing link*) – à l'échelle *phylogénétique*, ainsi que sur la propagation à l'échelle *historique* de ce nouveau mode de compréhension du congénère?

Arbib (2005) propose un scénario retraçant l'évolution phylogénétique du langage et dans lequel l'imitation aurait joué un rôle fondamental pour le passage d'un système de communication manuelle basé sur l'usage de proto-signes (*proto-sign*) vers le développement d'un système plus complexe orienté vers la communication vocale appelé proto-langage (*proto-speech*), précurseur du langage moderne de l'homme. D'après lui, les premiers *Homo sapiens* possédaient une aptitude au proto-langage de par les spécificités de leur cerveau (*language-ready-brain*), mais ne disposaient pas pour autant du langage complexe de

l'homme moderne (Arbib, 2005, p. 9). Arbib décompose cette évolution en sept étapes que nous avons synthétisées dans le Tableau ci-après.

STADES PRE-HOMINIDES	Evolution du système de neurones miroirs (F5)
Mécanisme de saisie	<p data-bbox="1072 421 1305 510">Système primitif de neurones miroirs dédiés à la saisie</p>
observation-exécution d'actions simples de saisie	
<p data-bbox="280 474 922 609">Système d'imitation primitif, i.e. imitations manuelles simples, orientées vers l'objet et sous-tendues par une exposition prolongée. Rôle de l'expérience dans la puissance du système.</p>	<div data-bbox="1050 645 1327 833" style="border: 1px solid black; padding: 5px; margin: 10px auto; width: fit-content;"> <p data-bbox="1066 654 1311 824">Contrôle progressif des neurones miroirs vers la reproduction de gestes orientés vers la communication orofaciale</p> </div> <div data-bbox="1066 1438 1321 1563" style="border: 1px solid black; padding: 5px; margin: 10px auto; width: fit-content;"> <p data-bbox="1114 1447 1279 1554">Processus de récursivité du système de neurones miroirs</p> </div>
PRIMATES (great apes)	
<p data-bbox="280 663 976 824">Mise en place d'un système d'imitation complexe : capacité de reconnaissance des actions d'autrui comme un ensemble d'unités familières, avec l'ajout éventuel de nouvelles actions pouvant être réalisées avec des variantes de ces unités familières.</p> <p data-bbox="280 824 954 860">Utilisation de l'imitation à des fins de communication.</p>	
<p data-bbox="280 869 960 1003">Emergence du proto-signe : système de communication essentiellement manuel permettant le passage d'un ensemble restreint de vocalisations à un répertoire plus conséquent.</p>	
<p data-bbox="280 1012 986 1106">Evolution vers le proto-langage : cette étape se manifeste par l'acquisition progressive du contrôle de l'appareil vocal pour une extension des vocalisations</p> <ul style="list-style-type: none"> <li data-bbox="280 1106 673 1137">- système d'imitation complexe <li data-bbox="280 1137 992 1200">- création de symboles liés à un éventail ouvert d'actions et de situations <li data-bbox="280 1200 970 1263">- équilibre entre ce qui compte pour le locuteur et pour le récepteur. Possibilité de partage <li data-bbox="280 1263 954 1326">- communication intentionnelle : le message véhiculé a pour but de produire un effet sur le récepteur <li data-bbox="280 1326 970 1397">- structuration hiérarchique des actions en petites unités et organisation temporelle appropriée de ces actions 	
HOMO SAPIENS via l'évolution culturelle et historique	
<p data-bbox="280 1451 986 1550">Emergence du langage : modification de la communication orientée d'une relation objet-action vers un système de langage basé sur des structures verbe-argument</p> <p data-bbox="280 1550 954 1581">Apparition d'une organisation temporelle des actions</p> <ul style="list-style-type: none"> <li data-bbox="280 1581 986 1644">- création de symboles devenant des mots porteurs de sens <li data-bbox="280 1644 992 1706">- apparition des structures syntaxiques, sémantiques et de la récursivité <li data-bbox="280 1706 992 1769">- développement des notions de temps liées aux verbes pour la création de repères temporels <li data-bbox="280 1769 794 1800">- faculté de transmission et d'apprentissage 	

Tableau 13 : Récapitulatif des sept étapes proposées par Arbib (2005) pour une évolution du langage.

Ce dernier défend l'idée que le *proto-langage* aurait évolué vers le langage grâce à l'évolution du contrôle de l'activité cérébrale par le biais d'un long processus d'expansion

historique et culturel des modes de communication et de compréhension des actions et non en raison d'un développement anatomique du cerveau et de l'appareil vocal conférant à l'homme la capacité de créer le langage.

« *Only protolanguage co-evolved with the brain, and that the full development of linguistic complexity was a cultural/historical process that required little or no further change from the brains of early Homo sapiens.* » (Arbib, 2005, p. 11)

Nous ne souhaitons pas ici rendre compte du débat sur les possibilités vocales de l'*Homo sapiens*, mais il nous semble néanmoins important de rappeler que des analyses attentives de crânes Néanderthalien menées conjointement en linguistique et en anthropologie depuis plus d'une décennie confortent la proposition d'Arbib (Heim, 1989 ; Heim *et al.*, 2002). En effet, en dépit d'assertions contraires, communément trouvées dans la littérature et encore enseignées à travers le monde, l'homme de Neandertal bénéficiait de tous les pré-requis anatomiques pour la production du langage. Nous savons désormais que la thèse proposée par Lieberman *et al.* (1972) sur une émergence du langage fortement liée aux transformations du larynx et du pharynx, formulée à partir des observations pionnières de Boule (celui-ci ne disposant pas à l'époque de données comparatives) est invalidée par le placement erroné du centre de gravité de la tête. Les proportions du crâne de Neandertal ayant été depuis rétablies par Heim *et al.* (2002), celles-ci démontrent que :

« *La position de la tête et de la base du crâne au dessus de la colonne vertébrale ne différaient guère fondamentalement de celle de l'homme (Heim 1989, 1990)¹⁷.* » (Heim *et al.*, 2002, p. 2)

Nous ne discuterons pas des implications ontogénétiques de la proposition de Lieberman et Crelin quant à l'anatomie « inhospitalière » du nouveau-né vis-à-vis de la parole, mais celles-ci ont également été infirmées par de nombreux travaux de modélisation démontrant la capacité du nouveau né à articuler les sons de la parole, notamment les voyelles cardinales [a, i, u]. Pour une revue de ces études et des modélisations, voir Maeda (1989), Boë (1999) et Ménard (2002).

Pour en finir sur les possibilités vocales du primate, les différentes mesures crâniennes suggèrent donc que les limites du langage, du temps de l'*Homo sapiens*, ne sont très probablement pas liées à des contraintes anatomiques pesant sur l'appareil vocal mais peut-être plutôt à une évolution de l'activité neurale :

¹⁷ Ces nouvelles conclusions ont été confortées par Arensburg en 1989 à partir de la découverte d'un os hyoïde Néandertalien dont les dimensions sont identiques à celle de l'homme moderne.

« *If monkeys do not talk, according to present evidence, this is due to a lack of appropriate cortical equipment (perhaps through lack of differentiation of control between the larynx and articulators), and a lack of sufficient cognitive capabilities. The brain is entirely capable of controlling a vocal instrument with a somewhat longer or shorter pharynx: these differences do not actually change the capacity of contrasting maximally vowels.* » (Boë, 1999, p. 2504)

La découverte récente d'une coordination orolaryngée contrôlée dans F5 chez Charlie et Patty (plus âgés de presque vingt millions d'années par rapport à Neandertal) est d'autant plus essentielle qu'elle suggère non seulement que ces derniers disposent d'un équipement anatomique conforme à la production de vocalisations s'apparentant à une voyelle (*coo call*), mais qu'ils disposent également de cette capacité de *différentiation de contrôle entre la position du larynx et les mouvements des articulateurs supra-laryngés* (dont le siège neural est identique chez l'homme).

Que manque-t-il réellement à Charlie et Patty pour acquérir le langage de l'homme moderne ? Hauser *et al.* (2002) suggèrent que l'absence de structure hiérarchique de la communication, désignée autrement par la capacité de *récurtivité* du langage, est une caractéristique fondamentale qui n'est retrouvée dans aucune autre forme de communication chez les non humains. Toutefois, ces auteurs prennent en considération les apports récents des neurosciences et proposent le processus d'évolution suivant :

« *One possibility, consistent with current thinking in the cognitive sciences, is that recursion in animals represents a modular system designed for a particular function (e.g., navigation) and impenetrable with respect to other systems. During evolution, the modular and highly domain-specific system of recursion may have become penetrable and domain-general. [...] This change from domain-specific to domain-general may have been guided by particular selective pressures, unique to our evolutionary past, or as a consequence (by-product) of other kinds of neural reorganization.* » (Hauser *et al.*, 2002, p. 1578)

Une autre proposition intéressante est formulée par Philps (2000, 2003) à travers sa théorie d'un langage reposant sur le signe (*language sign theory*). Reprenant les conclusions de Corballis (2002) sur la filiation du langage à partir du contrôle des gestes de la main, puis de la bouche, et celles de Gentilucci *et al.* (2001) sur l'influence réciproque des gestes de préhension et des mouvements de la main chez l'homme, Philps postule que les gestes articulatoires émergeant des gestes manuels auraient bénéficié d'une nouvelle fonction *déictique*, c'est à dire de référence à ces gestes manuels ou plus généralement aux mouvements du corps (*body-naming strategy*).

« *If open-close hand gestures were indeed transformed into open-close articulatory gestures, then the latter could have served to refer back to these hand gestures*

deictically, and to stand for them symbolically by means of unconscious, neurophysiologically grounded, cognitive body-naming strategy. The processes involved in this putative strategy appear to include self-reference (Philps 2000: 217), vocomimesis (Donalds 2001: 291) and conceptual mapping (Lakoff 2003: 246). » (Philps, 2006, p. 248)

A partir de l'analyse d'un corpus de langues Proto-Indo-Européennes et Indo-Européennes, Philps rend compte d'un processus de vocomimesis des signes du langage, organisés autour de groupes de consonnes dont la forme et les propriétés sémantiques sont récurrentes.

« The language sign was originally configured vocomimetically during a period in the evolution of H. sapiens when the oral apparatus, originally used for purposes of nutrition, respiration and visuofacial communication, began to be employed additionally for articulatory purposes. One major assumption of this theory is that the initial conditions of a system largely determine its subsequent conditions, though not exclusively so. Moreover, whereas the linguistic sign is arbitrary by definition and by conception [en référence à Saussure], the language sign is envisaged as having become arbitrary. » (Philps, 2006, p. 248)

Ces groupes de consonnes correspondent à des préfixes consonantiques appelés des 'marqueurs sublexicaux' (*sublexical markers*) tels que 'gr-' ou 'sn-' en anglais et qui déterminent le champ sémantique des mots composés de ces préfixes. Par exemple, pour le préfixe 'gr', les mots 'grasp', 'grip', 'grope' ont un lien sémantique avec la notion de saisie. Philps propose de diviser l'ensemble des marqueurs sublexicaux contenus dans des langues Proto-Indo-Européennes en deux grands domaines sémiologiques : l'oralité (avec des préfixes évoquant les mouvements de la bouche) et la manualité (ceux des mains). Par ailleurs, il défend que l'existence des occlusives, dans la majorité des langues Proto-Indo-Européennes, découlerait de gestes articulatoires d'occlusions, initialement réalisés pour des buts d'ingestion, puis détournés vers la communication pour constituer les cadres syllabiques *invariants* sur lesquels viennent se greffer les gestes dynamiques de contenu. Sa théorie rejoint donc les propositions de MacNeilage (1998) sur l'évolution du langage.

« By this view, human speech seen as a motor function emerged when the cyclical, open-close alternations characteristic of ingestive activities underwent a series of sequences adaptations resulting in them being employed in content-modulated syllabic frames for purposes of visuofacial and phonatory communication. » (Philps, 2006, p. 251)

« Such a [hominid] protosyllable can be viewed as a gesture, that is as constriction and release of one of the vocal organs, set in the context of an overall vocal tract posture and combined with phonatory action. » (Studdert-Kennedy & Goldstein, 2003, p. 252)

III.1.5. *L'attunement ou l'appariement progressif de soi avec le monde*

Notre objectif n'était pas de valider ou infirmer les hypothèses relatées dans ce qui précède. Il va de soi que l'état de l'art dans l'approche évolutionnaire du langage, que ce soit en linguistique ou en psychologie est encore tributaire de bien plus de spéculations que de théories testables chez le primate non-humain et l'enfant. Il s'agissait et il s'agit toujours pour nous de faire émerger une composante commune aux différents travaux, une composante qui nous paraît tout à fait cruciale pour le développement du contrôle vocalique, et plus particulièrement de l'anticipation d'arrondissement vocalique. C'est parmi les paramètres de l'apprentissage, la faculté de réglage de son comportement en fonction de celui d'autrui, autrement dit la faculté d'*accorder* son comportement sur celui de l'autre, en un mot l'*attunement*. En effet, que l'on analyse les différentes formes d'imitations rapportées chez les chimpanzés ou que l'on s'intéresse aux vocalisations bimodales des macaques, ces comportements sont le fruit d'un processus de réglage moteur progressif vers un appariement entre soi et le monde des congénères.

Dans l'étude de Whiten *et al.* (2005) portant sur l'extraction de la nourriture, ce processus se manifeste par l'acquisition progressive du contrôle de la coordination manuelle requise pour l'exécution de la stratégie enseignée en modèle. Les auteurs rapportent que chacun des deux modèles a tout d'abord été entraîné à maîtriser l'une des deux méthodes. Puis le reste du groupe a été exposé à la stratégie dispensée par le modèle durant sept jours avant de pouvoir eux-mêmes réaliser l'exercice.

« [...] *the 'two alternatives' methodology shows that learning involves not merely the facilitation of an existing competence, but a capacity to acquire particular local variants of the technique, precisely as required if the behavioural variants identified in wild populations are indeed socially transmitted.* » (Whiten *et al.*, 2005, p. 738)

Ce processus d'*attunement* permet de passer de l'*émulation du but* à l'*imitation appariée en but comme en moyens*. La recherche du principe de parité (cf. section suivante) correspond à la recherche du comportement reconnu comme la norme au sein de la communauté sociale, puisque ceux qui avaient découvert les deux stratégies ont abandonné celle qui n'était pas employée par le reste de la colonie. Dans l'étude de Coudé *et al.* (2006), ce processus d'*attunement* est également implicitement évoqué à travers la phase d'apprentissage proposée par les expérimentateurs. L'analyse des vocalisations produites au fur et à mesure des sessions témoigne d'une volonté de faire concorder son comportement avec celui qui est requis par le conditionnement.

A partir des études de primatologie citées précédemment, la Figure 71 tente d'illustrer le rôle de l'*attunement* par rapport au processus global d'imitation d'un mouvement orofacial ou gestuel.

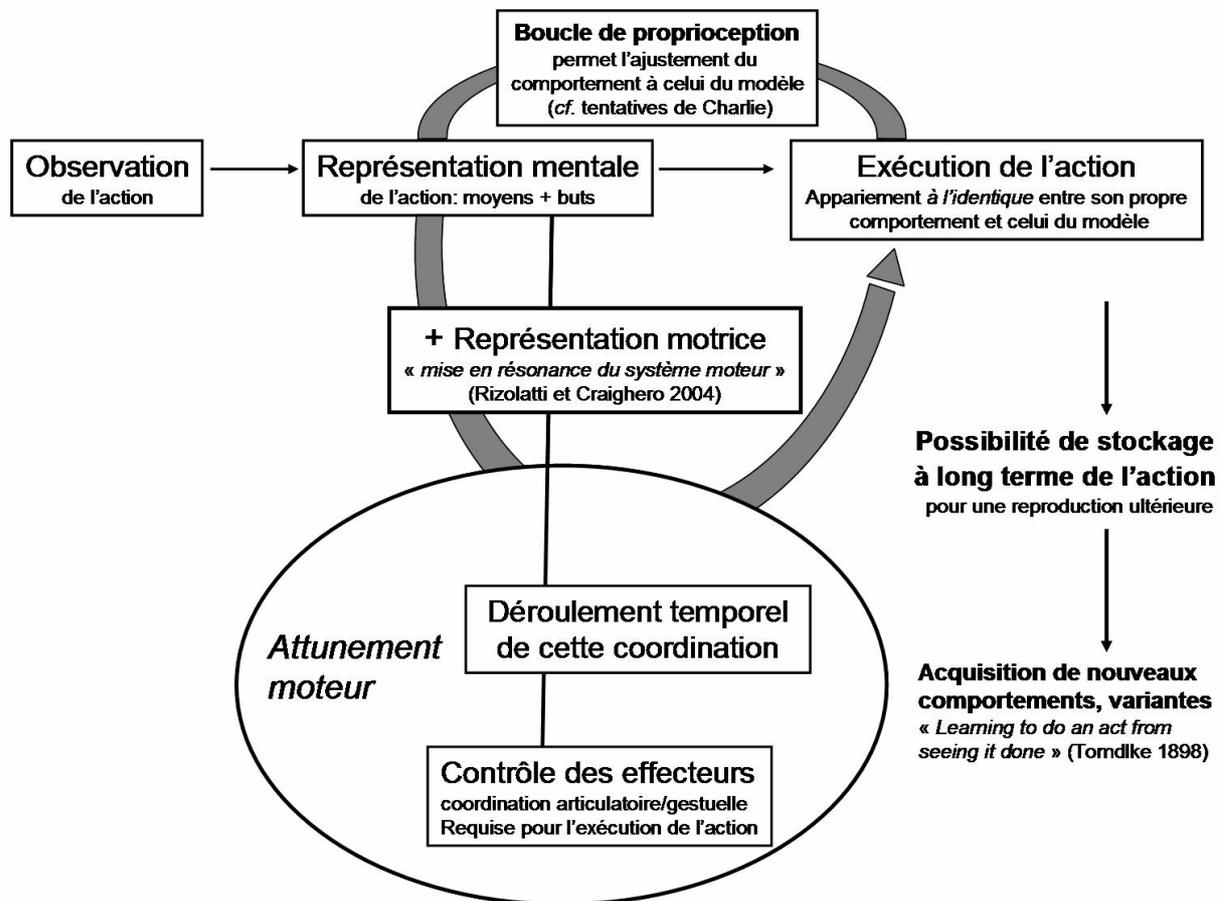


Figure 71 : Illustration du processus d'attunement dans le cadre du mécanisme de perception de l'action et de reproduction de l'action chez le primate d'après les études présentées dans ce chapitre.

Après observation de l'action, une première étape consiste en une représentation mentale de l'action dans laquelle sont déterminés les moyens et les buts à atteindre pour la reproduction de l'action. Cette représentation est accompagnée d'une représentation motrice qui suppose le contrôle des effecteurs nécessaires à la production de l'action ainsi que leur coordination temporelle. Lors de l'exécution de l'action, une boucle de proprioception permet de rectifier son comportement si celui-ci ne concorde pas avec celui du modèle. Dans cette perspective, l'*attunement* correspond au réglage du comportement sur celui du modèle par l'acquisition de contrôles moteurs spécifiques. En d'autres termes, nous pensons que ce processus d'*attunement* permet non seulement l'intégration progressive de nouveaux comportements moteurs, mais également une intégration générale de soi avec l'environnement, soit l'intégration sociale, telle que l'ont décrite Horner et Whiten (Horner & Whiten, 2005 ; Horner *et al.*, 2005 ; Horner *et al.*, 2006).

Le terme d'*attunement* a également été mentionné par Gallese (2006b) comme *intentional attunement* pour expliquer le processus de compréhension de l'autre aussi bien chez les primates non-humains que chez l'humain. D'après lui, l'homme aurait la faculté de partager ses intentions grâce à l'élaboration d'un espace interpersonnel multidimensionnel de communication des intentions (*shared manifold*). Au niveau fonctionnel, cet espace prendrait la forme d'une expérimentation de l'action perçue chez l'autre, c'est-à-dire d'une représentation mentale mais également motrice de cette action (*embodied simulation*).

« *There is also an experiential dimension of interpersonal relationships, which enables a direct grasping of the sense of the actions performed by others, and of the emotions, and sensations they experience. This dimension of social cognition is embodied in that it mediates between the multimodal experiential knowledge we hold of our lived body and the experience we make of others.* » (Gallese, 2006b, p. 2)

Enfin, au niveau neural, le système de neurones miroirs permettrait l'unification entre l'observation de l'action et sa représentation motrice. Nous proposons de représenter schématiquement les différents concepts proposés par Gallese (2006) pour caractériser les interactions sociales chez l'homme (Figure 72). En ce qui concerne plus particulièrement les primates, Gallese défend l'existence d'un espace interpersonnel *rudimentaire* de partage des intentions mais qui, contrairement à l'homme, serait restreint à l'optimisation de situations plus basiques (liées par exemple à la nourriture).

Nous avons souhaité dans cette partie faire état des hypothèses récemment émises dans le domaine des neurosciences. Celles-ci nous fournissent en effet des pistes de réflexions intéressantes quant au développement phylogénétique et ontogénétique de la parole. Elles s'appuient en outre sur les théories de l'acquisition du langage par le jeune enfant pour les confronter avec les comportements observés chez les primates non-humains.

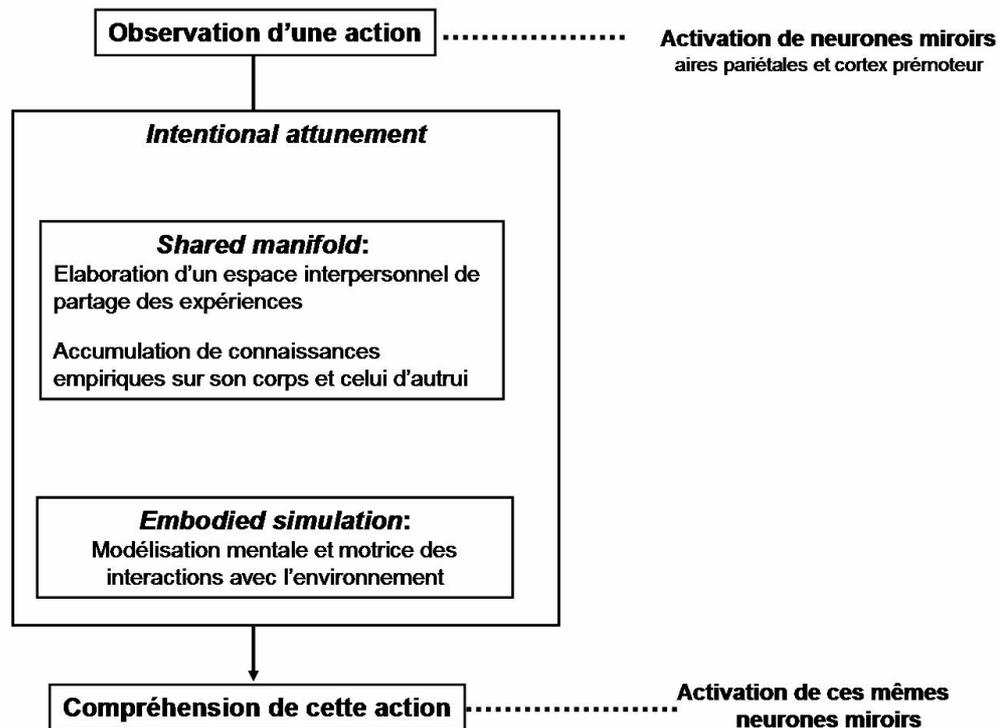


Figure 72 : Représentation schématique des concepts d'*Intentional attunement* et d'*Embodied simulation* élaborée à partir des propositions de Gallese (2001, 2005, 2006) pour rendre compte des interactions sociales chez l'homme.

En particulier, les études venues des neurosciences explorant les implications du système de neurones miroirs vis-à-vis du langage chez les primates n'en sont encore qu'à leurs débuts. Mais nous pensons que les vocalisations volontaires bimodales de Charlie et Patty fournissent un début de réponse encourageant à la question formulée par Arbib (2005), deux ans avant Coudé *et al.* (2006), au sujet d'une possible parenté entre les vocalisations primitives des primates non-humains et le langage articulé de l'homme.

« It could be tempting to hypothesize that certain species-specific vocalizations of monkeys (such as snake and leopard calls of vervet monkeys) provided the basis for the evolution of human speech, since both are in the vocal domain. However, these primate vocalizations appear to be related to non-cortical regions as well as the anterior cingulate cortex (see, e.g., Jürgens, 1997) rather than the F5, the homologue of Broca's area. » (Arbib, 2005, p. 4)

Les vocalisations enregistrées récemment chez Charlie et Patty se distinguent donc fortement des vocalisations recueillies chez d'autres animaux en ce sens qu'elles sont : 1) *volontaires*, c'est-à-dire obtenues après entraînement et non en situation d'interactions avec d'autres primates d'une même colonie ; 2) directement reliées à l'activité neurale de l'aire F5 du cortex prémoteur.

De fait, ces résultats montrent clairement que le contrôle des mouvements orofaciaux peut mener à la production de vocalisations proches de celles de l'homme, et même fonder le timing de base de la production vocalique chez l'homme en fournissant le déroulement temporel approprié des gestes coordonnés des poumons, du larynx et du conduit vocal.

III.1.6. La parité entre production bimodale et perception bimodale de la parole : organisation temporelle d'une coordination audiovisuelle

Comme on peut le lire dans le débat d'Interdisciplines sur le web (www.interdisciplines.org/coevolution/papers/11/7), *Coevolution of language and theory and mind*, dès le début de sa contribution « The Mirror System Hypothesis. Linking Language to Theory of Mind », Michael Arbib rappelle que la parité est la première de ses hypothèses pour utiliser le système des neurones miroirs dans le fonctionnement du langage :

« The original Mirror System Hypothesis (Arbib & Rizzolatti, 1997; Rizzolatti & Arbib, 1998) states that:

H1. The parity requirement for language in humans [that what counts for the speaker must count approximately the same for the hearer] is met because Broca's area evolved atop the mirror system for grasping with its capacity to generate and recognize a set of actions. »

L'exigence de ce principe de parité avait déjà été rappelée dans l'article posthume de Liberman cosigné par Whalen (Liberman & Whalen, 2000) rappelant en premier parmi les grands principes de la théorie motrice de la perception de la parole.

« How was the necessary parity between speaker and listener established in evolution, and how maintained ? »

La découverte de similarités dans le système de traitement de l'information motrice chez certains singes et chez l'homme ne s'arrête pas à l'imitation de mouvements buccaux ou manuels ou encore à la production de vocalisations volontaires bimodales. D'après l'étude de Kohler *et al.* (2002), il existerait dans l'aire F5 des neurones miroirs spécialisés dans le traitement *audiovisuel* du son. Ces neurones répondent uniquement en condition de perception d'un son associé à une action (lorsque l'expérimentateur déchire par exemple une feuille de papier, Figure XX) et non en condition de bruit blanc ou encore de son non associé à une action.

Il ressort également de travaux expérimentaux plus récents, comparant l'activation neurale de macaques rhésus en condition de vocalisations produites par leur congénères, ou de sons biologiques (Gil-da-Costa *et al.*, 2006), ou avec un éventail plus large de sons (parole humaine, vocalisations bruitées, *cf.* Poremba *et al.*, 2004), que:

« *species-specific vocalizations in non-human primates seem to be processed by homologs of the same core regions that process spoken language in humans* » (Gilda-Costa et al., 2006).

D'autres études telles que Barraclough *et al.* (2005) et Keysers *et al.* (2003) ont rapporté un traitement similaire de l'information dans le sulcus temporal supérieur (STS) dont on sait que, chez l'homme, il est activé en condition de parole bimodale (MacSweeney *et al.*, 2000).

Ghazanfar *et al.* (2005) ont récemment testé l'intégration multisensorielle de l'information chez deux singes Rhésus. Une des originalités de cette étude est que les stimuli audiovisuels correspondaient à des vocalisations produites par leurs congénères et non à des stimuli artificiels, comme c'est le cas dans la plupart des études similaires (Stein & Meredith, 1993). Le corpus, composé de vocalisations unimodales *versus* bimodales correspondait soit à des *grunts* soit à des « *coos* ».

Durant les tests, c'est l'activité neurale LFP (*local field potential activity*) des électrodes qui a été mesurée : cette méthode a été choisie par les auteurs en raison de sa comparabilité avec des données similaires obtenues par imagerie par résonance magnétique chez l'homme. Les électrodes ont été placées dans le sulcus temporal supérieur (STS) ainsi que dans les zones de la *core belt* et de la *lateral belt* du cortex auditif (*cf.* Figure 73).

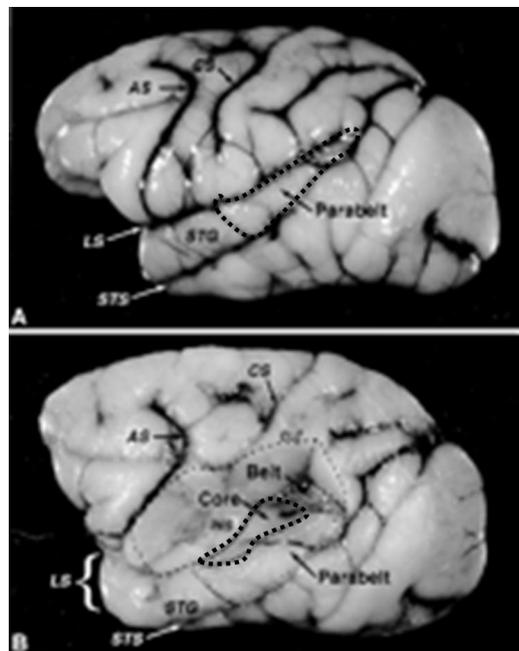


Figure 73 : Vue latérale gauche du cortex cérébral d'un macaque (Kass & Hackett, 2000). En A : le long du gyrus temporal supérieur (STG) la région *parabelt* est entourée d'un pointillé. En B : vue du même cerveau une fois enlevé le cortex pariétal recouvrant la rive ventrale du sulcus latéral et de l'insula. Sont indiquées les localisations approximatives de la région *core* et des portions caudale et latérale de la région *belt* ainsi que de la région *parabelt*. AS : sulcus arqué ; CS : sulcus central ; INS : insula ; LS : sulcus latéral ; STS : sulcus temporal supérieur.

Rappelons que pour ces deux zones, l'existence d'une intégration audiovisuelle n'était pas encore démontrée. Ghazanfar *et al.* (2005) ont observé une intégration multisensorielle en condition d'appariement audiovisuel des stimuli, caractérisée par une forte activation dans le cortex auditif, se manifestant soit par de fortes réponses de renforcement (*enhancement*) ou de suppression de l'activité (Figure 74). En revanche, lorsque les singes étaient soumis à des stimuli unimodaux, auditifs ou visuels seuls, de telles réponses n'ont pas été enregistrées.

Ghazanfar *et al.* (2005) ont cherché quels pouvaient être les paramètres influençant la réponse neurale des deux macaques rhésus testés et ont remarqué que le type d'appel (« *grunt* » ou « *coo call* ») ne constituait pas un facteur de variabilité au niveau des réponses observées. Nous verrons ci-dessous que l'anticipation est plus importante pour les « *coos* ».

Dans la littérature, une des hypothèses discutées est que le type de réponse neurale est déterminé par la congruence entre les stimuli auditifs et visuels mais les opinions sont divergentes à ce sujet (Calvert *et al.*, 2000). Or, à la lueur des résultats obtenus dans leur étude, Ghazanfar *et al.* (2005) proposent une autre interprétation. Des mesures de délai d'établissement du voisement (VOT), réalisées sur les stimuli audiovisuels ont révélé que la relation temporelle entre les signaux visuels (la configuration labiale) et auditifs pouvait prédire le type de réponse observée. Plus précisément, l'intégration multisensorielle des informations bimodales se manifeste majoritairement par des réponses d'augmentation de l'activité neurale lorsque le délai entre le début du mouvement labial et le signal acoustique est court (inférieur ou égal à 100 ms) alors que les réponses de suppression apparaissent lorsque l'intervalle entre les deux signaux est grand, au-delà de 200 ms¹⁸.

Par conséquent, l'activation neurale serait conditionnée par la *congruence temporelle* entre les signaux visuels et auditifs et non la congruence au niveau de l'identité de ces signaux. Le système de traitement bimodal des vocalisations chez les macaques Rhésus semble donc supporter une certaine plage de décalage temporel entre les signaux (entre 100 et 200 ms) au-delà de laquelle l'incongruence est trop forte pour permettre une bonne intégration des informations multisensorielles.

¹⁸ Toutefois, aucune corrélation n'a été décelée entre les VOT (compris entre 66 et 332 ms) et l'amplitude de la réponse observée.

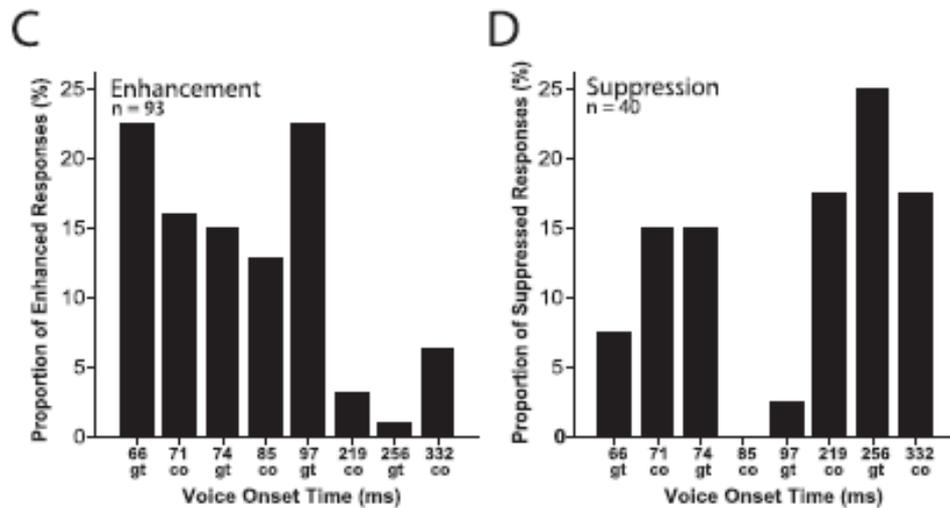


Figure 74 : Proportion des réponses d'augmentation (graphique C) et de suppression d'activité (graphique D) selon les différentes catégories de VOT. D'après Ghazanfar et al., 2005, p. 5008.

Ces nouveaux résultats sont très intéressants car ils illustrent l'existence d'un système de traitement de l'information sensorielle très proche de celui de l'homme. On retrouve d'ailleurs d'autres similarités au niveau de l'organisation temporelle des vocalisations.

« Rhesus monkey vocalizations are characterized by temporal dynamics common to human speech. In particular, the onset of facial movement related to articulation occurs before the onset of the auditory signal. Thus, the use of species-typical vocal signals allows the assessment of neural homologies between nonhuman primate vocal communication systems and human speech. » (Ghazanfar et al., 2005, p. 5004)

Une illustration de ce décalage entre les signaux visuels et auditif est proposée dans la Figure ci-dessous. Celui-ci est plus net dans les « coos » où l'on observe une plus forte avance de la composante visuelle sur le début acoustique de la vocalisation que dans les « grunts ».

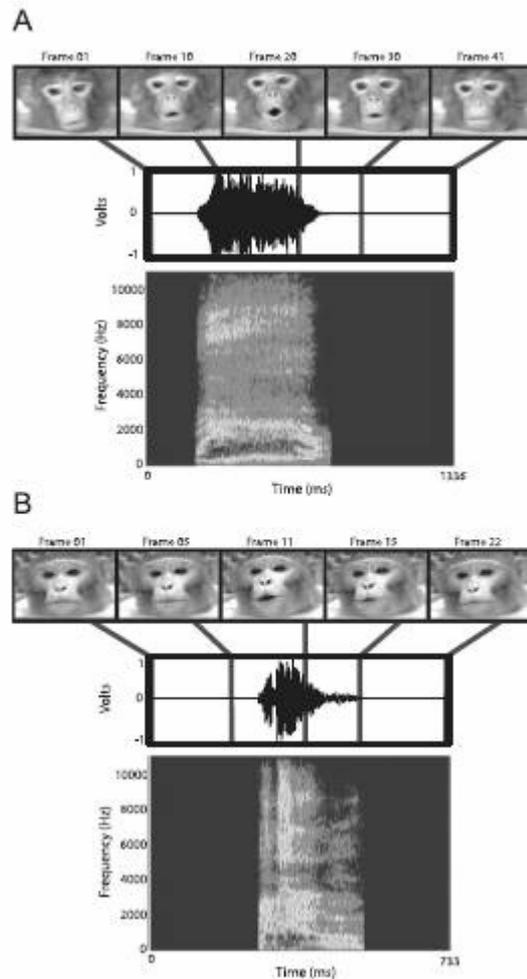


Figure 75 : Illustration des signaux audio et visuels des vocalisations proposés aux singes Rhésus. La figure A correspond au 'coo', l'image B au 'grunt'. Les fenêtres du haut présentent 5 trames vidéo sélectionnées à partir du début du mouvement de la bouche jusqu'à la fin du mouvement. La fenêtre du milieu présente le signal acoustique de la vocalisation sur lequel sont représentées les 5 trames (lignes verticales), accompagné de son spectrogramme (fenêtre du bas). Pour les deux vocalisations, le début du geste labial précède le début acoustique de la vocalisation. D'après Ghazanfar *et al.*, 2005, p. 5006.

Rappelons que Ghazanfar *et al.* (2005) ont su tirer partie de cette coordination oro-laryngée, pour leur mesure de l'intégration neurale bimodale super-additive de ce mouvement d'anticipation des lèvres sur le signal laryngé chez le macaque rhésus, notamment pour ses « coos » (la même anticipation de mise en place du conduit vocal, que nous avons déjà observé chez Charlie et Patty, avant de lancer l'excitation acoustique laryngée de ce conduit). Ils citent pour ce point l'article où Abry et collègues donnent leur conception du timing naturel de la coordination de la configuration du geste supra-glottal vocalique, chez l'homme, en relation avec son excitation glottique (*cf.* aussi Cathiard *et al.*, 1996) et la testent perceptivement :

« Coos are long-duration, tonal calls produced with the lips protruded (Fig. 1A). Grunts are short-duration, noisy calls produced with a more subtle mouth opening

with no lip protrusion (Fig. 1B). Both call types are spectrally rich and wideband. » (Ghazanfar et al., 2005, p. 5007)

Et d'ajouter:

« As in human speech (Abry et al., 1996), the onsets of mouth movements during production of Rhesus monkey vocal signals precede the auditory component. »

C'est donc bien la présence de la même coordination de base qui a été testée dans la fusion des sens, selon le titre du beau livre des pionniers en la matière, Stein & Meredith (1993) : *The Merging of the senses*. (Pour une revue récente de la parcellation du cortex multisensoriel, cf. Ghazanfar & Schroeder, 2006, qui donnent aussi l'état des connaissances sur les connexions neuro-anatomiques nécessaires, notamment celle qui manquait, soit du visuel vers l'auditif, par Cappé & Barone, 2005, après que Falchier *et al.*, 2002, aient découvert l'inverse.)

Nous allons maintenant prendre la place de citer l'important paragraphe où Ghazanfar *et al.* (2005) soulèvent la question essentielle du pourquoi de plusieurs échecs à intégrer supra-additivement l'audio et le visuel, sachant que les principes de la *Gestalt*, présents naturellement dans les structures de la parole, ont été violés par plusieurs de ces expériences.

« Previous neuroimaging studies of multimodal speech suggested that suppression is especially prominent when the speech tokens from the two modalities are incongruent in identity (Calvert et al., 2000). The present data, along with recent human neuroimaging data (Wright et al., 2003; Besle et al., 2004; van Wassenhove et al., 2005), suggest that identity [catégorielle ou mieux configurationnelle, pour ce dernier point cf. Abry et al., 1994] incongruence is not a requirement for response suppression. Recently, two human evoked-potential studies have reported that face plus voice integration is represented only by suppressed auditory N100 responses (Besle et al., 2004; van Wassenhove et al., 2005). This is not supported by our LFP [local field potential] data, in which we found both suppression and enhancement (in fact, more frequently enhancement) to our congruent face plus voice stimuli relative to voice alone. We suggest that the consistently suppressed responses in the N100 component in these human studies are attributable to the very long time interval between the presentation of the face and the voice signal (Besle et al., 2004; van Wassenhove et al., 2005). In both studies, the time between the appearance of the face and the onset of the auditory signal typically exceeded 500 ms. In our data, enhanced responses were primarily seen when this time interval was <100 ms, and suppressed responses were primarily seen at intervals >200 ms. » (Ghazanfar et al., 2005, p. 5010)

En ce qui concerne tout spécifiquement ces limites de violation du principe gestaltiste ou de la loi de groupement temporel, c'est un *caveat* qu'avaient déjà rappelé les pionniers en matière d'intégration neuronale multisensorielle, Wallace, Wilkinson et Stein, il y a déjà une dizaine d'années de cela:

« Generally, multisensory interactions were evident when pairs of stimuli were separated from one another by <500 ms, and the products of these interactions far exceeded the sum of their unimodal components. » (Wallace et al., 1996)

Qu'en est-il de l'intégration multisensorielle chez l'homme ? En comparant les résultats obtenus chez les macaques Rhésus avec les connaissances réunies chez l'homme, on retrouve les mêmes seuils de tolérance à l'intégration des signaux bimodaux de parole.

De fait on a pratiquement les mêmes gammes de valeurs de cohérence gestaltiste naturelle en ms telles que Cathiard *et al.* (1996) les donnent à partir de leurs expériences sur l'humain, soit une frontière perceptive à 210 ms au maximum. Cathiard *et al.* (1992, 1996) ont testé l'étendue de la perception de l'anticipation d'arrondissement vocalique en français dans des séquences du type [i#y] (avec une pause de 160 ms ou 460 ms) selon que les sujets bénéficient seulement d'informations unimodales (visuelles) ou bimodales. Lorsque l'anticipation naturelle des informations visuelles se situe entre 100 et 200 ms, l'individu intègre les deux informations et bénéficie de l'avance des indices visuels (relatifs à la configuration labiale) pour la perception de la cible vocalique suivante.

Pour explorer la non-congruence temporelle au niveau le plus général, plusieurs études de désynchronisation ont été réalisées, recourant à des tâches de détection ou d'intelligibilité (pour une revue, se référer à Summerfield, 1992, Cathiard, 1994). Dans les deux cas, ces études révèlent pour la parole que l'individu est moins enclin à récupérer les indices acoustiques fournies par une avance du son sur le visuel et tolère donc moins cette configuration que l'inverse (Smeele & Sittig, 1991, Smeele *et al.* 1992). En effet, lorsque la désynchronisation du son correspond à la plage [-80 ms, +80 ms], l'identification des mots est équivalente à une condition de production normale de parole, ce qui n'est pas du tout le cas lorsque la désynchronisation atteint -160 ms.

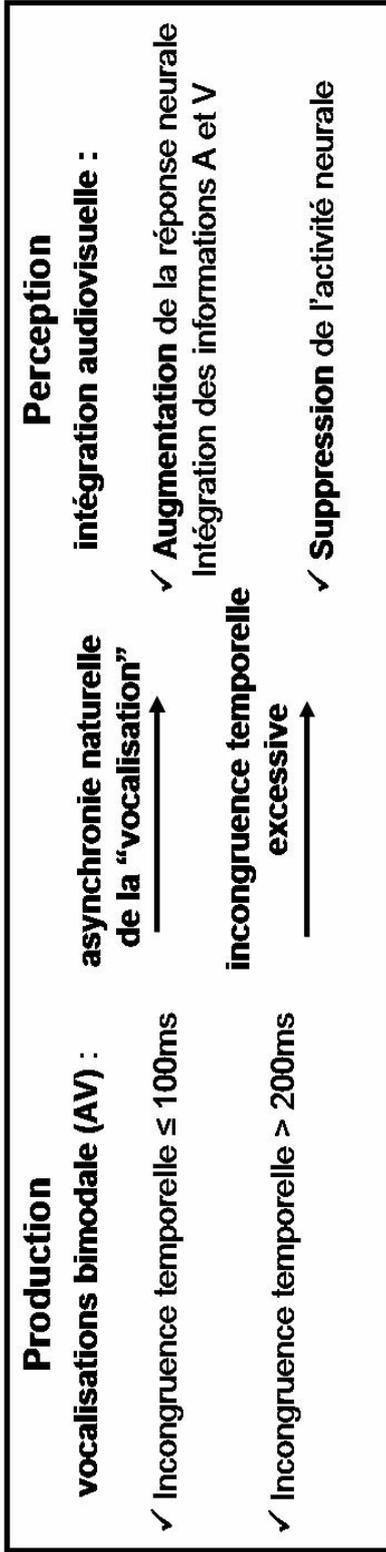
Pour explorer de manière moins globale la robustesse de la coordination audiovisuelle, plusieurs résultats peuvent être utilisés. Tout d'abord, lorsque le seuil de 200 ms dans l'anticipation naturelle est franchi, c'est-à-dire que les informations visuelles précèdent de plus de 200 ms le début acoustique de la mise en route du voisement, il n'est plus possible de relier la configuration labiale observée avec le son correspondant. Dans l'étude de Cathiard *et al.* (1996), ce phénomène de décrochage perceptif est observé lorsque le démarrage du geste d'arrondissement labial est initié dès la voyelle [i] ; celui-ci coïncide avec des informations acoustiques contradictoires puisque le sujet entend [i] et voit l'amorce visuelle de la voyelle [y].

Il existe donc bien une fenêtre temporelle d'intégration audiovisuelle au-delà de laquelle la contradiction visuo-auditive perturbe l'intégration des signaux. Il semble que cette relation temporelle entre les informations visuelles et acoustiques soit cruciale pour le traitement de la parole chez l'homme, mais également chez le macaque. Ghazanfar *et al.* (2005) qui ont utilisé la même asynchronie naturelle chez les singes Rhésus, ont montré qu'elle était exploitée en perception (renforcement de l'activité neurale dans le cortex auditif primaire) jusqu'au point où l'asynchronie est trop forte pour permettre l'intégration des signaux (suppression des réponses neurales).

La Figure 76 récapitule les similitudes observées quant à l'intégration multimodale de la parole (ou vocalisation) chez l'homme et le macaque Rhésus. Le système de traitement bimodal de l'information (perception) et le système d'encodage bimodal des vocalisations (production) sont donc tous deux liés chez l'homme et le singe Rhésus par des contraintes temporelles fortes. De telles similitudes dans l'organisation de la communication vocale n'ont pas pour l'instant été étendues à d'autres familles de primates. Toutefois, les vocalisations bimodales obtenues par Coudé *et al.* (2006) chez un macaque Mestrina semblent corroborer ces résultats. D'autres études similaires sont, par conséquent, nécessaires pour étendre ces premières conclusions.

Chez les singes

Ghazanfar et al., 2005



Chez les humains

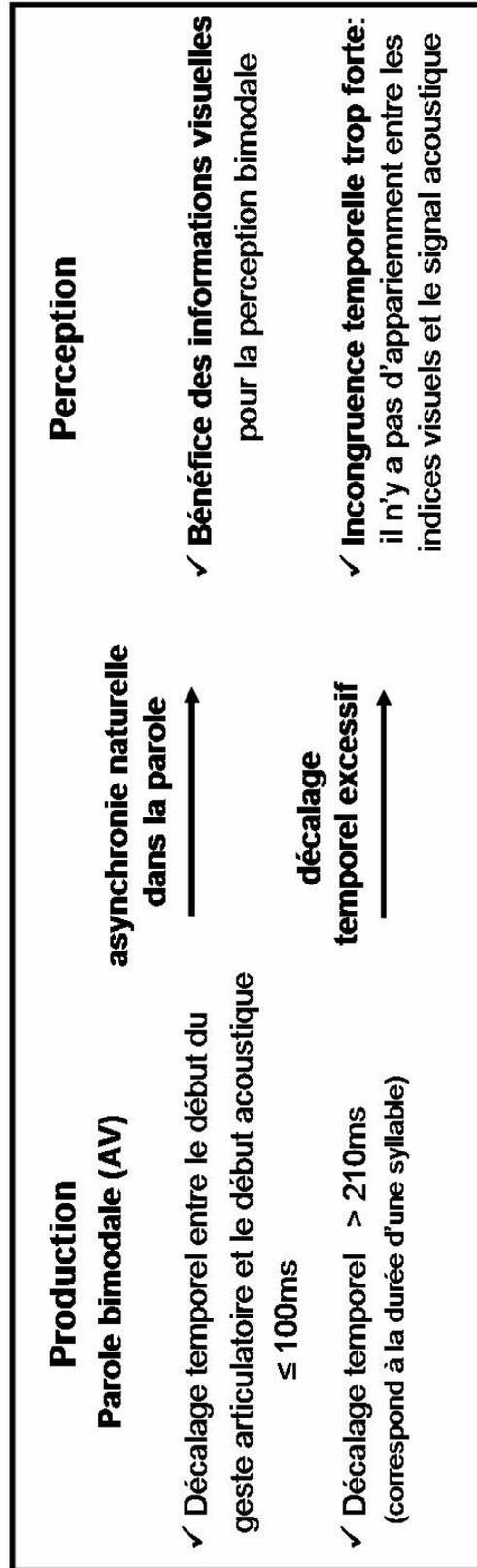


Figure 76: Comparaison du système d'intégration multimodale de la parole chez les primates non humains et chez l'homme.

* *

*

Ce que nous avons tenté de montrer tout au long des sections précédentes est qu'il existe, chez certains primates, un mécanisme d'imitation dont la fonction est d'établir un lien entre soi et le monde des congénères. Ce mécanisme offre en effet à l'observateur la possibilité de s'approprier l'action d'autrui par la reproduction en parité de cette action. Nous avons vu à travers les différentes études présentées que cette capacité s'applique à une gamme variée d'actions motrices : reproduction de coordinations gestuelles et, pour notre étude, d'une coordination orolaryngée. Au niveau de l'organisation neurale d'un tel mécanisme, les neurosciences nous apprennent que ce processus est sous-tendu par l'activation de neurones spécialisés dans l'unification de l'observation de l'action et de sa reproduction. Découverts dans un premier temps chez le macaque, puis chez l'homme, le système de neurones miroirs offre un ancrage moteur à la cognition sociale, et sans doute un ancrage linguistique, de par son rôle envisagé dans la communication interindividuelle. La mise en évidence de neurones de *l'imitation*, spécialisés dans la communication orofaciale dans des zones associées à la motricité, aussi bien chez le singe que chez l'homme, a en effet des résonances intéressantes au niveau de la parole, si l'on considère le caractère résolument bimodal de la parole chez l'homme et l'ubiquité de la modalité gestuelle (orofaciale) dans la communication interpersonnelle chez les primates.

Par ailleurs, Charlie et Patty nous ont montré que l'aire F5 est également le siège de vocalisations volontaires bimodales. Il existe une parité – comme principe nécessaire pour la communication – entre la production et le traitement perceptif de la parole. Et c'est sans doute dans le prolongement de la parité pour les vocalisations, dans « un néocortex essentiellement multisensoriel » (Ghazanfar & Schroeder, 2006), sachant que les régions neurales concernées par l'intégration multimodale de la parole sont identiques chez l'homme et le macaque Rhésus, notamment le sulcus temporal supérieur et même le cortex auditif, sans oublier les connexions avec le néocortex frontal et pariétal qui contiennent les neurones miroirs.

Ces différents enjeux font actuellement l'objet de nombreux travaux collaboratifs ; citons notamment le projet *Cog Speech* (« *Du Contrôle Orofacial des Gestes dans la communication chez les primates jusqu'à la parole humaine* ») réunissant des chercheurs de disciplines variées, telles que les sciences du langage, la psychologie et les neurosciences, dans le programme de l'European Science Foundation : *Origin of Man, Language, and Languages*. L'échange croissant des savoirs pluridisciplinaires contribue en effet à caractériser plus précisément la relation entre neurones miroirs et imitation et permet également d'évaluer

quelles sont les répercussions d'une telle rencontre au niveau de l'émergence phylogénétique et ontogénétique du langage. Ce dernier point nous intéresse tout particulièrement, notamment le processus d'*attunement* que l'on rencontre chez le primate non-humain et l'homme pour l'apprentissage de comportements moteurs et de manière cruciale chez l'homme pour l'apprentissage du langage.

III.2. Esquisse des premières habiletés imitatives de l'enfant

Tout comme le primate non-humain, le petit d'homme, le nouveau-né, doit acquérir les normes de la société dont il est membre. En l'occurrence pour le versant sonore du langage, il s'agira de la *norme phonologique* d'intégration à sa communauté linguistique lui permettant de communiquer avec ses membres. Après avoir évoqué précédemment le processus d'émulation et d'imitation chez le primate non-humain en soulignant ses implications au niveau de l'apprentissage, nous souhaitons aborder ce processus dans l'acquisition du langage et plus particulièrement pour la parole. En effet, nous pensons qu'une des étapes de l'ontogenèse du langage correspond au passage d'un processus d'émulation des *buts acoustiques linguistiquement définis* à l'imitation complète de la coordination oro-laryngée et de son déroulement temporel.

Nous insisterons sur l'organisation temporelle des gestes de parole car il nous semble que c'est au travers de son contrôle que les deux macaques de Parme ont réussi à produire leurs premières vocalisations bimodales *volontaires*. Du côté de la perception, le déroulement temporel approprié des signaux auditifs et visuels semble également la clef de l'intégration multisensorielle des vocalisations chez le chimpanzé, tout comme chez l'adulte.

Depuis plus de trente ans, de nombreux travaux en psychologie développementale explorent les capacités initiales du nouveau-né. Sous l'expression « capacités initiales » sont regroupés des domaines aussi vastes que l'étude des aptitudes motrices, l'organisation du système attentionnel, la perception multisensorielle etc. On sait désormais que le bébé dispose d'habiletés motrices et cognitives variées, contrairement aux assertions longtemps répandues dans la littérature sur la « vacuité de l'esprit humain à la naissance » ou encore de « l'état naturellement autiste du nouveau-né » et l'absence de toute compétence linguistique. Notons à ce propos qu'une grande partie des postulats théoriques de l'époque étaient essentiellement basés sur l'observation non expérimentale de jeunes enfants. La recherche développementale en étant à ses premiers pas, les chercheurs n'avaient pas encore élaboré de méthodologies

expérimentales rigoureuses et ne bénéficiaient pas de tous les outils techniques et apports méthodologiques que nous exploitons aujourd'hui.

C'est un fait acquis, l'homme est capable dès la naissance de produire des comportements spécifiques. Des études ont montré par exemple que le nouveau-né oriente préférentiellement son attention vers des stimuli visuels rappelant un visage humain plutôt que vers d'autres objets.

Ainsi les études de Goren *et al.* (1975) et Johnson *et al.* (1991) nous apprennent que le nouveau-né est équipé d'un système de reconnaissance des visages lui permettant de reconnaître prioritairement le visage de sa mère par rapport à un autre visage ou encore de distinguer la forme du faciès humain d'un autre stimulus visuel.

« La préférence néonatale pour les stimuli humains est le premier indice de leur traitement différentiel. Elle se manifeste dès la naissance au niveau de la perception visuelle, auditive, olfactive ainsi que dans l'imitation, restreinte aux mouvements humains. » (Nadel & Potier, 2002, p. 84)

Aux premières heures de sa vie, on observe l'apparition de comportements imitatifs, notamment de gestes orofaciaux et manuels dont le répertoire deviendra plus large que celui des primates non-humains ou encore d'autres animaux, tels que les dauphins ou certains oiseaux (mainates, perroquets). Notons que des travaux expérimentaux ont mis en évidence des conduites imitatives similaires de gestes orofaciaux chez des nouveaux-nés macaques ou chimpanzés (par exemple des protrusions de langue, Bard & Russel, 1999).

III.2.1. Les études pionnières de Meltzoff et Moore et le Mapping Intermodal Actif (AIM)

Concernant spécifiquement le bébé, nous rappellerons l'étude de Meltzoff & Moore (1977) qui a rendu célèbre la précocité des capacités imitatives du nouveau-né. Les auteurs ont en effet montré que les bébés sont capables d'imiter des expressions faciales ou des mouvements de parties du corps faits par des adultes devant eux : mouvements d'ouverture de bouche, de protrusion de la langue ou des lèvres (Figure 77).

Ces résultats ont largement été répliqués chez des nourrissons encore plus jeunes (Field *et al.*, 1982 avec des nourrissons âgés de 36 heures ; ou encore Meltzoff & Moore, 1989, avec des nourrissons de 42 minutes à 72 heures).



Figure 77 : Photos de trois imitations de mouvements buccaux: protrusion de la langue, ouverture de la bouche et protrusion des lèvres imités par des nouveaux-nés âgés de 2 à 3 semaines. D'après Meltzoff & Moore, 1977, p. 75.

L'hypothèse proposée par Meltzoff et Moore pour rendre compte de ce phénomène d'imitation précoce repose sur l'idée d'une cartographie intermodale active (« *Active Intermodal Mapping* ») permettant au nourrisson d'apparier son comportement moteur avec celui du modèle, par l'utilisation d'un schéma corporel du corps humain.

Plus précisément, ce processus consiste en les étapes suivantes.

- I. Dans une première phase, le nourrisson observe l'action d'un individu (par exemple une expression faciale telle qu'une protrusion de langue). La modalité visuelle est donc sollicitée pour la perception du comportement de l'autre.
- II. Afin de reproduire le geste perçu, c'est-à-dire imiter, le nouveau-né doit mettre en correspondance les mouvements du modèle avec son propre appareil musculaire. D'après Meltzoff et Moore, celui-ci dispose d'un schéma abstrait, prototypique du corps humain, lui permettant au-delà d'une représentation strictement visuelle de se représenter *proprioceptivement* le geste perçu sous une forme commune avec l'individu observé: la modalité motrice.
- III. Le nouveau-né se crée une représentation supramodale du comportement (représentation mentale stockée) et détermine quels sont les effecteurs à mobiliser selon la cartographie mentale du corps dont il dispose ;
- IV. Il unifie son comportement avec celui du modèle et compare les mouvements de son corps avec ceux du modèle, par un détecteur d'équivalence.
- V. Grâce à une boucle de proprioception (*feedback loop*), l'enfant bénéficie en temps réel d'informations perceptives sur son comportement par rapport à sa représentation du modèle et peut ainsi ajuster son comportement, le corriger si celui-ci n'est pas parfaitement apparié avec celui du modèle.

La Figure ci-dessous récapitule ces différentes étapes.

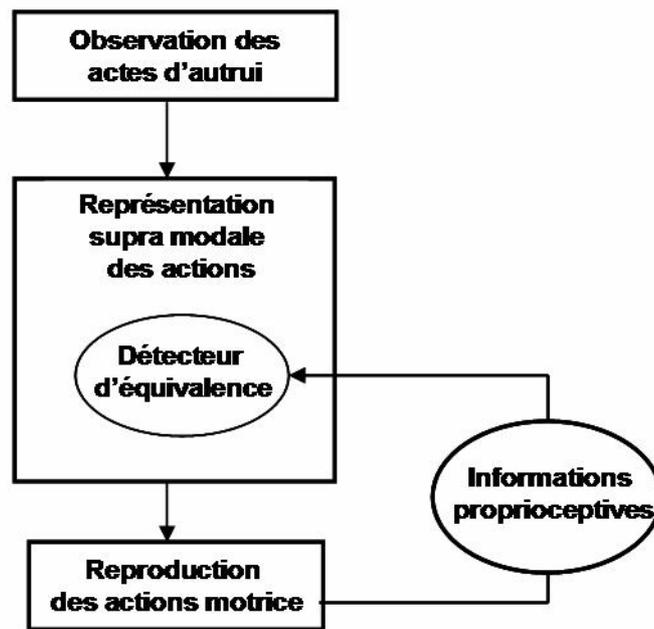


Figure 78 : Illustration de l'hypothèse de l'Active Intermodal Mapping (AIM) pour expliquer le phénomène d'imitation. D'après Meltzoff, 2005, p. 73.

Au-delà d'une explication cognitive et motrice du processus d'imitation précoce, le système AIM de Meltzoff & Moore suggère qu'il existe, chez le nourrisson, une certaine forme de compréhension naturelle de son environnement puisque celui-ci est capable d'interagir avec un individu en interprétant ses actes en termes d'actions motrices reproduites avec son propre corps.

Les auteurs se sont posé la question de la portée d'un tel système car il paraît étonnant que le nouveau-né soit capable de reproduire parfaitement une action sans avoir jamais vu son propre corps. S'agit-il alors simplement de réflexes ? D'une traduction automatique des informations visuelles en informations motrices n'impliquant ni distinction entre soi et l'autre, et ni reproduction *volontaire* de l'action perçue ?

En testant la capacité du nouveau-né à différer l'imitation, Meltzoff & Moore (1977) ont trouvé le moyen de répondre à ces interrogations. Une nouvelle étude a été réalisée, auprès de 12 nouveaux-nés âgés de 16 à 21 jours, reproduisant le même paradigme expérimental que précédemment, à l'exception d'une modification: pendant que l'expérimentateur exécutait un geste orofacial face au nouveau-né, on lui mettait une tétine en bouche afin qu'il ne puisse pas reproduire simultanément le stimulus visuel observé.

Les résultats ont montré que l'enfant possède la faculté de stocker la représentation supra-modale qu'il s'est faite du geste cible et peut le reproduire même lorsque le geste a disparu de

son champ perceptif. L'imitation du comportement cible n'est donc pas soumise à l'observation simultanée du modèle, elle n'est pas fixée dans le temps (« *not rigidly time-locked* », Meltzoff, 2000, p. 177) mais peut être différée jusqu'à 24 heures.

L'élaboration d'une telle théorie, totalement contraire aux hypothèses des psychologues constructivistes, offre une nouvelle lecture des capacités initiales du nouveau-né. Ces données comportementales montrent que les réponses imitatives du nourrisson ne constituent pas des stéréotypes ou encore ne découlent pas d'un suivi éphémère (*perceptive tethering*) des actions d'autrui. Non seulement le nouveau-né a la faculté de reproduire à l'identique de nouveaux comportements (avec une possibilité de correction et d'ajustement de celui-ci grâce à la boucle de proprioception), mais il peut également mémoriser ces comportements pour les reproduire ultérieurement. Ce processus conduit donc progressivement à la création d'un répertoire gestuel et à des possibilités d'interactions plus vastes avec le monde. Ces interactions seraient, dans un premier temps, essentiellement gestuelles (orofaciales), puis serviraient comme vecteur d'apprentissage des relations sociales.

« The fundamental idea is that the distinctive behavior and special interactive games of people serve as markers of their identity. If infants are uncertain about the identity of a person, they will be motivated to test whether this person has the same behavioural properties as the old one by imitating her behaviour and re-creating the previous social interaction. » (Meltzoff & Moore, 2000, p. 179)

Par l'imitation, le nourrisson bénéficie d'un outil privilégié pour la perception des actions d'autrui et leurs représentations. Il associe d'ailleurs les personnes qui l'entourent à des entités semblables à lui, pouvant réaliser des actions et avoir des états mentaux comme lui (théorie du « *Like me* »). Meltzoff (2005, 2007) croient d'ailleurs fortement que l'imitation et la compréhension de l'autre sont reliées par une relation causale.

Nous proposons de rendre compte de ce processus par le schéma suivant:

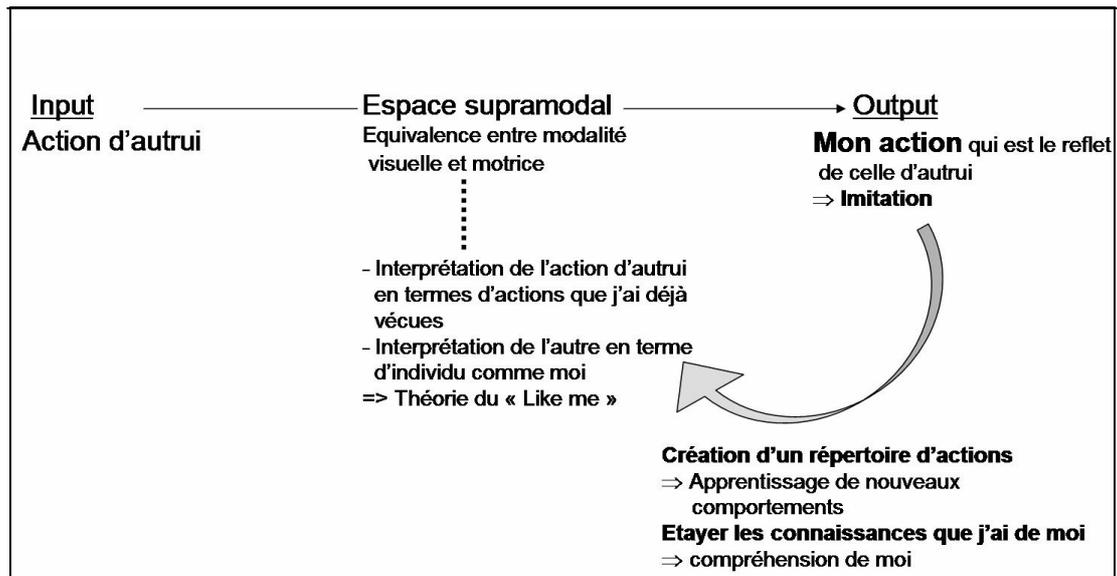


Figure 79: Représentation du processus d'imitation en tant qu'outil pour la compréhension d'autrui, d'après la théorie du « *Like Me* » de Meltzoff (2005).

L'enfant établirait une connexion entre les actions effectuées par son environnement et l'exécution de ces actions, par l'imitation. Parallèlement, l'accumulation des expériences le conduirait progressivement à associer les états de son propre corps avec ses états mentaux. De la même manière qu'il établit une équivalence entre son corps et celui d'autrui, il pourrait attribuer des états mentaux à cet autre corps, lorsque celui-ci exécute des actions que lui-même a déjà vécues (Meltzoff, 2005).

Comme dans l'étude de Horner *et al.* (2006) et celles de Whiten *et al.* (1999, 2005) sur le rôle de l'imitation dans la transmission culturelle chez les chimpanzés (transmission qui peut se faire sur plusieurs générations), l'imitation s'avère être ainsi un vecteur d'intégration sociale chez le nourrisson. Ce dernier règle volontairement son comportement sur celui de l'autre et se l'approprie. Cet ancrage social de l'imitation est fondamental car il conduit vers la communication intentionnelle ou le concept d'*intentional attunement* (mise en phase intentionnelle), évoqué par Gallese (2005).

« Nous ne sommes pas normalement éloignés des actions, des émotions et des sensations des autres, parce que nous sommes mis en phase avec les relations intentionnelles des autres. » (Gallese, 2004, traduit par Anne-Marie Varigault)

A travers la théorie de Meltzoff et Moore, nous retrouvons les concepts proposés par Gallese (2006) pour expliquer le rôle de l'imitation dans la cognition sociale. L'agent et l'observateur partagent un espace de représentation commun: le *schéma corporel* qui tisse un lien entre leur corps et celui de l'autre selon Meltzoff et Moore ou encore *l'espace*

interpersonnel de sens (Shared manifold) de Gallese où se mêlent les connaissances expérientielles de notre corps et celles que nous accumulons par l'expérience du corps de l'autre. L'imitation a donc une fonction expérientielle en ce sens qu'elle permet d'interagir avec le monde. La définition de l'imitation néonatale de Meltzoff et Moore comme un processus intentionnel, orienté vers un but d'appariement intermodal (« *Early imitation is intentional, goal-directed intermodal matching.* », Meltzoff & Moore, 2000, p. 177), fait écho aux idées de Gallese. Dès la naissance, l'enfant disposerait ainsi d'un espace interpersonnel primitif, l'espace « nous-centrique » (*we-centric space*) (Gallese, 2006b, p. 2), qui se traduirait notamment par des formes d'imitations orofaciales avec la mère.

« [...] *mothers and infants systematically engage in mutually coordinated activities during which their movements, facial expressions, and voice intonation synchronize in time (Reddy et al. 1997). [...] From 4 months of age onwards, infants and mothers show protodialogic behaviors in which they time their behavior in a bidirectional coordinated way (Trevarthen, 1979). [...] This evidence suggests that such protodialogic behaviours enable mother and infants to establish an affective attunement by means of which inner feeling states can be shared.* » (Gallese, 2006b, p. 2)

De fait, ce premier lien affectif très fort entre la mère et son enfant pourrait contribuer à l'orientation de l'enfant vers la communication interpersonnelle et plus particulièrement vers le langage parlé. Le rôle de la mère ou encore des nourrices a été plusieurs fois évoqué dans la littérature sur le développement du langage. Locke (1999) note par exemple que l'interaction mère-enfant offre un substrat affectif pour l'acquisition du langage :

« *The environment has a role to play, but that role is not fulfilled merely by providing "input" to the child. Research reveals a positive relationship between the quality of maternal attachment and the rate of language development (Van IJzendoorn, Dijkstra, & Bus, 1995). An emotionally secure mother-infant relationship provides a framework for language learning, but may also promote vocal learning by enhancing the infant's orientation to vocal and facial behaviours, vocal turn taking, and various incorporative and accommodative processes.* » (Locke, 1999, p. 377)

Arbib (2005) souligne également le rôle de la mère ou de la nourrice dans l'acquisition des habiletés imitatives notamment de comportements complexes, rôle qui contribue, d'après lui, à la différence marquée entre la rapidité d'acquisition de compétences nouvelles chez l'enfant, contrairement aux chimpanzés.

III.2.2. Sur le chemin de l'imitation des gestes de la parole

Le modèle de Meltzoff et Moore nous apprend que le bébé bénéficie d'un système d'intégration multimodale des informations, lui permettant de générer des correspondances entre ses différents sens: transfert de la modalité visuelle en modalité motrice (imitation d'une action vue) ou encore de la modalité tactile en modalité visuelle comme nous le montre l'expérience bien connue de Meltzoff & Borton (1979) sur la reconnaissance visuelle d'objets (des tétines) à partir de sensations buccales haptiques (sur ces questions de transfert intermodal visuo-tactile, cf. plus récemment Streri *et al.*, 2004 ; Gentaz & Mazens, 2006).

La disposition des nouveaux-nés à imiter des gestes orientés vers la communication vocale est appuyée également au niveau du traitement perceptif de ces informations. L'étude d'Aronson & Rosenbloom (1971) a montré qu'à un mois, le bébé bénéficie d'un *espace audiovisuel commun* qui lui permet de mettre en correspondance les informations auditives et visuelles et détecter d'éventuelles incohérences spatiales entre ces informations. L'étude de Kuhl & Meltzoff (1982) sur la perception bimodale de la parole conforte cette hypothèse en testant des bébés sur leur perception de la cohérence configurationnelle audiovisuelle. Ils montrent qu'à 4 mois et demi, le bébé a une préférence pour des stimuli audiovisuels congruents plutôt que pour des stimuli incohérents. Plus précisément, lorsqu'on lui présente en même temps les deux visages articulant silencieusement la voyelle [i] et [a], tout en lui faisant écouter simultanément l'une des deux voyelles, le bébé montre une préférence – marquée par un temps de fixation visuelle plus important – pour le visage congruent avec le son, ce qui révèle que le bébé est tout à fait capable d'établir une correspondance entre les gestes articulatoires vus sur les lèvres et le son entendu (ces données ont été étendues et confirmées par Legerstee, 1990, chez des enfants de 3 mois).

La spécificité des prédispositions pour la parole se manifeste donc par la sensibilité des bébés à la congruence des informations auditives et visuelles qui résulte de l'articulation des sons de la parole. On remarque d'ailleurs que cette préférence disparaît lorsque l'on échange la voix avec des tons synthétiques s'approchant des voyelles : l'intégration multisensorielle ne se fait que lorsque les mouvements orofaciaux (plus particulièrement les configurations labiales) sont appariés avec les informations auditives naturelles (Kuhl & Meltzoff, 1982).

Le bébé dispose donc très tôt d'un système de traitement des informations bimodales de la parole lui permettant d'associer une configuration articulatoire avec un son donné. Cette habileté, couplée à celle d'imitation, devrait à priori conduire rapidement l'enfant à acquérir les différentes unités de sa langue.

Nous allons voir à travers les travaux de Kuhl & Meltzoff (1996) sur l'imitation vocale, que le processus d'acquisition du langage place le bébé devant des défis plus complexes que la reproduction purement acoustique des unités de parole. Nous comprendrons à travers les propositions de Goldstein (2003), et les travaux récents de modélisation articulatoire de Serkhane (2005), que l'enfant doit acquérir des gestes dynamiques et pas seulement des produits acoustiques perceptivement identifiables à ceux du modèle.

III.2.2.1. Premiers pas vers la parole de l'adulte: l'émulation des buts acoustiques

Parmi les études développementales explorant les capacités initiales du bébé, nous nous arrêterons sur l'étude de Kuhl & Meltzoff (1996), considérée comme référence dans la littérature traitant de l'imitation en parole. Ces auteurs se sont intéressés à la production des trois voyelles cardinales [a], [i] et [u] par des bébés âgés de 12, 16 et 20 semaines. Ils leur ont fait visionner des films préalablement enregistrés dans lesquels ils pouvaient voir une femme produisant chacune des trois voyelles durant 5 minutes. Les films ont été projetés aux bébés à trois reprises (avec un intervalle d'une journée entre chaque écoute). Pendant leur projection, les auteurs ont enregistré les vocalisations des bébés. Ils ont noté une évolution importante entre les vocalisations produites à 12 semaines et celles produites un peu plus tard à 16 semaines, et enfin à presque 5 mois. La Figure 80 présente les différentes productions vocaliques réalisées par les enfants en fonction de leur âge. Ces productions sont représentées sous la forme d'une relation entre F1 et F2 afin d'observer l'évolution de leur répartition spectrale.

Un premier constat intéressant est que, dès 3 mois, le bébé tente de reproduire les voyelles perçues pendant les enregistrements. Il y a donc une volonté de la part du bébé de reproduire les stimuli audiovisuels. On retrouve ici l'idée d'une correspondance entre soi et l'autre qui se manifeste par le réglage de son comportement avec celui de l'autre. Les enfants ne parviennent pas à produire systématiquement des vocalisations correspondant aux modèles car ils n'ont pas encore acquis la maîtrise complète de la coordination oro-laryngée. A travers l'illustration des propriétés acoustique de ces voyelles, on distingue une évolution au cours des semaines: on voit apparaître entre 16 et 20 semaines l'amorce d'un triangle vocalique et une tendance à se diriger vers les patrons acoustiques de l'adulte.

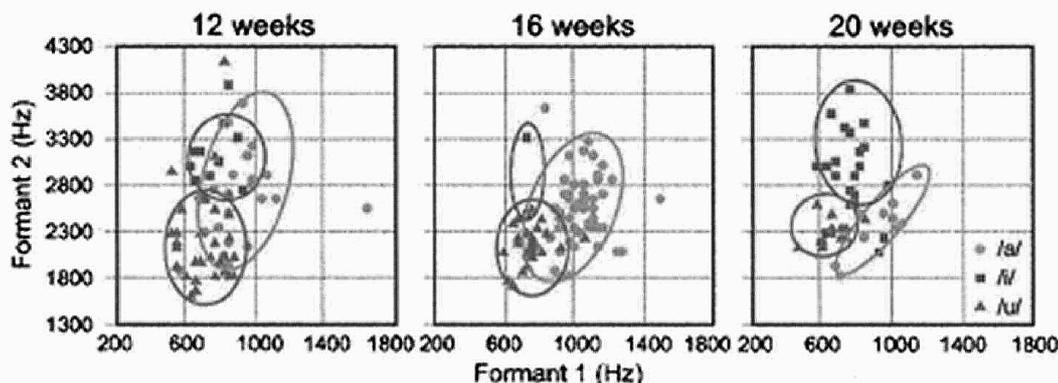


Figure 80 : Distribution des productions vocaliques proches des voyelles [a] (ronds), [i] (triangles) et [u] (carrés) à 12, 16 et 20 semaines. D'après Kuhl & Meltzoff, 1996, p. 2431.

Au vu de ces résultats, les auteurs formulent la conclusion suivante:

«The results demonstrate developmental change between 12 and 20 weeks—by 20 weeks, there is clear separation between the three vowel categories for infants. At this age, infants clearly imitate the model, and their vowels have appropriate formant frequency values in relation to one another, even though infants' vowels occur in a much higher frequency range.» (Kuhl, 2000, p. 11853)

Les auteurs concluent donc qu'un bébé de 4 et 5 mois peut produire trois voyelles distinctes avec des profils acoustiques distincts. Nous souhaiterions cependant proposer une autre interprétation des données obtenues chez ces bébés. Même si ces derniers ont réalisé différents tests perceptifs à partir des voyelles obtenues (tests de discrimination ou de catégorisation, Kuhl & Meltzoff, 1995, p. 149), il nous semble contestable d'associer ces voyelles – bien que *perçues* comme des voyelles – à de réelles *productions* de voyelles, phonologiquement distinctives, dans le cadre d'une adaptation du comportement moteur aux exigences phonologiques de la langue (*language-specific phonological attunement*, selon nos termes). Comme le rappelle Kent (1992), une attestation perceptive ne nous renseigne pas sur la performance motrice réalisée pour produire ces voyelles.

« A proper understanding of motor events may be clouded by a premature phonetic attribution to infant's sound productions. » (Kent, 1992, p. 84)

Il ne nous semble pas que les voyelles articulées par ces bébés à 4 mois soient des produits vocaliques distincts résultant d'un processus d'imitation, c'est-à-dire de reproduction du *chemin* (réalisation de la coordination articulaire appropriée ainsi que de son déroulement temporel) et du *but*, tels que formulés précédemment pour la description des comportements imitatifs chez les primates non-humains.

D'après nous, il s'agirait plutôt de *pseudo-voyelles*, témoignant d'une intention de reproduire le but acoustique final (*intentional attunement*), mais sans en maîtriser encore les moyens. On peut remarquer, sur la Figure 80, qu'à 5 mois, certaines des productions attribuées respectivement aux voyelles [i] et [u] se trouvent dans l'ellipse de la voyelle [a], de même que des productions [u] se retrouvent dans l'ellipse de la voyelle [i].

III.2.2.2. Modélisation articuloire des productions acoustiques des bébés

La problématique de l'émergence des gestes vocaliques a été récemment abordée par Serkhane (2005; Serkhane *et al.* (2005) lors d'un travail de modélisation des mécanismes d'exploration vocale et d'imitation orofaciale chez le bébé. Les recherches ont été centrées sur deux corpus: celui des voyelles produites dans l'étude de Kuhl & Meltzoff (1996) par des bébés de 4 mois et celui qui lui a été transmis par Barbara Davis et Christine Matyear, contenant un large échantillon de sons vocaliques sélectionnés à partir du répertoire de babillage de bébés âgés de 7 mois.

Serkhane *et al.* (2005) ont utilisé la plateforme *Speech Maps Interactive Plant* (SMIP), développée par Boë et Perrier à partir du modèle articuloire de Maeda (1989) et de son dérivé GROWTH, permettant l'ajustement de différences morphologies du conduit vocal en fonction de la croissance (GROWTH repris de Maeda par Boë, 1999, intègre les critères de développement non uniforme du conduit vocal). Ce SMIP permet ainsi d'explorer les capacités articuloires et acoustiques compatibles avec un conduit vocal réel. La modélisation articuloire a pu ainsi être adaptée aux productions enfantines des données fournies, en recréant la spécificité des conduits vocaux des enfants. Les auteurs ont ensuite superposé les productions vocaliques acoustiques des bébés de 4 mois, ainsi que celles des bébés de 7 mois, sur l'espace vocalique maximal (EVM) généré à ces âges par le modèle GROWTH, en activant toutes les commandes motrices disponibles dans l'appareil de production de parole (mandibule, larynx, corps de la langue, dos de la langue, apex, aperture labiale, protrusion labiale) (Figure 81).

« Les données réelles ne recouvrent pas tout l'espace qu'elles pourraient occuper si elles étaient le produit d'un contrôle mature. En particulier, les vocalisations de 4 mois sont relativement centrales et mi-hautes: les productions les plus en avant, arrière et ouvertes ne semblent pas être explorées. » (Schwartz et al., 2003, p.3)

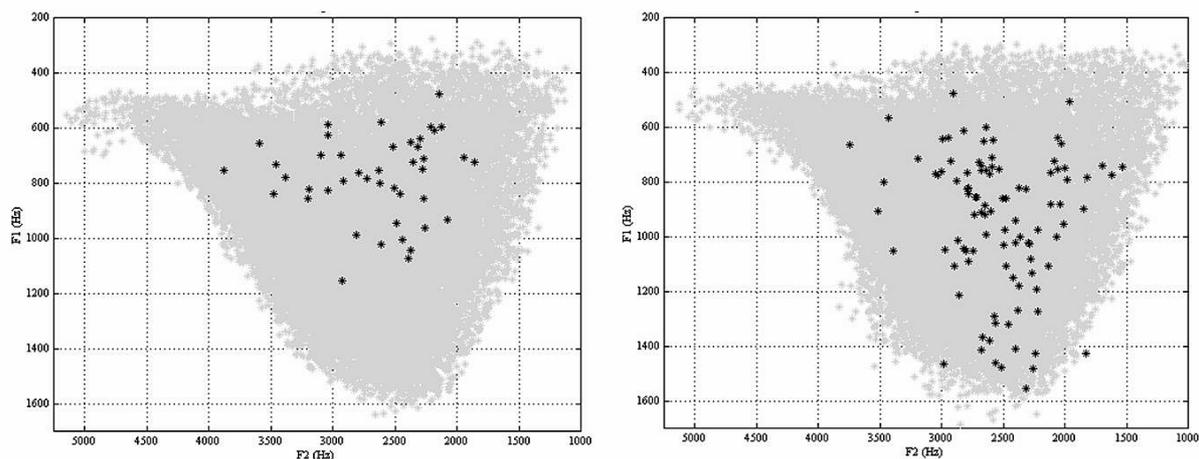


Figure 81: Cadrage acoustique des productions vocaliques à 4 mois (graphique de gauche) et à 7 mois (graphique de droite) sur le plan F1/F2. Données réelles en noir, vocalisations issues du modèle en gris. D'après Serkhane, 2005, p. 140.

Comme les données réelles ne correspondaient donc qu'à une sous-partie de l'espace vocalique maximal, Serkhane a réalisé un cadrage articulatoire, afin de réduire le jeu des commandes motrices disponibles dans le modèle à ceux strictement nécessaires pour la description des productions vocaliques à 4 et 7 mois (Figure 82). L'application de contraintes motrices au modèle a permis de prouver, par le test d'hypothèses formulées dans un cadre d'apprentissage bayésien, que les voyelles produites à 4 mois ne nécessitent pas de coordination impliquant la mandibule. Ces voyelles recueillies par Patricia Kuhl sont en réalité le produit de constrictions supra-glottiques des lèvres ou de la langue, qui n'ont pas eu besoin de recruter l'action de la mandibule, contrairement aux productions de babillage des bébés de 7 mois de Barbara Davis.

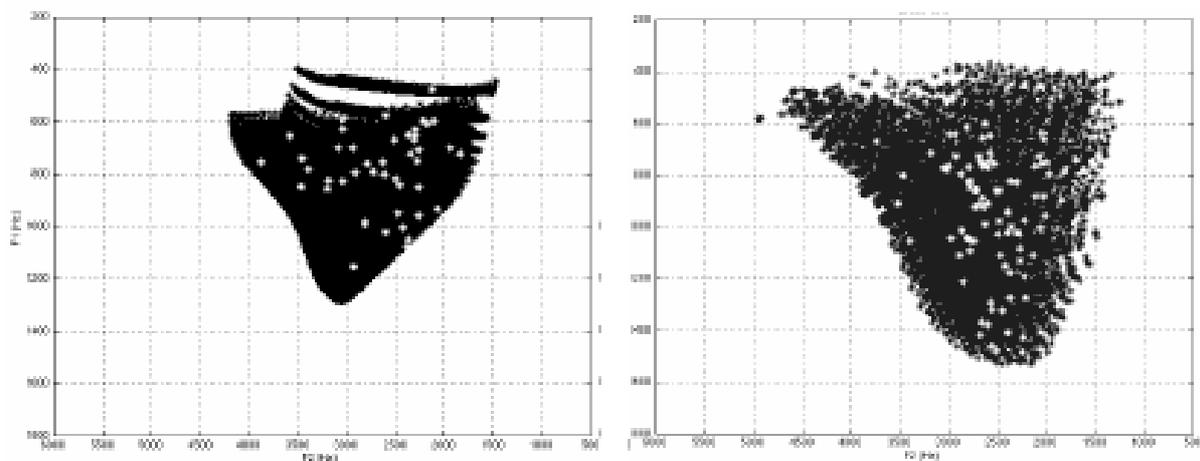


Figure 82 : Cadrage articulatoire des productions vocaliques à 4 mois (graphique de gauche) et à 7 mois (graphique de droite) sur le plan F1/F2. Données réelles en blanc, vocalisations issues du modèle en noir. D'après Serkhane et al., 2002, p. 4.

Ces résultats de simulation sont congruents avec les nombreuses études exploratoires menées ces vingt dernières années (Stoel-Gammon & Herrington, 1990 ; Oller, 1978). Nous prendrons pour référence les différents stades d'acquisition de la parole, proposés par Koopmans-van Beinum & van der Stelt (1986), qui proposent une lecture des productions vocales du bébé en rapport avec l'évolution des contrôles rythmiques ou discrets, laryngés et/ou oraux (supra-laryngés, oro-faciaux) de l'appareil de production de la parole. Le Tableau 14 offre une illustration de ces 5 stades d'acquisition et indique les principales modifications intervenant au niveau des productions vocales du bébé ainsi que des contrôles articulatoires. On remarque que les productions enregistrées à 4 mois par Kuhl et Meltzoff correspondent au stade III de la classification de Koopmans-van Beinum et van der Stelt (1986), c'est-à-dire à la période entre 2 et 4 mois et demi, au cours de laquelle le bébé effectue des constrictions labiales (audibles comme [aBu]) ou pharyngales (audibles comme [aR↔]) sans coordination mandibulaire.

« One articulatory movement per breath unit combined with continuous or interrupted phonation. » (Koopmans-van Beinum & van der Stelt, 1986, p. 42)

Ces productions sans coordination mandibulaire peuvent se continuer plus tard encore, comme en témoigne Kent (1992), d'après ses propres données développementales:

« Another stage, beginning at about three to four months and extending up to about six months, is one marked accomplishment in supraglottal or articulatory actions. In this stage, phonation often is accompanied by supraglottal constrictions and releases [...] and an impression [souligné par nous] of consonant-vowel sequence. » (Kent, 1992, p. 69)

Le contrôle *rythmique* de la mandibule n'apparaît, lui, qu'aux environs de 7 mois, à l'époque du babillage canonique, stade où les constrictions (dont les contacts de type labial [bababa] ou coronal [dadada]) peuvent être le simple sous-produit de la phase d'élévation de la mandibule.

Age en mois	0	1	2	3	4	5	6	7	8	10	12
	Stade 1		Stade 2		Stade 3		Stade 4		Stade 5		
Koopmans van Beinum & van der Stelt 1986	Phonation continue sans articulation	Phonation interrompue dans un cycle respiratoire sans articulation [aʔa ʔ a ʔ]	Phonation continue ou interrompue avec une articulation [aRa] [aβu] Constrictions labiales et pharyngales					Variations phonatoires avec ou sans articulations		Phonation continue ou interrompue avec articulations répétitives Babillage canonique	
	Contrôle de la position des cordes vocales	Contrôle de la phonation, du rythme, du cycle respiratoire	Articulation oro-laryngée					Contrôle de la qualité de voix, régimes phonatoires		Contrôle de la coordination oro-laryngée	

Tableau 14 : Stades de développement de la parole proposés par Koopmans-van Beinum & Van der Stelt (1986).

Un premier constat émerge de ces études. Les productions vocales du bébé traduisent l'état des capacités de contrôle moteur à une période donnée. D'après les tests réalisés par Serkhane *et al.* (2005), dans le cadre d'un apprentissage bayésien, les productions des bébés de 7 mois nécessitent une action mandibulaire. Les prédictions de l'apprentissage sur le modèle GROWTH, rapportées par Serkhane, concordent donc avec les stades de Koopmans-van Beinum & van der Stelt (1986), formulés à partir d'analyses de corpus de productions d'enfants à différents âges.

La maîtrise d'une coordination mandibulaire est indispensable à la production des voyelles prototypiques de l'adulte. Elle est l'aboutissement d'un long processus de réglage ou ajustement de coordinations motrices, c'est-à-dire d'un *attunement*, permettant l'accord du comportement de l'enfant avec les normes phonologiques de sa communauté linguistique, par le biais des adultes et des pairs.

Dans le Tableau 15, nous proposons de résumer ce passage de l'émulation à l'imitation à partir des données, que nous venons de reconsidérer à la lumière de leur modélisation en apprentissage, celles de Kuhl, celles de Matyear et Davis, et au-delà, quand viendra l'acquisition des coordinations labio-linguo-mandibulaires des voyelles après un an (section III.3.3.2). Nous avons choisi de représenter ces données selon trois niveaux de description: le niveau des configurations articulatoires (produit des constrictions ou des apertures du conduit vocal), les produits acoustiques et enfin, selon les contrôles neuraux auxquels les productions font appel.

Niveau articulatoire	Niveau acoustique	Niveau du contrôle neural
Configurations sans coordination mandibulaire	Pseudo-voyelle	Emulation du but
Configurations avec coordination mandibulaire	Cibles vocaliques complètes « attuned »	IMITATION de la coordination articulatoire adéquate Chemin + but

Tableau 15 : Illustration des produits articulatoire-acoustiques résultant du processus d'émulation du but vs. d'imitation.

La réalisation d'un certain contraste entre les produits ou buts acoustiques [a], [i] et [u] ne nécessite pas l'acquisition de coordinations impliquant la mandibule. Ces coordinations existent chez l'adulte et l'enfant, ainsi que l'ont montré les expériences de perturbations de la mandibule de Gibson & MacPhearson (1980) et de Oller & MacNeilage (1983). Elles ont démontré qu'en dépit d'une stratégie de compensation, la mâchoire concourt à l'élévation de la langue pour la voyelle [i] et [u] et à son abaissement lors de la production de la voyelle [a]. Il s'agit là de la coordination linguo-mandibulaire requise pour la production globale des voyelles.

Une étape majeure pour le bébé consiste donc à acquérir cette coordination spécifique pour la réalisation des gestes vocaliques. Dans cette nécessité de connaître le fonctionnement de contrôles plus ou moins fins, on notera, en ce qui concerne le sujet de notre étude, la précision requise pour ne pas faire de la petite aire aux lèvres d'une voyelle comme [u] ou [y] (souvent bien en-dessous de 80 mm², à comparer avec les 500 mm² d'un [i] ou d'un [a]) une occlusion bilabiale comme [b] à 0 mm². Une étape majeure pour l'enfant consiste donc à acquérir, dans une coordination spécifique de la mandibule et des deux lèvres, le contrôle de l'arrondissement vs. celui de l'occlusion labiale.

Commençons tout d'abord par l'acquisition du contrôle de la mandibule.

III.2.2.3. Explication des données de Kuhl & Meltzoff (1996) par la Phonologie Articulatoire

Dans un article récent, Goldstein (2003) expose sa vision de l'acquisition du langage en termes d'émergence des gestes *discrets* de la parole (*emergence of discrete gestures*). Ces notions font référence à la théorie de la *Phonologie Articulatoire* élaborée par Browman & Goldstein (1986) qui propose que la parole soit décomposée en unités dynamiques, discrètes d'action, *les gestes*, plutôt qu'en unités linguistiques classiquement statiques comme les phonèmes ou les segments.

« While the products of speech production (articulator movements, airflows, acoustics) are continuous and context-dependent, the act of speech production itself can be decomposed into discrete, dynamical units of contrast and combination. » (Goldstein, 2003, p. 86)

D'après lui, pour que les constrictions exercées au niveau du conduit vocal soient associées à des unités phonologiques contrastives dans une langue donnée, il est nécessaire que ces gestes soient *discrets*, c'est-à-dire uniques, ne pouvant pas être amalgamés les uns avec les autres. Cette propriété est rendue possible par le contrôle de deux processus. Tout d'abord le contrôle des effecteurs orofaciaux permettant la génération de ces constrictions (labiale et linguale via les différentes parties fonctionnelles de la langue). Ce premier contrôle est facile à acquérir puisque nous avons vu, avec les travaux de Meltzoff & Moore (1977), que le bébé est capable dès la naissance d'imiter un geste orofacial produit en sélectionnant, dans ses parties du corps, le ou les effecteurs à mobiliser parmi l'ensemble des effecteurs dont il dispose, ceux-là mêmes utilisés par le modèle. D'après Goldstein, ce contrôle correspond à la faculté de *partitionner* l'appareil vocal en différents effecteurs que l'on va pouvoir utiliser indépendamment. Au vu des preuves d'imitations très précoces (quelques heures après la naissance), cette faculté semble donc innée.

En revanche, un second contrôle, beaucoup plus complexe à acquérir, correspond à la capacité de différenciation des actions possibles d'un organe dans le cadre de la réalisation d'une constriction (par exemple la différence de degré de constriction). Selon l'auteur, ce contrôle est spécifique à la langue et requiert l'accumulation d'expériences, le réglage mutuel des comportements articulatoires (phase d'*attunement*).

« When gestures are used as part of a communication system, there must be agreement among the members of the speech community as to what the gestures are. Since every member has the same organs and an innate ability to recognize the identity of those organs in the self and others, this agreement is guaranteed in the case of the organ employed for a given gesture. As regards the values of the gesture's constriction goals, [...] the members of the community must attune their constriction actions to one another's. » (Goldstein, 2003, p. 86)

D'un point de vue développemental, Goldstein (2003) avance que le contrôle indépendant des effecteurs de la parole doit être complètement achevé avant de pouvoir entrer dans la phase de différenciation. Cette première étape serait donc assez longue et non uniforme, avec certains articulateurs contrôlés plus tôt que d'autres. La seconde phase correspondrait à l'affinement des actions possibles d'un organe pour la réalisation d'une constriction vocalique par exemple. Nous pourrions représenter ces actions sur un continuum par la production de toutes les valeurs possibles que peut prendre cette constriction.

Au niveau de l'acquisition de cette *différentiation* permettant le contraste phonologique, Goldstein présente les résultats d'une étude menée par Bernstein-Ratner (cité dans Goldstein, 2003) sur les productions de 6 bébés âgés entre 1;1 an et 1;9 an. L'auteur a extrait les productions de mots qui s'approchaient des cibles des adultes et a réalisé des tests perceptifs, où les sujets devaient transcrire les consonnes, puis les comparer avec les productions adultes. Le calcul des pourcentages d'erreurs révèle que les bébés recrutent le bon articulateur pour réaliser la constriction mais qu'ils ne parviennent pas à reproduire le degré de constriction de l'adulte (processus d'*attunement* non achevé).

Pour revenir aux données de Kuhl & Meltzoff (1996), les vocalisations considérées par les auteurs comme des *imitations de voyelles* témoignent d'un processus de *différentiation* non encore acquis. Vu l'âge des bébés (5 mois), on peut même avancer que ces derniers sont à peine entrés dans la phase d'acquisition du premier contrôle de *discrétisation* des articulateurs.

« *Within-organ contrasts require attunement to the language environment and thus should take longer to emerge.* » (Goldstein, 2003, p. 87)

Goldstein ne remet pas en question la possibilité d'exploiter les capacités précoces d'imitation orofaciale pour l'apprentissage du langage (il défend d'ailleurs la plausibilité d'un tel processus) mais il reste que, pour imiter les constriction vocaliques de l'adulte, le jeune enfant doit d'abord acquérir le contrôle indépendant des organes recrutés pour la réalisation des *primitives* de ces constriction.

Pour reprendre les mots de Studdert-Kennedy (2005), l'acquisition des consonnes et des voyelles sera plutôt une question d'acquisition d'unités dynamiques (les *gestes* dans le cadre de la phonologie articuloire), constituant un ensemble d'actions coordonnées spatialement et temporellement pour la réalisation des contrastes phonologiques d'une langue.

« *To see consonants and vowels as dynamic units of phonetic action, spatially and temporally coordinated gestural structures, rather than as static 'beads on a string', as the saying goes.* » (Studdert-Kennedy, 2005, p. 57)

III.3. De l'émulation des buts acoustiques à l'imitation d'une coordination articulatoire : le développement du contrôle vocalique

III.3.1. Ce que nous apportent les études acoustiques

Lorsque l'on s'intéresse à l'émergence du contrôle moteur, on constate rapidement en compilant la littérature qu'il existe peu d'études articulatoires au regard des multiples études acoustiques réalisées depuis plus de trente ans, ce déséquilibre étant probablement dû aux contraintes techniques et méthodologiques sous-jacentes aux études articulatoires.

De nombreuses études acoustiques ont traité de l'acquisition des stratégies coarticulatoires, considérant les caractéristiques spectrales et temporelles des coproductions CV chez le jeune enfant (Sussman *et al.*, 1999; Lee *et al.*, 1999; Nittrouer *et al.*, 1996 ; Tingley & Allen, 1975 , pour n'en citer que quelques unes). Elles apportent un éclairage sur le développement du contrôle moteur de la parole depuis les premières expériences de babillage jusqu'à la parole fluide de l'adulte. La plupart de ces études se sont orientées vers l'acquisition des consonnes au détriment de celle des voyelles (une large revue des études acoustiques menées dans ce domaine peut être trouvée dans James *et al.*, 2001). D'après cette revue, il semble que le contrôle des consonnes soit acquis plus tôt que celui des voyelles. A un an, l'enfant disposerait d'un inventaire varié de consonnes fricatives, plosives, de même que des consonnes labiales, approximantes et linguales.

L'erreur souvent commise est de croire que les voyelles sont largement répandues dans le babillage canonique et globalement plus faciles à acquérir, ce qui en ferait une problématique de recherche moins intéressante. Pourtant, d'après Davis & MacNeilage (1990), la maîtrise des voyelles a des implications essentielles dans le développement phonologique de l'enfant. Dans la partie précédente, nous avons pu également apprécier le rôle des gestes vocaliques dans la mise en forme du conduit vocal lors des séquences CV chez l'adulte. Le contrôle des voyelles nous semble donc être un passage obligé vers le contrôle de la coarticulation.

L'état des travaux développementaux menés sur le contrôle vocalique met en évidence certaines divergences d'opinions quant à l'âge où l'ensemble des voyelles sont supposées être correctement produites par l'enfant. Certains chercheurs comme Wellman *et al.* (1931), Vihman (1996) et Selby *et al.* (2000) défendent l'idée que les voyelles cardinales [i, A, u] sont maîtrisées vers 2 ans, parfois même à un an pour [i] et [a] selon Buhr (1980) (mais non à

4 mois comme le suggéraient Kuhl & Meltzoff, 1996) et l'ensemble des voyelles à 3 ans. D'autres avancent que le contrôle précis des articulateurs supra-laryngés nécessaire à la production des voyelles n'est définitivement acquis qu'à la puberté (Kent, 1976 ; Sanders, 1972).

Ces divergences d'opinions sont étroitement liées aux approches expérimentales adoptées pour attester du contrôle des voyelles. Elles peuvent être imputées à des méthodologies de collectes de données et d'analyses très variées: recueil de voyelles dans le cadre de mots intelligibles *vs.* dans tout type de contextes, transcriptions auditives *vs.* analyses acoustiques. Il en ressort que toutes les voyelles de la langue ambiante sont globalement *bien perçues* autour de 3 ans avec, pour certains auteurs, la certitude qu'une validation perceptive ne suffit pas à attester de la maîtrise de leur coordination articuloire. L'enfant parvient à *l'émulation du but* avec les stratégies dont il dispose, mais ne semble pas avoir acquis les coordinations articuloires fines de l'adulte pour la production de cibles vocaliques bien formées. Il semble essentiel de dissocier les études qui restreignent l'inventaire des voyelles à celles produites dans des mots intelligibles (Stoel-Gammon & Herrington, 1990 ; Davis & MacNeilage, 1990), de celles qui privilégient la performance de l'enfant et considèrent toutes les productions vocaliques, quel que soit le contexte d'apparition de ces voyelles (Buhr, 1980 ; Selby *et al.*, 2000).

« [...] *the accuracy of motor control improves with age until adult-like performance is achieved at about 11 or 12 years, somewhat after the age at which speech sound acquisition usually is judged to be completed.* » (Kent, 1976, p. 421)

Durant la première année de vie et même au-delà, les profils acoustiques parfois très instables de ces voyelles sont la preuve de l'immaturité du système moteur du bébé. Les résultats d'une étude longitudinale conduite par Lieberman (1980) chez 5 enfants anglais montrent par exemple que, durant le babillage, toutes les pseudo-voyelles produites par les bébés sont caractérisées par des fréquences formantiques très élevées au regard des profils de l'adulte. Les voyelles enregistrées appartiennent essentiellement à la partie basse et centrale du triangle vocalique et ne contiennent que très peu d'exemplaires de voyelles postérieures arrondies telle [u] (il note seulement une occurrence à 6 mois, puis une à 17 mois). Des tendances similaires ont été rapportées par Buhr (1980), suite à l'analyse des patrons formantiques des voyelles produites chez un seul enfant entre 4 et 15 mois. L'apparition des voyelles dans l'évolution ontogénétique du langage n'est donc en rien un hasard et reflète l'évolution des contraintes motrices du système de production de parole (le caractère universel des sons du babillage en est une preuve). Ces contraintes favorisent l'apparition de

certaines articulations au détriment d'autres. On note par exemple au début du babillage une prédominance des voyelles centrales et basses ne nécessitant pas de coordination fine entre les articulateurs: [a, ə, A, E, ø] (Davis & MacNeilage, 1990 ; Kent & Bauer, 1985 ; Stoel-Gammon & Herrington, 1990 ; Lieberman, 1980).

Kent (1976) récapitule les grandes lignes de ce parcours en soulignant que, si dès 2-3 ans, l'enfant est déjà bien avancé dans son apprentissage du langage, il est encore loin de reproduire parfaitement les patrons articulatoires prototypiques de l'adulte. Comme nous venons de le mentionner, cette instabilité se reflète dans les profils acoustiques des productions des bébés. Kent s'intéresse notamment à l'évolution de 3 paramètres acoustiques qui, selon lui, permettent de décrire les propriétés de la parole:

- I. les patrons formantiques apportent des indications sur la position des articulateurs pour la production des gestes vocaliques ;
- II. la fréquence fondamentale renseigne sur l'activité du larynx lors de ces voyelles ;
- III. les caractéristiques temporelles, tel que le délai d'établissement du voisement (VOT), attestent de la coordination orolaryngée.

A partir de ces trois paramètres (repris de Egushi & Hirsch, 1969), Kent (1976) conclut que les patrons articulatoires se stabilisent entre 5 et 8 ans, voir même plus tard, au moment de la puberté.

« The variability of speech motor control progressively diminishes until the age of eight-12 years, when adultlike stability is achieved. The exact age at which minimum variability is attained probably depends upon several factors, but of certain importance are the child's individual pattern of motor development. [...] The declining variabilities for the three types of measurements appear to reflect an increasing precision of motor control over a five-to-eight years. » (Kent, 1976 p. 440)

D'autres études acoustiques corroborent ces conclusions. Tel est le cas de Lee *et al.* (1999) dont les résultats montrent une stabilisation de la durée intrinsèque des voyelles à partir de 5 ans. L'une des raisons de cette variabilité viendrait de la stratégie employée par l'enfant. D'après la revue des données de Tingley & Allen (1975), Kent (1976) dresse deux grandes étapes dans le développement du contrôle moteur de la parole. Jusqu'à l'âge de 5 ans, l'enfant est soumis à des contraintes de coordinations fortes, si bien qu'il privilégie souvent les coordinations motrices déjà acquises au détriment d'autres plus complexes pour la réalisation des buts acoustiques. Les productions de l'enfant reposent donc essentiellement sur des patrons articulatoires favorisés non conformes à la langue ambiante, correspondant à des stratégies motrices *pré-planifiées* (selon Kent), sans possibilité de bénéficier réellement de

retour proprioceptif ou d'effectuer des ajustements en temps réel, comme on peut le constater dans le cas de l'imitation de gestes orofaciaux. Les voyelles articulées correspondent à des *pseudo-voyelles*, c'est-à-dire à des approximations de formes canoniques de l'adulte. Le degré de conformité de ces pseudo-voyelles avec les patrons de l'adulte dépendrait de différents facteurs tel que l'âge, le locuteur, ou encore le contexte phonétique. Enfin, elles seraient marquées par une instabilité au niveau de l'organisation temporelle propre à la langue.

« *Motor patterns for speech are largely preplanned rather than feedback controlled even at the age of five years. Thus, it appears that five-year-old children have learned strategies for the preprogramming of motor control, even though the precision of motor control continuously improves for a period of at least four more years (until the age of nine or 11).* » (Kent, 1976, p. 439)

Par conséquent, selon Kent, la maîtrise finale des voyelles et leur réglage temporel par rapport aux patrons de la langue requiert des habiletés motrices très complexes qui ne seront réellement maîtrisées qu'au cours de l'enfance, soit sur une période pouvant aller jusqu'à l'âge de 11 ans.

III.3.2. Quelques clés pour l'acquisition du contrôle moteur

Nous avons donné un bref aperçu des études acoustiques portant sur l'acquisition du contrôle vocalique. Il ressort de ces travaux que les patrons acoustiques des voyelles se stabilisent tardivement. Ces résultats signifient qu'il y a une évolution des performances articulatoires jusqu'à la maturation du système moteur permettant de produire les voyelles canoniques de l'adulte. Participent bien entendu à cette évolution d'autres facteurs tels que l'accumulation des expériences de productions vocales ou encore l'immersion constante dans la langue offrant ainsi au bébé des modèles de voyelles bien formées.

Dans le cadre de notre étude, nous nous placerons du côté de ceux qui cherchent à attester du contrôle moteur de la parole et non seulement des buts perceptivement intelligibles. Nous souhaitons ici défendre l'idée que la production des voyelles requiert, bien entendu, non seulement le contrôle du régime pulmonique et des régimes laryngés, mais aussi celui des couplages labiaux et linguaux avec la mandibule (le contrôle des pressions intra-orales résultant des contrôles précédents et du contrôle du velum a pu être regroupé, par Christian Abry, sous le terme général de contrôle de la *closance*, cf. la thèse en cours de Claire Lalevée qui s'adresse à cette question chez l'enfant entre 6 et 15 mois). Il va de soi que chacun des contrôles de ces coordinations peut être ajusté (*attuned*), tout particulièrement en ce qui concerne leur organisation temporelle (*timing*), à ceux de la langue ambiante.

III.3.2.1. Du contrôle de la mandibule: La théorie *Frame then Content* de MacNeilage (1998)

La théorie de MacNeilage (1998) propose que la mâchoire (la mandibule) soit le premier articulateur supralaryngé contrôlé par le bébé autour de l'âge de 7 mois, les autres articulateurs demeurant passifs, portés par la mandibule. Le contrôle émergent de la mandibule favorise la répétition de mouvements d'ouverture et de fermeture, donnant ainsi un rythme aux vocalisations du bébé et modifiant totalement l'allure de ses vocalisations. L'oscillation rythmique de la *porteuse* comme point de départ de la production des voyelles et des consonnes offre un cadre à la parole. L'abaissement de la mandibule favorise en effet l'émergence de voyelles tandis que sa remontée donne lieu aux gestes consonantiques. L'alternance des ouvertures et fermetures de la mandibule impose ainsi le *cadre syllabique (Frame)* : consonne-voyelle, que l'on retrouve en outre majoritairement dans les langues du monde (« *a universal unit in speech* », MacNeilage, 1998, p. 501). L'inventaire de ces voyelles et consonnes est bien entendu très limité au début du babillage puisque directement lié à l'oscillation *mécanique* de la mandibule : il s'agit principalement d'occlusives labiales ou coronales auxquelles s'associent des voyelles plutôt antérieures ou centrales et ouvertes. L'élaboration progressive de modulations articulatoires (notamment de la langue) apporte un *contenu* à ces cadres primitifs (*then Content*). Par ailleurs, les auteurs notent qu'il existe des contraintes coarticulatoires mutuelles entre voyelles et consonnes, qui favorisent certaines combinaisons de consonnes-voyelles (par exemple l'association des voyelles antérieures avec les consonnes coronales ou de voyelles centrales avec des consonnes labiales). Ces combinaisons, validées dans différentes langues, se retrouvent également dans la parole de l'adulte (Davis *et al.*, 1999).

La théorie « *Frame then Content* » de MacNeilage définit donc l'acquisition ontogénétique du langage en terme d'évolution biomécanique de l'appareil vocal. A partir de la pseudo-voyelle primitive [a], on peut envisager que l'émergence des autres voyelles corresponde à l'exploration de l'espace des constriction possibles au niveau articulatoire, formant ainsi un espace acoustique vocalique défini par la langue ambiante, de même que la diversification des contacts aboutit à la création de nouveaux phonèmes.

III.3.2.2. La théorie *Frame then Content* serait-elle une ontogénie venue de la phylogénie du langage ?

Nous avons largement insisté, au cours de ce manuscrit, sur les trois composantes fondamentales du système de production de parole: l'appareil respiratoire, la position du

larynx et les articulations supra-laryngées. Nous les avons évoquées dans le cadre des vocalisations de certaines familles de primates et il apparaît d'autant plus évident de rappeler leur importance dans le développement ontogénétique de la parole, puisque l'étape du babillage marque un gain de contrôle de ces trois composantes. Ces dernières sont illustrées sous forme de schéma par MacNeilage (1998).

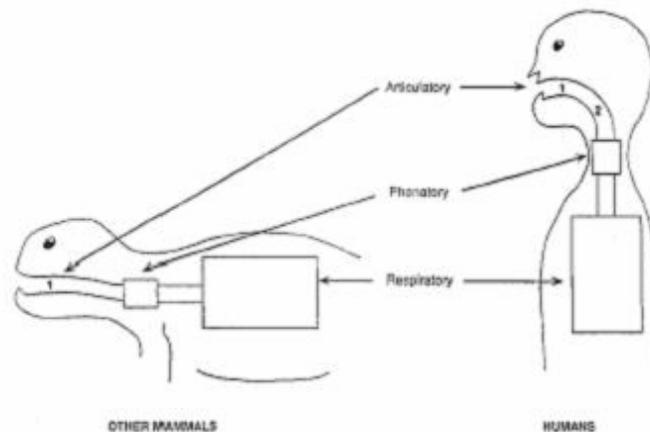


Figure 83 : Illustration des trois composantes du système de production de parole. D'après MacNeilage, 1998, p. 500.

Pour la production du langage, l'enfant doit acquérir le contrôle de chacune de ces composantes ainsi que leur coordination temporelle. Selon MacNeilage, la différence entre le langage articulé produit par l'homme et les vocalisations limitées des primates non-humains résulte de l'usage alternatif des contacts articulatoires réalisés par nos articulateurs supralaryngés lors de la production de syllabes CV.

« Thus, human speech is distinguished from other mammalian vocal communication, in movement terms, by the fact that a third, articulatory, level of modulated cyclicity continuously coexists with the two levels present in other mammals. » (MacNeilage, 1998, p. 501)

Comme les primates non-humains et autres animaux ne bénéficient pas de cette capacité à multiplier les types d'articulations permettant, par exemple, la génération de différents lieux d'articulations et plus généralement la création de phonèmes – variés, en bref, de créer le contenu (*Content*) – il semble que pour ces derniers, la théorie de MacNeilage se transforme en *Frame but no Content*. Du point de vue phylogénétique, on revient alors à la question cruciale de l'évolution du langage posée par les anthropologues, les neurolinguistes, les psychologues, les linguistes, dont nous avons brièvement exprimé précédemment les hypothèses. Nous ajouterons donc celle de MacNeilage :

« The likelihood that this cyclicity, associated with the syllable, evolved from ingestive cyclicities (e.g., chewing) is indicated by the fact that much of the new development of the brain for speech purposes occurred in and around Broca's area, in a frontal perisylvian region basic to the control of ingestive movements in mammals. An evolutionary route from ingestive cyclicities to speech is suggested by the existence of a putative intermediate form present in many other higher primates, namely, visuofacial communicative cyclicities such as lipsmacks, tonguesmacks, and teeth chatters. The modification of the frontal perisylvian region leading to syllable production presumably made its other ingestion-related capabilities available for use in modulation of the basic cycle in the form of different consonants and vowels (content). More generally, it is suggested that the control of speech production evolved by descent with modification within two general purpose primate cortical motor control systems, a medial system, associated with vocalization control in all primates, and a lateral system, including Broca's area, that has the necessary emergent vocal learning capacity. » (MacNeilage, 1998, p. 499)

III.3.3. Première étape vers l'attunement phonologique: l'émergence de la coordination supra mandibulaire

Nous tenterons dans cette section de déterminer quelles sont les coordinations articulatoires impliquées dans la production de la parole.

III.3.3.1. L'apport des travaux de Munhall & Jones (1998) sur la coarticulation

La théorie de MacNeilage a également été évoquée par Munhall & Jones (1998) à partir du suivi des mouvements labiaux d'un bébé de 8 mois ainsi que d'un adulte. Les déplacements verticaux de la lèvre supérieure et inférieure ont été mesurés grâce à l'application des diodes infra-rouge (système Optotrak) sur les zones médianes des deux lèvres près de la zone vermillon. On remarque, sur la Figure 84, que chez le bébé, contrairement à l'adulte, il n'existe pas de mouvement indépendant de la lèvre supérieure. Il existe bien, chez le bébé, un mouvement léger d'élévation de la lèvre supérieure lors de l'occlusion labiale pour le [b] mais qui est, en réalité, dû à la remontée de la lèvre inférieure elle-même entraînée par l'action d'élévation de la mandibule. Chez l'adulte en revanche, la lèvre supérieure est contrôlée et résiste à l'action de remontée du couple lèvre inférieure/mandibule. En revanche, pour la réalisation du [a], les deux lèvres sont contrôlées simultanément pour augmenter l'aperture, avec écartement maximal par montée de la lèvre supérieure et abaissement maximal de la mandibule. Des résultats similaires ont été observés par Tuller & Kelso (1984) sur des séquences de type [ba#'pab]. Ces derniers ont soustrait le mouvement de la mandibule à celui de la lèvre inférieure pour isoler le mouvement propre de la lèvre inférieure dans l'organisation de la coarticulation consonne-voyelle. Leurs résultats

démontrent qu'il n'existe pas de mouvement spécifique de la lèvre inférieure lors de la production du [a], les auteurs observant seulement un mouvement d'élévation spécifique de la lèvre inférieure lors du [b].

D'après Munhall & Jones (1998), ces données concordent avec la théorie de MacNeilage sur l'évolution des contrôles articulatoires de la parole.

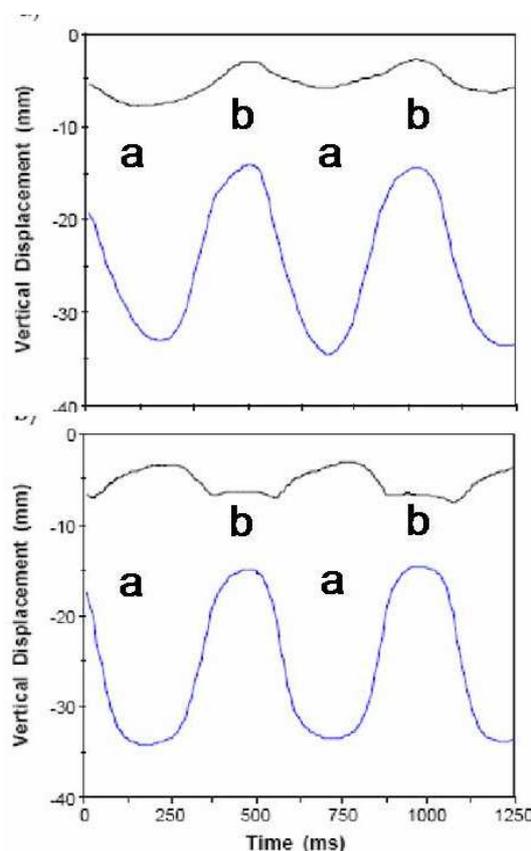


Figure 84 : Comparaison des productions de séquences [babab] chez un bébé de 8 mois (figure du haut) et un adulte (figure du bas) en fonction des déplacements verticaux en mm de la lèvre supérieure (tracé supérieur dans chaque figure) et de la mandibule (tracé inférieur). D'après Munhall & Jones, 1998, pp. 2-3.

L'enfant contrôle tout d'abord l'articulateur *proximal* de la parole: la mandibule dont l'oscillation rythmique donne lieu aux cadres syllabiques purs sans contenu. Puis, l'enfant acquiert le contrôle des effecteurs distaux, c'est-à-dire des articulateurs supra-laryngés jusque là passivement portés par l'articulateur dominant (Figure 84).

« This pattern is consistent with the proposal that initial babbling primarily involves mandibular motion; the lower lip rides on the jaw and deforms the upper lip on contact. [...] As MacNeilage suggests, early babbling could be produced by a simple "frame" style of speech motor control using a repetitive jaw cycle. » (Munhall & Jones, 1998, p. 4)

III.3.3.2. La modélisation des degrés de liberté des articulateurs comme outil pour la compréhension de la coarticulation

De nombreux chercheurs se sont intéressés aux modifications des effecteurs de la parole dans le processus de coarticulation. Une étude de Vilain (2000) s'appuie sur la modélisation des degrés de liberté des articulateurs à partir de radiographies du conduit vocal pour décrire les effets de coarticulation dans des séquences V_1-C-V_2 (où $V_1 = V_2 = [a, i, y]$ et $C = [b, d, g, v, Z]$).

En observant un échantillon des contours sagittaux générés dans cette étude, il ressort que la voyelle influence nettement la forme *globale* du conduit vocal, tandis que le geste consonantique consiste seulement en une modification *locale* du conduit vocal (dans le cas de la séquence [aba] (Figure 85), l'occlusion labiale associée à la consonne [b]).

Ce phénomène se manifeste de la même manière pour la séquence [ubu] (Figure 86). La voyelle impose une certaine configuration du conduit vocal et influence même l'articulation de la consonne intervocalique, le point de contact aux lèvres est en effet modifié par rapport à la séquence [aba], en raison de la protrusion des lèvres requise pour la constriction labiale du [u] qui est maintenue durant l'occlusion labiale. La production des segments successifs de parole, telle que proposée dans le modèle de *coproduction* des gestes d'Öhman (1967), semble donc s'organiser autour de gestes vocaliques dominants sur lesquels les gestes consonantiques sont surimposés.

Comment interpréter ces données en termes de contrôles articulatoires ?

On peut apprécier le contrôle de la langue durant la phase d'élévation de la mandibule:

Dans la séquence [aba], on peut remarquer que la position basse et centrale de la langue (obtenue par l'action du muscle hyoglosse et le génioglosse antérieur) pour l'articulation de la voyelle [a] est maintenue durant la totalité de la séquence, même lors de la phase d'occlusion labiale pour la consonne [b]. La langue est activement contrôlée pour résister à l'élévation de la mandibule lors de l'occlusion ; elle maintient la position vocalique par un mouvement compensatoire d'abaissement du dos de la langue.

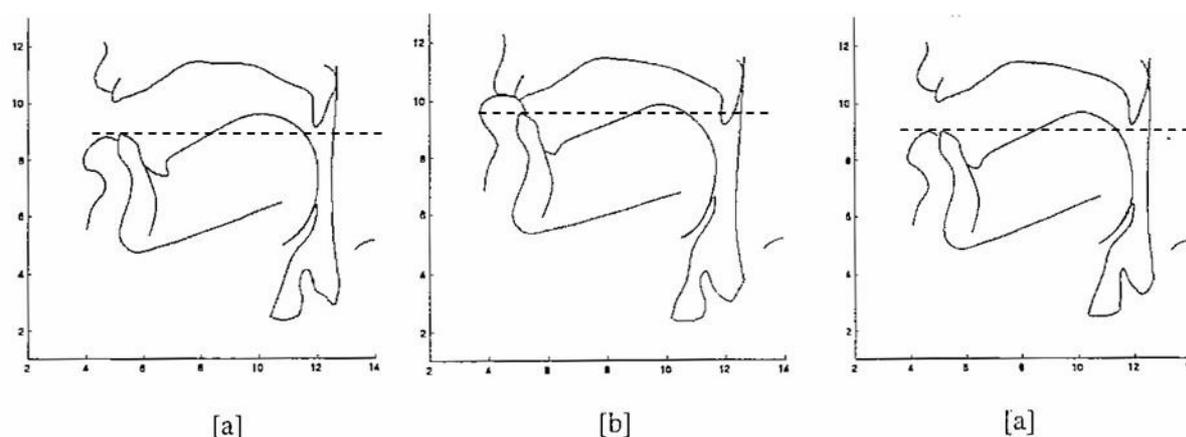


Figure 85 : Coupes sagittales du conduit vocal pour la production d'une séquence [aba]. La configuration globale du conduit vocal pour la voyelle [a] est maintenue durant toute la séquence. La consonne [b] est produite localement par l'occlusion labiale. Remarquer que si l'on référencie le point le plus haut de la langue par rapport à la hauteur de l'incisive inférieure, ce point aura baissée pendant l'occlusion de la consonne. D'après Vilain, 2000, p.177.

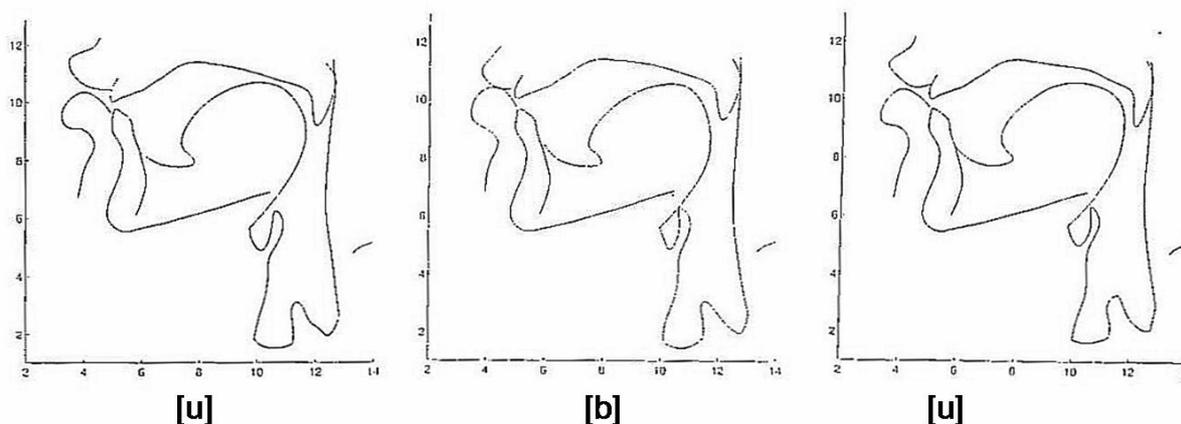


Figure 86 : Coupes sagittales du conduit vocal pour la production d'une séquence [ubu]. La configuration globale du conduit vocal pour la voyelle [u] est maintenue durant toute la séquence. La consonne [b] est produite localement par l'occlusion labiale. D'après Vilain, 2000, p. 179.

Ce geste est visible sur les coupes sagittales ci-dessus par la hauteur entre le sommet de la langue et les incisives inférieures: celle-ci est moindre durant l'occlusion labiale. Lors du relâchement de l'occlusion, l'aperture buccale augmente et la langue est déjà positionnée pour réaliser la cible vocalique. Si la langue n'effectuait pas ce mouvement contrôlé d'abaissement lors de la remontée de la mâchoire, les voyelles produites ressembleraient à des voyelles [ə] et non à des [a] bien formées.

Ce type de contrôle linguo-mandibulaire dans le cas de [ba] est donc crucial car il permet de préparer le conduit vocal pour la réalisation rapide et fluide de la voyelle suivante. Chez l'enfant, l'enjeu consistera à acquérir ce contrôle spécifique de la langue afin de maintenir la constriction du conduit vocal adéquate à la production de la cible vocalique pendant la phase d'occlusion consonantique.

En résumé de cette section, nous reprendrons le point de vue intégrateur proposé par Abry *et al.* (2007) qui propose de revisiter la conception de la coarticulation d'Öhman en l'intégrant dans le cadre développemental que nous venons d'exposer à partir des travaux de MacNeilage (1998) et de Vilain (2000).

Trois contrôles successifs ont été clairement distingués :

- I. le contrôle de l'oscillation mandibulaire, premier contrôle à être acquis vers 7 mois, correspondant au contrôle proximal de la porteuse de la parole.
- II. puis le contrôle indépendant des articulateurs portés (lèvre inférieure et langue). Ceux-ci vont permettre, durant la phase de fermeture du conduit vocal, le contrôle de constriction locale par le rapprochement vers des zones de contact du conduit vocal (lèvre supérieure, palais osseux, velum ou paroi pharyngale).
- III. enfin, durant la phase d'ouverture, le contrôle global du conduit vocal qui pourra aboutir à des patrons de coarticulation proches de ceux de l'adulte.

Les deux premiers contrôles sont clairement hérités de la théorie *Frame-then-Content* de MacNeilage (1998). Le troisième contrôle, modélisé sur le plan articulaire par Vilain (2000), correspond en fait à l'héritage le plus significatif de la conception de la parole selon Öhman, soit le contrôle des transitions vocaliques. Mais replacé ainsi dans une perspective développementale, ce contrôle de l'ensemble du conduit vocal n'apparaît plus comme une primitive, puisqu'il se trouve en réalité surimposé sur le cycle mandibulaire de babillage.

On notera en outre que le maintien du contrôle en position basse de la langue, pendant la fermeture des lèvres dans [aba], peut faire croire à un effet de *trough*. En effet si, comme Lindblom *et al.* (2002), on référence le point le plus haut de la langue par rapport à la mandibule, lorsque celle-ci monte pour aider les lèvres à réaliser la closure du [b], et si la langue est maintenue dans une position basse pour le [a] par l'action du hyoglosse, alors le tracé du point le plus haut de la langue, par rapport à la mandibule, présentera un creux ou une auge. C'est ce qui, sans doute, a fait croire à Lindblom & Sussman (2002, Figure 2) que le contrôle des deux voyelles [a] était désactivé pendant la closure. Sans préjuger des données

sur les voyelles hautes comme [i] (cf. Lindblom & Sussman, 2002, Figure 1 et Lindblom *et al.*, 2002, Figure 6), où le *trough* peut être dû à des phénomènes de pression intraorale pendant la closure, il reste en tous cas que, pour la voyelle la plus basse [a], il n'est pas donné, contrairement à ce qu'affirment Lindblom *et al.*, (2002), que le modèle d'Öhman, contrôlant un geste diphtongal ou une transition (ici stable de [a] vers [a]), ait besoin de faire appel à la notion de désactivation revisitée de Joos (1948), ce qui aboutit en fin de compte à prendre pour argent comptant un modèle de superposition des gestes à la Haskins.

III.4. Développement de la coarticulation chez l'enfant

III.4.1. Indices acoustiques de l'acquisition du contrôle de la coarticulation

Les conséquences acoustiques du phénomène de coarticulation ont été décrites par Sussman *et al.* (1999) à partir du suivi longitudinal, mois par mois, d'un bébé depuis ses premières productions de babillage jusqu'à l'âge de 40 mois (Figure 87). Les auteurs ont choisi de rendre compte des effets coarticulaires consonne-voyelle via l'équation du locus, couramment employée dans les études acoustiques chez l'adulte.

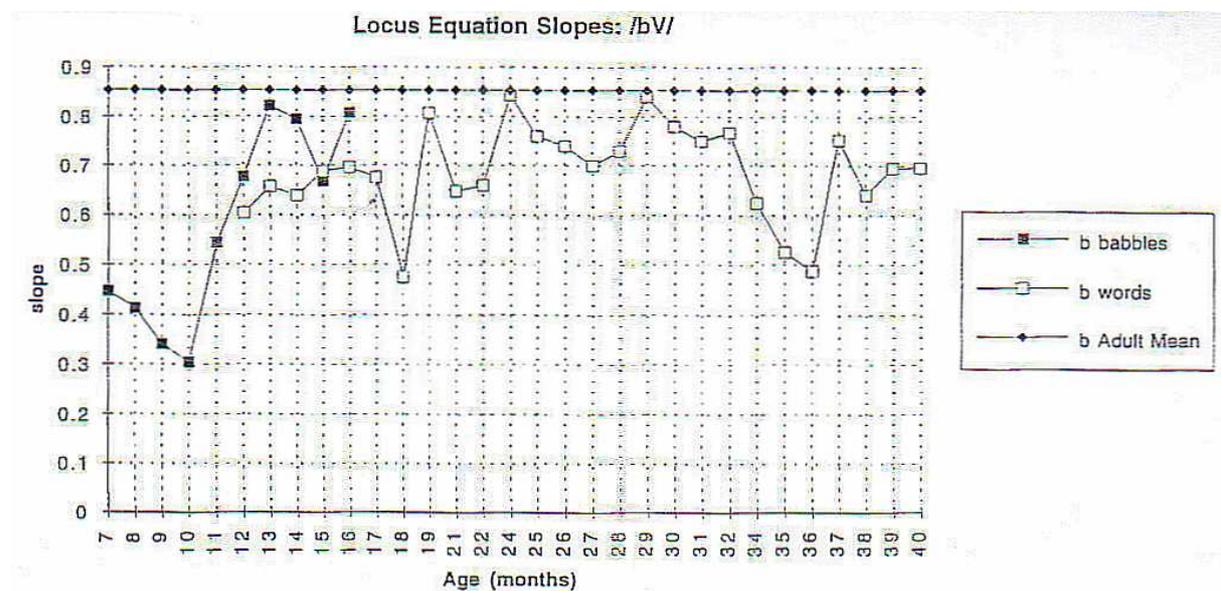


Figure 87 : Pente de l'équation du locus pour des séquences [bV] produite par un enfant chaque mois entre 7 et 40 mois. Les productions de babillage sont représentées par des carrés noirs, les mots par des carrés blancs. Les productions des adultes sont données pour référence (losanges noirs). D'après Sussman *et al.*, 1999, p. 1086.

Leurs analyses des séquences CV comportant l'occlusive [b] montrent que le contrôle de la coarticulation dans des séquences [ba] ne débute pas avant l'âge de 10 mois et l'affinement de ce contrôle semble se poursuivre même après l'apparition des premiers mots (p. 1085).

Celui-ci se manifeste par une influence acoustique de la voyelle sur le profil de la consonne suivante, tout comme l'avaient déjà remarqué Davis et MacNeilage (1995).

Comme le rappellent Sussman *et al.* (1999), l'enfant a globalement besoin d'acquérir deux contrôles moteurs pour la production d'une séquence [ba]: le contrôle de l'occlusion labiale ainsi que celui de la position de la langue

« [...] *the child must learn to integrate tongue position for the vowel concurrently with, independent of lip closure for the stop.* » (Sussman *et al.*, 1999, p. 1093)

Après 10 mois,

« [...] *the child began to produce independent tongue placements for the vowel portion of the CV. Once the motoric independence between lip closure and tongue positioning was achieved it was maintained by the child.* » (Sussman *et al.*, 1999, p. 1093)

A partir de 10 mois, les mouvements articulatoires impliqués pour la production des voyelles se répercutent sur le contexte consonantique. Les consonnes labiales, majoritaires, permettent une coarticulation maximale car la langue peut se positionner librement dans le conduit vocal, préparant la position vocalique indépendamment de ce qui se passe au niveau des lèvres. Avant cet âge, les données collectées indiquent que le bébé privilégie dans un premier temps des associations voyelles-consonnes proches d'un point de vue articulatoire, telles que la voyelle centrale [ə] suivie de l'occlusive labiale [b]. Mais il est important de souligner que ce type d'articulation est le résultat *mécanique* de l'oscillation rythmique de la mandibule, entraînant avec elle les autres articulateurs passifs tels que la langue. Les données de la figure 87 reflètent clairement ce type de comportement à 10 mois.

Le contrôle de la *coproduction*, tel que formulé par Öhman (1967), est donc observé dès la fin de la première année de vie par Sussman *et al.* (1999) en ce qui concerne l'abaissement de la langue.

Qu'en est-il du contrôle de l'occlusion labiale ?

III.4.2. A propos du développement d'un contrôle spécifique: celui des lèvres

Nous avons successivement évoqué l'indépendance de la mandibule se produisant autour des 7 mois, suivie de celle de la langue (acquise vers 1 an). Si la langue est un articulateur essentiel pour la création de constrictions vocaliques, le contrôle des lèvres sert également au contraste vocalique, notamment entre les voyelles dites étirées, telle que [i], et les voyelles arrondies telles que [u, y] en français.

Il nous reste donc à évoquer l'acquisition de l'indépendance des lèvres pour achever le parcours qui mène du contrôle de l'articulateur mandibulaire jusqu'au contrôle indépendant des trois effecteurs recrutés pour la coordination linguo-labio-mandibulaire utilisée en parole. Nous n'évoquerons pas, dans ce chapitre, les études portant sur l'indépendance du velum qui ne sont pas directement liées à notre problématique (Lalevée & Vilain, 2006).

III.4.2.1. L'indépendance graduelle des articulateurs supra-laryngés

Les études décrites précédemment suggèrent un contrôle graduel des articulateurs supra laryngés. Les études cinématiques de Green *et al.* (2000, 2002) confortent cette hypothèse. Ces derniers ont réalisé deux études en production chez des enfants de langue anglaise âgés de 1, 2 et 6 ans et chez des adultes. Par des mesures des mouvements de la lèvre inférieure et supérieure ainsi que de la mandibule, ces derniers ont essayé de dater l'indépendance de chacun de ces articulateurs. Il s'agit de l'une des études articulatoires développementales les plus quantitatives puisqu'au total, 34 locuteurs ont été enregistrés: 5 enfants âgés entre 11 et 13 mois, 9 enfants âgés de 23 à 29 mois, 10 enfants de 6 à 7 ans et enfin 10 adultes âgés entre 27 à 35 ans. La tâche consistait en la répétition des séquences [baba], [papa] et [mama]. Les mouvements de la mâchoire, des lèvres inférieure et supérieure ont été enregistrés avec un système de détection automatique récupérant les positions de trois pastilles éclairées par lumière infrarouge et collées sur le menton, la ligne médiane des lèvres supérieures et inférieures au niveau de la zone vermillon ainsi que sur le front et le nez, pour obtenir des mesures de référence. Le mouvement indépendant de la lèvre inférieure a été calculé en soustrayant celui de la mandibule. Les auteurs ont ensuite comparé les déplacements des articulateurs pour les trois séquences.

Nous avons souhaité présenter la Figure 88 qui ne concerne que les productions de l'adulte, pour pouvoir par la suite comparer plus aisément son patron articulatoire avec celui des enfants. Ce premier graphique rappelle amplement celui de Munhall & Jones (1998) : on retrouve un patron similaire au niveau du déplacement de la mandibule et de la lèvre supérieure. Le mouvement de la lèvre inférieure ayant été isolé, on remarque qu'il est largement tributaire des mouvements mandibulaires.

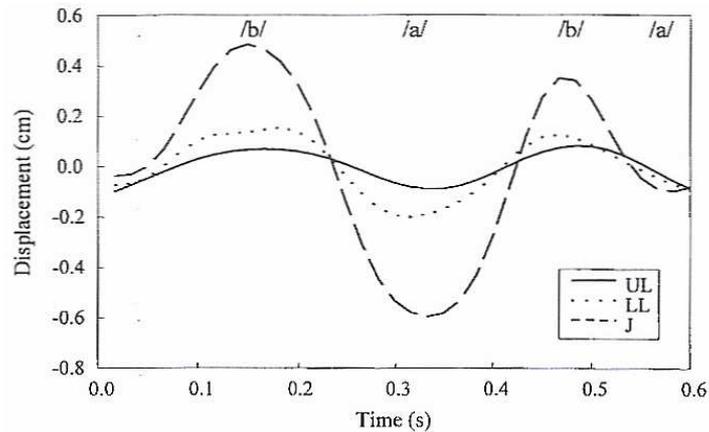


Figure 88 : Moyenne du déplacement vertical en cm de la lèvre supérieure (UL ; ligne pleine), lèvre inférieure (LL ; ligne pointillée) et mandibule (J ; points) pour la séquence [baba] produite par un adulte. Le tracé du mouvement de la lèvre supérieure a été inversé pour faciliter la lecture des signaux. D'après Green *et al.*, 2002, p. 69.

La Figure 89 propose cette fois une comparaison des déplacements verticaux de la mandibule, de la lèvre inférieure et de la lèvre supérieure chez un adulte, un enfant de 6 ans, et deux bébés de 1 et 2 ans. Ces données sont issues d'une étude de Green *et al.* (2000) avec la même méthodologie que celle décrite précédemment.

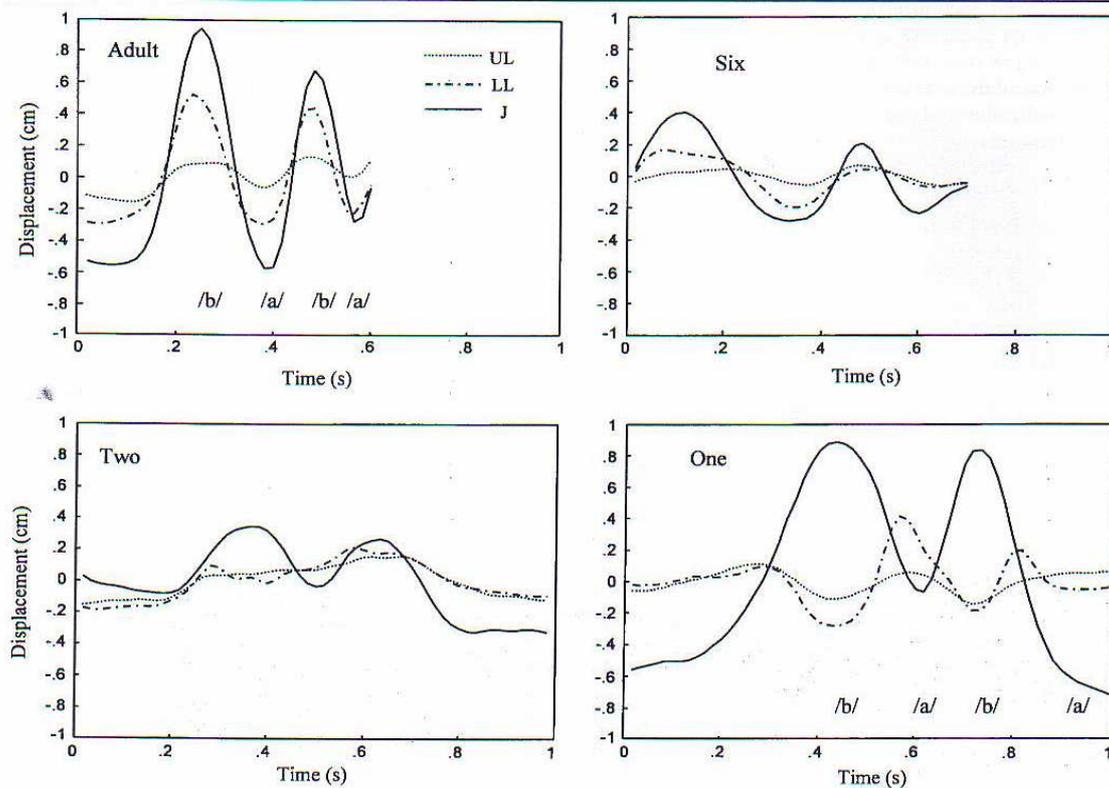


Figure 89 : Déplacements verticaux en cm de la mandibule (UL; ligne pleine), de la lèvre inférieure (LL; ligne pointillée) et de la lèvre supérieure (J; points) pour un adulte (en haut, à gauche), un enfant de 6 ans (en haut, à droite), un bébé de 2 ans (en bas, à gauche) et un bébé de 1 an (en bas, à droite) pour la séquence [baba]. D'après Green *et al.*, 2000, p. 246.

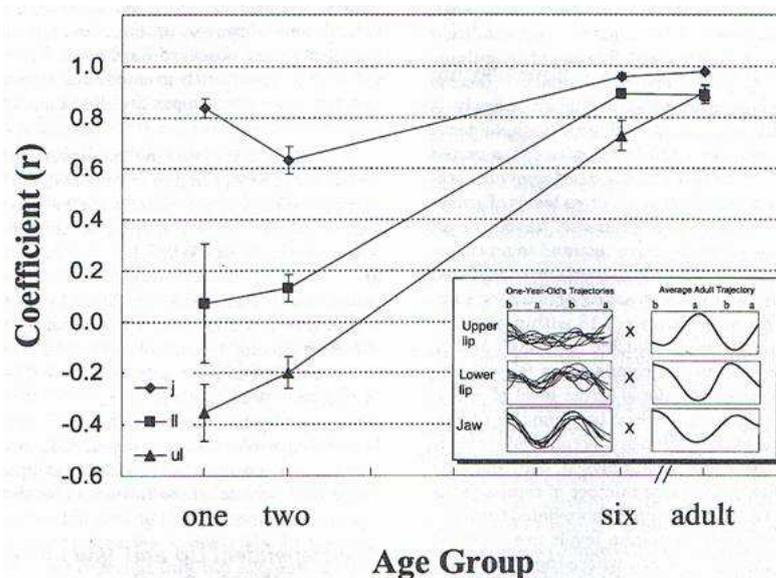


Figure 90 : Comparaison inter-âge du comportement de la lèvre supérieure (triangles), de la lèvre inférieure (carrés) et de la mandibule (losanges). Les points représentent le coefficient moyen r obtenu pour l'ensemble des locuteurs de chaque groupe d'âge. Pour les trois articulateurs, les tracés cinématiques de chaque locuteur ont été mis en relation avec une trajectoire moyenne calculée à partir de tous les tracés de l'adulte. D'après Green *et al.*, 2002, p. 73.

Au vu de la Figure 90, on note très clairement des périodes distinctes de contrôle des trois articulateurs. En ce qui concerne tout d'abord la mandibule, on s'aperçoit que, dès 1 an, le coefficient de corrélation est très fort et le reste à 2 ans, tout comme à 6 ans. Ceci témoigne de la stabilité précoce de la mandibule dont les déplacements sont très proches de ceux de l'adulte.

En revanche, les coefficients obtenus pour les lèvres supérieure et inférieure sont très différents. Jusqu'à 2 ans et encore au-delà, les bébés présentent des mouvements labiaux très variables. On observe une évolution très lente entre 1 et 2 ans (surtout en ce qui concerne la lèvre inférieure), plus nette à partir de 2 ans. Ce n'est qu'à 6 ans que ces mouvements semblent se stabiliser pour la lèvre inférieure rejoignant les patrons de l'adulte. Pour la lèvre supérieure, on note que, même à cet âge, les trajectoires ne sont toujours pas complètement conformes à celles des adultes. Les auteurs expliquent cette instabilité en faisant référence aux travaux de Abbs & Gracco (1984) et Blair (1986) décrivant la plasticité des lèvres et leur contrôle neural complexe.

D'après ces données, le contrôle fin de la closance, tel que réalisé par l'adulte en coordonnant les deux lèvres et la mâchoire dans un déroulement temporel spécifique à une communauté linguistique, n'est pas installé avant 6 ans, voire même plus tard, si l'on considère les profils variables de la lèvre supérieure.

« *The earliest stages of speech motor development are constrained by the non uniform development of articulatory control [...]. The observed developmental patterns suggest that the properties of the oral motor system significantly influence the pattern of speech sound acquisition* ». (Green et al. 2002, p. 66)

Au vu de leurs résultats, Green et al. proposent l'hypothèse suivante:

« *During this period of development [jusqu'à 4 ans], instability in articulatory movement patterns might be anticipated as children move away from "biologically preferred" articulatory patterns toward those required of their ambient language, whereas the stability observed in the mandibular movement at age 1 year in the present study may reflect a state of "prelinguistic equilibrium", with some oromotor movements being sharply constrained by biomechanic and neuromotor factors.* » (Green et al., 2002, p. 76)

D'après eux, au moment du babillage canonique, le bébé entre dans une phase d'*intégration* durant laquelle les mouvements des lèvres sont assimilés aux mouvements contrôlés de la mandibule ; les syllabes produites correspondent à des stéréotypes résultant uniquement de l'action de l'oscillation mandibulaire. La seconde étape consiste en une *différentiation* des contrôles articulatoires, période durant laquelle les articulateurs supralaryngés gagnent progressivement leur indépendance, apportant ainsi des modulations phonétiques aux cadres syllabiques purs (*pure Frames*) et se conformant aux normes de la langue. Cette phase débute autour de 2 ans. L'enfant tente alors de coordonner des *postures* vocaliques et des articulations consonantiques mais avec très peu de succès jusqu'à 6 ans, comme en témoignent les Figures 89 et 90.

L'intégration du contrôle des lèvres pour la production des voyelles nécessite donc une nouvelle étape. Pour la réalisation d'une voyelle arrondie telle que [u], en plus d'un mouvement de la langue opéré par l'action des muscles styloglosse et styloïde, le bébé doit effectuer un mouvement d'arrondissement et d'avancement des lèvres pour modifier la constriction labiale et produire ainsi la voyelle. Ce type de voyelle requiert donc des mouvements plus complexes et semble maîtrisé bien plus tard.

« *A new movement [...] achieved by the cooperation of two muscles inserting on the styloid process of the cranium, the styloglossus and stylohyoid muscles*», et « *another new movement, that of lip rounding, is also necessary for high back vowels. [...] It is also of interest to note that back vowels, which, according to our analysis, make more demand on new movements, are acquired later than high front vowels.* » (MacNeilage & Davis, 1990, p. 470)

Sharkey & Folkins (1985) ont également observé une grande variabilité des profils labiaux chez leurs enfants de 4 ans, par rapport aux enfants de 7 ans et aux adultes, alors que les mouvements de la mandibule sont déjà accordés avec ceux de l'adulte à 4 ans. La période entre 2 et 6 ans semble donc cruciale pour le développement du contrôle des articulateurs

supra laryngés et l'*attunement phonologique* à la langue. Malheureusement, Green *et al.* (2000) n'ont pas réalisé de mesures durant cette période, ce qui nous apporte peu d'informations sur la nature de cette évolution. Ces données confortent toutefois les conclusions des études précédentes, à savoir que les contrôles moteurs de la parole se développent séquentiellement.

De nombreuses études ont montré que durant la période du babillage – période marquante au niveau de l'émergence de la parole – les enfants de toutes communautés linguistiques produisent le même ensemble de sons (Locke, 1983). Il semble donc que les étapes d'acquisition du langage, et plus particulièrement les étapes d'acquisition des contrôles moteurs de la parole soient identiques pour tous les enfants, quelle que soit leur langue d'appartenance. Les études cinématiques de Green *et al.*, (2002), Green *et al.* (2000) et Sharkey & Folkins (1985) suggèrent que cette acquisition soit marquée par un développement graduel des articulateurs supra laryngés.

« The full, species-typical capacity for language is realised gradually, in phases – coherent ontogenetic periods during which cognitive and neural capabilities associated with some aspect of language, or the path to it, “ come on line”. » (Locke, 1999, p. 379)

L'enfant apprend donc en premier lieu à réaliser les contacts labiaux, avec la mandibule, puis les postures vocaliques, par des stratégies dans un premier temps pré-planifiées (« *the short-term strategies adopted for approximating adult speech* » Green *et al.*, 2002, p. 67), enfin par *l'imitation de la trajectoire articuloire adaptée à réalisation de la cible acoustique* avec le déroulement temporel de cette coordination (« *the long-term course of speech motor development* ») qui renvoie à la phase d'attunement phonologique à la langue.

III.4.2.2. L'anticipation d'arrondissement vocalique

Chez l'adulte, les stratégies coarticulatoires ont fait l'objet de nombreuses études acoustiques et articuloires et ont d'ailleurs été largement discutées dans le cadre des différents modèles de la coarticulation anticipante (*cf.* nos parties I et II). Chez l'enfant, on peut regretter que ce contrôle soit encore peu abordé dans le cadre d'études articuloires, les contraintes méthodologiques et techniques liées au recueil de données cinématiques enfantines rendant une fois de plus l'expérimentation très difficile. Toutefois, comme pour la caractérisation du contrôle vocalique, les analyses acoustiques des productions CV peuvent nous fournir des indices sur les changements anatomiques de l'enfant. Une large revue des

études acoustiques ayant exploré l'évolution de l'espace vocalique chez le jeune enfant est fournie par Rvachew *et al.* (2007) ainsi Ménard *et al.* (2007).

De manière générale, la littérature sur la coarticulation anticipante chez l'enfant fait émerger deux points de vue divergents. Certains pensent que la parole de l'enfant est produite plus par segments par rapport à celle des adultes (Kent, 1983 ; Katz *et al.*, 1991). Ce phénomène reflèterait un processus d'acquisition graduel du contrôle moteur, dans lequel les habiletés motrices d'organisation temporelle des sons sont acquises en premier, tandis que les détails plus fins de cette coordination temporelle (*attunement*) seraient acquis plus tard. Dans cette perspective, l'enfant montrerait une plus grande variabilité dans les patrons articulatoires que les adultes.

Un autre point de vue considère que la parole des enfants est organisée de manière syllabique et montre, de ce fait, un degré de coarticulation ou de chevauchement temporel des gestes articulatoires plus important que chez l'adulte (Nittrouer & Whalen, 1989 ; Browman & Goldstein, 1986, qui observe une diminution du chevauchement temporel à partir de 29 mois, surtout en ce qui concerne les voyelles cardinales [a, i, u]). Dans cette perspective, la maturation du système moteur se traduit par un processus de différenciation des contrôles articulatoires et une diminution des influences contextuelles.

Le débat sur l'organisation des unités de parole est ainsi toujours d'actualité et l'exploration de la coarticulation anticipante constitue donc un terrain de recherche fécond pour définir cette organisation des unités de parole.

« If one could find a boundary over which coarticulation did not occur, these boundaries would provide clues indicating units of production. » (Smith, 2006, p. 337)

Peu d'études existent sur la coarticulation anticipante chez l'enfant. Nous détaillerons celle d'Abelin *et al.* (1980) qui propose un ensemble important de données cinématiques de protrusion chez 6 enfants suédois âgés entre 7 et 10 ans ainsi qu'un adulte pris pour référence (il s'agit du locuteur enregistré dans l'étude de Lubker, 1981).

Les auteurs ont enregistré l'activité électromyographique (EMG) de l'orbiculaire supérieur des lèvres lors de productions de mots, phrases et groupes de mots sans sens sous la forme de séquences $V_1C_nV_2$ où V_1 correspondait aux voyelles [a], [i] ou [o], V_2 était choisie parmi les voyelles arrondies [u:, o:, y:, A:, II:, ∞:] et où l'intervalle consonantique C_n pouvait atteindre 5 consonnes. L'activité électromyographique a été traitée afin de repérer le début du geste d'arrondissement des lèvres.

Les Figures 91 et 92 présentent les comportements d'arrondissement pour chacun des 6 enfants et l'adulte de référence. Les auteurs ont mis en relation la durée du geste d'arrondissement (depuis le début du geste jusqu'au début acoustique de la voyelle arrondie) et la durée de l'intervalle consonantique entre les deux voyelles. Une analyse de régression linéaire a été réalisée pour l'ensemble des sujets, mais les coefficients de corrélation et pentes n'ont pas été précisés par les auteurs.

Un des jeunes locuteurs, âgé de 8 ans, présente un profil d'arrondissement de type *Look-ahead* avec une durée du mouvement d'arrondissement relativement proportionnelle à l'allongement de l'intervalle consonantique, se traduisant sur la Figure 91 par une pente proche de 1. Son comportement d'arrondissement vocalique est similaire à celui de l'adulte pris pour référence. Pour ces deux locuteurs, le mouvement d'arrondissement était généralement initié dès la fin de la voyelle non arrondie.

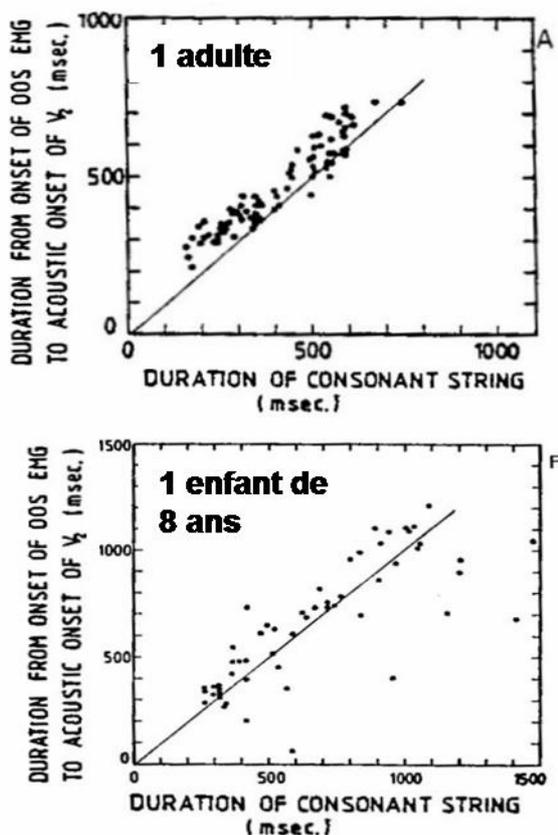


Figure 91 : Relation entre la durée du geste d'arrondissement (depuis le début du geste jusqu'au début acoustique de la voyelle arrondie) et la durée de l'intervalle consonantique entre les deux voyelles pour un adulte et un enfant de 8 ans. D'après Abelin *et al.*, 1980.

En ce qui concerne les cinq autres jeunes locuteurs (Figure 92), Abelin *et al.* (1980) interprètent leurs données comme des comportements de type *Time-locked*.

« Taken together, these observations suggest that for the children in this experiment, the onset of labial emg activity is temporally locked to the onset of the rounded V2. If the data are plotted in terms of the number of intervocalic consonants vs. emg onset, the impression of temporal locking is even more marked. » (Abelin *et al.*, 1980, p. 10)

Cette interprétation des résultats nous semble cependant très discutable dans la mesure où les données sont très dispersées pour chacun des 5 locuteurs (Figure 92). Contrairement à ce que prétendent les auteurs, la durée du geste d'arrondissement n'est pas constante, quelle que soit la longueur de l'intervalle consonantique, comme le prédit le modèle *Time-locked*. Au contraire, les données sont représentées dans une plage temporelle assez large pouvant varier jusqu'à 500 ms, parfois même. Une analyse statistique par régression linéaire révélerait certainement des coefficients non significatifs pour la totalité des sujets.

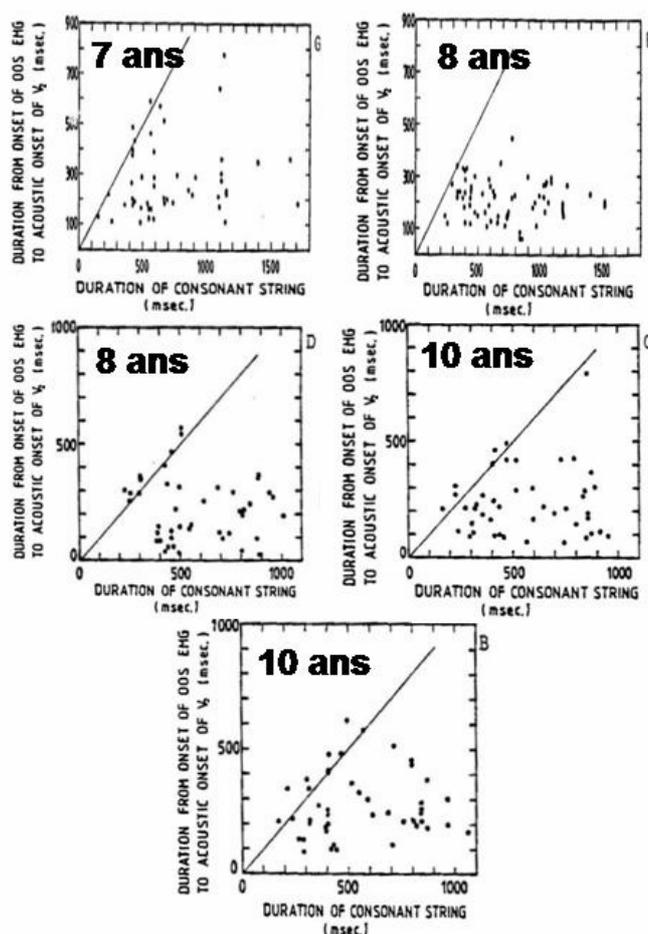


Figure 92 : Relation entre la durée du geste d'arrondissement (depuis le début du geste jusqu'au début acoustique de la voyelle arrondie) et la durée de l'intervalle consonantique entre les deux voyelles pour les 5 autres enfants suédois. D'après Abelin *et al.*, 1980.

En outre, la mise en relation de la durée du geste d'arrondissement par rapport au nombre de consonnes intervocaliques (Figure 93) ne conforte en rien les conclusions des auteurs. L'apparence *Time-Locked* du comportement de ces 5 enfants résulte d'un biais de moyenne

des réalisations en fonction du nombre de consonnes qui serait supprimé si les auteurs avaient fait apparaître les écarts-types. Et il est toujours délicat, même avec cette nouvelle présentation des données, de parler de comportement invariant, étant donné que le début de l'activité EMG, lié au geste d'arrondissement labial, varie, là encore, entre 100 et 350 ms chez les 5 locuteurs. Au mieux pourrait-on dire que ces comportements s'inscrivent dans une plage temporelle comprise entre 100 et 350 ms.

Par conséquent, pour la majorité de ces jeunes locuteurs, s'il n'est effectivement pas possible de rejeter l'hypothèse d'un comportement *Time-locked*, ce n'est pas parce que les locuteurs initient leur geste d'arrondissement de manière invariante par rapport au début acoustique de la voyelle arrondie, mais plutôt parce que ces données sont statistiquement non significatives et ne permettent pas d'infirmer l'*Hypothèse Nulle*.

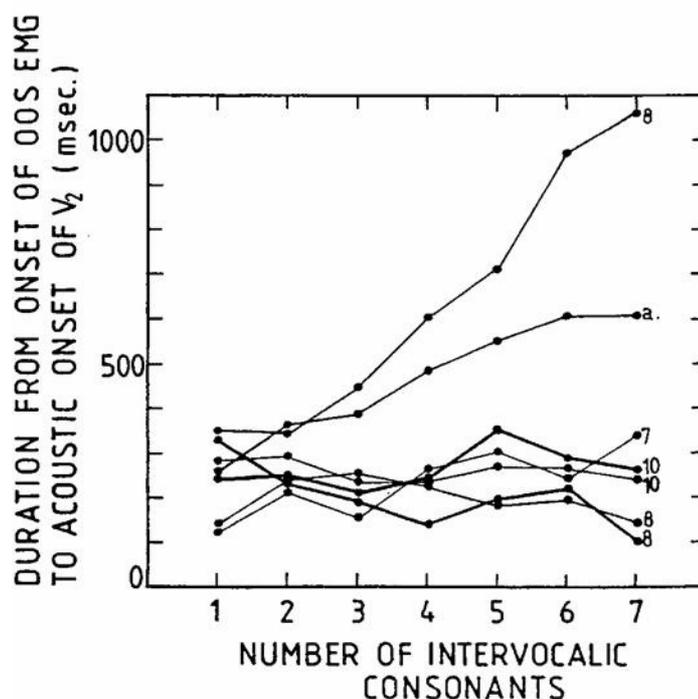


Figure 93 : Relation entre la durée du geste d'arrondissement en msec et le nombre de consonnes intervocaliques pour les 5 enfants Suédois et l'adulte de référence. Les 5 courbes groupées du bas correspondent aux 5 sujets dits '*Time-locked*'. D'après Abelin *et al.* 1980, p. 11.

Ainsi, dans cette partie, nous avons tenté de rendre compte du débat qui existe sur l'acquisition du contrôle vocalique. Il semble que la conclusion de ce débat soit largement une question de point de vue. Si notre objectif est simplement d'attester du niveau d'intelligibilité des voyelles, alors les études perceptives nous apprennent que la plupart de voyelles sont acquises dès un an. Si au contraire, notre approche consiste à rendre compte des contrôles

moteurs régissant la production de ces voyelles, alors il nous faut nous tourner du côté des études acoustiques ou articulatoires qui tentent, par l'analyse de productions enfantines, d'extraire les indices de la maturation de l'appareil vocal.

Les premières études nous semblent pointer vers le processus d'*émulation* des buts acoustiques, mais ces *productions vocaliques* sont encore éloignées de la réalisation de vraies voyelles, telles que celles décrites chez l'adulte. Pour parvenir à cette étape, l'enfant devra acquérir les contrôles séparés, puis coordonnés des effecteurs de la parole pour la reproduction des voyelles de l'adulte (*principe d'imitation*). Ce processus de maturation que nous désignons par *attunement* est calé sur les contraintes phonologiques de chaque langue.

Le but de notre étude ne sera pas d'étudier l'acquisition du contrôle vocalique *per se* mais plutôt de décrire l'acquisition du contrôle de l'organisation temporelle des gestes vocaliques, en tant que norme imposée par la langue. Plus précisément, nous nous intéresserons à l'acquisition ou plutôt l'*attunement* d'un contrôle moteur de la parole, qui est celui de l'anticipation vocalique en français. La description d'un tel contrôle implique d'avoir préalablement saisi les étapes d'acquisition du contrôle vocalique, pour ensuite apprécier leur intégration dans le mécanisme de coarticulation anticipante.

Notre exploration des stratégies anticipatoires chez l'adulte a révélé qu'il s'agit d'un contrôle paramétré par locuteur, en ce sens que chaque individu finit par adopter sa paramétrisation, son coefficient d'expansion du mouvement d'anticipation le plus adéquat possible au vu de sa communauté (plus généralement *de auditu* et *de visu*) pour la production fluide des segments successifs de sa parole.

La question que l'on peut alors se poser se résume de la manière suivante. Est-ce que le développement de ce contrôle de la coarticulation est un phénomène d'acquisition gradient, qui se développe graduellement avec de fortes variations individuelles ? Ou est-ce qu'il existe un âge critique pour l'acquisition de ce contrôle moteur spécifique à la langue, avec sa paramétrisation personnelle ?

**Partie IV. Extension du *MEM* de constriction
à des enfants de 3 à 8 ans :
l'attunement à la langue à travers l'anticipation
vocalique**

L'étude présentée dans ce chapitre vise à tester, sur le domaine encore bien méconnu qu'est celui de l'acquisition du contrôle de l'anticipation d'arrondissement, le modèle *MEM* de constriction qui, à l'issue de notre chapitre II, reste le seul commun à des adultes de différentes langues comme le français et l'anglais. Ce qui signifie que, quel que soit le pas d'acquisition de ce modèle (l'ontogénèse de son *attunement*), l'enfant de l'une ou l'autre langue n'a pas d'autre choix que de converger sur *ce* modèle. Précisons bien que dans ce cadre, la seule question qui reste essentiellement pendante, est uniquement une question de date, ce qui n'enlève rien à la réponse que ce modèle peut apporter à la question de l'émulation du but et à celle de l'imitation du but et du chemin. Autre précision, il n'est pas question de mesurer prioritairement la protrusion chez l'enfant, pour mettre en évidence un modèle qui repose dans son extension la plus large sur le contrôle de la constriction aux lèvres.

Dans le cadre de notre étude, nous nous sommes attachés à décrire le comportement d'anticipation vocalique chez de jeunes enfants français en essayant de couvrir une période cruciale pour le développement du langage. Au-delà de la maîtrise de voyelles statiques isolées (prolongées, Ménard, 2002), notre étude porte sur l'acquisition du contrôle vocalique dans le déroulé temporel de ses coordinations oro-laryngées, ici labiolaryngées. Dans le cadre déjà exposé, qui est celui d'une distinction entre *émulation* et *imitation* en acquisition, nous pouvons rappeler ce qui s'est déjà passé en-deçà de cette période, celle qui correspondra au plus jeune de nos enfants, où selon Kent (1976) débute véritablement l'usage extensif du langage chez l'enfant :

« [...] *the age of two or three years is assumed as the beginning of a child's extensive speech usage on a daily basis.* » (Kent, 1976 p. 440)

C'est à savoir le passage d'un stade de simple *émulation* des buts acoustiques, tel que nous pensons que les données de Kuhl & Meltzoff (1996) le montrent, à un stade d'*imitation* du trajet d'anticipation : l'*attunement*. Les données de la littérature donnent les premiers indices d'un début de maîtrise en ce qui concerne la coproduction du [a] vers un an (Sussman *et al.*, 1999).

Alors que le contrôle de la closure des lèvres n'est, selon Green *et al.* (2002), pas vraiment établi à deux ans, mais seulement à 6 – sachant que leurs données sont lacunaires entre ces deux âges –, l'acquisition à la fois d'un contrôle de la langue et d'un nouveau contrôle des lèvres, non plus seulement la closure mais aussi l'arrondissement, représente un nouveau défi

pour l'enfant. C'est ce que nous avons signifié en rappelant d'entrée la position de MacNeilage & Davis (1990) :

« *A new movement [...] achieved by the cooperation of two muscles inserting on the styloid process of the cranium, the styloglossus and stylohyoid muscles*», et « *another new movement, that of lip rounding, is also necessary for high back vowels. [...] It is also of interest to note that back vowels, which, according to our analysis, make more demand on new movements, are acquired later than high front vowels.* » (MacNeilage & Davis, 1990, p. 470)

D'après la revue de la littérature que nous avons donnée, ce dernier type de contrôle, en ce qui concerne son organisation temporelle dans l'anticipation, semble être acquis encore bien plus tardivement (cf. exemplairement Abelin *et al.*, 1980). Des changements importants se produiraient entre 5 et 8-9 ans, voire jusqu'à 11-12 ans, toujours selon Kent (1976).

« [...] *the variability of speech motor control progressively diminishes until the age of eight-12 years, when adultlike stability is achieved.* » (Kent, 1976, p. 440)

Il rapporte que Tingley & Allen (1975) ont suggéré plus fondamentalement, en ce qui concerne le contrôle :

« [...] *motor patterns for speech are largely preplanned rather than feedback controlled even at the age of five years.* » (Kent, 1976, p. 439)

Quoi qu'il en soit, en utilisant des indices indirects sur l'articulation comme les fréquences de formants, les patrons de Fo et les patrons de timing, il constate que:

« [...] *the declining variabilities for the three types of measurements appear to reflect an increasing precision of motor control over a five-to-eight years.* » (Kent, 1976, p. 440).

De fait, avant cet âge de 8 ans, les productions vocaliques par les enfants ne correspondraient pas vraiment aux patrons adultes. On pourrait les considérer comme des gabarits individuels variables qui sont plus ou moins bien achevés, ceux-ci dépendant de l'enfant, de son âge et du contexte phonétique, réalisés à l'intérieur d'une fenêtre temporelle qui n'est pas encore finement *tunée* : de notre point de vue, sans patron d'anticipation stabilisé.

Il est évidemment difficile, ainsi que nous l'avons déjà vu d'après les données de la littérature, de déterminer, à l'intérieur de cette large période de 2 à 12 ans, des étapes bien définies dans l'acquisition du contrôle moteur de la parole. Mais l'on peut avancer sans risque que notre étude expérimentale, portant sur des enfants de 3 à 8 ans, se situe dans une période

au cours de laquelle des changements importants se produisent notoirement dans les habiletés motrices de l'enfant.

Le véritable enjeu de cette seconde partie est donc pour nous – après l'enjeu de l'extension du *Modèle d'Expansion du Mouvement* aux adultes franco-québécois et anglo-américains – d'étendre le même modèle au jeune enfant. Ces enfants seront français car dans cette thèse, nous nous limitons par priorité à la proposition que si le modèle est, dans sa formulation précise, un succès pour la comparaison des comportements de langues aussi phonétiquement différentes dans le contrôle des voyelles que l'anglais et le français, ce n'est plus qu'une question empirique de savoir quand ce *modèle à principe et paramètre* va être acquis par l'enfant de telle ou telle langue. C'est ce que nous montrerons uniquement pour l'enfant français, à charge de continuer, nous ou d'autres chercheurs, pour d'autres communautés, ce qui est, dans sa structure, le même programme de recherches que celui des extensions à d'autres langues que le français et l'anglais. Dans ce sens, la partie II s'est tout aussi limitée par stratégie et contingence de recherche au minimum requis pour l'extension : deux langues.

Rappelons que la seule étude sur l'enfant testant spécifiquement l'expansion de l'anticipation d'arrondissement est l'étude sur le suédois d'Abelin *et al.* (1980), que nous avons présentée en détail à la section III.4.2.2. Il s'agissait du repérage de l'anticipation par l'activité EMG de l'orbiculaire supérieur chez 6 enfants âgés de 7 à 10 ans. Nous avons largement expliqué en quoi cette étude ne permettait pas de conclure, pour différentes raisons, à un comportement régulier, hormis celui d'un enfant de 8 ans, comparable à l'adulte de référence.

Notre étude se démarque de l'étude d'Abelin *et al.* (1980) sur plusieurs points :

- I. nous suivrons l'évolution de l'aire aux lèvres – par la technique du maquillage en bleu des lèvres – puisque ce paramètre acoustiquement crucial s'avère en outre le plus robuste, comme nous l'avons amplement démontré chez l'adulte, pour décrire l'anticipation d'arrondissement ;
- II. nous testerons des enfants aussi jeunes que possible, compte tenu des contraintes d'élicitation et de procédure expérimentale (Green *et al.*, 2003), qui sont des contraintes lourdes, que jusqu'à présent nous n'avons pas réussi à contrôler avant 3;5 ans, âge du plus jeune de nos petits locuteurs;
- III. nous adopterons une approche longitudinale pour les plus jeunes de nos locuteurs, soient deux sessions à un an d'intervalle, ce qui n'avait pas été entrepris dans cette

seule étude comparable, celle d'Abelin *et al.* (1980). Nous en attendons éventuellement une chance d'observer un changement de comportement sur un enfant qui n'aurait pas démontré la régularité (*lawfulness*) de celui-ci.

Tout à la fois plus généralement et plus en détail, nos principaux objectifs sont :

- I. d'étendre nos connaissances sur l'acquisition du contrôle vocalique par le biais de données articulatoires, rarement obtenues jusqu'ici pour les enfants;
- II. d'évaluer s'il existe, chez l'enfant, une période critique pour l'acquisition du contrôle temporel des gestes nécessaires à la réalisation du comportement d'anticipation, tel que rencontré chez l'adulte ;
- III. de déterminer si l'adulte et l'enfant ont des stratégies coarticulatoires similaires ou, au contraire, si celles-ci sont différentes, avec l'idée que l'enfant adopterait progressivement le patron de l'adulte par *attunement*.

IV.1. Protocole expérimental

IV.1.1. Sujets

Sept enfants de langue maternelle française (cinq garçons NP : 3;5 ; KV : 4;5 ; TB : 4;5 ; JB : 7 ; DV : 8 et deux filles CA : 4 et enfin EC : 4) ont été enregistrés au sein de l'ICP (Tableau 16). Etant donné le jeune âge de certains enfants et la réserve que peut susciter auprès des parents ce type d'enregistrement articulatoire, nous avons préféré recruter nos petits locuteurs par connaissance, soit au sein du laboratoire soit au sein de notre entourage. Ce choix a également été motivé par le fait que nous souhaitions adopter une approche longitudinale avec nos locuteurs les plus jeunes. En recrutant les enfants au sein de notre entourage, il nous était ainsi plus facile de maintenir le contact avec les familles.

Au cours du recrutement de nos locuteurs, nous nous sommes assuré qu'aucun enfant ne présentait de troubles articulatoires ou de l'audition (ces informations étant attestées par une petite anamnèse recueillie auprès des parents ainsi que par les évaluations scolaires) et qu'aucun d'entre eux n'avait d'accent régional.

Afin d'analyser le comportement d'anticipation des enfants par rapport à celui de l'adulte, nous avons considéré les données labiales de deux adultes de sexe féminin (CL et MA), préalablement enregistrées dans les mêmes conditions au sein du laboratoire (Ménard *et al.*, 2004).

Initiales des jeunes locuteurs français	Age (1 ^{ère} session)	Age (2 ^{ème} session)	Sexe
NP	3;5	4;5	H
CA	4	5	F
EC	4	/	F
KV	4;5	5;5	H
TB	4;5	5;5	H
JB	7	/	H
DV	8	/	H

Tableau 16 : Ages des 7 jeunes locuteurs indiqués en année;mois.

IV.1.2. Corpus

L'exercice consistait, pour chaque petit locuteur, en la production, en ordre aléatoire, de petits énoncés dépourvus de sens présentés aux enfants comme des noms de peluches. Ces noms ont été imaginés sous la forme [iC_ny] où C correspondait à un nombre n de consonnes intervocaliques variant entre 0 et 3.

Les noms [isy], [iky], [iksy], [ikRy], [itkRy], [iskRy], [iksty] ainsi que la transition minimale [iy] ont été intégrés à des phrases porteuses de type « Le toutou Iku est rouge » afin de donner du sens à ces productions et favoriser l'intérêt des jeunes locuteurs.

Nous avons obtenu entre 20 et 61 séquences selon le locuteur. Cette variabilité interlocuteur est liée au fait que les plus jeunes enfants ont parfois éprouvé des difficultés à produire les énoncés composés d'une longue chaîne consonantique, en particulier [iskRy] et [itkRy] comportant la consonne vibrante [R]. D'autre part, des séquences ont été supprimées lors du traitement des données audio-visuelles en raison d'aléas techniques que nous expliciterons plus tard. Contrairement au corpus recueilli chez l'adulte, nous avons choisi de tester le contraste vocalique [i] vs. [y] afin de pouvoir tout simplement comparer nos résultats avec les données existantes chez l'adulte français (Abry *et al.*, 1996).

IV.1.3. Dispositif d'enregistrement des données audiovisuelles

Nos locuteurs ont été enregistrés dans la chambre sourde de l'ICP. Etant donné le jeune âge des enfants, chaque enregistrement a été réalisé avec la présence d'un parent.

En ce qui concerne nos quatre plus jeunes locuteurs (CA, KV, NP, TB), nous avons réalisé deux sessions d'enregistrement, séparées par une année d'intervalle, ceci nous permettant de suivre l'évolution de leur comportement d'anticipation. Afin de minimiser la fatigue et la lassitude des enfants vis-à-vis de l'exercice, nous avons pris soin d'élaborer un

corpus d'étude relativement court, n'excédant pas 15 minutes d'enregistrement. Le temps de préparation restait important (environ 30 à 45 minutes afin que les enfants se sentent pleinement en confiance durant l'enregistrement).

Un soin particulier a donc été apporté à la mise en scène ludique de ces enregistrements afin que les jeunes locuteurs soient motivés par la tâche, mémorisent facilement les items sélectionnés et produisent confortablement les énoncés. Nous leur avons expliqué qu'il s'agissait d'un enregistrement de voix pour les personnages d'un dessin animé. Nous avons attribué les 8 items cibles à 8 peluches disposées dans une salle annexe à la chambre sourde. Nous souhaitons, non pas que l'enfant mémorise parfaitement chaque item avec la peluche correspondante, mais qu'il associe simplement ces énoncés à un objet réel, tel qu'une peluche, afin de donner un sens à l'exercice. Chaque enregistrement était précédé d'une phase de jeux, de sorte que l'enfant se familiarise avec les séquences étudiées mais également avec l'expérimentateur. Afin de créer un cadre agréable, nous avons également décoré la chambre sourde avec des guirlandes lumineuses, des étoiles lumineuses et des tissus brillants.

Durant l'enregistrement, l'enfant était confortablement installé dans un siège auto aux côtés de deux expérimentateurs. L'un d'eux était assis face à l'enfant à une distance moyenne de 1 mètre et lisait l'une des 8 séquences en adoptant une intonation neutre et un débit constant puis demandait à l'enfant de répéter la séquence en question. Les 64 séquences (8 séquences x 8 répétitions) étaient proposées en ordre aléatoire. L'enregistrement était ponctué par des encouragements réguliers de manière à stimuler l'enfant dans l'exercice.

Le second expérimentateur se tenait assis à côté de l'enfant. Il s'assurait de la tenue du maquillage et surveillait les éventuels mouvements de l'enfant. Il fallait éviter que celui-ci bouge les mains et qu'il les porte à son visage. De même, il lui était rappelé de ne pas faire de bruit avec ses mains ou ses pieds lorsqu'il répétait les séquences.

A la fin de l'exercice, un goûter était organisé et chaque enfant recevait un diplôme, ainsi qu'un petit cadeau en remerciement de sa participation.

Nous avons globalement appliqué le même dispositif expérimental que celui utilisé pour l'adulte. Comme il est évidemment plus délicat d'enregistrer de jeunes enfants, nous avons procédé à quelques ajustements quant à la procédure d'élicitation. Pour la collecte de nos données enfantines, nous avons donc respecté les critères suivants.

- I. Nous avons soigneusement maquillé les lèvres des enfants avec un bleu à lèvres, en insistant sur les commissures, pour ainsi favoriser l'acquisition des contours labiaux avec le système de détection automatique par Chroma-key développé dans

notre laboratoire. Ce maquillage est résistant à l'eau et s'avère particulièrement recommandé chez le jeune enfant car celui-ci a souvent les lèvres humides. Or, il est important de conserver un maquillage homogène pour une détection efficace des contours labiaux (surtout des contours internes qui permettent le calcul de l'aire intérolabiale).

II. De manière à minimiser tout mouvement du corps et du visage, nous avons utilisé un siège auto, muni d'une ceinture de sécurité, ainsi qu'un bloc de mousse comme repose tête (Figure 95). Ce dispositif vise à maintenir une assise droite pendant la durée de l'enregistrement.

III. Le système d'acquisition des contours labiaux nécessitant un éclairage important, nous avons utilisé deux lumières néons fluorescentes¹⁹ de 220 watts chacune, positionnées face à l'enfant de chaque côté de la caméra. Pour protéger les enfants de la lumière, nous leur avons fait porter des lunettes opaques, celles-ci nous servant également à coller deux pastilles bleues de calibration (Figure 94).

IV.



Figure 94 : Photo du dispositif expérimental pour l'enregistrement des enfants.

¹⁹ Il s'agit de deux néons recréant une lumière du jour par un flux lumineux de 2549 Lux ou lumens à un mètre de distance (notons que c'est la quantité de lumens qui définit le rendement lumineux de l'éclairage). L'emploi de ce type de lampes fluorescentes est intéressant dans la mesure où elles bénéficient d'un rendement beaucoup plus important que les lampes classiques (type halogènes ou lumières à incandescence).



Figure 95 : Vue globale du dispositif technique pour l'enregistrement des locuteurs.

Chaque enfant a été enregistré à l'aide du « poste Visage-Parole » de l'ICP (Lallouache, 1991). La caméra utilisée est une caméra 3CCD de marque JVC KY 15E, ayant une résolution de 638x582 pixels, et un rapport signal sur bruit de 55dB. L'image est codée au format RVB (quantité appropriée de rouge, vert, bleu composant chaque pixel de l'image). Le temps de pose de la caméra est réglable selon la qualité d'image que l'on souhaite obtenir. Pour les besoins de notre étude, nous avons réglé la vitesse d'obturation (ou temps de pose) à 1/250^{ème} de façon à obtenir une détection très fine des paramètres labiaux. La caméra était placée face au jeune locuteur et réglée en mode zoom pour obtenir une visibilité maximale au niveau des lèvres.

Avant le début de l'enregistrement, quelques réglages ont été réalisés au niveau de la caméra afin de garantir une qualité optimale des images recueillies. Il s'agit des mêmes réglages opérés lors du recueil des données audio-visuelles chez l'adulte :

- I. réalisation de la balance des blancs à l'aide d'un plan blanc placé devant le visage du petit locuteur ;
- II. plan de référence (carreaux de 1cm x1cm) pour la conversion des pixels en centimètres, nécessaire à la détection ultérieure des contours labiaux. Ce plan, positionné devant le visage du locuteur, était filmé durant quelques secondes.

Les séquences vidéo ont été enregistrées à une fréquence de 25 Hz et stockées sur un magnétoscope Betacam SONY SP UVW 1600 avec une résolution de 768x576 pixels²⁰.

Il s'agit de la norme PAL-SECAM en vigueur en France.

Le signal acoustique, synchrone à la vidéo, a été enregistré à une fréquence de 44.1 KHz (encodé en 32 bits et stockés en 24 bits), avec un micro AKV 1000S, dont la sortie était reliée au magnétoscope Betacam.

IV.1.4. Prétraitement des données audiovisuelles

Tout comme lors du recueil des données audiovisuelles chez l'adulte, nous avons procédé à différentes manipulations et traitements des signaux bruts, avant que ces derniers puissent être analysables avec les programmes Matlab.

Nous avons numérisé les séquences avec le logiciel CAPTURE (Audouy, 2000). Ce logiciel fonctionne à partir d'un fichier Report (fichier texte) dans lequel sont spécifiées les informations suivantes:

- les *Time Codes* de début et fin de chaque séquence (exprimée en heures : minutes : secondes : millisecondes);
- la fréquence d'échantillonnage du signal sonore;
- le répertoire dans lequel les images seront stockées.

A partir du fichier Report, le logiciel extrait automatiquement, pour chaque séquence, les images correspondantes au format Bitmap ainsi que le signal acoustique, synchrone à la vidéo.

Voici ci-dessous un exemple de ligne écrite dans un fichier Report pour la capture d'une séquence audiovisuelle correspondant à la transition cible [iy].

<u>Time codes (début/fin)</u>	<u>Freq. d'échantillonnage</u>	<u>Répertoire de transfert des images</u>
00:14:03:18 00:14:04:12	22050	D:\aude\petits_locuteurs \kilian_sess2\iy1\

Les contours labiaux de face (Figure 96) ont été automatiquement détectés pour chaque séquence avec le logiciel TACLE (Traitement Automatique des Contours des LEvres, Audouy, 2000). Une description complète de ses principes de fonctionnement et des paramètres de traitement est proposée en partie II. Nous rappellerons simplement que ce système effectue tout d'abord un désentrelacement des images, nous permettant ainsi d'obtenir, après interpolation des lignes manquantes, deux trames A et B ayant chacune une durée de 20. Puis, ce dernier opère par seuillage numérique des zones bleues (traitement numérique par Chroma-key) sur les images préalablement numérisées. Les valeurs ainsi obtenues sont sauvegardées dans un fichier (.tac) lisible avec le logiciel Excel ou peuvent être

visualisées sous la forme de courbes avec l'interface TRAP ou l'application ATAC (comme pour l'adulte).

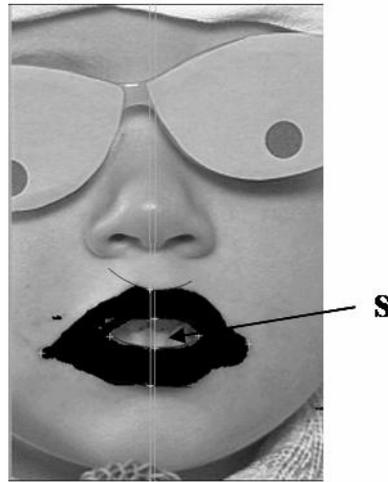


Figure 96 : Image obtenue après seuillage en noir par Chroma-key numérique. S représente l'aire intérolabiale calculée en cm^2 .

Nous avons procédé à une vérification systématique de l'ensemble des séquences audiovisuelles traitées avec le système TACLE, de manière à nous assurer qu'aucune erreur de détection ou liée à un dysfonctionnement technique ne fausse nos interprétations futures.

Le corpus ayant été enregistré en chambre sourde, nous avons obtenu un enregistrement sonore de très bonne qualité, ce qui a facilité la délimitation des cibles vocaliques.

Nous avons réalisé un étiquetage phonétique manuel (se présentant sous la forme de TextGrids) avec le logiciel de traitement de signaux de parole PRAAT. Les séquences sonores ayant été préalablement numérisées à partir du corpus vidéo avec le logiciel CAPTURE, nous avons directement construit nos TextGrid à partir de ces petites séquences et non à partir du signal sonore global, comme c'était le cas pour les adultes.

Chaque TextGrid ne comportait qu'un seul niveau hiérarchique de segmentation correspondant à l'étiquetage phonétique des transitions cibles du corpus.

Pour la délimitation manuelle des voyelles [i] et [y], nous nous sommes référés au début et à la fin de la structure formantique sur le spectrogramme. Nous avons ainsi défini l'*Intervalle d'Obstruence* (IO) entre les deux voyelles. Rappelons que cet étiquetage est crucial pour notre étude ; non pas pour la transcription des consonnes intervocaliques *per se* mais pour le calcul de la durée de la transition consonantique par rapport au geste d'arrondissement labial initié pour la voyelle arrondie [y].

IV.1.5. Mesures des événements et des phases sur le signal d'aire aux lèvres

Le comportement d'anticipation d'arrondissement de nos locuteurs a été analysé avec l'interface TRAP. Par le jeu d'une superposition de fenêtres graphiques, il était possible de visualiser l'évolution temporelle de l'aire intérolabiale (exprimée en centimètres carrés) pour chaque séquence sonore ainsi que sa vitesse. Notons que le signal d'aire aux lèvres a été préalablement lissé à l'aide d'un algorithme Matlab, par un filtre passe-bas de Tchebychev avec une fréquence de coupure à 7 Hz.

Pour le repérage et l'étiquetage des événements et phases définies dans le *Modèle d'Expansion du Mouvement (MEM)*, nous avons appliqué la même procédure que pour les signaux collectés chez l'adulte. Certains ont été sélectionnés manuellement par le biais de l'interface TRAP, d'autres selon une procédure semi-automatique. Un exemple des différents événements et des phases est proposé, dans la Figure 97, pour la séquence « Le dragon Issu est orange » produite par le jeune locuteur TB, alors âgé de 5 ans et demi.

A partir de l'étiquetage phonétique réalisé au préalable sur le signal acoustique, nous avons extrait automatiquement l'*Intervalle d'Obstruence (IO)* (exprimé en secondes) par un algorithme Matlab.

Sur la fonction d'aire intérolabiale, les étiquettes correspondant respectivement à la valeur maximale d'aire intérolabiale pour la voyelle [i] (événement 1), ainsi qu'à la valeur minimale pour [y] (événement 4), ont été apposées manuellement pour chaque séquence, par le repérage du passage par zéro de la vitesse (lignes pointillées, Figure 97).

A partir de l'identification de ces deux événements sur la fonction d'aire, nous avons calculé les deux phases de *Time Falling* et de *Hold*. Pour leurs calculs respectifs, nous avons considéré trois événements temporels : le temps correspondant à 10% de l'écart entre la valeur maximale et minimale d'aire (événement 2), le temps correspondant à 90% de cet écart (événement 3) et enfin le temps correspondant à une remontée de 10% de l'aire suivant l'aire minimale (événement 5). La phase de *Time Falling* (exprimée en millisecondes) représente ainsi l'intervalle temporel délimité entre les événements 2 et 3; la phase de *Hold* (également exprimée en millisecondes) est délimitée entre les événements 3 et 5. Ces trois repères ont été calculés automatiquement par un programme TRAP, en délimitant au préalable la zone à explorer.

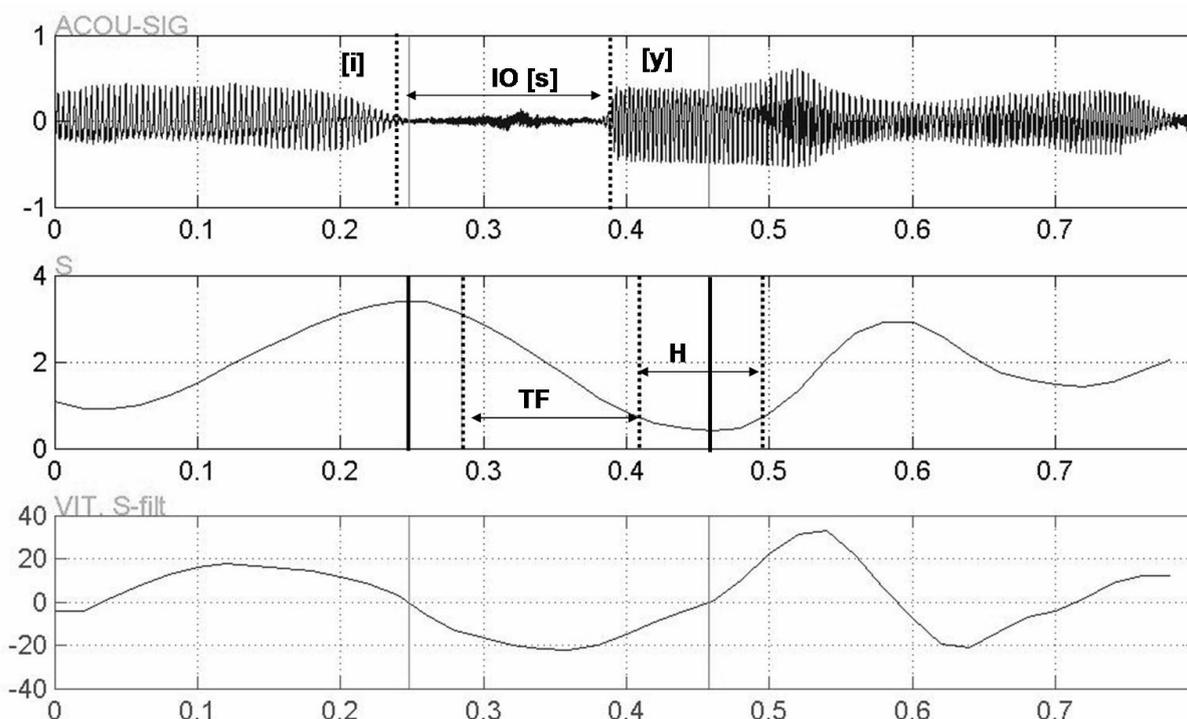


Figure. 97 : Visualisation du signal acoustique de la transition [isy] (signal du haut) à partir de la séquence « Le dragon Issu est orange » produite par le locuteur TB âgé de 5 ans et demi. Tracés de l'aire intérolabiale (signal du milieu ; en cm^2) et de sa dérivée vitesse (signal du bas). L'Intervalle d'Obstruence (IO) est délimité, sur le signal acoustique, par les lignes pointillées. Les étiquettes correspondent à 1 : l'aire maximale de [i] ; 2: 10% de la différence d'aire entre [i] et [y] ; 3: 90% de la différence d'aire entre [i] et [y] ; 4 : aire minimale de [y] ; 5 : 10% de la différence d'aire entre [i] et [y] suivant la valeur minimale de [y]. L'intervalle entre les événements 2 et 3 détermine la phase de *Time Falling*, celui entre les événements 3 et 5 la phase de *Hold*.

IV.1.6. Aléas techniques rencontrés au cours de l'étude

Lors de l'acquisition automatique des contours labiaux, nous avons parfois observé des mesures erronées de l'aire intérolabiale, liées à la détection de pixels bleus isolés à l'intérieur de la zone d'aire intérolabiale. La présence de ces pixels parasites résultait du reflet du maquillage sur les dents très claires de certains de nos très jeunes locuteurs. Lorsque ce type d'erreur apparaissait sur quelques images, nous procédions à une correction manuelle avec un utilitaire de retouches des contours labiaux (élaboré dans notre laboratoire par Christophe Savariaux) ou plus rarement avec le logiciel Paint Shop Pro. Lorsque ce type d'erreurs affectait la totalité d'une séquence (ou du moins la transition vocalique étudiée), nous éliminions la séquence de nos analyses.

IV.2. Extension du MEM de constriction à l'enfant francophone

Nous avons montré, au chapitre II, qu'à partir d'un modèle paramétrable par locuteur, il était possible de rendre compte du contrôle de l'anticipation d'arrondissement, lorsque celui-ci est maîtrisé au stade adulte (*tuned*). Nous testerons l'acquisition d'un tel contrôle chez l'enfant, en nous référant au *Modèle d'Expansion du Mouvement*.

Rappelons les questions qui ont guidé notre démarche :

- I. Existe-t-il un âge critique pour l'acquisition du comportement anticipatoire ? Peut-on attester tôt dans l'enfance d'un certain contrôle de la coarticulation anticipante, tel qu'observé chez l'adulte, ou est-ce que l'acquisition de ce contrôle moteur spécifique suit le processus d'acquisition graduel du contrôle vocalique, dont nous avons mentionné les étapes dans la partie III? Il s'agira de déterminer si l'on retrouve, dans l'acquisition du contrôle de l'arrondissement vocalique, ce passage de l'*émulation* du but vers l'*imitation* de la coordination articulaire ajustée temporellement aux besoins de la langue.
- II. Peut-on envisager d'appliquer un paradigme expérimental élaboré initialement pour l'adulte à des enfants en bas âge, en l'occurrence de 3 ans et demi à 8 ans ? Au niveau du corrélat articulaire susceptible de retracer efficacement l'acquisition ou plutôt l'*attunement* du comportement d'anticipation vocalique à la langue, nous avons choisi la constriction. Il s'agit d'un paramètre pertinent pour rendre compte de la dynamique labiale dont on connaît désormais les implications au niveau de la sortie acoustique du conduit vocal. Notre exploration des stratégies d'anticipation chez l'adulte français, franco-québécois et anglo-américain a montré que la constriction était le seul corrélat articulaire suffisamment robuste pour la description de l'anticipation vocalique.

Nous présenterons les résultats des deux sujets adultes de référence, puis successivement les résultats obtenus par les enfants lors des deux sessions d'enregistrement.

IV.2.1. Résultats des adultes de référence

Les résultats des deux adultes sont présentés à la Figure 98.

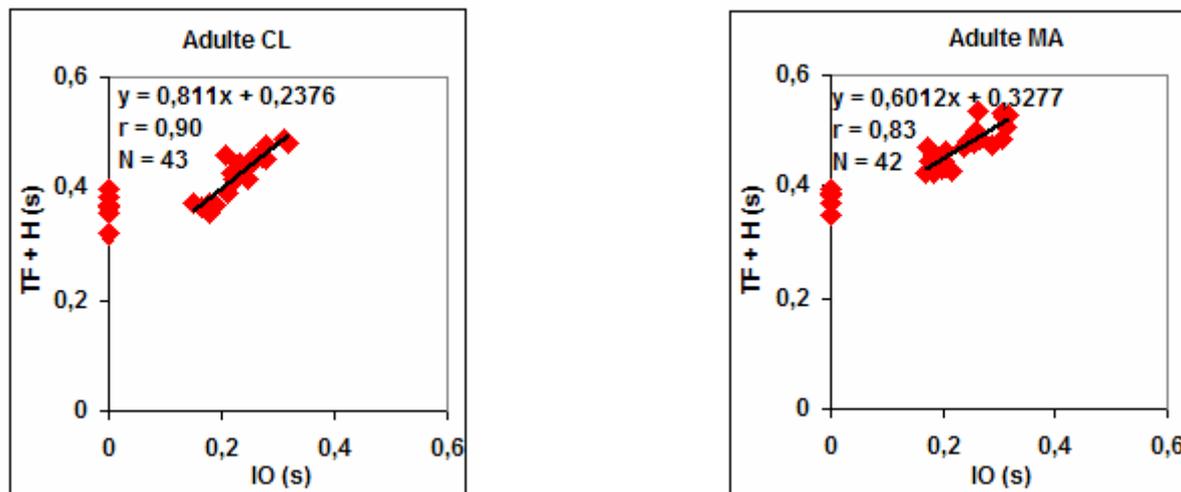


Figure 98 : Relation entre la durée totale du mouvement de constriction (TF+H) et IO chez les 2 adultes français de référence. Coefficients de corrélation r significatifs à $p < 0,00001$. N correspond au nombre de séquences considérées pour l'analyse par régression.

Ces deux sujets ayant bien prononcé les séquences minimales [iy] avec une valeur nulle pour IO (soit sans pause silencieuse), nous avons pu calculer leurs *constantes d'exécution* (Tableau 17). Nous constatons des valeurs relativement élevées pour [iy] et retrouvons, au moins pour un locuteur, la tendance à des transitions de gestes plus longues avec une consonne.

Sujet	[iy]	[iky]
CL	360	370
MA	380	450

Tableau 17 : Durées moyennes du mouvement d'arrondissement (TF+H) calculées pour les transitions minimales [iy] et les transitions [iky] (en ms).

A partir de [iCy], nos deux locuteurs adultes montrent des comportements d'arrondissement idiosyncratiques avec un coefficient d'expansion du mouvement de 0,81 pour CL et de 0,60 pour MA. Ces coefficients sont très conformes à ceux précédemment trouvés par Abry & Lallouache (1995b) pour leurs quatre locuteurs français (Figure 99) et à ceux trouvés pour nos locuteurs adultes franco-québécois et anglo-américains.

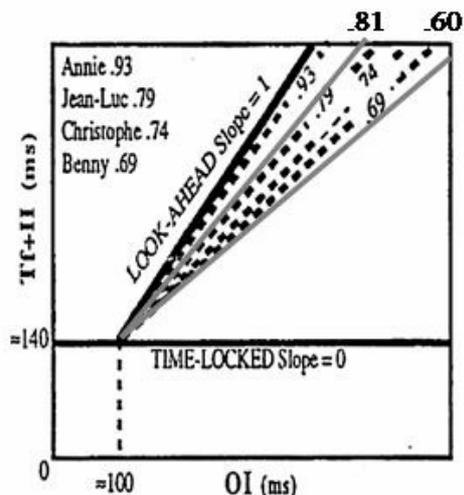


Figure 99 : Comparaison entre les pentes obtenues pour nos deux adultes français de références (lignes grises pleines) et celles obtenues pour les 4 locuteurs enregistrés par Abry & Lallouache (1995b).

IV.2.2. Résultats des enfants : 1^{ère} session d'enregistrement

IV.2.2.1. Comportement d'anticipation obtenu chez 6 des 7 de nos petits locuteurs

Nous présentons, à la Figure 100, les résultats obtenus pour 6 de nos 7 enfants enregistrés, qui vont présenter un comportement d'anticipation d'arrondissement déjà bien régulier. L'âge des sujets est indiqué à côté de leurs initiales. Dans l'ensemble, nous avons obtenu un continuum temporel depuis les séquences composées d'une consonne intervocalique jusqu'à celles en comportant 3. L'enfant DV ayant réalisé ses transitions [iy] avec une occlusive glottale, soit [iʔy], nous n'avons pas calculé sa *constante d'exécution* et avons réalisé la régression linéaire sur toutes les séquences [iC_ny] en incluant ces séquences [iʔy].

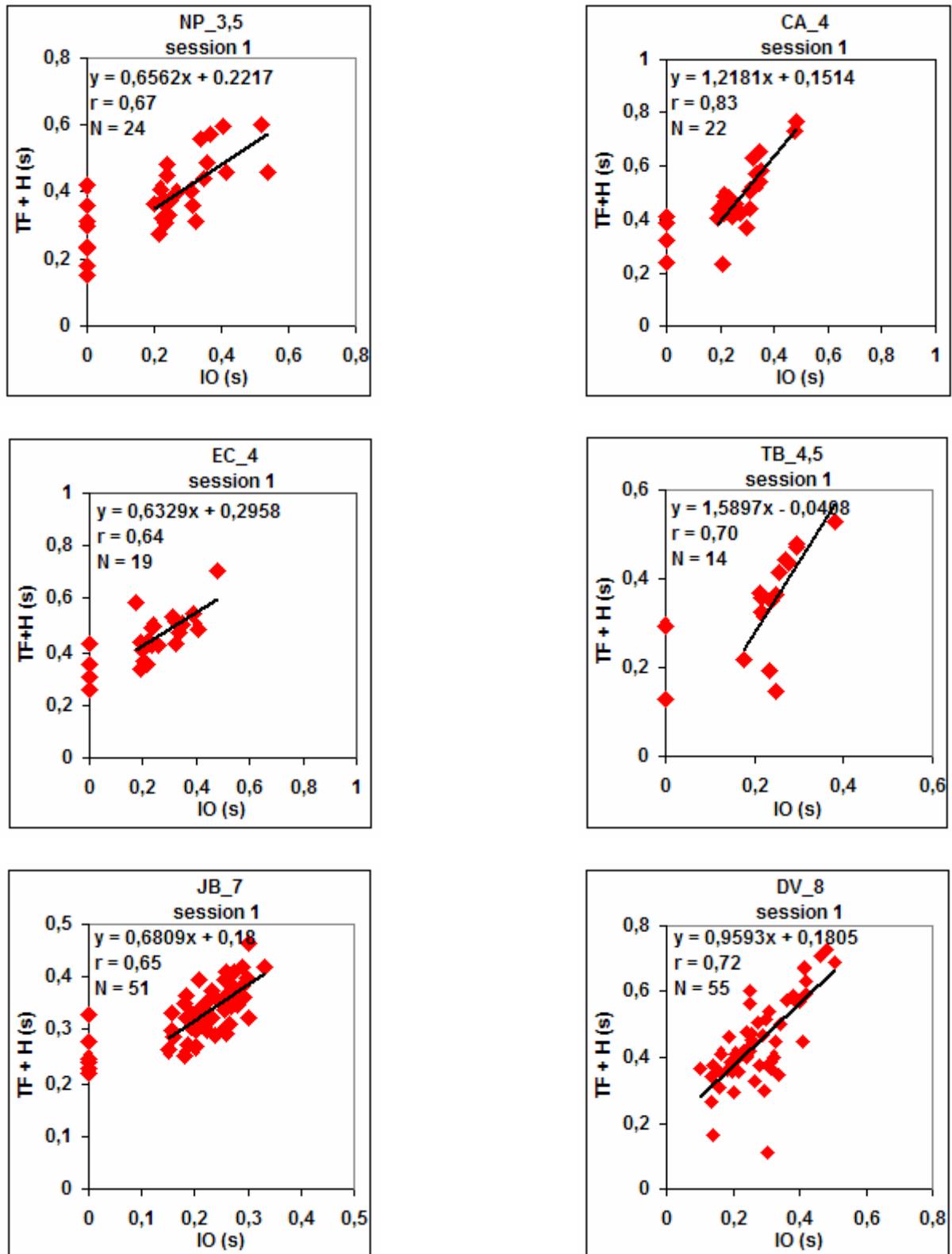


Figure 100 : Relation entre la durée totale du mouvement de constriction (TF+H) et IO (en sec.) chez 6 enfants. L'âge des sujets est indiqué à la suite de leurs initiales. Coefficients de corrélation r significatifs à $p < 0,01$ pour NP et EC, TB ; à $p < 0,00001$ pour CA, JB et DV. N correspond au nombre de séquences considérées pour l'analyse par régression.

Le Tableau 18 indique les durées moyennes obtenues par les 5 enfants ayant produit les transitions minimales sans interruption [iy], ainsi que les durées de [iky]. Les valeurs des deux adultes sont également rappelées.

Comparaisons des durées moyennes de mouvement TF + H pour les transitions minimales [iy] et les séquences [iky] (ms)		
- Session 1 -		
Enfants	[iy]	[iky]
CA_4	350	450
CE_4	340	400
NP_3,5	270	380
JB_7	270	280
TB_4,5	210	220
Adultes		
CL	360	370
MA	380	450

Tableau 18 : Durées moyennes du mouvement d'arrondissement (TF+H) calculées pour les transitions minimales [iy] et comparées aux séquences [iky] (ms).

Les *constantes d'exécution* des enfants varient de 210 à 350 ms et se situent donc toujours en-deçà de celles des adultes. Les durées pour les séquences à une consonne sont toujours supérieures pour un même sujet, et comprises entre 220 et 450 ms, englobant celles des adultes.

A partir de [iCy], le mouvement d'arrondissement peut s'étendre au delà de la *constante d'exécution* calculée quand IO est nul. Les graphes de la Figure 100 mettent en évidence une extension variable pour nos 6 locuteurs, puisqu'il ressort, de l'analyse par régression linéaire, des pentes comprises entre 0,63 et 1,59 ($p < 0,01$). Mais cette variabilité est toujours bien *lawful*, reflétant la loi propre au locuteur.

De manière générale, on remarque que l'organisation temporelle de ces énoncés diffère entre adultes et enfants. Alors que IO ne dépasse pas 370 ms chez les adultes, celui des enfants peut aller jusqu'à près de 600 ms (enfant NP). Pour les deux adultes MA et CL, les durées de mouvements varient dans une fenêtre temporelle de 350 à 550 ms, tandis que la gamme des enfants est plus grande, pouvant atteindre presque 800 ms (enfants CA et DV) lorsque IO est grand. Abelin *et al.* (1980) avaient également noté que la durée de l'intervalle

consonantique était globalement plus longue chez les enfants par rapport à l'adulte de référence (600-700 ms), à l'inverse de la durée du mouvement d'arrondissement. Les auteurs avancent que la durée de l'intervalle consonantique diminuerait progressivement avec l'âge tandis que le mouvement d'arrondissement se déploierait pour atteindre les valeurs de l'adulte. Si nous retrouvons une même variabilité sur les durées de IO, nous ne pouvons pas dire que les durées de mouvement (TF+H) observées entre adultes et enfants permettent d'affirmer une diminution avec l'âge.

IV.2.2.2. Un jeune locuteur qui n'a pas (encore) un comportement d'anticipation lawful

Nous présentons, à la Figure 101, les données obtenues pour notre 7^{ème} petit locuteur KV, âgé de 4 ans et demi. Notons que, comme l'enfant DV, cet enfant a réalisé ses transitions [iy] avec une occlusive glottale (soit [iʔy]). Nous ne pouvons donc pas donner sa *constante d'exécution* et avons réalisé la régression linéaire sur toutes les séquences [iC_ny] en incluant les séquences [iʔy]. L'enfant KV, à 4 et demi, n'a visiblement pas encore acquis le contrôle de l'arrondissement vocalique puisque, contrairement aux autres locuteurs, il présente des données très dispersées. Son comportement d'arrondissement ne suit donc pas une expansion linéaire du mouvement.

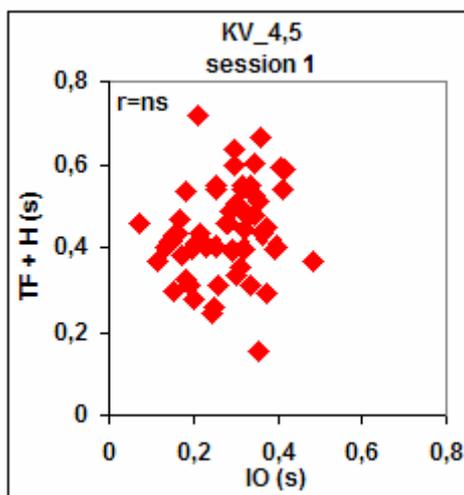


Figure 101 : Relation entre la durée totale du mouvement de constriction (TF H) et IO chez l'enfant KV. Le coefficient de corrélation r est non significatif. N correspond au nombre de séquences considérées pour l'analyse par régression linéaire.

Il semble ainsi qu'à 4 ans et demi, certains enfants puissent, comme c'est le cas de KV, ne pas avoir acquis la maîtrise parfaite du geste vocalique. La littérature dans ce domaine montre que l'acquisition des voyelles est non seulement variable d'un enfant à l'autre, mais qu'elle

diffère grandement en fonction de si l'on s'intéresse à l'émulation des buts acoustiques ou à l'imitation de la coordination motrice et à son déroulement conforme à la langue. Dans le cas de notre locuteur KV, le processus d'*attunement* phonologique à la langue n'est pas achevé à 4 ans et demi, dans la mesure où son comportement d'anticipation d'arrondissement vocalique ne présente encore aucune stabilité. Nous verrons qu'il y parviendra.

IV.2.3. Résultats des enfants : 2^{ème} session d'enregistrement un an plus tard

IV.2.3.1. Mise en évidence de profils d'anticipation vocalique stables

Nous présentons les résultats des enfants NP, CA et TB qui ont, tous les trois, pu être enregistrés une nouvelle fois, un an après la première session d'enregistrement (Figure 102). Pour le locuteur CA, l'analyse par régression linéaire a été réalisée sur la totalité des séquences en incluant la transition [iy] car celle-ci a été produite systématiquement avec une occlusion glottale entre les deux cibles vocaliques. Pour nos jeunes locuteurs TB et NP, il est possible de calculer une *constante d'exécution* : NP a une constante de 270 ms identique à celle obtenue lors de la session 1 ; quant à celle de TB, de 200 ms, elle est toute proche de celle de la session 1 (210 ms).

L'enfant NP a un coefficient d'expansion de 0,82 (0,65 à la première session), l'enfant CA un coefficient de 1,14 (1,21 en première session) et l'enfant TB un coefficient de 1,09 (1,59). Il n'y a pas de différence significative entre les deux sessions (NP : $t=-0,86$ ddl=63 ; CA : $t=0,31$ ddl=66 ; TB : $t=1,06$ ddl=67).

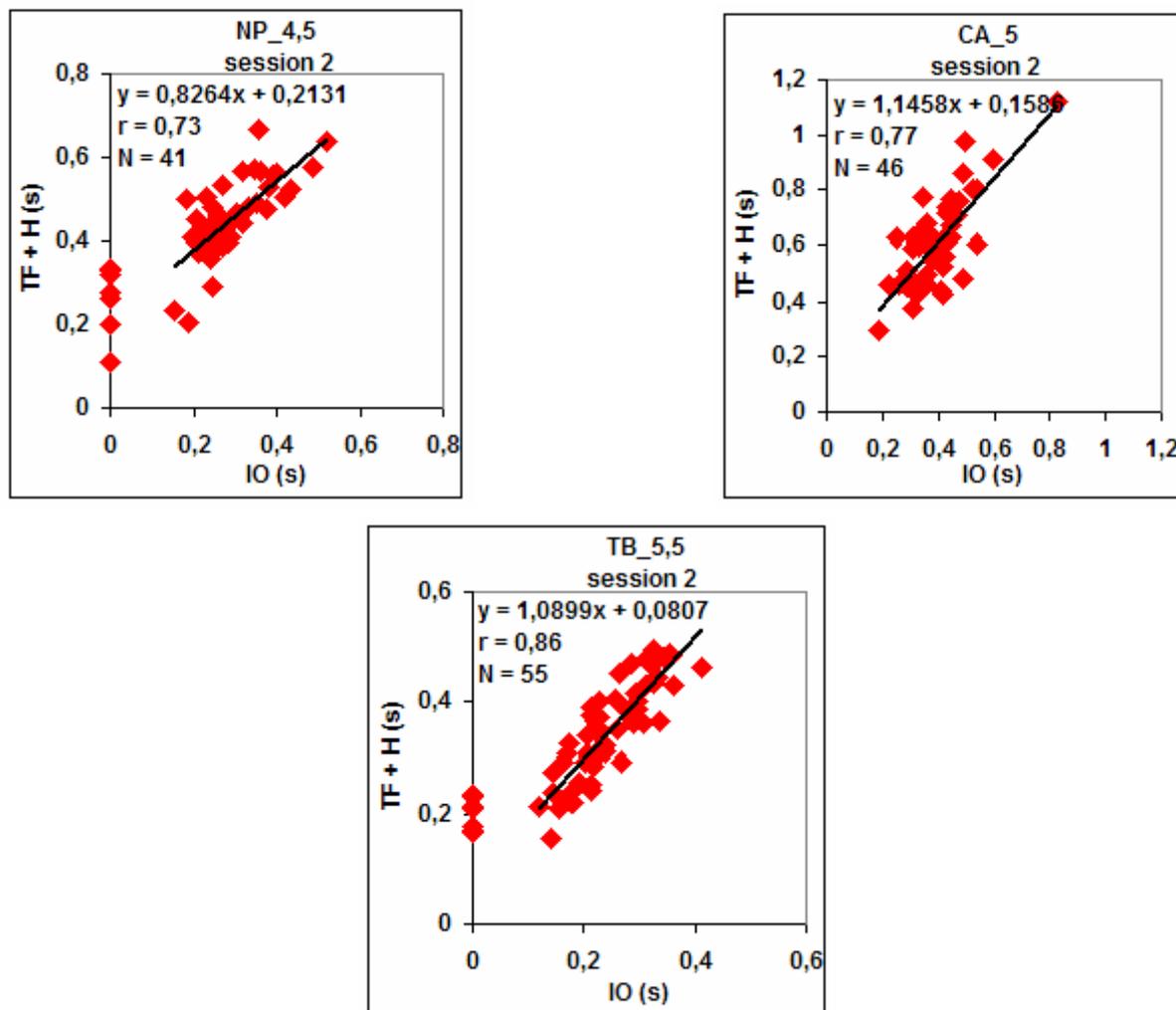


Figure 102 : Relation entre la durée totale du mouvement de constriction (TF+H) et IO chez 3 enfants lors de la deuxième session d'enregistrement. Coefficients de corrélation r significatif à $p < 0,0001$. N correspond au nombre de séquences considérées pour l'analyse par régression linéaire pour chaque session.

Nous proposons deux exemples de réalisation du locuteur TB, prononçant les prénoms 'Issu' et 'Iksu'. Ces exemples illustrent bien le comportement de ce jeune locuteur. Dans « issu » (extrait de la séquence « Issu est rouge »), le geste d'arrondissement débute après la fin acoustique de la voyelle [i] et la phase de *Time Falling* couvre une grande partie de l'*Intervalle d'Obstruence*. Le minimum d'aire aux lèvres est atteint au milieu du [y]. Lorsque IO est grand comme dans « Iksu », le mouvement d'arrondissement s'amorce dans la voyelle [i] et le minimum d'aire aux lèvres (repéré par le premier passage par zéro de la vitesse) est atteint assez tôt dans IO (au premier tiers). En comparant les deux séquences, on peut se rendre compte que, alors que la durée de l'*Intervalle d'obstruence* a plus que triplé, la phase de *Time Falling* est restée quasi constante. C'est donc la phase de *Hold* qui s'est expansée

(elle a quadruplé), donnant un plateau de constriction qui couvre largement l'intervalle consonantique et toute la durée de la cible acoustique du [y] (notons que dans ce plateau, où la vitesse tourne autour de zéro, il est sans intérêt de repérer le *minimum minimorum* comme date importante pour le timing : le choix du premier passage par zéro pour déterminer l'amplitude n'a aucune incidence notable sur la détermination des événements cadrant les phases TF et H, cf. notre protocole de mesure).

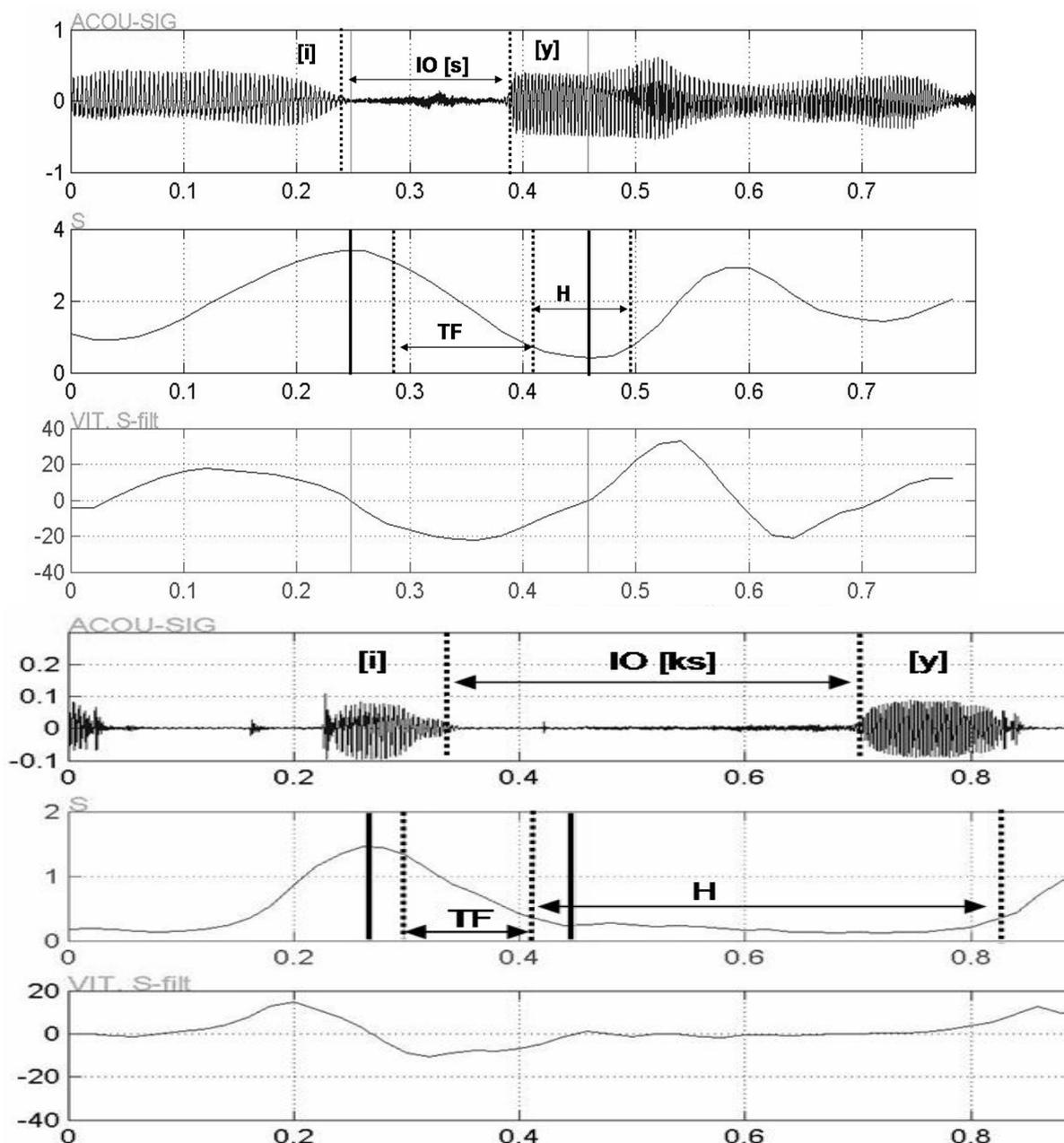


Figure 103 : Signal acoustique et évolution temporelle de l'aire interlabiale (en cm²) et de sa dérivée vitesse, lors de la réalisation des préfixes « Issu » (graphique du haut) et « Iksu » (graphique du bas), par l'enfant TB. Sur le signal acoustique, la fin de la voyelle [i] et le début de [y] sont indiqués par des lignes pointillées. Sur le signal de constriction, le maximum et le minimum de constriction sont indiqués par des lignes verticales continues et les phases *Time Falling* (TF) et *Hold* (H) sont situées entre les lignes pointillées.

Le locuteur NP n'a pas la même stratégie d'expansion de son mouvement. C'est ce que l'on peut voir à partir de ses réalisations « Issu » et « Itkru », dans la Figure 104, puisqu'avec l'augmentation de l'*Intervalle d'obstruence*, les deux phases du geste augmentent largement toutes les deux.

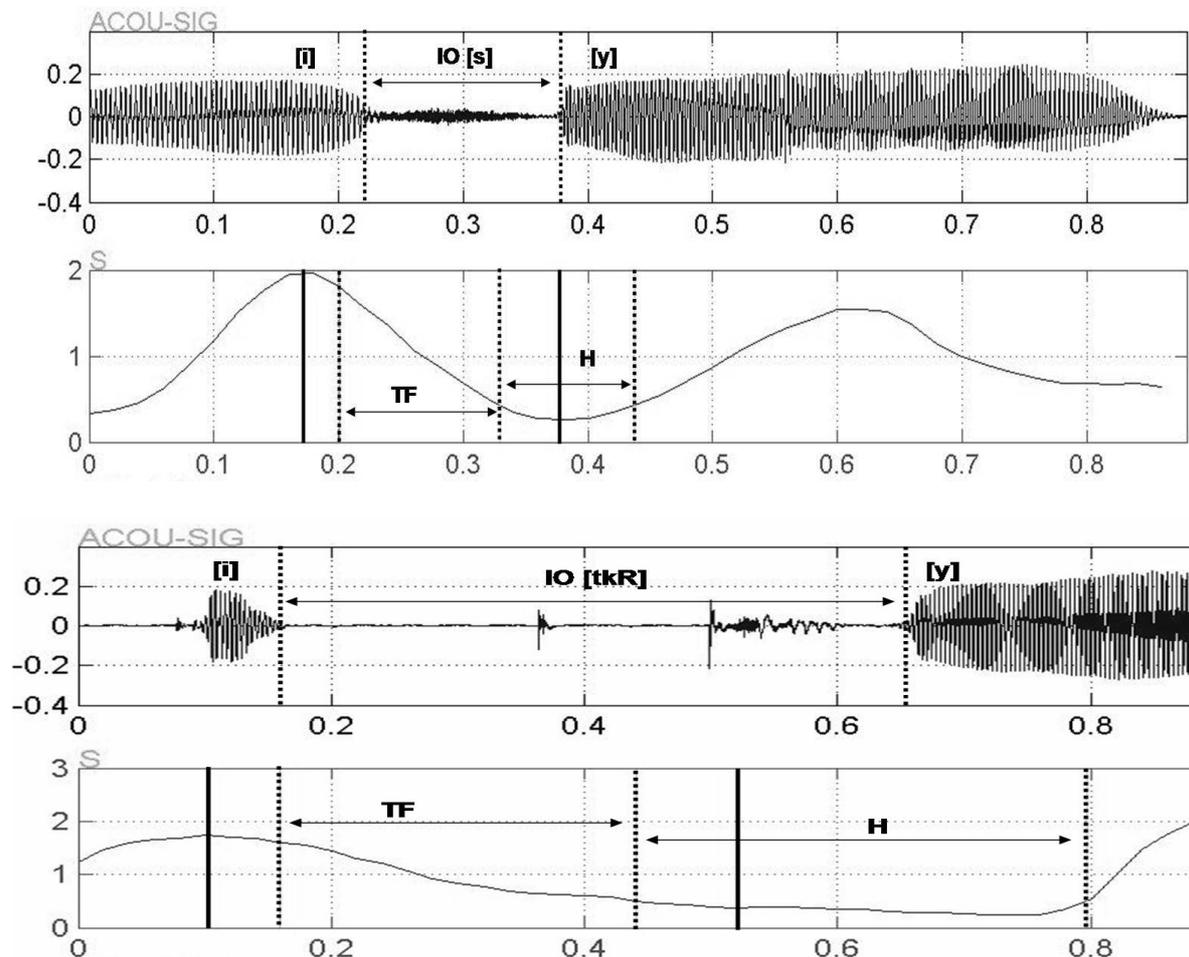


Figure 104 : Signal acoustique et évolution temporelle de l'aire interlabiale (en cm^2), lors de la réalisation des prénoms « Issu » (graphique du haut) et « Itkru » (graphique du bas), par l'enfant NP. Sur le signal acoustique, la fin de la voyelle [i] et le début de [y] sont indiqués par des lignes pointillées. Sur le signal de constriction, le maximum et le minimum de constriction sont indiqués par des lignes verticales continues et les phases *Time Falling* (TF) et *Hold* (H) sont situées entre les lignes pointillées.

IV.2.3.2. Notre 7ème petit locuteur sait maintenant anticiper comme les autres enfants

Intéressons-nous maintenant au cas de notre jeune locuteur KV qui, à 4 ans et demi, présentait un comportement d'arrondissement très variable par rapport à celui des autres locuteurs, y compris ceux de son âge (EC, CA, TB) ou même d'enfants plus jeunes que lui (NP). L'analyse par régression linéaire avait révélé des données trop dispersées pour être

interprétables par le modèle *MEM* (Figure 105). En examinant son patron d'anticipation un an plus tard c'est-à-dire à 5 ans et demi, nous obtenons des données d'arrondissement totalement différentes. Ce jeune locuteur semble avoir enfin mis en place une stratégie d'anticipation d'arrondissement se manifestant par une pente spécifique de 0,82 ($r = 0,85$, $p < 0,00001$).

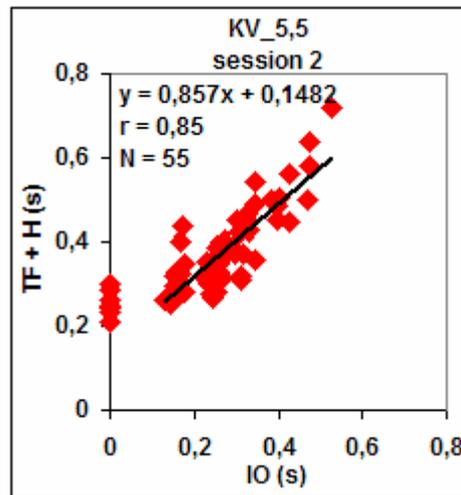


Figure 105 : Relation entre la durée totale du mouvement de constriction (TF+H) et IO pour le locuteur KV. Le coefficient de corrélation r est significatif à $p < 0,00001$.

D'après les résultats de la première session, le comportement d'anticipation vocalique de l'enfant KV, à 4 ans et demi, n'était pas stabilisé en termes de coordination temporelle de ces gestes articulatoires. On peut penser qu'à ce stade, la coordination de ses articulateurs supra-laryngés n'était pas encore finement contrôlée, si bien qu'il ne pouvait pas reproduire le pattern d'anticipation d'arrondissement typique de sa langue. Il était certes capable d'atteindre un but acoustique, autrement dit de reproduire la cible vocalique arrondie, mais pas encore capable de régler le déroulement temporel du geste d'arrondissement approprié à sa langue. A 5;5 ans, il semble avoir acquis ce déroulement *lawful*.

Il nous semble que la période entre 4 et 5 ans correspond pour ce locuteur à une phase essentielle d'*attunement* phonologique, durant laquelle l'enfant gagne en contrôle de ses articulateurs, affine leur coordination pour la coproduction des gestes de parole et se conforme progressivement aux normes d'organisation temporelle de la langue.

IV.2.3.3. Contributions respectives des phases de *Time Falling* et *Hold*

Nous avons, comme pour les adultes, testé la contribution des phases de *Time Falling* et de *Hold* au geste global. A partir de la mise en relation séparée de chacune de ces phases avec l'*Intervalle d'Obstruence*, nous avons calculé la droite de régression pour chaque ensemble de données. Nous présenterons les valeurs des coefficients d'expansion obtenues pour chaque

locuteur dans le Tableau 19 (les graphes correspondant sont tous donnés dans les Annexes V et VI).

Comparaison des pentes obtenues pour les phases <i>Time Falling + Hold, Time Falling, Hold</i> - Enfants français -				
Locuteurs	Age lors de chaque session	<i>Time Falling + Hold</i>	<i>Time Falling</i>	<i>Hold</i>
CA	4 ans	1,22	1,07	0,15 (ns)
	5 ans	1,15	0,21 (ns)	0,94
KV	4,5 ans	0,27 (ns)	0,24	0,03 (ns)
	5,5 ans	0,86	0,49	0,37
NP	3,5 ans	0,66	0,29 (ns)	0,36 (ns)
	4,5 ans	0,83	0,48	0,35 (ns)
TB	4,5 ans	1,59	0,22 (ns)	1,37
	5,5 ans	1,09	0,16 (ns)	0,93
EC	4 ans	0,63	0,17 (ns)	0,46 (ns)
DV	8 ans	0,96	0,90	0,05 (ns)
JB	7 ans	0,68	0,42	0,26 (ns)
Adultes				
CL		0,81	0,42 (ns)	0,40 (ns)
MA		0,60	0,25	0,35

Tableau 19 : Comparaison des pentes obtenues pour le geste global de constriction (TF+H) et séparément pour les phases de *Time Falling* (TF) et *Hold* (H) par rapport à IO, pour les 7 enfants et les 2 adultes. Les pentes sont validées par des coefficients de corrélations à $p < 0,01$, à l'exception de celles accompagnées du symbole ns.

Nous n'observons, dans ce tableau, aucune régularité concernant la contribution de chaque phase à l'extension du geste d'arrondissement. Certains enfants varient davantage leur phase de *Hold* (comme TB ; CA en session 2), d'autres davantage la phase de *Time Falling* (CA en session 1 ; NP en session 2 ; DV et JB). D'autres encore modulent les deux phases en même temps (KV en session 2 ; l'adulte MA). On peut même trouver trois cas, celui de NP en session 1, EC et l'adulte CL, où aucune des deux phases ne peut être mise en corrélation avec l'*Intervalle d'Obstruence*, alors que la corrélation existe bien, si le geste global est pris en compte.

Nous pouvons ainsi renforcer la conclusion déjà émise pour les adultes (section II.2.2.3.) à partir d'une analyse des mêmes phases, que c'est bien le geste global d'arrondissement, dans sa composante d'établissement associé à sa composante de tenue, qu'il faut modéliser si l'on veut mettre à jour la loi de son expansion spécifique à chaque locuteur.

IV.3. Discussion générale

Au travers de cette étude expérimentale sur l'anticipation vocalique, nous nous sommes intéressés à l'*attunement* phonologique de jeunes enfants français. D'après l'ensemble des données enfantines recueillies (Tableau 19), nous pouvons avancer qu'il existe bien une organisation *lawful* de l'anticipation vocalique, qui se traduit par une coordination à la fois spatiale et temporelle des mouvements articulatoires supra-laryngés. Elle peut l'être dès 3 ans et demi (c'est le cas de NP). Nous avons vérifié que, chez les plus jeunes sujets qui avaient déjà un comportement régulier d'anticipation, celui-ci se retrouvait un an plus tard sans différence significative (Figure 106). Pour les sujets qui avaient été enregistrés plus tard que 4;5 ans – ceux de 7 et 8 ans –, nous n'avons pas eu besoin de vérifier que la régularité de l'anticipation était évidemment bien acquise. Comme nous l'avons observé chez un de nos plus jeunes sujets qui n'avait pas encore maîtrisé ce comportement à 4;5 ans, cet enfant KV aura, un an plus tard, rejoint le comportement régulier de ses pairs de 4;5 à 5;5 ans.

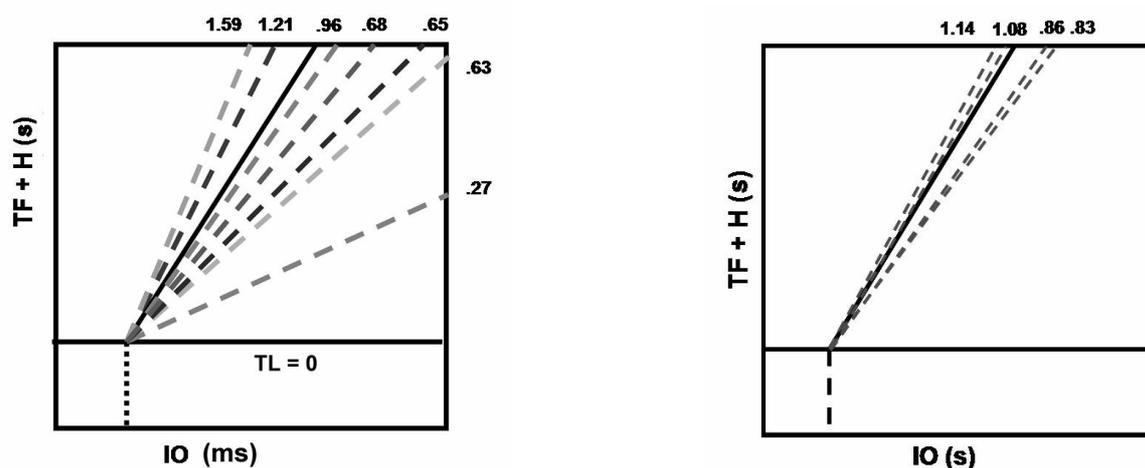


Figure 106 : Récapitulatif des pentes obtenues pour les enfants lors de la première session (figure de gauche) et les 4 enfants lors de la deuxième session (figure de droite). NB. La droite de l'enfant KV n'est bien évidemment donnée que pour la deuxième session.

Les données longitudinales nous apprennent donc que lorsque le profil d'anticipation est acquis (*tuned*), celui-ci l'est pour de bon. Cela ne veut pas dire que les enfants ne vont pas continuer à évoluer dans leur pattern d'anticipation d'arrondissement vocalique, sans qu'il

existe pour l'instant de données en ce sens, testées dans le même paradigme du MEM d'anticipation de constriction. Nous ne voudrions pas donner l'impression, par cette formulation prudente, de rejoindre les généralités qui sont encore souvent données. Ainsi, rejoignant Kent (1976), que nous avons déjà abondamment cité, James *et al.* (2001) dans une étude de productions de mots mono-, bi- et poly-syllabiques pour 354 enfants de 3 à 8 ans, renforcent d'entrée plusieurs points de Kent (1976), en divisant de manière classique les aspects paradigmatique (traduisons : contrôle des cibles vocaliques) et syntagmatique (traduisons : coarticulation dont *sequencing* and *phasing* ; pour ces derniers concepts repris de Gallistel, 1980, *cf.* Kent, 1983). A commencer par le « consensus »:

« The consensus that vowel development is complete by the age of three years is one reason why vowels are not a focus of the discipline (Bauman-Waengler, 2000; Vihman, 1998). » (James et al., 2001, p. 127)

Précisant plus loin qu'il s'agit, selon la distinction paradigmatique/syntagmatique, du contrôle de réalisations statiques :

« By the age of three years, children appear to have acquired the first aspect, the paradigmatic aspect of vowel production, meaning the static, articulatory gestures for individual vowels are mastered. » (James et al., 2001, p. 128)

Le Tableau 20, que nous insérons ici, prouverait s'il en était besoin que tous les enfants de notre étude se sont révélés tous capables d'obtenir de petites aires aux lèvres contrôlées – c'est-à-dire un *goal* sans closion –, comme en témoignent les valeurs minimales observées.

Constriction minimale pour la cible [y] (cm ²)																	
Enfants français												Adultes français					
CA		KV		NP		TB		EC		DV		JB		CL		MA	
Sess 1	Sess 2	Sess 1	Sess 2	Sess 1	Sess 2	Sess 1	Sess 2	Sess 1	Sess 2	Sess 1	Sess 1	Sess 1	Sess 2	Sess 1	Sess 2	Sess 1	Sess 2
0,14	0,09	0,07	0,02	0,12	0,07	0,09	0,05	0,18		0,14		0,11		0,09		0,08	

Tableau 20 : Valeurs d'aire aux lèvres minimale (en cm²) pour les 7 enfants dans les sessions où ils ont été enregistrés.

Toujours selon James *et al.* (2001), le versant syntagmatique se développe à partir de ces 3 ans jusqu'à l'adolescence (la puberté ou les 12 ans de Kent):

« From the age of three years to the adolescent years, the second aspect, the syntagmatic aspect, of vowel production is mastered (Kent, 1976; Kent & Forner, 1980). During this time the vowels undergo refinement as children better handle the sequencing demands of speech and incorporate appropriate prosodic patterns, culminating in adult rate and rhythm of speech. This fluency of speech reflects the acquisition of the dynamic, co-articulatory movements of speech as well as the

mastery of the stress markers. In this period, some manifestations of the immature vowel system are perceptually apparent and others are so subtle that they are only apparent when using acoustic analyses. » (James et al., 2001, p. 128)

Perception, acoustique... Rien n'est dit des analyses articulatoires encore plus subtiles mais, o combien difficiles. La méthode retenue, qui peut être utilisée sur une large population, s'est en fait limitée à l'analyse auditive de transcriptions d'erreurs sur la production de voyelles en syllabe faible (*weak* dans la terminologie de la *hiérarchie prosodique*). Ainsi les résultats de James et al. (2001) montrent que les enfants de 3-4 ans ont des erreurs importantes sur la production des voyelles dans des mots de plus de 2 syllabes, alors que ceux de 5-6-7 ans semblent les avoir déjà maîtrisées.

Notre approche teste un patron complet *path-goal*, cible articulatoire-acoustique et chemin d'anticipation. Nous n'avons de fait aucune raison de penser, à partir de nos données, que des enfants entre 3;5 à 5;5 ans ne peuvent pas avoir acquis tout bonnement la maîtrise de ce patron complet. Ce résultat est donc bien moins vaguement général que la conclusion de James et al. (2001), qui vient s'inscrire dans la ligne de Kent (1976):

« The significant age, syllable number and interaction effects indicate that children are still mastering aspects of vowel production up to the age of six years. » (James et al., 2001, p. 132)

Ajoutant en résumé:

« More specifically, the results imply that mastery of the paradigmatic aspects of vowels is achieved by three years of age and it is the syntagmatic aspects of vowels that continue to develop after the age of three years. » (James et al., 2001, p. 132)

C'est une conclusion certes peu spécifique par rapport au résultat de Green et al. (2002), qui ont montré que l'indépendance du contrôle des lèvres dans la *closure* labiale est devenue comparable à celle de l'adulte à 6 ans. Et à l'évidence tout à fait moins spécifique que la nôtre pour le contrôle *path-goal* des deux lèvres dans l'*arrondissement* des voyelles.

On peut dire que, même par rapport aux études sur un grand nombre d'enfants d'âges échelonnés, comparables à leur étude par transcription auditive (de deux « *qualified speech pathologists* »), celle-ci est en régression par rapport au fameux *survey* de Kent (1976) qui s'attaquait à l'acoustique, avec durées, formants et fondamental.

James et al. (2001, p. 132) ne lui ont donc fait qu'un hommage par précaution, un simple salut de la main, sans la plus petite acoustique au bout du compte:

« *In this period [from 3 to 6 and beyond], some manifestations of the immature vowel system are perceptually apparent and others are so subtle that they are only apparent when using acoustic analyses.* » (James et al., 2001, p. 128)

Nos contraintes expérimentales nous classent d'entrée dans le genre d'un modèle quantitatif, qui peut être très limité en nombre de locuteurs, comme celui testé sur 4 locuteurs de Perkell & Matthies (1992; et encore avaient-ils été sélectionnés en éliminant les candidats qui avaient trop peu de protrusion trackable), un nombre qui, dans ce domaine, n'a pas nuit à la publication de ce travail expérimental dans une revue prestigieuse pour la parole, comme le JASA. Nos contraintes sur un suivi goal+path pourraient nous laisser croire que les généralisations sont définitivement hors de portée de notre approche. A moins que ne se développent dans le futur des méthodes fiables de suivi embarqué de l'évolution des formes labiales (pas seulement des points); et nous avons été bien placés dans un laboratoire de pointe sur ce sujet, l'ICP, pour savoir que ce n'est pas pour demain.

En réalité le paradigme du modèle *MEM* nous offre une réfutation possible au fur et à mesure de la saisie de nouveaux sujets : c'est le test du *principe* de variabilité individuelle régulière (*individual lawful variability*) et la mesure du *paramètre* propre au locuteur, son coefficient d'expansion de l'anticipation (*speaker-specific expansion rate*). Il n'est donc pas besoin d'attendre le grand nombre de locuteurs pour généraliser. Dans une épistémologie classiquement poppérienne : un nouveau locuteur ne peut que nous mettre devant l'hypothèse nulle à son sujet. Dans le cas des enfants, un nouveau test, qui sera alors réalisé un an ou moins plus tard, nous dira si le sujet s'est régularisé (*attuné*), et selon quel pas d'expansion. Mieux : le changement de langue n'ayant pas démenti les prédictions de notre modèle, en passant du français à l'anglais (en dépit de plus de 40 ans de résultats contradictoires sur les deux langues), et toujours en ce qui concerne spécifiquement les enfants, ceux-ci n'ont d'autre choix que d'avoir acquis le modèle *MEM* d'anticipation de constriction, à un moment ou à un autre de leur trajet linguistique. Cette date devient une simple question empirique, sans faire appel à un *principe* d'une fenêtre ou période critique concernant l'âge de cette acquisition, cet âge n'étant ici aussi qu'un *paramètre* individuel.

En conclusion, nous avons réussi à repousser les limites expérimentales de ce type d'étude articulatoire jusqu'à l'âge de 3 ans et demi, en transposant un paradigme expérimental élaboré initialement pour l'analyse du décours temporel de l'arrondissement vocalique chez l'adulte à l'enfant. Ce paradigme consiste en la production de séquences comportant des groupes de consonnes variés (jusqu'à 3 consonnes) parfois complexes à prononcer pour de jeunes enfants

(cas de 'Iskru', ou 'Itkru'). Il pourrait nous être reproché d'avoir utilisé des énoncés dépourvus de sens pour la constitution de notre corpus. Certains auteurs (Abelin *et al.*, 1980; Lubker & Gay, 1982) ont en effet avancé que les profils d'anticipation vocalique étaient plus variables lorsque le corpus était composé de mots sans sens par rapport à des mots avec sens. Pour notre part, nous avons délibérément créé nos énoncés afin d'avoir la possibilité de manipuler la longueur de l'*Intervalle d'Obstruence* et la nature des consonnes formant cet intervalle. Nous avons donc pris soin d'attribuer un sens à nos énoncés en les associant à des noms de peluches afin que nos jeunes locuteurs ne soient pas perturbés par la tâche. Ce qui s'est révélé plutôt ludique. La difficulté ne tient en fait pas à la sémantique, mais à la maîtrise phonétique des groupes de consonnes (pour la phonologie de ces groupes pour l'enfant, *cf.* entre autres Lleo & Prinz, 1997 et Kirk & Demuth, 2005). En effet nous avons tenté d'enregistrer des enfants en-dessous de cet âge de 3 ans et demi. Mais ces derniers ne parvenaient tout simplement pas à réaliser la tâche de production demandée. On peut s'en consoler en faisant appel au fameux consensus de 3 ans, rappelé plus haut, pour la maîtrise des cibles (*goal*) des voyelles, qui semble bien être un pré-requis pour acquérir le chemin (*path*) d'anticipation. L'avenir nous dira si d'autres chercheurs munis des mêmes moyens, ou de suivis plus perfectionnés (*light* ou *smart*) de l'aire aux lèvres, descendront en-dessous de qui a été pour nous une limite.

Conclusions et perspectives

Le travail de recherche mené durant cette thèse, dans le contexte d'une rivalité scientifique entre différents modèles d'anticipation pour la parole, était porté par un double enjeu. Le premier enjeu était de valider pour l'anglais – pour lequel un droit à la différence de comportement par rapport au français avait été revendiqué par les Laboratoires Haskins depuis plus de 10 ans – notre *Modèle d'Expansion du Mouvement*. Ce modèle qui repose sur le *principe* d'un comportement d'anticipation régulier (*lawful*) – croissant avec la distance temporelle entre les voyelles imposée par la durée des consonnes intervocaliques – offre en outre la possibilité d'un *paramètre* spécifique du locuteur, avec son propre coefficient d'expansion de l'anticipation. Déjà réputé robuste pour le français de France, pourvu qu'on utilise la *constriction* et pas la *protrusion*, nous l'avons à nouveau éprouvé pour le français du Québec, en neutralisant de possibles effets de la mandibule, une suspicion qui interférait jusqu'alors de manière indécidable dans l'interprétation des résultats des tests successifs. Pour cela nous avons eu recours à une condition *bite-block*. Première conclusion : les francophones, mais aussi les anglophones, enregistrés à Montréal, se sont rangés sous notre *MEM* de constriction. Le second enjeu principal était celui de tester l'acquisition d'un tel modèle chez l'enfant, avec pour objectif de rendre compte de l'*attunement* de ce comportement moteur ambiant dans l'environnement linguistique, dans cette thèse celui du français. Nous pouvons maintenant conclure que la période entre 3 et 6 ans ne contient pas une fenêtre temporelle étroite, critique, mais que l'âge d'acquisition de ce contrôle est lui aussi paramétrable par sujet, après 3 ans comme avant 6 ans selon l'enfant, tout comme est individualisé le coefficient d'expansion du modèle *MEM*.

L'extension du MEM du français à l'anglais

Nous avons donc choisi pour le recueil de nos propres données labiales d'emprunter le paradigme expérimental élaboré par Abry & Lallouache (1991) et de lui faire traverser l'Atlantique pour le tester sur des lèvres de la belle Province, francophones et anglophones. Pour départager les modèles, nous avons recueilli simultanément deux paramètres labiaux en compétition, largement connus pour leurs implications acoustiques dans l'arrondissement vocalique : la protrusion de la lèvre supérieure et l'aire intérolabiale ou constriction. La détection de la protrusion, paramètre classique dans la littérature sur la coarticulation anticipante, se fait principalement par suivi de points cutanés (*flesh points*) là où sont posés des *sensors* suivis par système magnétométrique ou optoélectrique, en l'occurrence le système Optotrak. En ce qui concerne la constriction, la tâche est d'une tout autre complexité. A notre connaissance, aucune exploitation empirique du paramètre de constriction ne semble avoir été relatée dans la littérature, si ce n'est celle menée dans notre laboratoire. Rappelons qu'il s'agit d'une détection de contours, de formes, et non d'un suivi par points. Comme l'ont reconnu Perkell & Matthies dès 1992 (p. 2913), cela nécessite un dispositif de saisie, de prétraitement et de traitement d'images appropriés, que l'ICP est seul au monde à avoir conçu (Lallouache, 1991). L'alliance de l'Optotrak de McGill et du *Lip-Shape-Tracker* de l'ICP a été pour nous une première. Le laboratoire de Phonétique de l'UQAM, dirigé par Lucie Ménard, a permis de rendre opérationnelle cette rencontre.

L'analyse des données labiales nous a conduit à plusieurs conclusions.

1. Tout d'abord, la protrusion n'est clairement pas un paramètre descripteur de l'activité d'arrondissement vocalique suffisamment robuste, quel que soit le statut phonologique de l'arrondissement dans les deux langues. En effet, pour une langue comme le français où l'arrondissement est une dimension articulatoire réputée distinctive, permettant de différencier les voyelles étirées ([-rond]) des voyelles arrondies ([+rond]), nous avons remarqué que la protrusion n'était pas systématiquement le corrélat articulatoire le plus sûr pour rendre compte de l'arrondissement vocalique. Cela est vrai pour le français de France (Abry & Lallouache, 1991, 1995a) mais également pour le franco-québécois, puisque l'un de nos locuteurs a aussi montré une très faible dynamique de protrusion. En ce qui concerne nos locuteurs anglo-américains, ce constat est applicable à l'ensemble des sujets enregistrés, à l'exception d'un seul locuteur ayant montré une amplitude de protrusion suffisante.

2. L'étude de la constriction nous a fourni des résultats décisifs. Contrairement à la protrusion, nous avons en effet réussi à dégager des patrons d'anticipation réguliers pour nos 8 locuteurs, francophones comme anglophones. Le suivi de la constriction intérolabiale traduit donc plus fidèlement le contrôle de l'arrondissement vocalique. Il s'agit là sans doute d'un paramètre universellement exploité pour marquer l'arrondissement vocalique, réalisé visuellement et acoustiquement. On sait simplement que ce contraste [i]/[u] n'a pas autant de contraintes sur les lèvres dans plusieurs langues d'Extrême-Orient, dont le japonais. Pour les sujets anglophones en tous cas, la constriction a permis de faire émerger des profils d'anticipation d'arrondissement vocalique réguliers.

3. Pour ces 4 sujets anglophones, ainsi que pour nos 4 adultes franco-québécois, comme pour les 4 Français enregistrés auparavant dans notre laboratoire (Abry & Lallouache, 1995b), la constriction nous a permis en outre de faire émerger des patrons d'anticipation *spécifiques à chaque locuteur*. Ainsi il n'existe pas un type de profil d'anticipation généralisable à l'ensemble des sujets, tel que proposé par les modèles classiques *Look-ahead*, *Time-locked* et *Hybride*. Le mouvement d'anticipation n'est pas plus dépendant pour son initiation de la fin acoustique de la voyelle non arrondie qu'il n'est fixement lié au début de la voyelle arrondie. De fait il n'y a pas non plus un principe de propagation régressive des traits distinctifs dans tout l'intervalle consonantique, comme le proposaient Benguerel & Cowan (1974) pour le français, suivant en cela le modèle général *Look-ahead* de Henke (1966). De même, l'hypothèse d'un mouvement d'arrondissement invariant quelle que soit la durée de l'intervalle consonantique, se produisant systématiquement à date fixe par rapport au début de la voyelle arrondie (modèle *Time-locked* ou *Frame-model* des Laboratoires Haskins), n'est pas du tout représentatif des patrons observés dans notre étude, que ce soit pour la protrusion ou la constriction. Nous avons donc pu montrer que chaque locuteur organise ses mouvements d'anticipation de manière régulière, suffisamment non-bruitée pour présenter un coefficient de corrélation significatif entre le mouvement et l'intervalle consonantique. Et ceci selon un coefficient d'expansion qui lui est spécifique. Notons que l'hypothèse d'un comportement *Time-locked* invariant (même pour un style d'élocution donné dans le *Frame-model*) ne nous semble pas statistiquement fondé : accepter l'invariance en *timing absolu* reviendrait à accepter l'*Hypothèse Nulle*. Les données présentées comme des preuves d'un contrôle invariant de l'anticipation d'arrondissement (se manifestant statistiquement par une pente nulle) ne sont en effet jamais soutenues par des coefficients de corrélation significatifs. A l'inverse, les principes soutenus par le modèle *MEM* semblent tout à fait correspondre à la réalité des comportements observés, en français, et par extension en anglais et en franco-

québécois, puisque nous avons systématiquement obtenus des coefficients de corrélation significatifs.

4. Soulignons que la démonstration que nous avons faite (qui n'avait jamais été faite jusqu'à présent) que les patrons de protrusion ou de constriction n'étaient pas dépendants des perturbations (consonantiques) venues de la mandibule, a été concluante puisque le seul cas où les deux conditions, avec *bite-block* et sans *bite-block*, se sont trouvées significativement différentes, correspondait pour un sujet francophone à une réduction de sa dynamique de protrusion *et* de sa dispersion en condition avec *bite-block*.

5. Le succès de notre entreprise tient sans doute au choix d'événements robustes, contrairement aux événements dérivés, tels les pics de vitesse et d'accélération, car ceux-ci se révèlent (i) être facilement bruités ; (ii) et ils peuvent apparaître en trop grand nombre sur les signaux cinématiques (même après un filtrage raisonnable), ce qui pose alors la question du choix du repère à retenir. Nous n'avons ainsi retenu les passages par zéro de la vitesse que pour déterminer l'amplitude des mouvements, dans la gamme de leur minima et maxima, et pas pour leurs *dates*. Les événements qui servent à délimiter nos phases, comme le 10% et 90% d'avancée de protrusion ou de chute d'aire intérolabiale, nous permettent de définir un intervalle décrivant bien, d'une part, l'amorce du geste d'arrondissement (*Time Falling*), tandis que la phase de tenue ou *Hold* nous garantit, surtout dans le cas de la constriction, l'atteinte de la cible acoustique associée à un minimum d'aire aux lèvres pour le [u]. Ensemble, ces deux phases illustrent l'expansion du mouvement par rapport à l'*Intervalle d'Obstruence* sans tomber dans l'artefact statistique de la corrélation « *tout-partie* », puisqu'elles n'ont pas de bornes communes.

6. Par ailleurs, nous avons montré que ces comportements d'anticipation trouvent une cohérence lorsque les *deux* phases *Time Falling* et *Hold* sont considérées dans leur ensemble. Lorsque l'on essaie de décomposer l'anticipation, en ces deux sous-phases, on obtient souvent des données très dispersées, des corrélations non significatives ; et même lorsqu'elles le sont, le comportement n'est pas forcément cohérent, y compris chez un même sujet. La prise en compte de ces deux phases groupées (*TF+H*) nous livre au contraire la cohérence du geste global d'anticipation, pour tous les sujets.

On notera pour conclure sur les résultats notables de cette partie que : (i) tous les sujets ne peuvent être examinés avec un modèle de protrusion ; (ii) que ceux qui le sont présentent un comportement régulier, avec des corrélations significatives ; (iii) que lorsque ce comportement connaît une expansion d'anticipation, bien régulière, mais de taux faible,

(c'est le cas d'un locuteur québécois), ceci ne vaut que pour la protrusion. Car quelle que soit la langue, tous les sujets se retrouvent au bout du compte pratiquer un modèle *MEM* pour leur anticipation de constriction, avec un taux élevé d'expansion.

Ceci est donc une véritable réussite dans la généralisation d'un comportement moteur, qui avait jusqu'à présent échoué spécifiquement dans son extension au comportement humain par excellence qu'est le langage. D'autant plus que les études sur la protrusion avaient permis aux Laboratoires Haskins de soutenir, après une période de lutte des modèles (MIT contre Haskins), un compromis empirique: soit un modèle pour l'anglais, et un pour le français. Ce qui est toujours le plus facile à accepter, étant donnée la diversité, pour l'instant toujours non réduite substantiellement à l'universel, des comportements linguistiques (par rapport à la locomotion, par exemple). La proposition que Byrd et Saltzman avaient finalement énoncée, et qui dépassait le contrôle de l'anticipation d'arrondissement, sera à partir de cette thèse devenue caduque :

« [...] *the anticipation intervals of gestures may be defined as the time between gestural onset and target attainment. These intervals have been shown to display different degrees of temporal elasticity across languages (e.g. Abry & Lallouache, 1995; Saltzman et al., 2000). [...] In English, for example, anticipation intervals are relatively fixed regardless of the length of the preceding no-conflict intervals; in French, however, the anticipation periods stretch in proportion to the length of the no-conflict intervals.* » (Byrd & Saltzman, 2003, p. 157)

De cette proposition, nous approuvons bien volontiers que le mouvement d'arrondissement montre des degrés d'élasticité différents : la preuve est donnée par les différentes pentes que nous avons obtenues pour chacun de nos 8 locuteurs. Mais au vu des résultats à la fois pour les francophones et anglophones, nous pensons qu'il est inexact de faire référence à un ajustement à la langue, puisque l'anglais se comporte visiblement comme le français ou le franco-québécois.

Il faut constater un ajustement par individu, paramètre du *MEM*. Le déroulement temporel de l'anticipation d'arrondissement est en effet régi par la propre loi du locuteur autour d'une variabilité intra-locuteur dite *lawful*, non dispersée, contrairement à ce qu'on a pu voir dans les études sur la protrusion tout particulièrement en anglais. Cette variabilité des données suit la loi propre du locuteur et non celle de sa langue d'appartenance. De nouveau, les pentes spécifiques à chaque locuteur reflètent des stratégies d'exécution du mouvement d'arrondissement distinctes d'un sujet à l'autre.

L'*élasticité temporelle* suggérée par Byrd et Saltzman peut être tout à fait expliquée et prédite chez des locuteurs de même communauté linguistique, dans le cadre d'un modèle dont

les paramètres de mesures ainsi que les événements et phases proposés sont suffisamment robustes pour épouser la nature dynamique et variable de tels comportements.

Nos résultats ont confirmé que le modèle *MEM* est capable de rendre compte du comportement *tunérisé* d'un individu, qu'il soit de langue française ou anglaise. Son originalité est de proposer un *principe*, s'attachant à décrire l'anticipation d'arrondissement en tant que *phénomène* variable régulièrement (avec une *lawful variability*), de longue portée (*long range anticipation*). Mais dans la mesure où il s'agit d'un modèle paramétrable par locuteur, il permet le *tuning* ou *attunement* personnel. Lorsque l'on observe les données individuellement, on remarque effectivement que chaque locuteur, quelle que soit sa langue, adapte son mouvement d'anticipation à la longueur de ses gestes consonantiques, qui sont intégrés de manière fluide dans les coordinations oro-laryngées. Celles-ci se déroulent sur plusieurs consonnes intervocaliques, si la langue n'est pas restreinte à la syllabe universelle CV, mais comprend des *clusters* finaux et initiaux de mots, ce qui donne lieu au bout du compte, dans la jonction des mots en parole continue, à des *clusters* encore plus complexes, comme c'est le cas, entre autres, en français et en anglais.

De l'adulte à l'enfant

Le second objectif de notre projet visait à attester de la mise en place du comportement d'anticipation d'arrondissement vocalique chez l'enfant. Nous avons réalisé un enregistrement audiovisuel de 7 enfants de langue française, âgés entre 3 ans et demi et 8 ans, en incluant un suivi longitudinal pour les 4 plus jeunes enfants. L'exercice consistait à faire produire à ces enfants une série de courtes séquences comportant des noms, tous présentés comme des noms de peluche, débutant toujours par la voyelle [i], se terminant toujours par la voyelle [u] et comportant un nombre variable de consonnes entre les deux. Comme pour les études portant sur l'anticipation d'arrondissement chez l'adulte, leurs lèvres maquillées de bleu nous ont permis d'extraire efficacement la constriction intérolabiale.

Au lieu de parler d'*acquisition* du comportement d'anticipation d'arrondissement, nous avons préféré parler d'*attunement* ou d'ajustement du comportement moteur vers le patron de l'adulte. Ceci pour plusieurs raisons. Tout d'abord, nous avons montré, dans notre partie III., que l'acquisition du contrôle vocalique est un processus relativement long.

«[...] the accuracy of motor control improves with age until adult-like performance is achieved at about 11 or 12 years, somewhat after the age at which speech sound acquisition usually is judged to be completed. » (Kent, 1976, p. 421)

Dans ce long processus, les études cumulatives menées ces trente dernières années n'ont pas définitivement dégagé les grandes étapes (*milestones*) dans l'acquisition de la parole. Suivant que l'on se place du côté du produit acoustique ou de la coordination articulaire requise pour la réalisation *tunée* des voyelles, la littérature nous offre des points de vues divergents. A partir du travail de thèse de Serkhane (2005) sur les données de Kuhl & Meltzoff (1996), nous avons suggéré que l'enfant est en mesure de produire des cibles vocaliques perçues comme telles par des auditeurs adultes très tôt, dès 4-5 mois selon les auteurs. Nous ne répétons pas ici le peu que nous savons, à travers la littérature, sur l'émergence du contrôle vocalique. Nous souhaitons simplement renforcer l'idée qu'il existe une différence certaine entre les voyelles des enfants produites à 5 mois, par rapport à celles produites par l'adulte, et même par l'enfant de 4 ans. Dans le premier cas, le processus cognitif et moteur mis en œuvre par le jeune enfant relève de l'*émulation* du but. Ce dernier tente avec les moyens dont il dispose de produire la cible acoustique (le but). L'*imitation* d'une cible vocalique, telle que produite par un adulte, requiert une coordination articulaire et temporelle. Kent (1976) insiste d'ailleurs sur le fait que jusqu'à environ 8 ans, voire plus, les patrons acoustiques des voyelles reflètent un contrôle moteur encore immature.

« Motor patterns for speech are largely preplanned rather than feedback controlled even at the age of five years. [...] The variability of speech motor control progressively diminishes until the age of eight-12 years, when adultlike stability is achieved. [...] The declining variabilities for the three types of measurements [Fo patterns, formant frequencies and temporal patterns], appear to reflect an increasing precision of motor control over a five-to-eight years. » (Kent, 1976, p. 439-440)

En ce qui concerne plus particulièrement le contrôle de l'arrondissement requis pour la transition d'une voyelle étirée de type [i] à une voyelle arrondie telle que [y], celui-ci semble solliciter une coordination particulièrement élaborée des différents articulateurs supra-laryngés.

« A new movement [...] achieved by the cooperation of two muscles inserting on the styloid process of the cranium, the styloglossus and stylohyoid muscles. [...] Another new movement, that of lip rounding, is also necessary for high back vowels. [...] It is also of interest to note that back vowels, which, according to our analysis, make more demand on new movements, are acquired later than high front vowels. » MacNeilage & Davis (1990 p. 470)

Dans ce contexte de développement encore relativement flou, notre but n'était pas de déterminer un âge critique pour l'acquisition du contrôle de la cible vocalique [y] (par exemple, telle qu'elle est prononcée de manière isolée). Bien sûr, l'étude de l'anticipation d'arrondissement vocalique nous renseigne indirectement sur le degré de contrôle de la coordination supra-laryngée : un comportement d'anticipation d'arrondissement instable

témoignera en effet d'une coordination motrice et/ou temporelle non maîtrisée (non *tunée*). Notre propos était plutôt d'attester du contrôle de la coordination temporelle appropriée pour la réalisation d'une voyelle arrondie telle que [y] dans le timing d'une séquence coarticulée. Nous souhaitions évaluer si l'enfant montre très tôt un comportement d'anticipation d'arrondissement *lawful*, comme observé chez les adultes francophones et anglophones, ou si celui-ci s'ajustait progressivement.

On notera que nous avons choisi d'emprunter le terme d'*attunement phonologique* plus approprié à notre étude, plutôt que de nous référer simplement à celui d'*acquisition*. On trouve ce concept d'*attunement* (ou *ajustement*) dans des disciplines très diverses : en neurosciences et en psychologie cognitive (*intentional attunement*), en linguistique (*phonological attunement*) ou encore dans certains ouvrages de diagnostic pour repérer les pathologies ou retards du langage (*phonological disorders, speech disorders*) tels que le « *Diagnostic and statistical manual of mental disorders* ».

L'analyse de nos données labiales chez l'enfant français nous a, comme chez l'adulte, conduit à plusieurs constatations :

1. Lors du premier enregistrement, nous avons observé pour la plupart de nos jeunes locuteurs des comportements d'anticipation similaires à ceux rencontrés chez l'adulte, c'est-à-dire suivant une évolution linéaire déterminée par un coefficient d'expansion spécifique au locuteur. Pour 6 de nos petits Français, le modèle *MEM* parvient donc à rendre compte des profils d'anticipation de chacun, y compris de notre plus jeune locuteur, âgé de 3 ans et demi lors du premier enregistrement. Il est intéressant de constater qu'avec les mêmes paramètres de mesure que ceux appliqués pour l'adulte, il est possible de décrire la variabilité intra-individuelle rencontrée de manière *lawful*. Le modèle *MEM* nous permet donc d'attester chez l'enfant d'une coordination temporelle, dont la variabilité s'explique par la propre loi du jeune locuteur. D'après notre revue de la littérature sur le contrôle moteur de la parole, il est indéniable que les stratégies d'organisation de l'arrondissement vont s'affiner au cours des années, en même temps que la coordination spatiale des articulateurs supra-laryngés. Mais nous pouvons d'ores et déjà conclure que la coordination temporelle requise pour l'anticipation d'arrondissement vocalique – dans notre cas spécifique d'une voyelle [i] à son opposée arrondie [y] – peut être installée relativement tôt, entre 3 et 5 ans.

2. Seul un de nos plus jeunes enfants, KV, sort des sentiers battus puisqu'à 4;5 ans, ses données de mouvement sont très dispersées et ne sont pas validées statistiquement. En effet,

la variabilité des données observées ne correspond pas à une variabilité cohérente (*lawful*). Pour ce locuteur, nous pouvons suggérer que le processus mis en œuvre relève de *l'émulation du but*, comme c'était le cas des voyelles produites par les enfants de Kuhl & Meltzoff (1996). La cible acoustique [y] est atteinte, mais on ne peut conclure à la reproduction du patron temporel d'anticipation de la voyelle, telle que produite par l'adulte avec une coordination articulatoire et un déploiement temporel approprié (*imitation : du but comme du chemin*).

3. Pour les quatre plus jeunes locuteurs ayant été enregistrés deux fois, nous avons observé pour trois d'entre eux des profils d'anticipation stables d'une année sur l'autre (pas de différence significative entre les deux sessions). En revanche, le jeune locuteur KV, qui n'affichait pas un comportement dans la lignée régulière du *MEM* lors de la première session, rejoint maintenant, à 5;5 ans, ses petits camarades. La variabilité de ses données de constriction est désormais explicable par le modèle *MEM*. Le jeune KV déploie son geste d'arrondissement, de manière linéaire par rapport à l'intervalle acoustique disponible entre les deux voyelles, et son profil est validé par un coefficient de corrélation significatif.

Par conséquent, si le patron d'arrondissement vocalique peut être maîtrisé en français dès 3 ans et demi, il peut l'être aussi nettement plus tard. Il n'y a donc pas de fenêtre critique qui serait irrémédiablement manquée à cet âge: le petit KV nous a démontré que l'acquisition du patron peut attendre jusqu'à 5 ans et demi.

Il aurait été bien entendu intéressant d'enregistrer d'autres enfants de cet âge ou des enfants plus jeunes. Nous n'avons pas réussi à enregistrer des enfants en-dessous de 3 ans. Une de nos tentatives a été réalisée auprès d'une locutrice tout juste âgée de 3 ans : celle-ci n'est pas parvenue à accomplir la tâche de production proposée, soit des noms de peluches avec des groupes consonantiques, comme cela est facilement concevable, pour qui connaît la littérature sur l'acquisition de ces groupes chez l'enfant. Il semble donc bien que nous ayons atteint la limite expérimentale de notre tâche chez le jeune enfant.

Malgré ces quelques difficultés, nous pensons que cette étude développementale est un succès en ce sens que nous avons réussi à adapter le poste « Visage-Parole » ou *ICP-Lip-Shape-Tracker* conçu pour le recueil de données labiales chez l'adulte, à de jeunes enfants. Nous avons globalement appliqué le même paradigme expérimental que pour les adultes en respectant les contrôles expérimentaux inhérents à la bonne détection des données labiométriques. En outre, il s'agit de la première étude articulatoire traitant de l'anticipation

d'arrondissement chez des enfants aussi jeunes. Rappelons que notre seule référence de production était celle d'Abelin *et al.* (1980). Mais non seulement, les enfants enregistrés lors de cette étude étaient tous plus âgés (le plus jeune ayant 7 ans). Mais on a pu observer que leurs comportements d'anticipation étaient tous très dispersés (à l'exception d'un seul de 8 ans ayant un profil *Look-ahead* comme le locuteur adulte de référence) et ne pouvaient donc être associés à aucun des profils d'anticipation classiques que l'on trouve dans la littérature. Notre base de données labiales enfantines vient donc à propos augmenter décisivement les données rapportées dans la littérature sur le contrôle moteur et plus spécifiquement pour les études portant sur la coarticulation des gestes vocaliques et consonantiques dans le flux de la parole.

* *

*

Il serait désormais intéressant de nous tourner vers le versant perceptif de la coarticulation anticipante, en utilisant ces données de production enfantines, afin d'examiner si l'auditeur exploite au niveau perceptif les stratégies idiosyncratiques d'anticipation d'arrondissement vocalique mises en évidence en production. Cet axe de recherche est actuellement exploité par l'équipe de Lucie Ménard au Laboratoire de Phonétique de l'UQAM et chez l'adulte notamment par deux équipes, celle de Strasbourg par Rudolph Sock et au sein de notre laboratoire par Marie-Agnès Cathiard.

En outre, l'étude approfondie du contrôle de l'anticipation d'arrondissement peut être reliée à des perspectives et applications plus techniques. Nous pensons par exemple aux améliorations qui peuvent être apportées en robotique de la parole à travers les systèmes de synthèses audiovisuelles. La « Tête Parlante » de l'ICP fonctionne déjà avec des règles de type *MEM*, contrairement à la tête Baldi de Massaro, *MASSY* de Fagel et le système de synthèse *CASY* des Haskins. Enfin, on ne peut omettre les implications de notre travail au niveau de l'orthophonie puisque la connaissance des stratégies d'organisation de la coarticulation anticipante chez des enfants ne présentant aucune difficulté de langage peut aider à la mise en place de méthodes de rééducation pour les enfants rencontrant des difficultés d'articulation.

Enfin on aura compris que les processus de l'imitation, qui reviennent en force dans le champ des sciences cognitives, n'auront pu être qu'évoqués dans ce travail, notamment par

cette distinction productive entre *émulation* et *imitation*. Depuis la coordination oro-laryngée découverte chez le macaque, lorsqu'il est entraîné à produire *volontairement* un « *coo* » (cf. les tous récents travaux du groupe de Rizzolatti et Fogassi à Parme, déjà évoqués), et depuis l'organisation temporelle qui en découle pour l'intégration perceptive de cette communication oro-faciale bimodale, un long chemin reste sans aucun doute à parcourir pour trouver le passage jusqu'à l'anticipation de la configuration du conduit vocal naturellement en avance sur le déclenchement laryngé. Ce patron de base chez le macaque (les lèvres prennent le temps de mettre en place leur constriction, puis l'impulsion glottique excite le conduit correctement configuré), c'est dans son principe même la base de l'organisation temporelle du patron d'anticipation que nous avons mesuré pour les adultes francophones et anglophones, et chez l'enfant. Sans anticiper sur cette phylogénèse correspondant à l'ontogénèse du petit d'homme, il n'empêche que nous avons le sentiment d'avoir posé en perspective de notre travail cette direction, pour trouver les chemins d'une véritable imitation et pas d'une simple émulation dans l'*attunement* de la parole, de l'enfant à l'adulte.

Bibliographie

- Abbs, J.H., & Eilenberg, G.R. (1976). Peripheral mechanisms of speech motor control. In N.J. LASS (ed.) *Contemporary Issues in Experimental Phonetics*, New York: Academic Press, 139-166.
- Abbs J.H., & Gracco V.L. (1984). Control of complex motor gestures: Orofacial muscle responses to load perturbations of lip during speech. *Journal of Neurophysiology*, 51,705-723.
- Abelin, A., Landberg, I., & Persson, L. (1980). A study of anticipatory labial coarticulation in the speech of children. *Phonetic Experimental Research at the Institute of Linguistic*, Université de Stockholm, 2, 2-18.
- Abry, C., Boë, L.-J., Corsi, P., Descout, R., Gentil, M., & Graillot, P. (1980). Labialité et phonétique. Données fondamentales et études expérimentales sur la géométrie et la motricité labiales. Publication de l'Université des Langues et Lettres de Grenoble.
- Abry, C., & Boë, L.-J. (1986). Laws for lips. *Speech Communication*, 5, 97-104.
- Abry, C., & Lallouache, M.T. (1991). Audibility and stability of articulatory movements : deciphering two experiments on anticipatory rounding in French. In *Proceedings of the 12th Int. Congress of Phonetic Sciences*, Aix-en-Provence, France, 1, 220-225.
- Abry, C., & Lallouache, M.T. (1991). Does increasing representational complexity lead to more speech variability? On explaining some French data by current models of anticipatory rounding. In O. Engstrand & C. Kylander (eds.), *Current Phonetic Research Paradigms: Implications for Speech Motor Control*, Stockholm, *Perilus*, 14, 1-5.
- Abry C., Cathiard M.-A., Robert-Ribès J. & Schwartz J.-L. (1994). The coherence of speech in audio-visual integration. Commentaire de l'article-cible de M. Radeau, Auditory-visual spatial interaction and modularity. *Current Psychology of Cognition*, 13 (1), 52-59.
- Abry C., & Lallouache, T. (1995a). Le MEM : Un modèle d'anticipation paramétrable par locuteur. Données sur l'arrondissement en français. *Bulletin de la Communication Parlée*, 3, 85-99.

- Abry, C., & Lallouache, T.M. (1995b). Modeling lip constriction anticipatory behaviour for rounding in French with the MEM. In *Proceedings of ICPHS, 4*, Stockholm, Suède, 152-155.
- Abry, C., & Perrier, P. (1996). Le contrôle des mouvements audibles et visibles dans la parole. In *Fondements et perspectives en traitement automatique de la parole*, Aupelf-UREF, Université Francophone, 55-80.
- Abry, C., Lallouache, M.-T., & Cathiard, M.-A. (1996a). How can coarticulation models account for speech sensitivity to audio-visual desynchronization? In D. Stork & M. Hennecke (eds.), *Speechreading by Humans and Machines*, NATO ASI Series F, 150, Springer-Verlag, Berlin, Heidelberg, Tokyo, 247-255.
- Abry, C., Cathiard, M.-A., El Abed, R., Lallouache, M.-T., Leroy, M.-C., Perrier, P., & Poveda, P. (1996b). Silent speech production : Anticipatory behaviour for 2 out of the 3 main vowel gestures/features, while pausing. In *Proceedings of the 4th Speech Production Seminar, 1st ESCA Tutorial and Research Workshop on Speech Production Modeling: from control strategies to acoustics*, Autrans, France, 101-104.
- Abry, C., Cathiard, M.-A., Vilain, A., Laboissière, R. Loevenbruck, H., Savariaux, C., & Schwartz, J.-L. (2007). Some insights in bimodal perception given for free by the natural time course of speech production. In E. Vatikiotis-Bateson, P. Perrier & G. Bailly (eds.), *Advances in auditory-visual speech processing*. Cambridge: MIT Press
- Almé, A.-M., & McAllister, R. (1984). Experiments in speech processes. *Phonetic Research Seminar, Report IV*, 3-19.
- Arbib, M.A., & Rizzolatti, G. (1997). Neural expectations: A possible evolutionary path from manual skills to language. *Communication and Cognition*, 29, 393-423.
- Arbib, M.A., (2005). From monkey-like action recognition to human language: an evolutionary framework for neurolinguistics. *Behavioral Brain Sciences*, 28, 64p.
- Aronson, E., & Rosenbloom, S. (1971). Space perception in early infancy: perception within a common auditory-visual space. *Science*, 11, 1161-1163.
- Atal, B.S. (1983). Efficient Coding of LPC Parameters by Temporal Decomposition. In *Proceedings ICASSP*, 83, 81-84.
- Audouy, M. (2000). Logiciel de traitement d'images vidéo pour la détermination de mouvements des lèvres. Projet de fin d'études, option génie logiciel, ENSIMA, Grenoble.
- Badin, P., Motoki, K., Miki, N., Ritterhaus, D. & Lallouache, T.M. (1994). Some geometric and acoustic properties of the lip horn. *Journal of the Acoustical Society of Japan (English)*, 15(4), 243-253.
- Bard, K.A., & Russell, C.L. (1999). Evolutionary foundations of imitation: Social, cognitive and developmental aspects of imitative processes in non-human primates. In J. Nadel & G. Butterworth (eds.), *Imitation in Infancy*, Cambridge University Press, 89-123.

- Barraclough, N.E., Xiao, D., Baker, C.I., Oram, M.W., & Perrett, D.I. (2005). Integration of visual and auditory information by STS neurons responsive to the sight of actions. *Journal of Cognitive Neuroscience*, 17(3), 377-391.
- Barry, W.J. (1983). Some problems of interarticulator phasing as an index of temporal regularity in speech. *Journal of Experimental Psychology: Human Perception and Performance*, 9, 826-828.
- Bell-Berti, F. (1993). Understanding velic motor control: Studies of segmental context. In Academic/Huffman & Krakow (eds.), *Phonetics and phonology*, 5, *Nasals, nasalization and the velum*, Academic Press, Inc, 63-85.
- Bell-Berti, F., & Harris, K.S. (1974). More on the motor organization of speech gestures. *Haskins Status Report on Speech research*, SR-37/38, 73-77.
- Bell-Berti, F., & Harris, K.S. (1981). A temporal model of speech production, *Phonetica*, 38, 9-20.
- Bell-Berti, F., & Harris, K.S. (1982). Temporal patterns of coarticulation: lip rounding. *Journal of Acoustical Society of America*, 71(2), 449-554.
- Benguerel, A.P., & Cowan, H.A. (1974). Coarticulation of upper lip protrusion in French. *Phonetica*, 30, 41-55.
- Benoît, C. (1986). Note on the use of correlations in speech timing. *Journal of the Acoustical Society of America*, 80(6), 1846-184.
- Berthoz, A. (1998). Le secret du geste : l'anticipation. In A. Berthoz (ed.), *Sciences et Vie « Le cerveau et le mouvement »*, 68-76.
- Besle, J., Fort, A., Delpuech, C., & Giard, M.-H. (2004). Bimodal speech: Early suppressive visual effects in human auditory cortex. *European Journal of Neuroscience*, 20, 2225-2234.
- Bimbot, F., Chollet, G., Deleglise, P., & Montacié, C. (1988). Temporal decomposition and acoustic-phonetic decoding of speech. In *Proceedings ICASSP*, 88, 445-448.
- Blair, C. (1986). Interdigitating muscle fibers throughout orbicularis oris inferior : preliminary observations. *Journal of Speech and Hearing Research*, 29, 266-269.
- Boë, L.-J. (1999). Modelling the growth of the vocal tract vowel spaces of newly-born infants and adults: Consequences for ontogenesis and phylogenesis. In *Proceedings of the 14th International Congress of Phonetic Sciences*, 3, 2501-2504.
- Bonda, E., Petrides, M., Frey, S., & Evans, A.C. (1994). Frontal cortex involvement in organized sequence of hand movements: Evidence from positron emission tomography study. *Soc Neurosci Abstr*, 20, 353.
- Bonnot, J.-F.P. (1990). Production de la parole et coarticulation. Une analyse critique des principaux modèles. *Travaux de l'Institut de Phonétique de Stasbourg*, 20.

- Bonnot, J.-F.P, Chevrie-Muller, C., Arabia-Guidet, C., Maton, B., & Greiner, G.F. (1986). Coarticulation and motor encoding of labiality and nasality in CVCVCV nonsense Words. *Speech Communication*, 5 (1), 83-95.
- Boyce, S. E. (1988). *The influence of phonological structure on articulatory organization in Turkish and in English: Vowel harmony and coarticulation*. Ph.D. dissertation, Yale University, New Haven, CT.
- Boyce, S. E. (1990). Coarticulatory organization for lip rounding in Turkish and English. *Journal of the Acoustical Society of America*, 88 (6), 2584-2595.
- Boyce, S.E., Krakow, R.A., Bell-Berti, F., & Gelfer, C.A. (1990). Coverging sources of evidence for dissecting articulatory movements into core gestures. *Journal of Phonetics*, 18, 173-188.
- Boyce, S.E., Krakow, R.A., & Bell-Berti, F. (1991). Phonological underspecification and speech motor organisation. *Phonology*, 8, 216-236.
- Boysson-Bardies, B. de. (1996). *Comment la parole vient aux enfants*, éditions Odile Jacob.
- Browman, C.P., & Goldstein, L. (1986). Towards an articulatory phonology. In C. Ewan & J. Anderson (eds.) *Phonology Yearbook 3*. Cambridge: Cambridge University Press, 219-252.
- Browman, C.P., & Goldstein, L. (1992). Articulatory phonology: An overview, *Phonetica*, 49, 155-180.
- Buccino, G., Binkofski, F., Fink, G.R., Fadiga, L., Fogassi, L., Gallese, V., Seitz, R.J., Zilles, K., Rizzolatti, G., & Freund, H.-J., (2001). Action observation activates premotor and parietal areas in a somatotopic manner: An fMRI study. *Eur. J. Neurosci.*, 13, 400–404.
- Buhr, R.D. (1980). The emergence of vowels in an infant. *Journal of Speech, Language and Hearing Research*, 23, 73-94.
- Bush, N. (2001). Frequency effects and word-boundary palatalization in English. In J. Bybee & P. Hopper (eds.), *Frequency and the emergence of linguistic structure*, Amsterdam: Benjamins, 255-280.
- Bybee, J. (2001). *Phonology and language use*. Cambridge University Press.
- Byrd, D., & Saltzman, E. (2003). The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics*, 31, 149-180.
- Calvert, G.A., Campbell, R., & Brammer, M.J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Curr Biol*, 10, 649-657.
- Cappe, C., & Barone, P. (2005). Heteromodal connections supporting multisensory integration at low levels of cortical processing in the monkey. *Eur. J. Neurosci.*, 22, 2886-2902

- Cathiard, M.-A. (1994). *La perception visuelle de l'anticipation des gestes vocaliques : cohérence des événements audibles et visibles dans le flux de la parole*. Thèse de Psychologie Cognitive, Grenoble.
- Cathiard, M.-A., Attina, V., & Alloatti, D. (2003). Labial Anticipation Behavior during Speech with and without Cued Speech. In *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona, Spain, 1939-1942.
- Cathiard, M.-A., Tseva, A., & Lallouache, M.-T. (1992). Identification visuelle des gestes de protrusion et de rétraction des lèvres au cours des pauses acoustiques : Les performances de sujets français et grecs. *Colloque C1, supplément au Journal de Physique III, 2, 2ème Congrès Français d'Acoustique*, 319-322.
- Cathiard, M.-A., Lallouache, M.-T., & Abry, C. (1996). Does movement on the lips mean movement in the mind ? In D. Stork & M. Hennecke (eds.), *Speechreading by Humans and Machines*, NATO ASI Series F: Computer and Systems Sciences, 150, Springer-Verlag, Berlin Heidelberg New York London Paris Tokyo, 211-219.
- Chaminade, T., Meary, D., Orliaguet, J.P., & Decety, J. (2001). Is visual anticipation a motor simulation? A PET study. *NeuroReport*, 12, 3669-3674.
- Chomsky N., & Halle M. (1968). The sound pattern of English. In Harper & Row, cons. (traduction française : *Principes de phonologie générative*, Paris, Seuil, 1973).
- Cohen, M.A., Grossberg, S., & Stork, D.G. (1988). Speech perception and production by a self-organizing neural network. In Y.C. Lee (ed.), *Evolution, Learning, Cognition, and Advanced Architectures*, Hong Kong: World Scientific Publishers, 1-15.
- Corballis, M.C. (2002). From hand to mouth: The origins of language. In N.J. Princeton: Princeton University Press, *From mouth to hand: gesture, speech, and the evolution of handedness. Behavioral Brain Sciences*, 26, 199-260.
- Coudé, G., Ferrari, P.F., Rozzi, S.; Borelli, E., Bonini, L., Veroni, V., Rodà, F., Rizzolatti, G., & Fogassi, L. (2006). Motor control sensorimotor integration and cognitive functions of the cortical ventral motor areas of the macaque monkey: A mapping study. *Soc. Neurosci. Abs*, 194.2.
- Daniloff, R., & Moll (1968). Coarticulation of lip rounding. *Journal of Speech and Hearing Research*, 11, 707-721.
- Daniloff, R., & Hammarberg, R. (1973). On defining coarticulation. *Journal of Phonetics*, 1, 239-248.
- Davis, B.L., & MacNeilage, P. F. (1990). Acquisition of correct vowel production: a quantitative case study. *Journal of Speech, Language and Hearing Research*, 33, 16-27.
- Davis, B.L., & P.F. MacNeilage (1995). The articulatory basis of babbling. *Journal of Speech, Language and Hearing Research*, 38, 1199-1211.

- Davis, B.L., Macneilage, P.F., & Matyear, C.L. (1999). Acquisition of serial complexity in speech production: a comparison of phonetic and phonological approaches to first word production. *Phonetica*, 60p.
- Descout, R., Boë, L.-J., & Abry, A. (1980). Labialité vocalique et labialité consonantique. In C. Abry, L.-J. Boë, P. Corsi, R. Descout, M. Gentil & P. Graillot, *Labialité et phonétique. Données fondamentales et études expérimentales sur la géométrie et la motricité labiale*, Publication de l'Université des Langues et Lettres de Grenoble, 111-126.
- Dong, Y., Deng, L., & Acero, A. (2006). A lattice search technique for long-contextual-span hidden trajectory model of speech. *Speech Communication*, 48, 1214-1226.
- Eguchi, S., & Hirsh, I. J. (1969). Development of speech sounds in children. *Acta Oto-Laryngol. Suppl*, 257, 1-51.
- Eimas, P.D., Siqueland, E.R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, 171, 303-306.
- Engstrand, O. (1980). Acoustic constraints or invariant input representation? An experimental study of selected articulatory movements and targets. *Rapport Uppsala University Department of Linguistics*, 7, 67-95.
- Engstrand, O. (1989). Towards an electropalatographic specification of consonant articulation in Swedish. *Perilus* 10, 115-156.
- Fadiga, L., Fogassi, L., Pavesi, G., & G. Rizzolatti. (1995). Motor facilitation during action observation: a magnetic stimulation study. *Journal of Neurophysiology*, 73, 2608-2611.
- Falchier, A., Clavagnier, S., Barone, P., and Kennedy, H. (2002). Anatomical evidence of multimodal integration in primate striate cortex. *The Journal of Neuroscience*, 22(13), 5749-5759.
- Fant, G. (1968). Analysis and synthesis of speech processes, In B. Malmberg (ed.), *Manual of Phonetics*, In Amsterdam, North-Holland Publ. Co. Chapt. 8, 173-276.
- Fant, G. (1973). *Speech Sounds and Features*, MIT Press, Cambridge MA.
- Farnetani E. (1999). Labial coarticulation. In W.J. Hardcastle, & N. Hewlett (eds.), *Coarticulation: Theory, Data, and Techniques*, 144-163, Cambridge:Cambridge University Press.
- Farnetani, E., & Recasens, D. (1999). Coarticulation models in recent speech production theories. In W.J. Hardcastle, & N. Hewlett (eds.), *Coarticulation: Theory, Data, and Techniques*, 31-65, Cambridge:Cambridge University Press.
- Ferrari, P.F., Gallese, V., Rizzolatti, G., & 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 (8), 1703-1714.
- Field, M.M., Woodson, R.W., Greenberg, R., & Cohen, C. (1982). Discrimination and imitation of facial expressions by neonates. *Science*, 218, 179-181.

- Fowler, C.A. (1977). Timing control and speech production. PhD dissertation non publiée, Dartmouth College, Hannover. N.H.
- Fowler, C.A., & Saltzman, E. (1993). Coordination and coarticulation in speech production. *Language and Speech*, 36, 171-195.
- Fromkin, V. (1964). Lip positions in American English vowels. *Language and Speech*, 7(3), 215-225.
- Gallagher, S. (2004). Agency, ownership, and alien control in schizophrenia. In D. Zahavi, T. Grünbaum & J. Parnas (eds.), *Interdisciplinary perspectives on self-consciousness*, Amsterdam: John Benjamins Publishers, 89-104.
- Gallese, V. (2001). The shared manifold hypothesis: From mirror neurons to empathy. *Journal of Consciousness Studies*, 8, 33-50.
- Gallese, V. (2004). La mise en phase intentionnelle. Le système miroir et son rôle dans les relations interpersonnelles. *Interdisciplines*.
[<http://www.interdisciplines.org/mirror/papers/1>]
- Gallese V. (2005) Being like me: Self-other identity, mirror neurons and empathy. In S. Hurley & N. Chater (eds.), *Perspectives on imitation: From cognitive neuroscience to social science*, Boston, MA: MIT Press.
- Gallese, V. (2006a). Mirror neurons and intentional attunement: A commentary on David Olds. *Journal of American Psychoanalytic Association*, 54, 46-57.
- Gallese, V. (2006b). Intentional attunement: A neurophysiological perspective on social cognition and its disruption in autism. *Brain Research*, Research report, 35124, 10p.
[<http://www.elsevier.com/locate/brainres>]
- Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, 119, 593-609.
- Gallese, V., Fogassi, L., Fadiga, L., & Rizzolatti, G. (2002). Action Representation and the inferior parietal lobule. In W. Prinz & B. Hommel (eds.), *Atten. Perform.*, XIX, Oxford Univ. Press, Oxford, 247-266.
- Gallistel, C.R. (1980). *The organization of action: a new synthesis*. Erlbaum, Hillsdale NJ.
- Gay, T. (1977). Articulatory movements in VCV sequences. *Journal of the Acoustical Society of America*, 62, 183-193.
- Gay, T. (1978). Articulatory units: Segments or syllables? In A. Bell & J. Hooper (eds.), *Syllables and Segments*. Amsterdam: North-Holland, Cambridge: Cambridge University Press, 451-470.
- Gay, T. (1981). Mechanisms in the control of speech rate, *Phonetica*, 38, 148-158.

- Gelfer, C.E., Bell-Berti, F., & Harris, K.S. (1989). Determining the extent of coarticulation: Effects of experimental design. *Journal of the Acoustical Society of America*, 86(6), 2443-2445.
- Gentaz, E., & Mazens, K. (2006). Les nouveau-nés sont-ils capables de voir avec leurs mains ou de toucher avec leurs yeux ? *Devenir*, 18(3), 263-281.
- Gentil, M. & Boë, L.-J. (1980). Les lèvres et la parole. In C. Abry, L.-J. Boë, P. Corsi, R. Descout, M. Gentil & P. Graillot, *Labialité et phonétique. Données fondamentales et études expérimentales sur la géométrie et la motricité labiales*. Publication de l'Université des Langues et Lettres de Grenoble, 11-106.
- Gentilucci, M., Fogassi, L., Luppino, L.G., Matelli, M., Camarda, R., & Rizzolatti, G. (1988). Functional organisation of the inferior area 6 in the macaque monkey. *Exp. Brain Res.*, 71, 475-490.
- Gentilucci M., Benuzzi F., Bertolani L., & Gangitano M. (2001). The influence of stimulus color on the control of reaching-grasping movements. *Exp. Brain Res.*, 137, 36-44.
- Gentner, D.R., Grudin, J., & Conway, E. (1980). Finger movements in transcribing typing. *Tech. Rep. n°8001, La Jolla, CA: University of California, San Diego, Center for Human Information Processing*.
- Ghazanfar, A.A., Maier, J.X., Hoffman, K.L., & Logothetis, N.K. (2005). Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex. *J. Neurosci.*, 25, 5004–5012.
- Ghazanfar, A.A., & Schroeder, C.E. (2006). Is neocortex essentially multisensory ? *TICS*, 10(6), 278-285.
- Ghazanfar, A.A., & Miller, C.T. (2007). Acoustic communication. *Encyclopedia of Animal Behavior*, Greenwood Press, Westport, CT, 879-881.
- Ghazanfar, A.A. Turesson, H.K., Maier, J.X., van Dinther, R., Patterson, R.D., & Logothetis, N.K. (2007). Vocal-tract resonances as indexical cues in rhesus monkeys. *Current Biology*, 17, 425-430.
- Gibson, A., & McPhearson, L. (1980). Production of bite-block vowels by children. *PERILUS*, Université de Stockholm, Suède, 2, 26-43.
- Gil-da-Costa, R., Martin, A., Lopes, M.A., Munoz, M., Fritz, J.B., & A. R. Braun. (2006). Species-specific calls activate homologs of Broca's and Wernicke's areas in the macaque. *Nature*, 9(8), 1064-1070.
- Goldstein, L. (2003). Emergence of discrete gestures. In *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona, Spain, 85-88.
- Goldstein, L., & Rubin, P. (2007). Speech, dances of the vocal tract. *Odyssey magazine*.
- Goren, C., Sarty, M., & Wu, P. (1975). Visual following and pattern discrimination of face-like stimuli by newborn infants. *Pediatrics*, 56, 544-549.

- Green, J.R., Moore, C.A., Higashikawa, M., & Steeve, R.W. (2000). The physiologic development of speech motor control: Lip and jaw coordination. *Journal of Speech, Language, and Hearing Research*, 43, 239-56.
- 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*, 4, 66-79.
- Green, J.R., Moore, C.A., & Reilly, K.J. (2003). Methodological issues in studies of early articulatory development: A response to Dworkin, Meleca, and Stachler. *Journal of Speech, Language, and Hearing Research*, 46, 1020-1021.
- Guenther, F.H., Ghosh, S.S., Nieto-Castanon, A., & Tourville, J.A. (2006). A neural model of speech production. In J. Harrington & M. Tabain (eds.), *Speech Production: Models, Phonetic Processes, and Techniques*, London: Psychology Press, 27-39.
- Hammarberg, R. (1976). The metaphysics of coarticulation. *Journal of Phonetics*, 4, 353-363.
- Hardcastle, W.J., & Hewlett, N. (1999). *Coarticulation. Theory, data and techniques*. Cambridge University Press.
- Harris, J. (2006). Wide-domain r-effects in English. A paraître in *UCL Working Papers in Linguistics*, 18.
- Hast, M.H, Fischer, J.M, Wetzell, A.B, & Thompson, V.E. (1974). Cortical motor representation of the laryngeal muscles in *Macaca mulatta*. *Brain Research*, 73(2), 229-40.
- Hauser, M.D., Chomsky, N. & Fitch, W.T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science*, 298, 1569–1579.
- Hawkins, S., Heid, S., House, J., & Huckvale, J. (2000). Assessment of naturalness in the ProSynth speech synthesis project. *IEE: State-of-the-art in speech synthesis*, 13/1-13/6.
- Hawkins, S., & Nguyen, N. (2003). Effects on word recognition of syllable-onset cues to syllable-coda voicing. In J. Local, R. Ogden & R. Temple (eds.), *Papers in Laboratory Phonology VI*, Cambridge Univ. Press, Cambridge, 38-57.
- Heid, S., & Hawkins, S. (2000). An acoustical study of long-domain /r/ and /l/ coarticulation. In *Proceedings of the 5th Seminar on Speech Production: Models and Data (ISCA)*, Kloster Seeon, Bavaria, Germany, 77-80.
- Heim J.-L. (1989). La reconstitution du crâne de la Chapelle-aux-Saints. Film 16 mm, Paris, (ed.), SFRS, *Une nouvelle reconstitution du crâne néandertalien de la Chapelle-aux-Saints*. *C.R. Acad. Sc. Paris*, t. 308, série II, 6, 1187-1192.
- Heim, J.-L., Boë, L.-J., & Abry, C. (2002). La parole à la portée du conduit vocal de l'Homme de Neandertal : Nouvelles recherches, nouvelles perspectives, *CRAS*, 1-9.
- Henke, W.L. (1966). *Dynamic articulatory model of speech production using computer simulation*. PhD Dissertation MIT, Cambridge.

- Horner, V., & Whiten, A. (2005). Causal knowledge and imitation/emulation switching in chimpanzees (*Pan troglodytes*) and children (*Homo sapiens*). *Animal Cognition*, 8(3), 164-181.
- Horner, V., Whiten, A., Flynn, E., & de Waal, F.B.M. (2006). Faithful copying of foraging techniques along cultural transmission chains by chimpanzees and children. *Proc. Nat. Acad. Sci.* 103, 13878-13883.
- James, D.G.H., Van Doorn J., & McLeod, S. (2001). Vowel production in mono-, di- and polysyllabic words in children aged 3;0 to 7;11 years. In *Proceedings of Speech Pathology, Australia Conference*, 127-136.
- Jeannerod, M. (1994). The representing brain: Neural correlates of motor intention and imagery. *Behavioral and Brain Sciences*, 17, 187-245.
- Johnson, M.H., Dziurawiec, S., Ellis, H., & Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition*, 90, 1-19.
- Johnson, M.H., & Morton, J. (1991). *Biology and cognitive development: The case of face recognition*. Oxford, England: Blackwell.
- Joos, M. (1948). Acoustic Phonetics. *Language*, 24, Suppl.1., 1-136.
- Jordan, M.I. (1990). Motor Learning and the Degrees of Freedom Problem. In M. Jeannerod (ed.), *Attention & Performance XIII, Motor Representation and Control*, 796-836, Hillsdale, NJ: Lawrence Erlbaum Associates.
- Kandel, S., Orliaguet, J.-P., & Boë, L.-J. (2000a). Detecting anticipatory events in handwriting movements. *Perception*, 29, 953-964.
- Kandel, S., Orliaguet, J.-P., & Viviani, P. (2000b). Perceptual anticipation in handwriting: The role of implicit motor competence. *Perception & Psychophysics*, 62(4), 706-716.
- Kaas, J.H., & Hackett, T.A. (2000). Subdivisions of auditory cortex and processing streams in primates. *PNAS*, 97(22), 11793-11799.
- Katz, W.F., Kiple, C., & Tallal, P. (1991). Anticipatory coarticulation in the speech of adults and young children: acoustic, perceptual, and video data. *Journal of Speech and Hearing Research*, 34, 1222-1232.
- Keating, P.A. (1990). The window model of coarticulation: Articulatory evidence. In J. Kingston & M. Beckman (eds.), *Papers in Laboratory Phonology I*, Cambridge University Press, 451-470.
- Keele, S.W., Cohen, A., & Ivry, R. (1990). Motor programs: Concepts and issues. In M. Jeannerod (ed.), *Attention and performance XIII, Motor representation and control*, Hillsdale, NJ: Lawrence Erlbaum Associates, 77-110.
- Kent, R.D. (1976). Anatomical and neuromuscular maturation of the speech mechanism: Evidence from acoustic studies. *Journal of Speech and Hearing Research*, 19, 421-447.

- Kent, R.D. (1983). The segmental organization of speech. In P.F. MacNeilage (ed.), *The production of speech*, New York: Springer, 57-89.
- Kent, R.D. (1992). The biology of phonological development. In C.A. Ferguson, L. Menn, & C. Stoel-Gammon (eds.), *Phonological Development. Models, Research, Implications*, Timonium, Maryland, York Press, 65-91.
- Kent, R.D., & Forner, L.L. (1980). Speech segment durations in sentence recitations by children and adults. *Journal of Phonetics*, 8, 157-168.
- Kent, R.D., & Bauer, H.D. (1985). Vocalization of one-year olds. *Journal of Child Language*, 12(3), 491-526.
- Kirk, C., & Demuth, K. (2005). Asymmetries in the acquisition of word-initial and word-final consonant clusters. *Journal of Child Language*, 32 (4), 709-734.
- Keysers, C., & Perrett, D.I. (2004). Demystifying social cognition: A Hebbian perspective. *Trends in Cognitive Sciences*, 8, 501-507.
- Keysers, C., Kohler, E., Umiltà, M.A., Nanetti, L., Fogassi, L., Gallese, V. (2003). Audiovisual mirror neurons and action recognition. *Experimental Brain Research*, 153, 628-636.
- Köhler, W. (1927). *L'intelligence des singes supérieurs*. Paris, Félix Alcan.
- Kohler, E., Keysers, C., Umiltà, M.A., Fogassi, L., Gallese, V., & Rizzolatti, G. (2002). Hearing sounds, understanding actions: action representation in mirror neurons. *Science*, 297, 846-848.
- Koopmans-van Beinum, F., & Van der Stelt, J. (1986). Early stages in the development of speech movements. In L. Zetterström (ed.), *Precursors of early speech*, New-York, Stockton Press, 37-49.
- Kozhevnikov, V.A., & Christovich, L.A. (1965). Speech, articulation, perception. *Joint Publ. Res., Serv., US Department of Commerce*, 30, Washington.
- Kugler, P.N., Kelso, J.A.S., & Turvey, M.T. (1982). On coordination and control in naturally developing systems. In J.A.S. Kelso & J.E. Clark (eds.), *The development of human movement coordination and control*, New York, London: John Wiley, 5-78.
- Kuhl, P.K. (1992). Speech prototypes: Studies on the nature, functions, ontogeny and phylogeny of the « centers » of speech categories. In Y. Tohkura, E. Vatikiotis-Bateson & Y. Sagisaka (eds.), *Speech perception, production and linguistic structure*, Tokyo, Ohmsha, 239-264.
- Kuhl, P.K. (2000). A new view of language acquisition. In *Proceedings of the National Academy of Science*, 97, 11850-11857.
- Kuhl, P.K., & Meltzoff, A. (1982). The bimodal perception of speech in infancy. *Science*, 218, 1138-1141.

- Kuhl, P.K. & Meltzoff, A.N. (1995). Vocal learning in infants : development of perceptual-motor links for speech. In *Proceedings of the International Congress of Phonetic Sciences*, Stockholm, Suède, 1, 146-149.
- 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.
- Kühnert, B., & Nolan, F. (1999). The origin of coarticulation. In W.J. Hardcastle, & N. Hewlett (eds.), *Coarticulation: Theory, data, and techniques*, Cambridge: Cambridge University Press, 7-30.
- 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.
- Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Oxford: Blackwell.
- Lavee, C., & Vilain, A. (2006). What does it take to make a first word? The development of speech motor control during the first year of life. In *Proceedings of the 7th International Seminar on Speech Production*, Ubatuba, Brazil, 83-90.
- Lallouache, M.-T. (1991). *Un poste «Visage-parole » couleur. Acquisition et traitement automatique des contours des lèvres*. Thèse. Dissertation ENSERG, Grenoble.
- Lee, S., Potamios, A., & Narayanan, S. (1999). Acoustic of children's speech: Developmental changes of temporal and spectral parameters. *Journal of Acoustical Society of America*, 105(3), 1455-1468.
- Legerstee, M. (1990). Infants use multimodal information to imitate speech sounds. *Infant Behavior and Development*, 13, 343-354.
- Le Goff, B., & Benoît, C. (1997). A French-Speaking Synthetic Head. In *Proceedings of the ESCA Workshop on Audio-Visual Speech Processing (AVSP'97)*, Rhodes, Greece, 145-148.
- Lleo, C., & Prinz, M. (1997). Consonant clusters in child phonology and directionality of syllable structure assignment. *Journal of Child language*, 23, 31-56.
- Lieberman, P. (1980). On the development of vowel production in young children. In G.H. Yeni-Komshian, J.F. Kavanagh & C.A. Ferguson (eds.), *Child phonology, 1: Production*. New York: Academic Press, 23-42.
- Lieberman, P., Crelin, E.S., & Klatt, D.H. (1972). Phonetic ability and related anatomy of the newborn and adult human, Neanderthal man, and the chimpanzee. *American Anthropologist*, 74, 287-307.
- Liberman A.M., & Whalen, D.H. (2000). On the relation of speech to language. *Trends in Cognitive Sciences*, 4 (5), 187-196.

- Lindblom, B.E.F., & Studdert-Kenedy, M. (1967). On the role of formant-transitions in vowel recognition. *Journal of the Acoustical Society of America*, 42, 830-843.
- Lindblom, B.E.F., Sussman, H.M., Modarresi, G., & Burlingame, E. (2002). The trough effect: implications for speech motor programming. *Phonetica*, 59, 245-262.
- Lindblom, B., & Sussman, H.M. (2002). Principal components analysis of tongue shapes in symmetrical VCV utterances. *Quartely Progress and Status Report*, Department of Speech, Music and Hearing and Centre for Speech Technology, KTH, Stockholm, Suède, 44, 1-4.
- Linker, W. (1982). Articulatory and acoustic correlates of labial activity in vowels: A cross-linguistic study. *UCLA working papers in Phonetics*, 56 (monograph).
- Local, J., & Kelly, J. (1986). Long domain resonance patterns in English. In *Proceedings of IEE Conference on Speech Input/Output: Techniques and Applications*, London: Institution of Electrical Engineers, 304-309.
- Locke, J.L. (1983). *Phonological acquisition and change*, New-York, Academic Press.
- Locke, J.L. (1993). *The child's path to spoken language*. Cambridge MA and London: Harvard U. Press.
- Locke, J.L. (1999), Towards a biological science of language development. In M. Barrett (ed.), *The development of language*, Hove (East Sussex), UK: Psychology Press, 373-395.
- Löfqvist, A. (1990). Speech as Audible Gestures. In W. J. Hardcastle & A. Marchal, (eds.), *Speech Production and Speech Modelling*. Kluwer Academic Publishers, 289-322,.
- Löfqvist, A., & Gracco, V. L. (1997). Lip and jaw kinematics in bilabial stop consonant production. *Journal of Speech Language and Hearing Research*, 40, 877-893.
- Logothetis, N.K., Guggenberger, H., Peled, S., & Pauls, J. (1999) Functional imaging of the monkey brain. *Nat. Neurosci.* 2, 555-562.
- Lubker, J. (1979). The reorganization times of bite-block vowels. *Phonetica*, 36, 273-293.
- Lubker, J. (1981). Temporal Aspects of Speech Production: Anticipatory Labial Coarticulation. *Phonetica*, 38, 51-65.
- Lubker, J., & Gay, T. (1982). Anticipatory labial coarticulation: experimental, biological, and linguistic variables. *Journal of Acoustical Society of America*, 71(2), 437-448.
- MacNeilage, P.F., & DeClerk, T.L. (1965). On the motor control of coarticulation in CVC monosyllables. *Journal of Acoustical Society of America*, 45, 1217-1233.
- MacNeilage, P.F., Davis, B. (1990). Acquisition of speech production: Frame, then, content. In M. Jeannerod (ed), *Attention and performance XIII*, Lawrence Erlbaum Associates, Publishers, Hillsdale, New Jersey, 453-475.

- MacNeilage, P.F. (1998). The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences*, 21, 499-511.
- MacSweeney, M., Amaro, E., Calvert, G., Campbell, R., David, A.S., McGuire, P., Williams, S., Woll, B., & Brammer, M. J. (2000). Activation of auditory cortex by silent speechreading in the absence of scanner noise: An event-related fMRI study. *NeuroReport*, 11 (8), 1729-1733.
- Maeda, S. (1989). 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*, Academic Publishers, Kluwer, 131-149.
- Maestriperi, D. (2005). Gestural communication in three species of macaques (*Macaca Mulatta*, *M. nemestrina*, *M. arctoides*): Use of signals in relation to dominance and social context. *Gesture*, 5, 57-73.
- Matsuzawa T. (1994). Field experiments on use of stone tools by chimpanzees in the wild. In R.W. Wrangham, W.C. McGrew, F.B.M. De Waal & P.G. Heltne (eds.), *Chimpanzee Cultures*, Cambridge, Harvard Univ. Press, 351-370.
- McAllister, R. (1978). Temporal asymmetry in labial co-articulation. *Papers from the Institute of Linguistics*, Stockholm University, 35, 1-29.
- McAllister, R., & Engstrand, O. (1992). Interpretations of tongue movement patterns in VCV sequences. In *Proceedings of Fonetik 92: Papers from 6th Swedish Phonetical Conference*, Gothenburg, 115-119.
- Massaro, D.W. (1998). *Perceiving talking faces: Insights into Auditory Attention*. Cambridge, MA: MIT Press.
- Mehler, J., & Dupoux, E., (1990). *Naître humain*. Editions Odile Jacob.
- Meltzoff, A.N. (2005). Imitation and other minds: The “Like me” hypothesis. In S. Hurley & N. Chater (eds), *Perspectives on imitation: from neurosciences to social science*, 2, Cambridge, MA: MIT Press, 55-77.
- Meltzoff, A.N. (2007). “Like me”: A foundation for social cognition. *Developmental Science*, 10(1), 126-134.
- Meltzoff, A.N., & Moore, M.K. (1977). Imitation of facial and manual gestures by human neonates. *Science*, 18, 75-78.
- Meltzoff, A.N., & Borton, R.W. (1979). Intermodal matching by human neonates. *Nature*, 282, 403-404.
- Meltzoff, A.N., & 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., & Moore, M.K. (2000). Imitation of facial and manual gestures by human neonates. In D. Muir & A. Slater (eds.), *Infant development, the essential readings*, Blackwell Pub. Ltd., 165-181.
- Ménard, L., & Boë, L.-J. (2001). Perceptual categorization of maximal vowel spaces from birth to adulthood simulated by an articulatory model. In *Proceedings of Eurospeech 2001*, Aalborg Danemark, 167-170.
- Ménard, L. (2002). *Production et perception des voyelles au cours de la croissance du conduit vocal : Variabilité, invariance et normalisation*. Thèse en Sciences du langage, Institut de la Communication Parlée, 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., Cathiard, M.-A., & Savariaux C. (2004). Le développement de la coarticulation labiale anticipante en français : Une étude préliminaire. In *Actes des 25èmes Journées d'Etude sur la Parole*, Fès, Maroc, 373-376.
- Ménard, L., Schwartz, J.-L., Boë, L.-J., & Aubin, J. (2007). Production-perception relationships during vocal tract growth for French vowels: Analysis of real data and simulations with an articulatory model. *Journal of Phonetics*, 35 (1), 1-19.
- Menzerath, P., & A. de Lacerda (1933). *Koartikulation, Steuerung und Lautabgrenzung*. Berlin, Bonn: Ferd. Dümmlers Verlag.
- Mrayati, M., Carré, R., & Guérin, B. (1988). Distinctive region and modes: A new theory of speech production. *Speech Communication*, 7, 257-286.
- Munhall, K.G., & Jones, J.A. (1998). Articulatory evidence for syllabic structure, *Behavioral and Brain Sciences*, 21(4), 1-8.
- Nadel, J., & Decety, J. (2002). *Imiter pour découvrir l'humain. Psychologie, neurobiologie, robotique et philosophie de l'esprit*. Puf, Psychologie et sciences de la pensée.
- Nadel, J., & Potier, C. (2002). Imiter et être imité dans le développement de l'intentionnalité. In J. Nadel & J. Decety (eds.), *Imiter pour découvrir l'humain*. Psychologie, neurobiologie, robotique et philosophie de l'esprit. Puf, Psychologie et sciences de la pensée, 84-104.
- Nguyen, N., Fagyal, Z., & Boula de Mareüil, P. (2002). Art des mets, art d'aimer: A preliminary study on the acoustic aspects of vowel harmony in French. In *Proceedings of the 4th International Meeting of the GDR Phonology*, Grenoble, France.
- Nguyen, N., Fagyal, Z., & Cole, J. (2004). Perceptual relevance of long-domain phonetic dependencies. *Actes des Journées d'Etudes Linguistiques*, Nantes, 6p.
- Nguyen, N., & Fagyal, Z. (2006). Acoustic aspects of vowel harmony in French. *Journal of Phonetics*, accepté.

- Noiray, A., Ménard, L., Cathiard, M-A., Savariaux, C., & Abry, C. (2004). The development of anticipatory labial coarticulation in French: A pioneering study. In *Proceedings of INTERSPEECH, ICSLP*, Jeju, Corée du sud, 53-56.
- Noiray, A., Ménard, L., Cathiard, M-A., Savariaux, C., & Abry, C. (2006). Extending the Movement Expansion Model (MEM) for rounding from French to English. In *VIIth International Speech production Seminar*, Ubatuba, Brazil, 8p.
- Noiray, A., Ménard, L., Cathiard, M-A., Savariaux C., Aubain, J., & Abry, C. (2006). Emergence of rounding anticipatory control in French. *5th International Conference on Speech Motor Control*, Nijmegen, Holland, 111.
- Nittrouer, S., & Whalen, D. (1989). The perceptual effects of child-adult differences in fricative-vowel coarticulation. *Journal of the Acoustical Society of America*. 86(4), 1266-1276.
- Nittrouer, S., Neely, S.T., & Studdert-Kennedy, M. (1996). How children learn to organize their speech gestures: Further evidence from fricative-vowel syllables. *Journal of Speech Hearing Research*. 39, 379-389.
- Öhman, S.E.G. (1966). Coarticulation in VCV utterances: Spectrographic measurements. *Journal of Acoustical Society of America*, 39, 151-168.
- Öhman, S.E.G. (1967). Numerical model of coarticulation. *Journal of the Acoustical Society of America*, 41, 310-320.
- Oller, D.K. (1978). Infant vocalization and the development of speech. *Allied Health and Behavioral Sciences*, 1, 523-549.
- Oller, D.K (1980). The emergence of speech sounds in infancy. In G. Yeni-Komshian, J.F. Kavanagh & G.A. Ferguson (eds.), *Child Phonology, vol 1. Production*, New York: Academic Press; 93-112.
- Oller, D.K., & MacNeilage, P.F. (1983). Development of speech production: Perspectives from natural and perturbed speech. In P.F. MacNeilage (ed.), *The production of Speech*, New-York, Springer Verlag, 91-108.
- Orliaguet, J.P., Kandel, S., & Boë, L-J. (1997). Visual perception of motor anticipation in cursive handwriting: Influence of spatial and movement information on the prediction of forthcoming letters. *Perception*, 26, 905-912.
- Orliaguet, J.-P. (2004). Perception visuelle du mouvement humain : De l'anticipation motrice à l'anticipation perceptive. In R. Sock & B. Vaxelaire (eds.), *L'anticipation à l'horizon du Présent*, Bruxelles : Mardaga, 127-139.
- Ostry, D.J., Gribble, P.L., & Gracco, V.L. (1996). Coarticulation of jaw movements in speech Production: Is context sensitivity in speech kinematics centrally planned? *Journal of Neuroscience*, 16, 1570-1579.
- Perkell, J. (1974). A physiologically oriented model of tongue activity in speech production. PhD dissertation, Massachusetts Institute of Technology, Cambridge, MA.

- Perkell, J.S. (1986). Coarticulation strategies: Preliminary implications of a detailed analysis of lower lip protrusion movements. *Speech Communication*, 5, 47-68.
- Perkell, J.S. (1990). Testing theories of speech production: Implications of some detailed analyses of variable articulatory data. In W.J. Hardcastle & A. Marchal (eds.), *Speech production and speech modeling*, Kluwer Academic Publishers, 263-288.
- Perkell, J.S., & Chiang, C. (1986). Preliminary support for a "hybrid model" of anticipatory coarticulation. In *Proceedings of the 12th International Conference of Acoustics*, Toronto, Canada, A3-6.
- Perkell, J.S., & Matthies, M.L. (1992). Temporal measures of anticipatory labial coarticulation for the vowel /u/ : Within – and – cross-subject variability. *Journal of the Acoustical Society of America*, 87 (Suppl. 1), S123, 2911-2925.
- Perrett, D.I. (1989). Frameworks of analysis for the neural representation of animate objects and actions. *J. Exp. Biol.*, 146, 87-113.
- Perrier, P. (2004). Modéliser le physique pour comprendre le contrôle : Le cas de l'anticipation en production de parole. R. Sock & B. Vaxelaire (eds.). *L'anticipation à l'horizon du Présent*, Bruxelles : Mardaga, 159-177.
- Perrier, P., Ostry, D.J., & Laboissière, R. (1996). The Equilibrium Point Hypothesis and its Application to Speech Motor Control. *Journal of Speech and Hearing Research*, 39(2), 365-377.
- Perrier, P., Laboissière, R., Abry, C., Maeda, S. (eds.) (1997). Special issue on speech production: Models and data, *Speech Communication*, 22, 2-3.
- Philps, D. (2000). Le sens retrouvé ? De la nomination de certaines parties du corps: Le témoignage des marqueurs sub-lexicaux de l'anglais en <CN->. *Anglophonia / Sigma*, 8, 207-232.
- Philps, D. (2003). S- incrémental et régénération submorphémique en anglais. *Bulletin de la Société de Linguistique de Paris*, XCVIII, 1, 163-196.
- Philps, D. (2006). From mouth to hand. In A. Cangelosi, A. Smith & K. Smith (dir.), *The evolution of language*, New Jersey/Londres/Singapour: Word scientific Publishing, 247-254.
- Picq, P. (2003). *Au commencement était l'homme*. De Toumaï à Cro-Magnon, Paris, Odile Jacob.
- Pinker, S., & Jackendoff, R. (2005). The Faculty of Language: What's Special about it? *Cognition*, 95(2), 201-236.
- Plaut, D.C., & Kello, C.T. (1999). The emergence of phonology from the interplay of speech comprehension and production : A distributed Connectionnist approach. In B. MacWhinney (ed.), *The emergence of language*, 381-416, L.E.A. Publishers, London.

- Poremba, A., Malloy, M.M., Saunders, R.C., Carson, R.E., Herscovitch, P., & Mishkin, M. (2004). Species-specific calls evoke asymmetric activity in the monkey's temporal poles. *Nature*, 427, 448-451.
- Ramsay, J.O, Munhall, K.G., Gracco, V.L., & Ostry, D.L. (1996). Functional data analyses of lip motion. *Journal of the Acoustical Society of America*, 99(6), 3718-3727.
- Recasens, D. (1989). Long range coarticulatory effects for tongue dorsum contact in VCVCV sequences. *Speech Communication*, 8/4, 293-307.
- Repp, B. (1986). Some observations on the development of anticipatory coarticulation. *Journal of Acoustical Society of America*, 79(5), 1616-1619.
- Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, 3, 131-141.
- Rizzolatti, G., & Arbib, M.A. (1998). Language Within Our Grasp. *Trends in Neurosciences*, 21(5), 188-194.
- Rizzolatti G., Fogassi L., & Gallese V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience*, 2, 661-670.
- Roberts, M.E.L. (1972). A cinephotographic study of coarticulation of lip rounding in English and French. M. Sc dissertation, University of British Columbia.
- Rousselot, P.-J. (1897-1901). *Principes de phonétique expérimentale*. I-II. Paris : H. Welter.
- Rubin, P., Saltzman, E., Goldstein, L., McGowan, R., Tiede, M., & Browman, C. (1996). CASY and Extensions to the Task-Dynamic Model. In *Proceedings of the 1st ETRW on Speech Production Modeling: From Control Strategies to Acoustics & 4th Speech Production Seminar: Models and Data*, Autrans, France, 125-128.
- Rumelhart, D.E., & Norman, D.A, (1982). Simulating a skilled typist: A study of skilled cognitive-motor performance. *Cognitive Science*, 6, 1-36.
- Rvachew, S., Mattock, K., Polka, L., & Ménard, L. (2006). Developmental and cross-linguistic variation in the infant vowel space. *Journal of the Acoustical Society of America*, 120 (4), 2250-2259.
- Saltzman E. (1986). Task dynamic coordination of the speech articulators: A preliminary model. *Experimental Brain Research Series*, 15, 129-144.
- Saltzman, E. (1991). The Task Dynamic model in speech production. In H.F.M. Peters, W. Hulstijn & C.W. Starkweather (eds.), *Speech Motor Control and Stuttering*. Elsevier Science Publishers. *Excerpta Medica*, Amsterdam, 37-52.
- Saltzman, E. (1995). Dynamics and coordinate systems in skilled sensorimotor activity. In R. Port & T. van Gelder (eds.), *Mind as Motion*. Cambridge, MA: MIT Press, 150-173.
- Saltzman, E., & Kelso, J.A.S. (1987). Skilled Actions: A Task-Dynamic Approach. *Psychological Review*, 94(1), 84-106.

- Saltzman, E., & Munhall, K. (1989). A dynamic approach to gestural patterning in speech production. *Ecological Psychology*, 1, 333-382.
- Saltzman, E., & Byrd, D. (2000). Task-dynamics of gestural timing: Phase windows and multifrequency rhythms. *Human Movement Science*, 19, 999-526.
- Sanders, E. (1972). When are speech sounds learned? *Journal of Speech and Hearing Disorders*, 37, 55-63.
- 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., Besière, P., Boë, L.-J., Davis, B.L., & Serkhane, J. (2003). Un androïde doué pour la parole. Rapport Robea.
[http://www-laplace.imag.fr/ROBEA/androide/Robea_Androide.pdf]
- Selby, J.C., Robb, M.P., & Gilbert M.P. (2000). Vowels systems of normally developing and phonologically disordered children. *Clinical Linguistics and Phonetics*, 14, 255-265.
- Serkhane, J. (2005). Un bébé androïde vocalisant: Etude et modélisation des mécanismes d'exploration vocale et d'imitation orofaciale dans le développement de la parole. Thèse en Sciences Cognitives, INP, Grenoble.
- Serkhane, J., Schwartz, J.-L., Boë, L.-J., Davis, B., & Matyear, C. (2002). Motor specifications of a baby robot via the analysis of infants vocalizations. In *Proceedings ICSLP*, 45-48.
- Serkhane J., Schwartz J.-L., Boë L.-J., Davis B., Bessière P., & Mazer E. (2002). Etude comparative de vocalisations de bébés humains et de bébés robots. In *XXIVème Journées d'Etude sur la Parole*, Nancy, France, 1-4.
- Serkhane, J. Schwartz, J.-L., & Bessière, P. (2005). Building a talking baby robot : A contribution to the study of speech acquisition and evolution. *Interaction Studies*, 2(6), 253-256.
- Sharkey, S.G., & Folkins, J.W. (1985). Variability of lip and jaw movements in children and adults: Implications for the development of speech motor control. *Journal of Speech and Hearing Research*, 8-15.
- Sharma S. & Hermansky H. (1999). Recognition of speech from temporal patterns. *The Journal of the Acoustical Society of America*, 105(2), 1158.
- Smeele, P.M.T., & Sittig, A.C. (1991). The contribution of vision to speech perception. In *Proceedings of the 2nd European Conference on Speech Communication and Technology*, 3, Genes, Italie, 1495-1497.
- Smeele, P.M.T., Sittig, A.C., & Van Heuven, V.J. (1992). Intelligibility of audio-visually desynchronised speech: asymmetrical effect of phoneme position. In *Proceedings of the 2nd International Conference of Spoken Language Processing*, Alberta, Canada, 65-68.

- Smith, A. (2006). Speech motor development: Integrating muscles, movements, and linguistic units. *Journal of Communication Disorders*, 39(5), 331-349.
- Stein, B.E., & Meredith, M.A. (1993). *The merging of the senses*. Cambridge, MA: The MIT Press.
- Stoel-Gammon, C., & Herrington, P. (1990). Vowel systems of normally developing and phonologically disordered children. *Clinical Linguistics and Phonetics*, 4, 145-160.
- Streier, A., Gentaz, E., Spelke, E.S., & Van de Walle, G. (2004). Haptic perception of object unity in rotating displays. *The Quarterly Journal of Experimental Psychology*, 57, 523-538.
- Studdert-Kennedy, M. (2005). How did language go discrete? In M. Tallerman (ed.), *Language origins, perspectives on evolution*, Oxford University Press, 48-67.
- Studdert-Kennedy, M., & Goldstein L. (2003). Launching language: The gestural origin of discrete infinity. In M.H. Christiansen & S. Kirby (eds.), *Language evolution*, Oxford, Oxford University Press, 235-254.
- Summerfield, Q. (1992). Lipreading and audio-visual speech perception. In V. Bruce, A. Cowey, A.W. Ellis & D.I. Perrett (eds.), *Processing the facial image*, Proceedings of a Royal Society Discussion Meeting, Clarendon Press, Oxford, 71-78.
- Sussman, M., Westbury, J.R. (1981). The effects of antagonistic gestures on temporal and amplitude parameters of anticipatory labial coarticulation. *American Speech-Language-Hearing Association*, 24(1), 16-24.
- 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 and Hearing Research*, 42, 1080-1096.
- Sutton, D., Larson, C., and Lindeman, R. C. (1974). Neocortical and limbic lesion effects on primate phonation. *Brain Research*, 71, 61-75.
- Tingley, B.M., & Allen, G.D. (1975). Development of speech timing control in children. *Child Development*, 46, 186-194.
- Toda, M., Maeda, S., Carlen, A.J., & Meftahi, L. (2003). Lip protrusion/rounding dissociation in French and English consonants: /w/ vs. /s/ and /z/. In *Proceedings of the 15-th International Congress of Phonetic Sciences*, Barcelone, Espagne, 1763-1766.
- Trevarthen, C. (1979). Communication and cooperation in early infancy: A description of primary intersubjectivity. In M. Bullowa (ed.), *Before speech: The beginning of interpersonal communication*, London: Cambridge University Press, 321-347.
- Troille, E., Cathiard, M.-A., & Abry, C. (accepté). Consequences on bimodal perception of the timing of the consonant and vowel audiovisual flows. *International Conference on Auditory-Visual Speech Processing 2007*, Kasteel Groenendael, Hilvarenbeek, Pays-Bas.
- Tuller, B., Harris, K.S., & Gross, B. (1981). Electromyographic study of jaw muscles during speech. *Journal of Phonetics*, 9, 175-188.

- Tuller, B., & Kelso, J.A.S. (1984). The timing of articulatory gestures: Evidence for relational invariants. *Journal of the Acoustical Society of America*, 76, 1030-1036.
- Turvey, M. (1990). Coordination. *American Psychologist*, 45, 938-953.
- Van Ijzendoorn, M.H., Dijkstra, J.Y., & Bus, A.G. (1995). Attachment, intelligence, and language: A meta-analysis. *Social Development*, 4, 115-128.
- Van Wassenhove, V., Grant, K.W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *PNAS*, 102(4), 1181-1186.
- Vihman, M.M. (1996). *Phonological development*. The origin of language in the child. Blackwell Publishers Ltd, UK.
- Vihman, M.M. (1998). Later phonological development. In J.E. Bernthal & N.W. Bankson (eds.), *Articulation and phonological disorders*, Hardcover, 113-147.
- Vilain, A., Abry, C., Badin, P., & Brosda, S. (1999). From idiosyncratic pure frames to variegated babbling: Evidence from articulatory modelling, In *Proceedings of the International Conference of Phonetic Sciences'99*, 3, San Francisco, USA, 2497-2500.
- 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 Sciences du Langage, Université Stendhal, Grenoble.
- Wallace, M.T., Wilkinson, L.K., & Stein, B.E. (1996). Representation and integration of multiple sensory inputs in primate superior colliculus. *J. Neurophysiol.*, 76(2), 1246-1266.
- Wellman, B.L., Case, I.M., Mengert, J.B., & Bradbury, D.E. (1931). Speech sounds of young children. *University of Iowa Studies in Child Welfare*, 2(5).
- West, P. (2000). Long distance coarticulatory effects of British English /l/ and /r/: and EMA, EPG and acoustic study. In *Proceedings of the 5th Seminar on Speech Production: Model and Data*, Kloster Seeon, Allemagne, 105-108.
- Whalen, D. (1990). Coarticulation is largely planned. *Journal of Phonetics*, 18, 3-35.
- Whiten, A., Goodall, J, McGew, W.C., Nishida, T., Reynolds, V., Sugiyama, Y., Tutin, C.E.G., Wrangham, R.W., & Boesch, C. (1999). Cultures in chimpanzees. *Nature*, 399, 682-685.
- Whiten, A., Horner, V., & de Waal, F.B.M. (2005). Conformity to cultural norms of tool use in chimpanzees. *Nature*, 437(29), 737-745.
- Wright, T.M., Pelphey, K.A., Allison, T., McKeown, M.J., & McCarthy, G. (2003). Polysensory interactions along lateral temporal regions evoked by audiovisual speech. *Cerebral Cortex*, 13, 1034-1043.

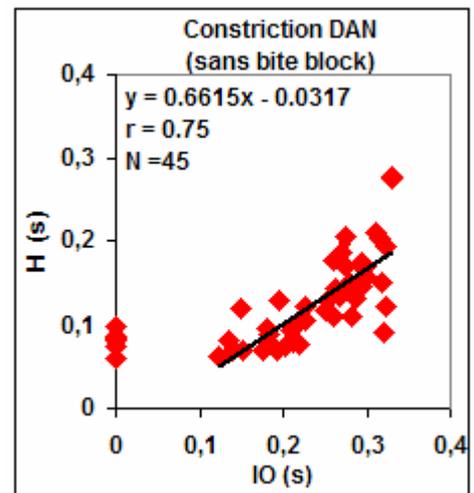
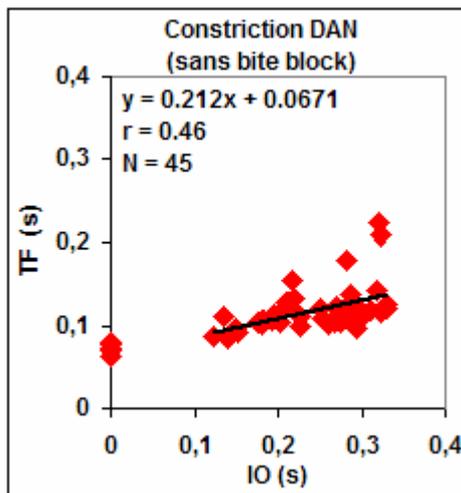
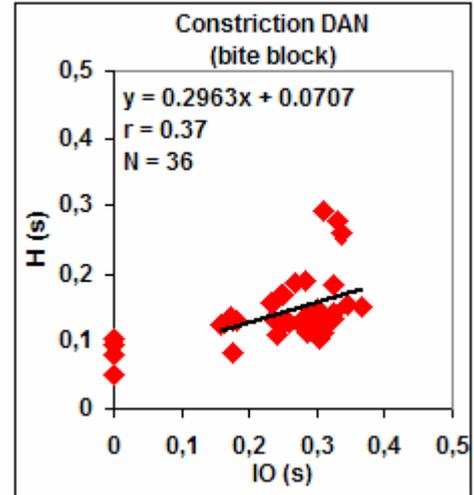
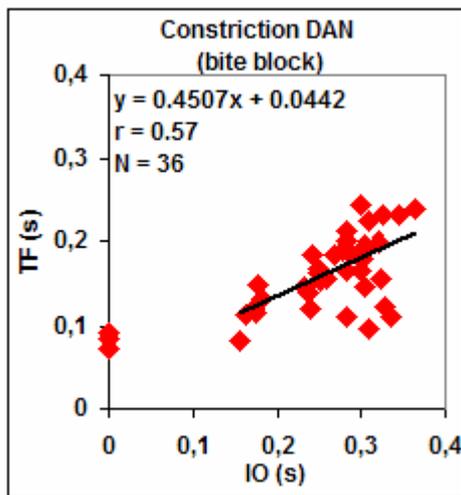
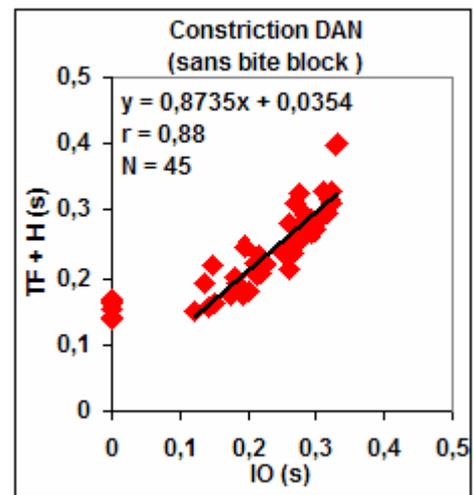
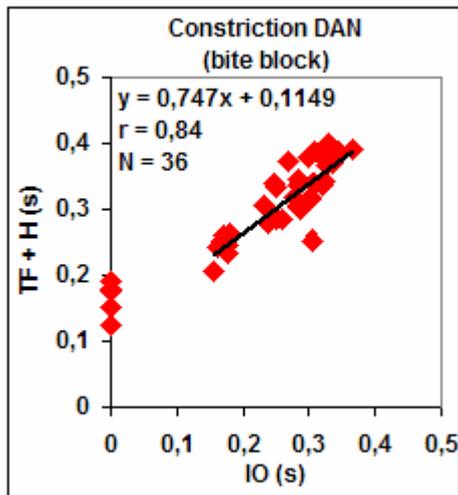
ANNEXES

ANNEXE I.

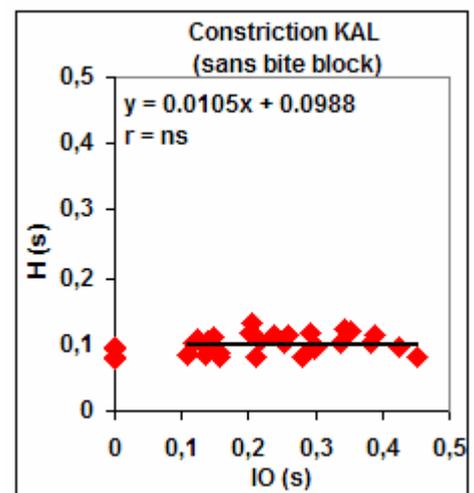
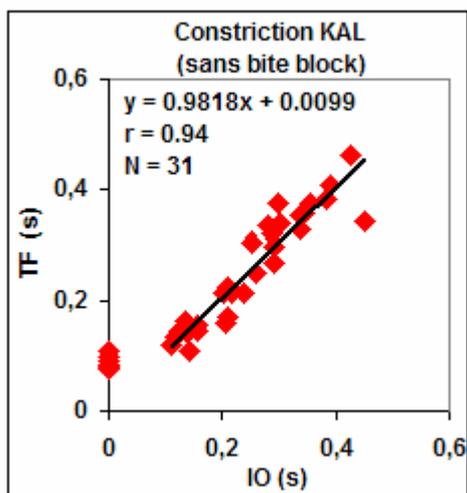
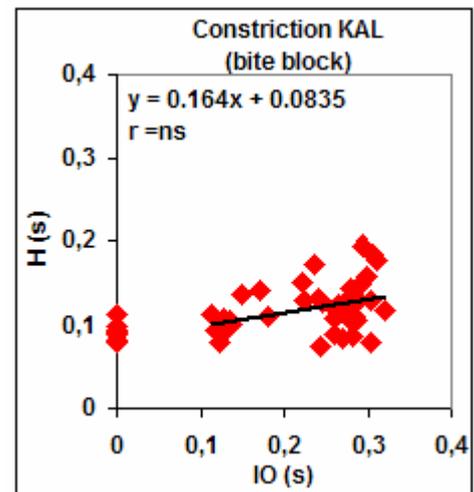
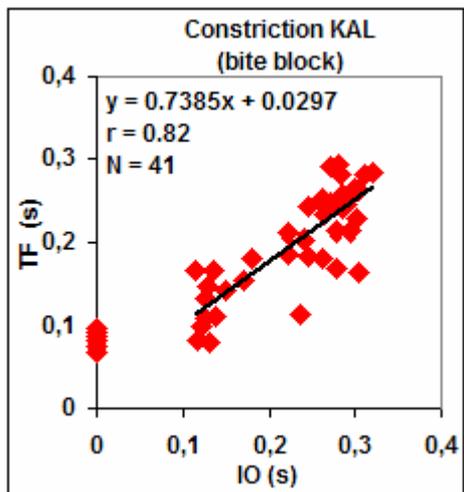
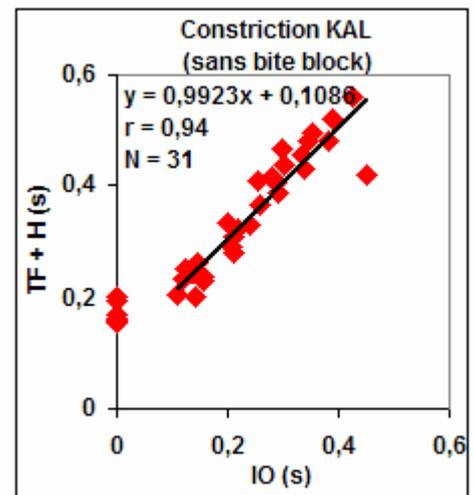
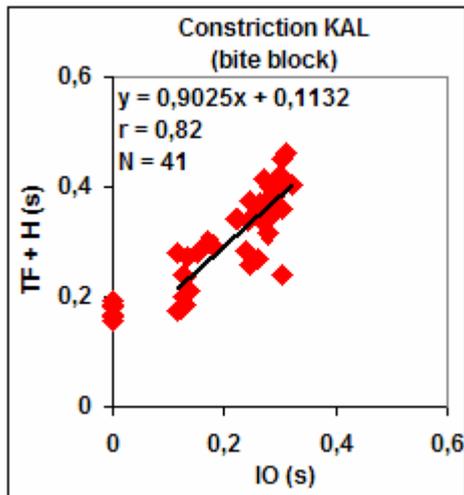
Relation entre les phases *Time Falling + Hold*, *Time Falling*, *Hold* et l'*Intervalle d'Obstruence* (en secondes) pour les quatre locuteurs anglo-américains

- Etude de la constriction -

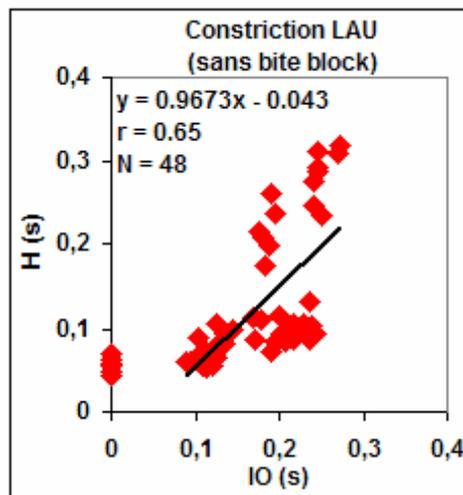
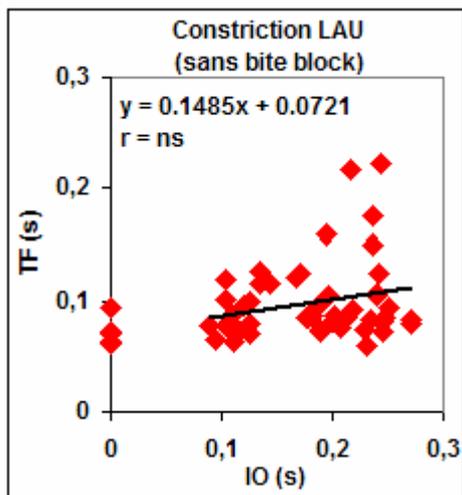
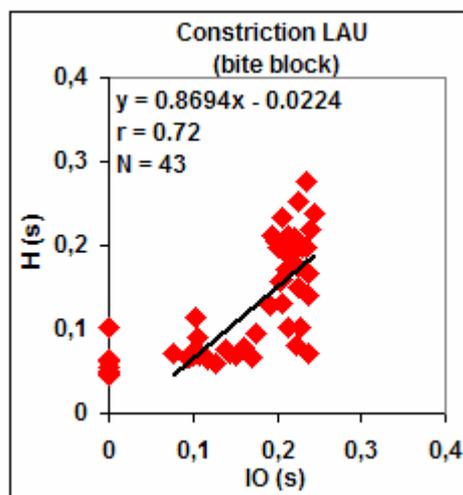
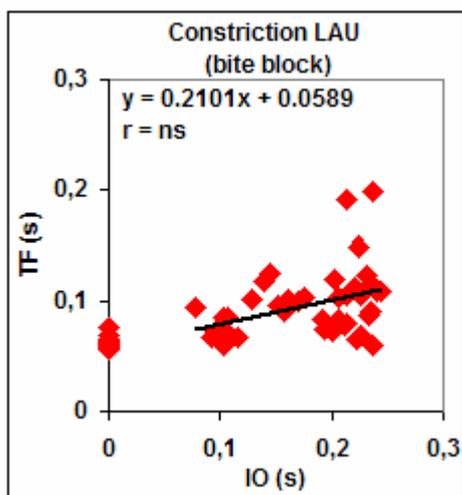
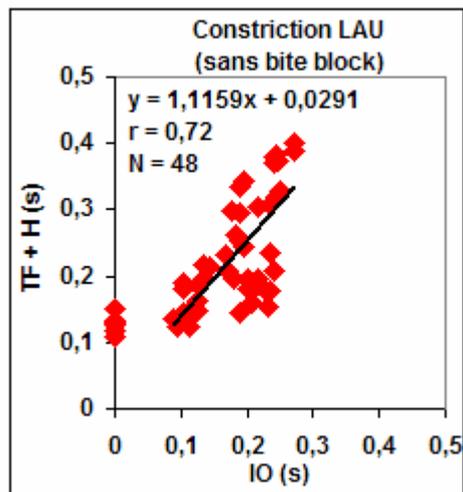
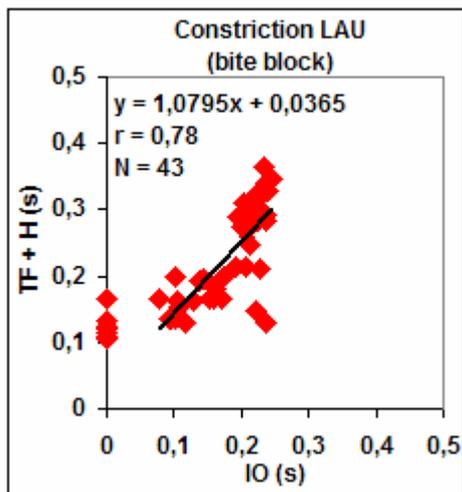
1. Locuteur DAN



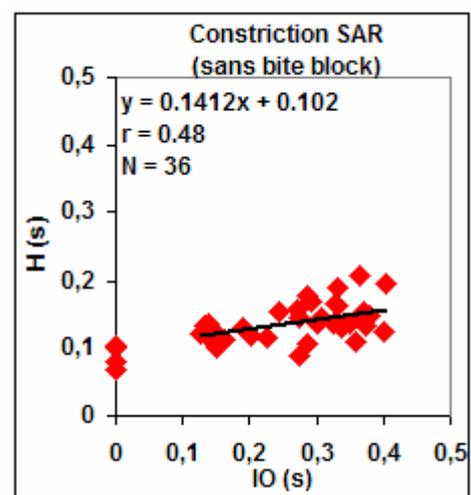
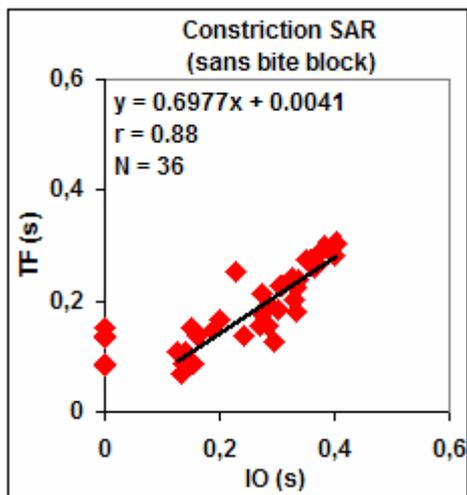
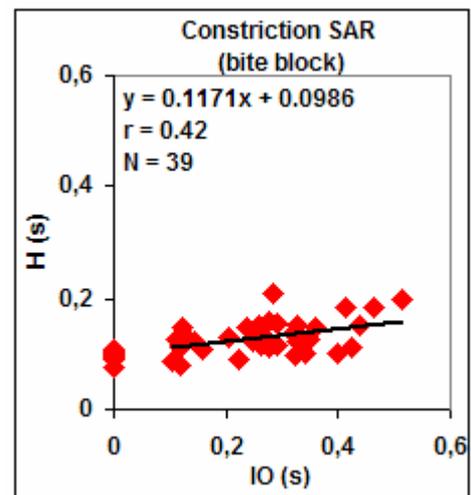
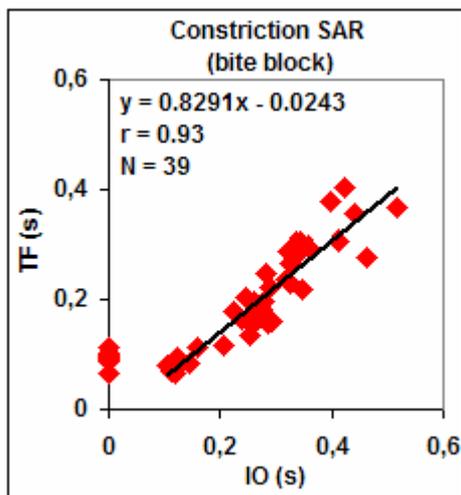
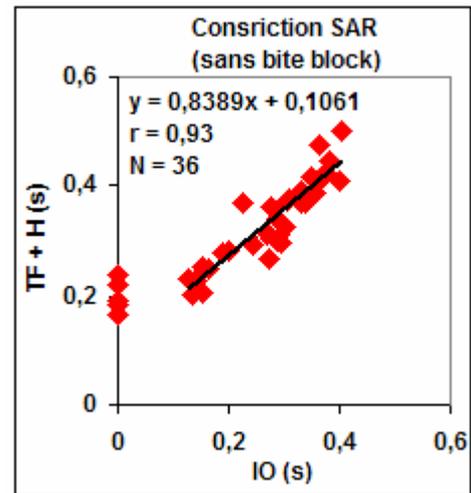
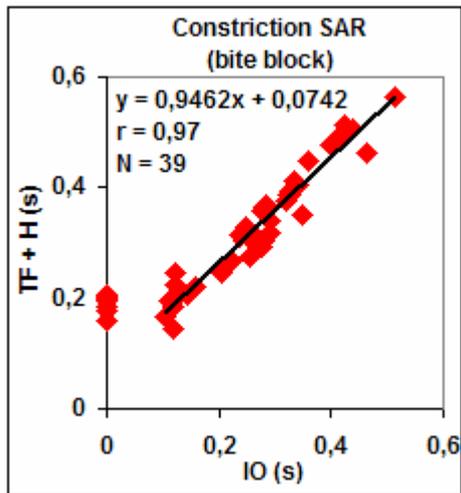
2. Locuteur KAL



3. Locuteur LAU



4. Locuteur SAR

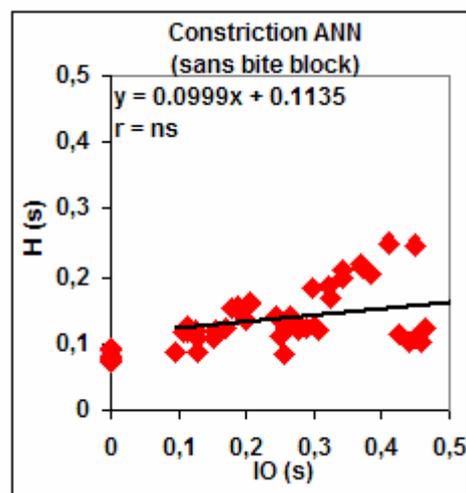
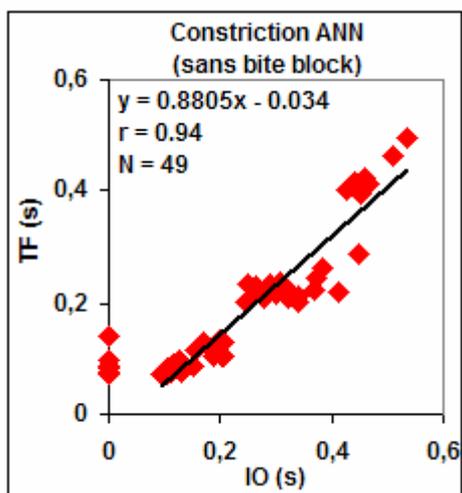
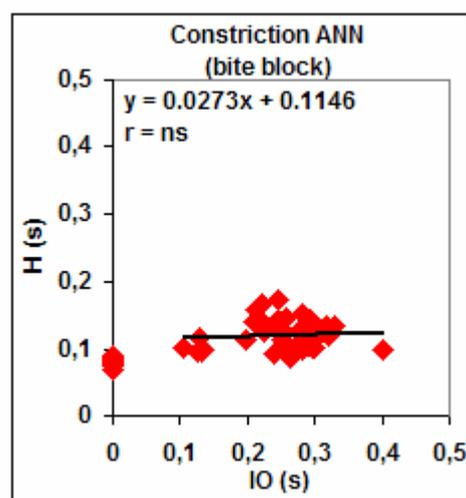
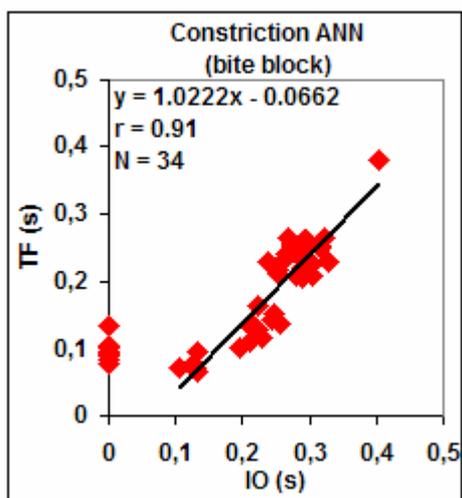
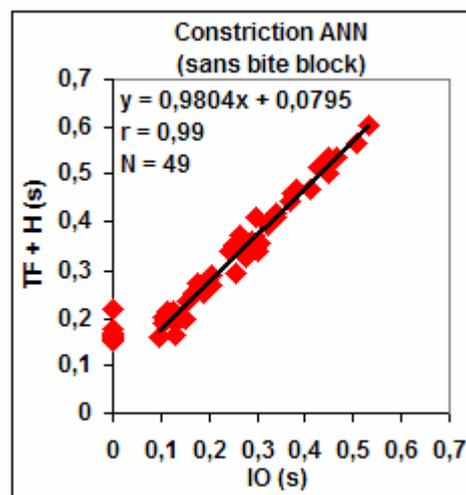
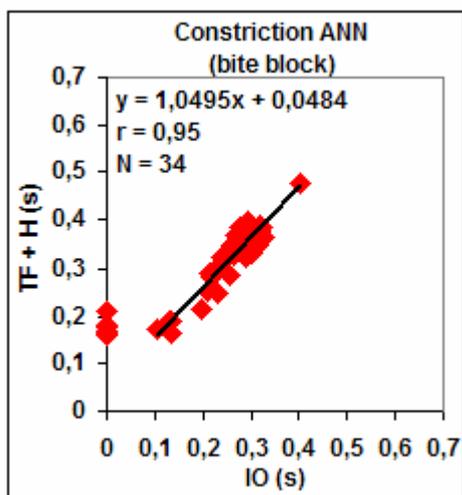


ANNEXE II

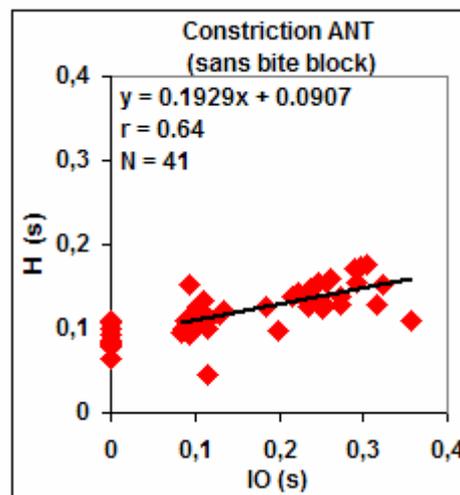
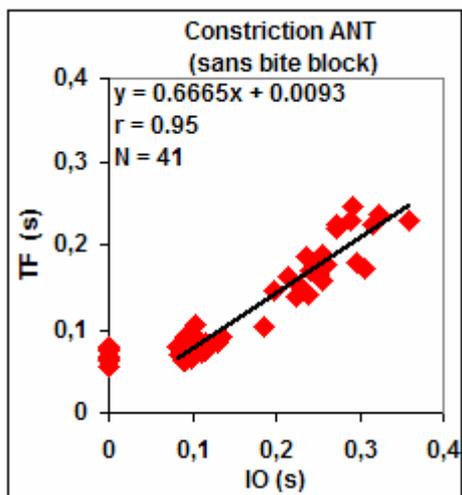
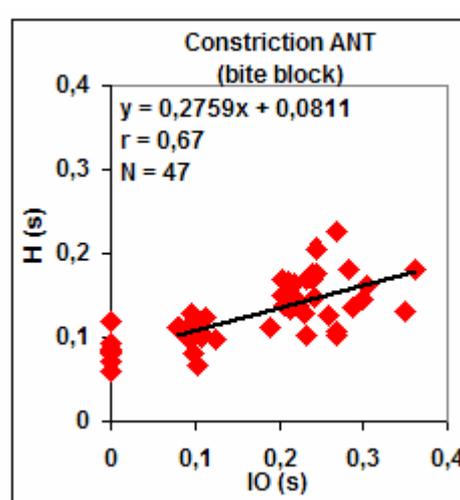
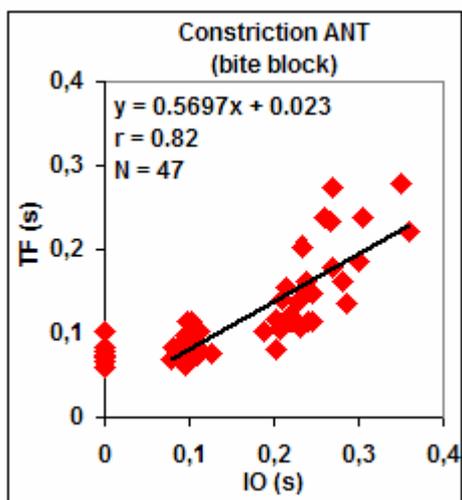
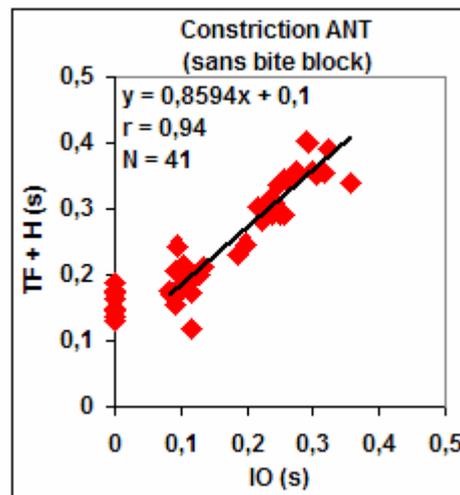
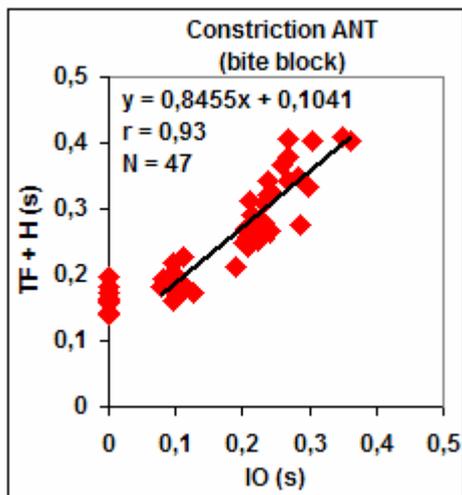
Relation entre les phases *Time Falling+ Hold*, *Time Falling*, *Hold* et l'*Intervalle d'Obstruence* (en secondes) pour les quatre locuteurs franco québécois

- Etude de la constriction -

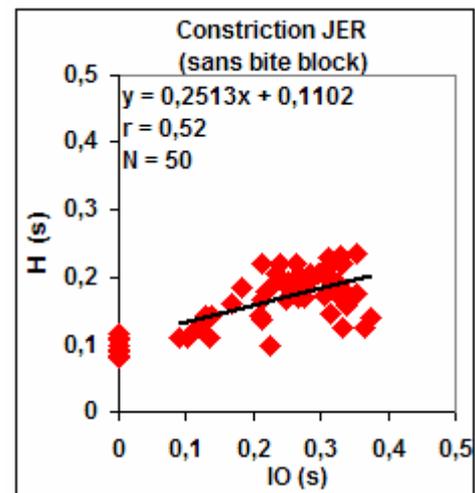
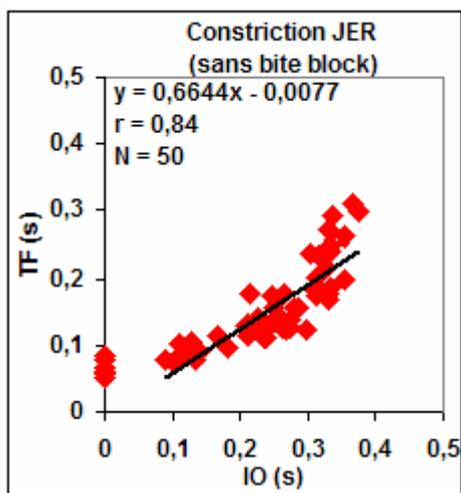
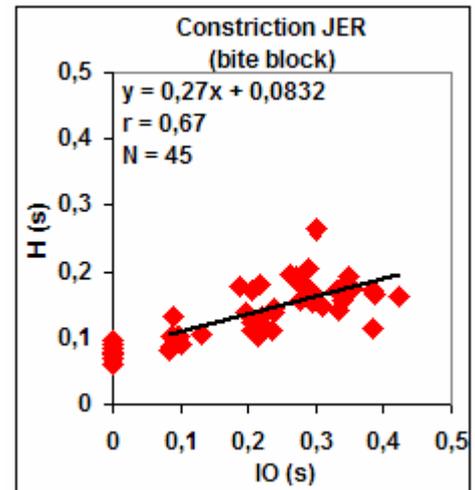
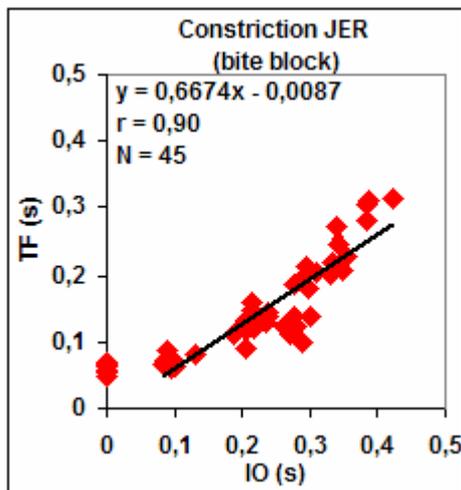
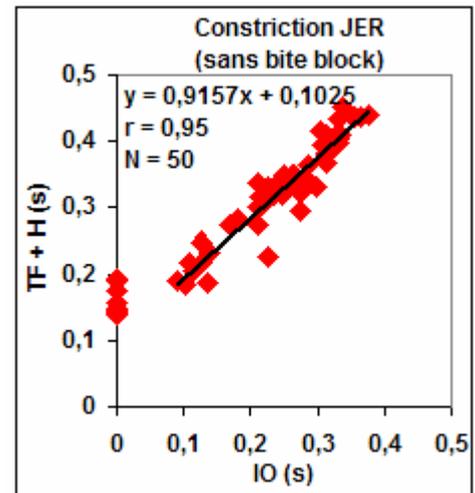
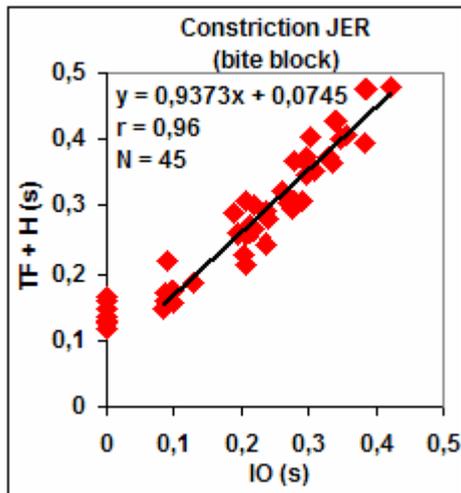
1. Locuteur ANN



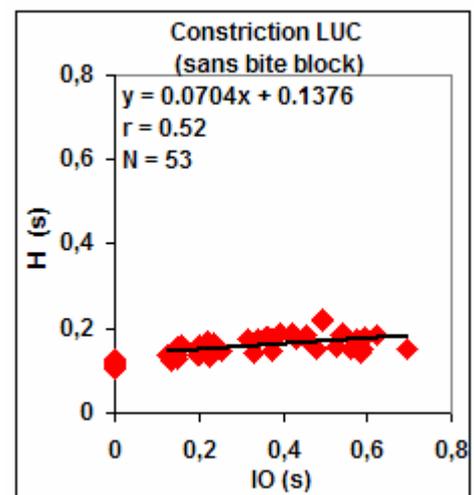
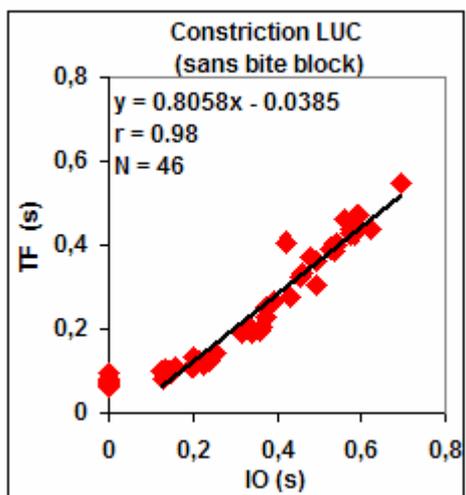
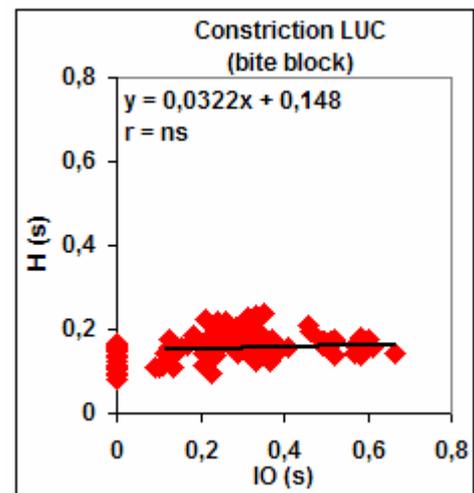
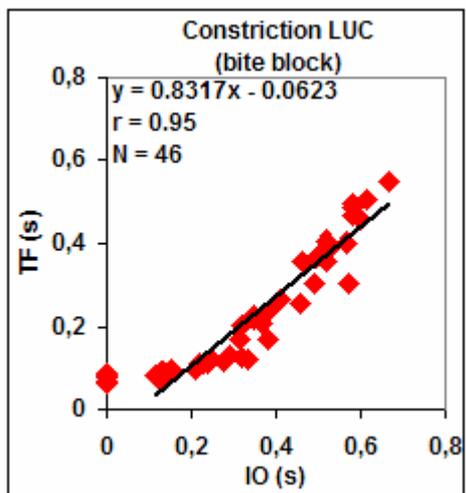
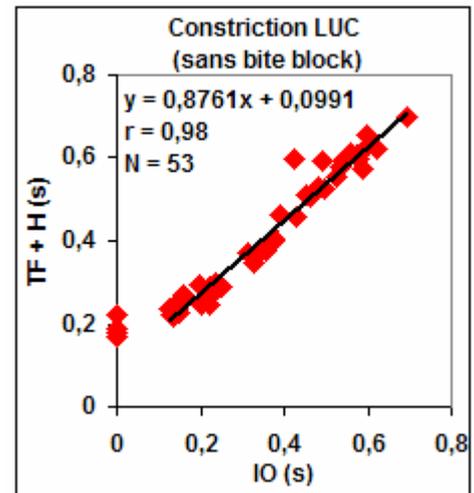
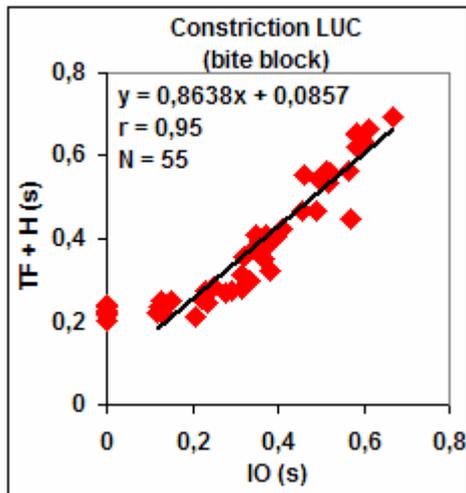
2. Locuteur ANT



3 .Locuteur JER



4. Locuteur LUC

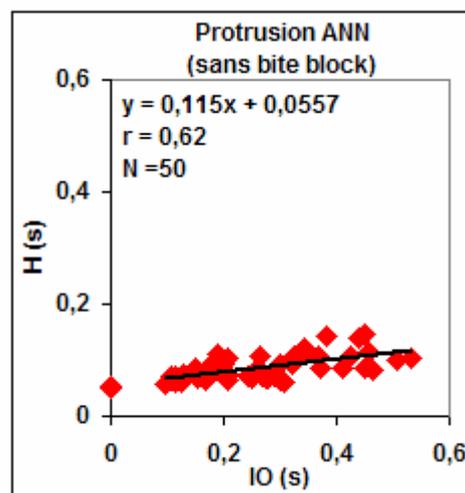
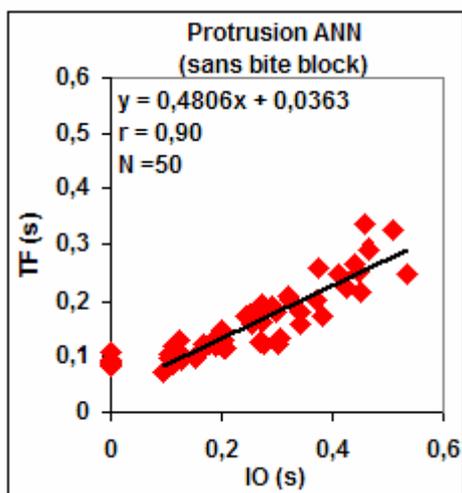
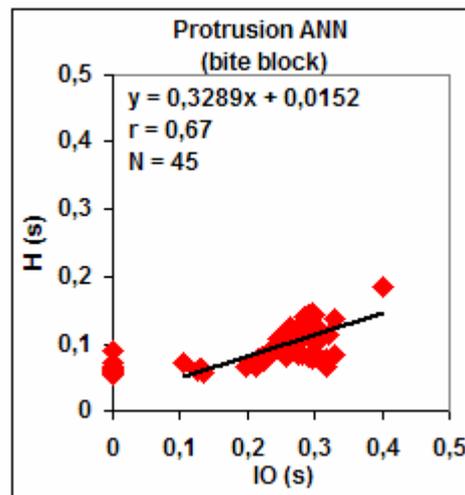
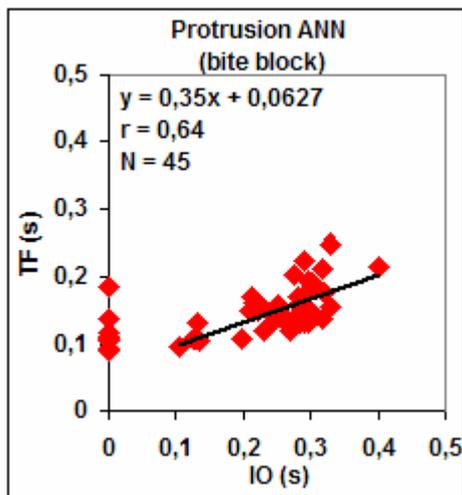
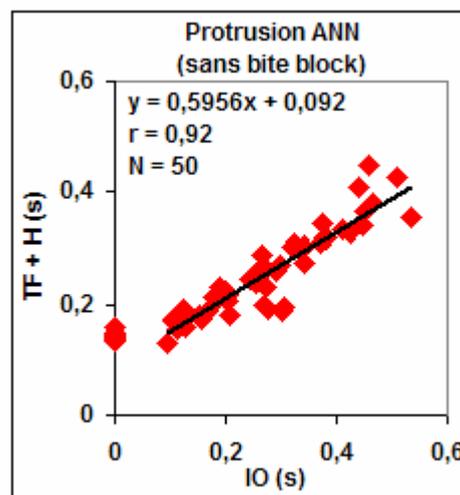
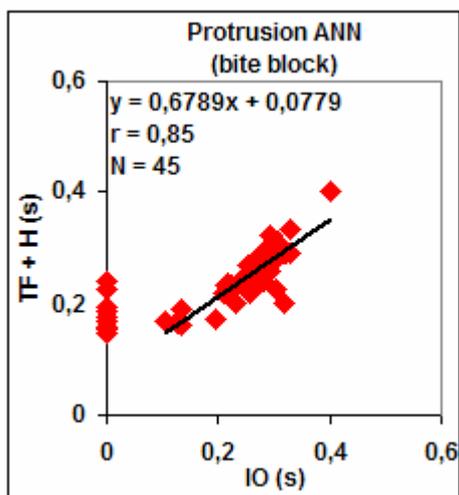


ANNEXE III

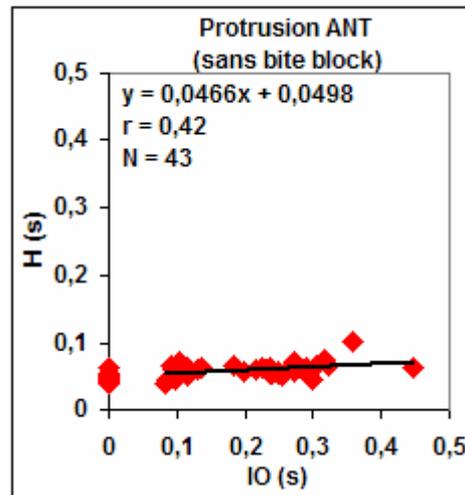
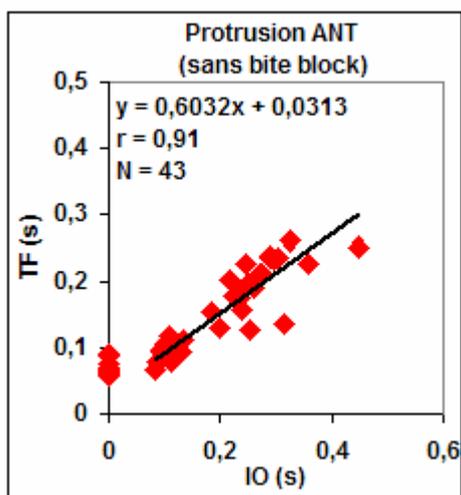
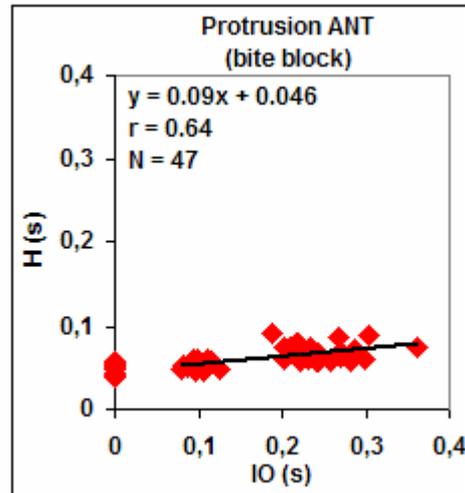
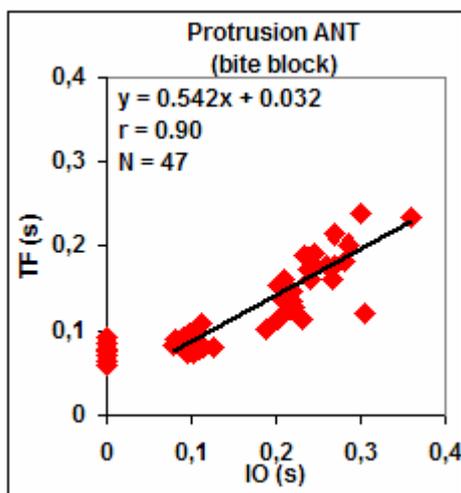
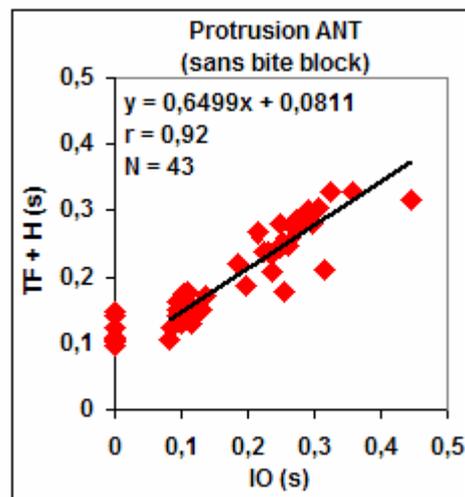
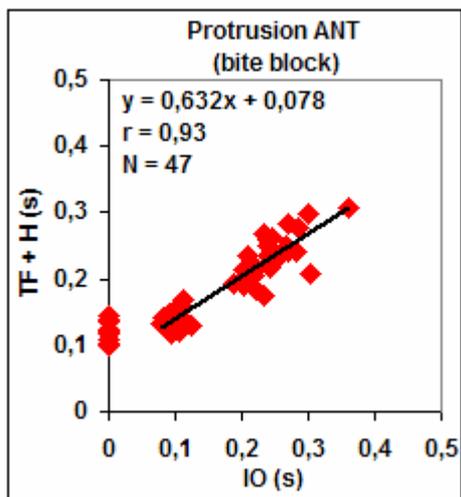
Relation entre les phases *Time Falling+ Hold*, *Time Falling*, *Hold* et l'*Intervalle d'Obstruence* (en secondes) pour les quatre locuteurs franco québécois

- Etude de la protrusion -

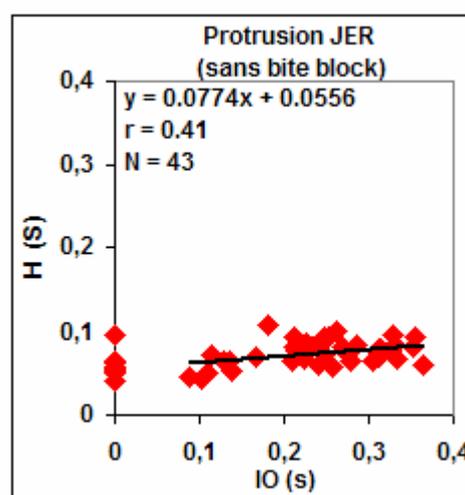
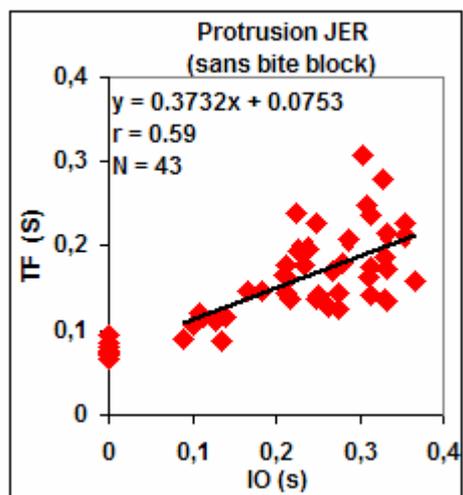
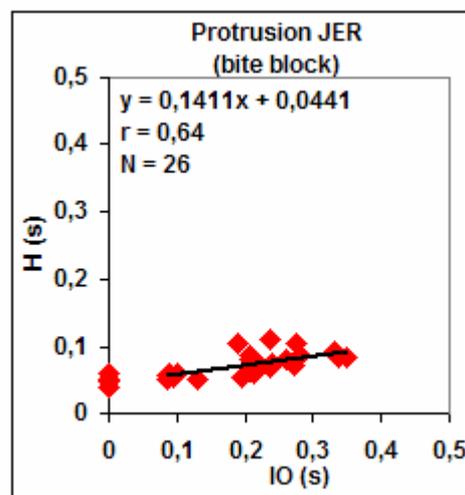
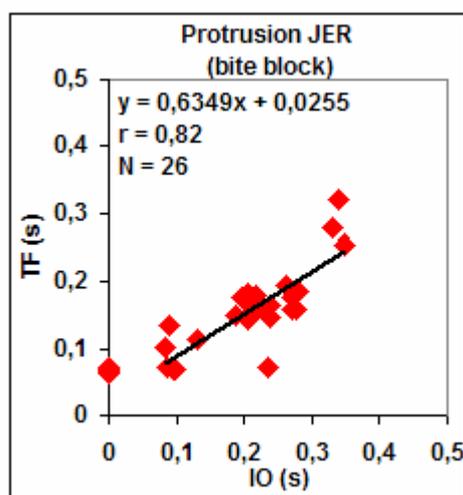
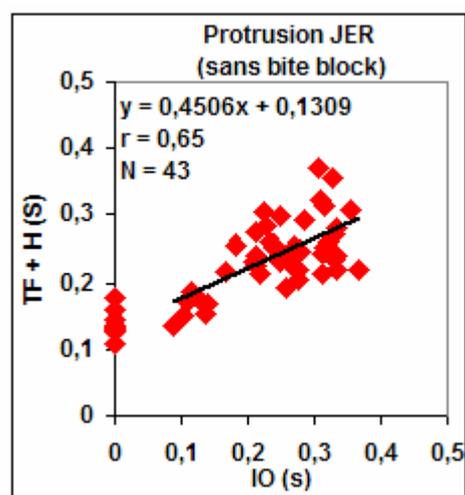
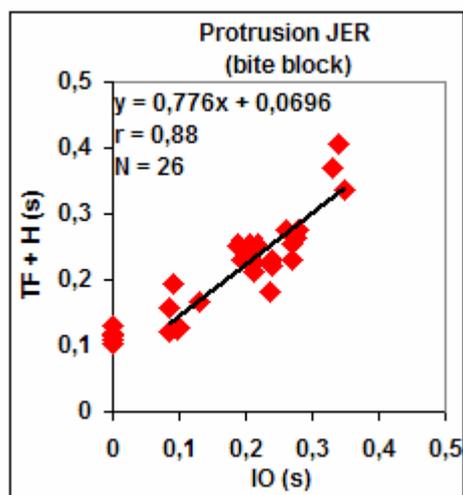
1. Locuteur ANN



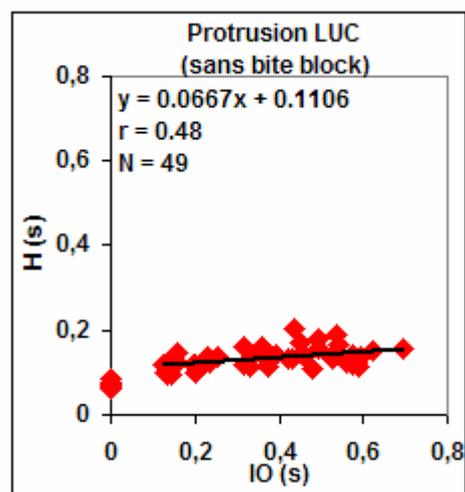
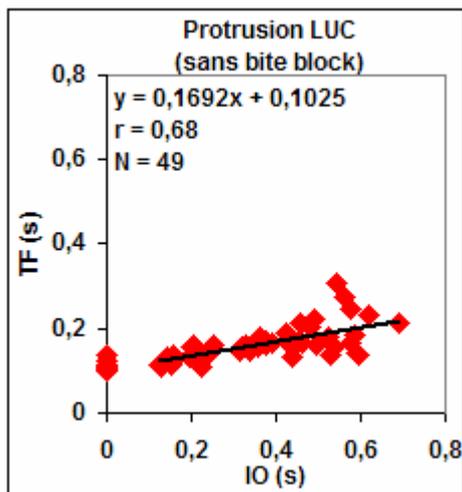
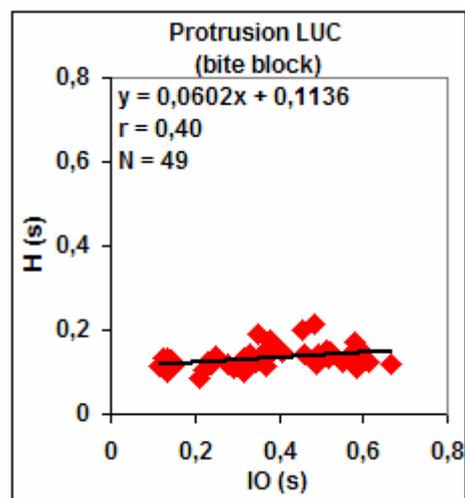
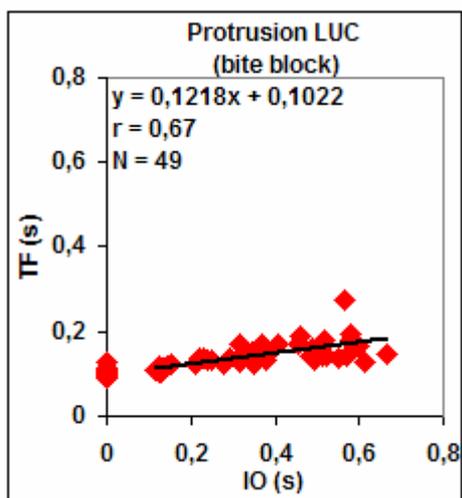
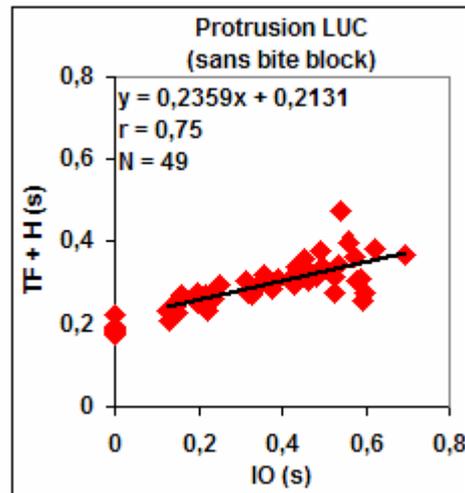
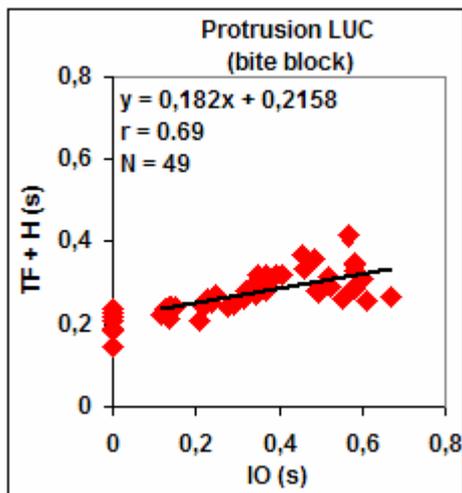
2. Locuteur ANT



3. Locuteur JER



4. Locuteur LUC

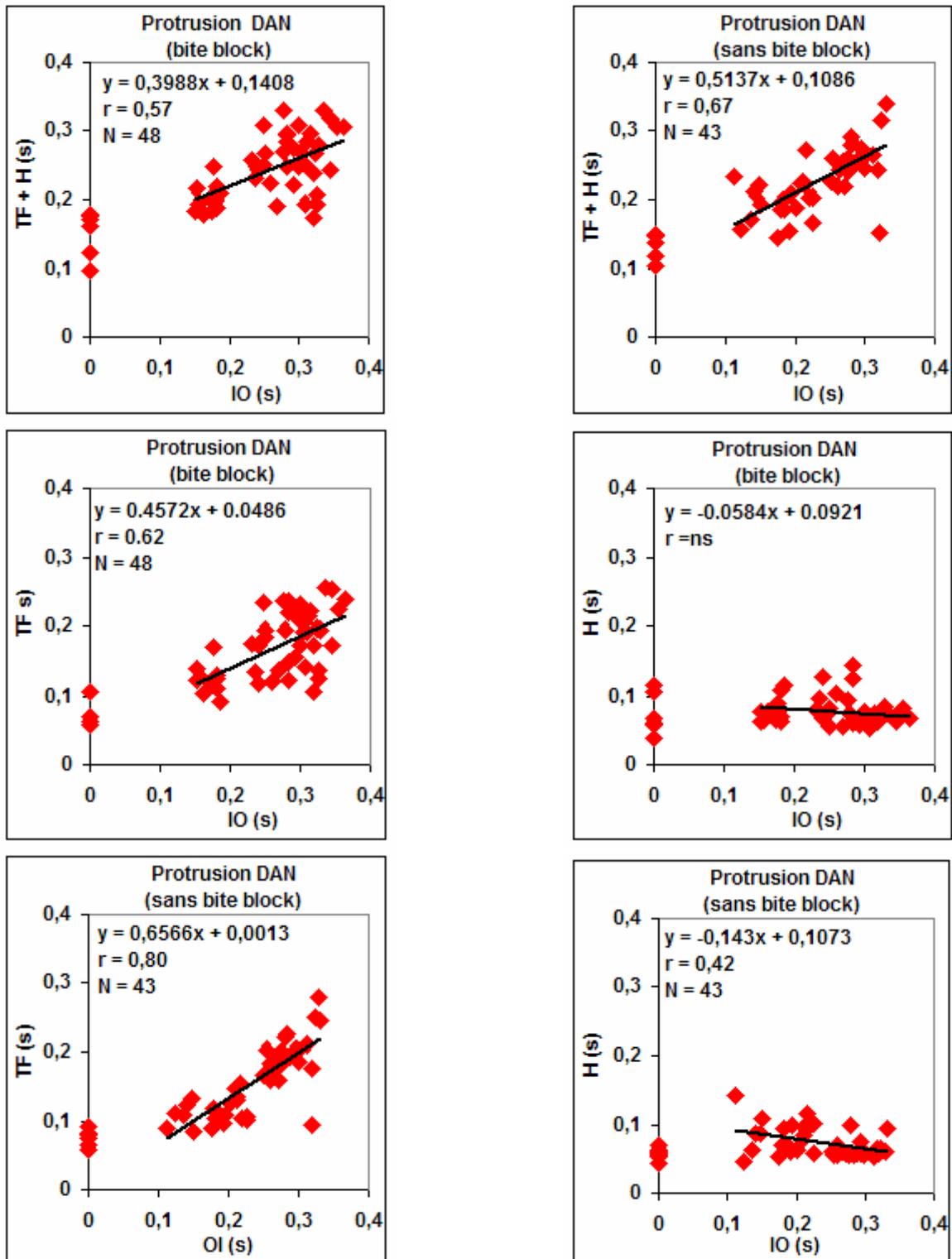


ANNEXE IV

Relation entre les phases *Time Falling + Hold*, *Time Falling*, *Hold* et l'*Intervalle d'Obstruence* (en secondes) pour un locuteur anglo-américain

- Etude de la protrusion -

1. Locuteur DAN

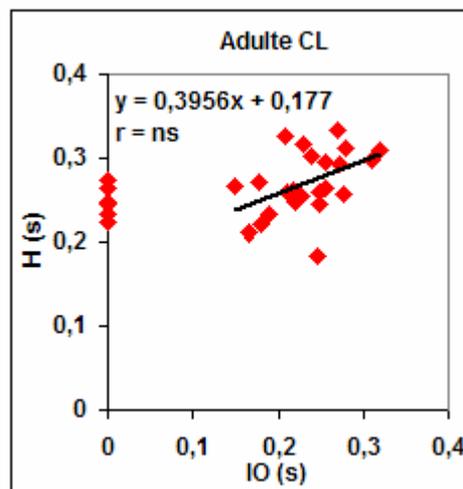
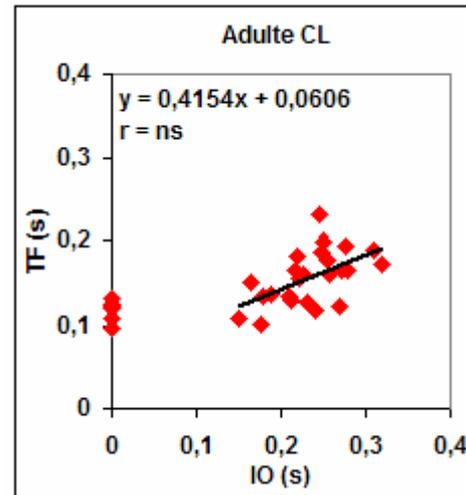
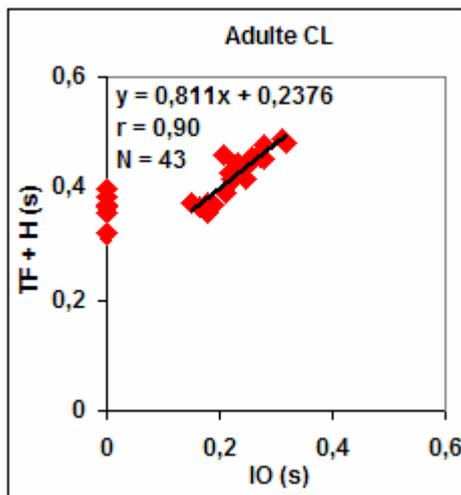


ANNEXE V

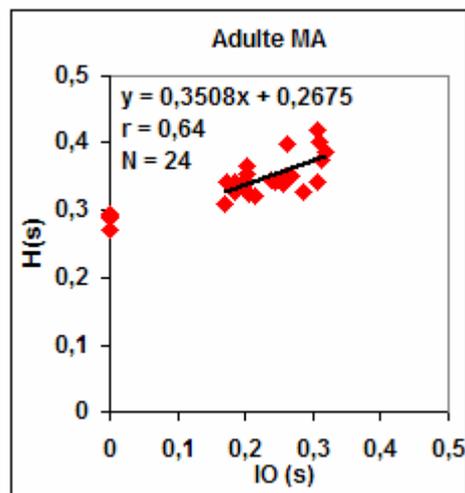
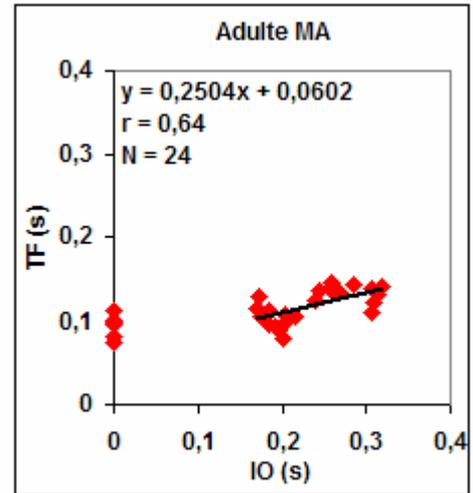
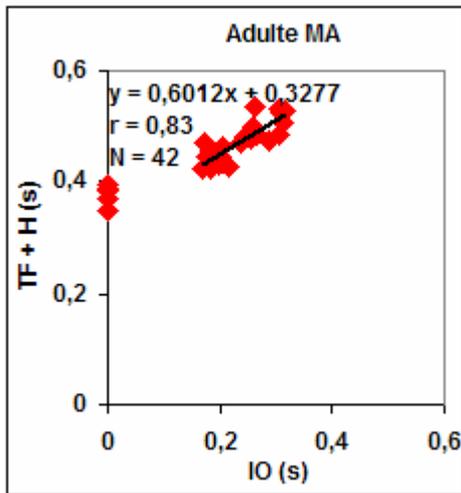
Relation entre les phases *Time Falling + Hold*, *Time Falling*, *Hold*
et l'*Intervalle d'Obstruence* (en secondes)
pour les deux adultes français de référence

- Etude de la constriction -

1. Locuteur CL



2. Locuteur MA



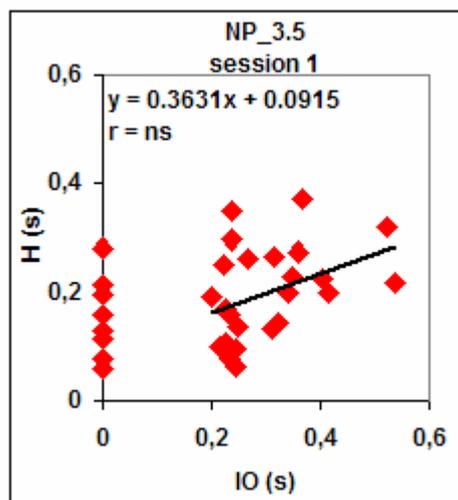
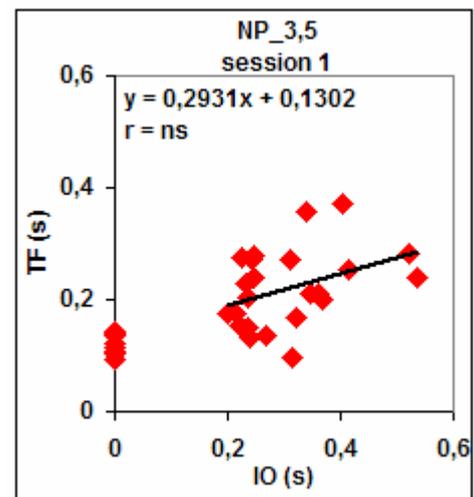
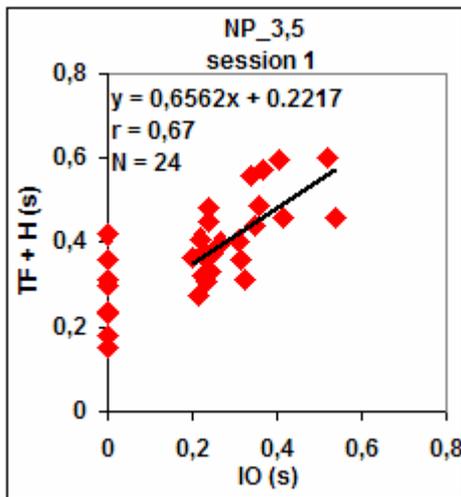
ANNEXE VI

Relation entre les phases *Time Falling + Hold*, *Time Falling*, *Hold* et l'*Intervalle d'Obstruence* (en secondes)
pour les enfants

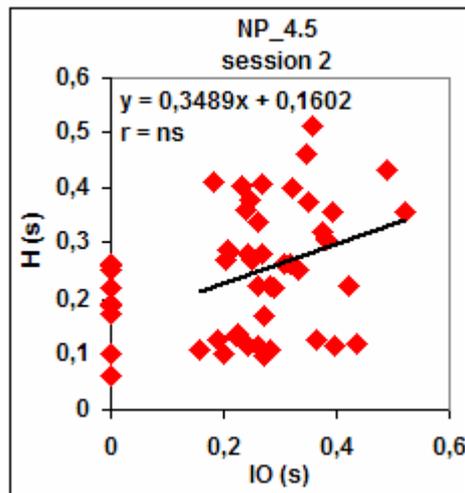
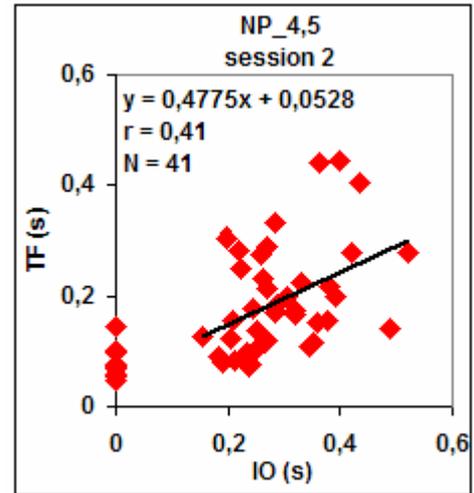
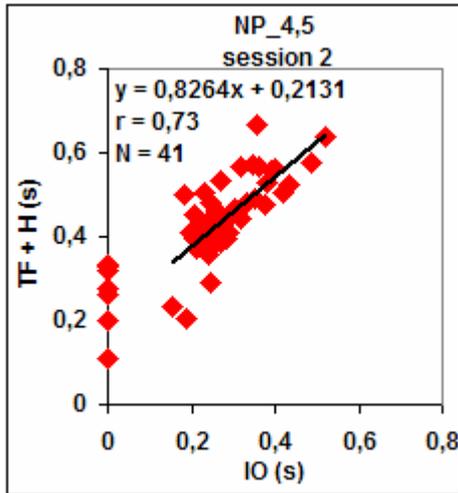
- Etude de la constriction -

1. Locuteur NP

1.1. Première session : 3 ans et demi

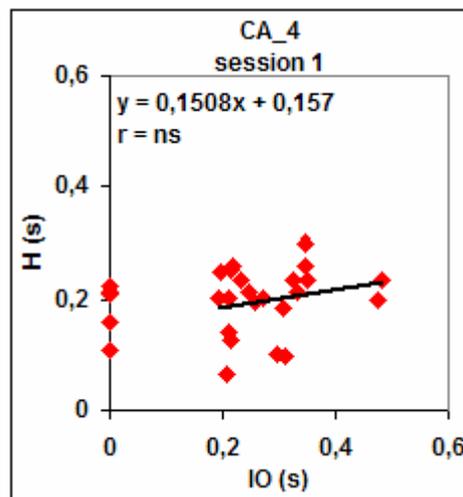
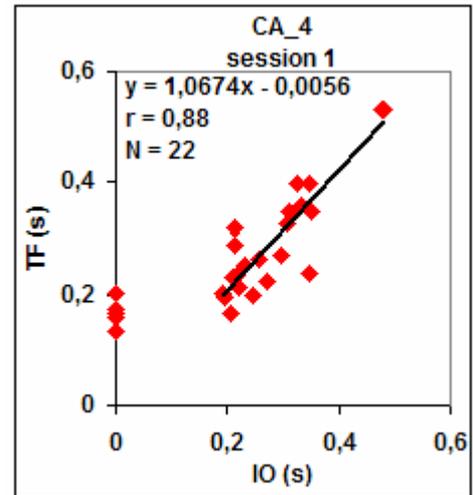
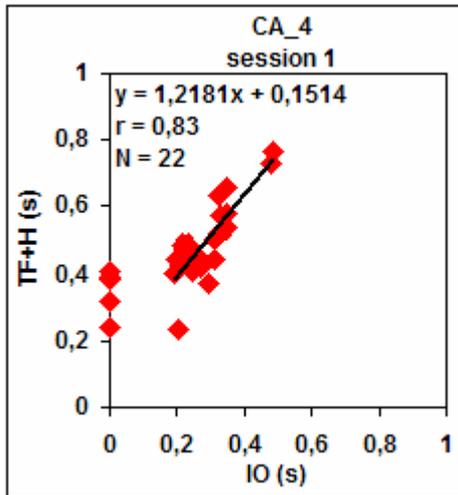


1.2. Deuxième session : 4 ans et demi

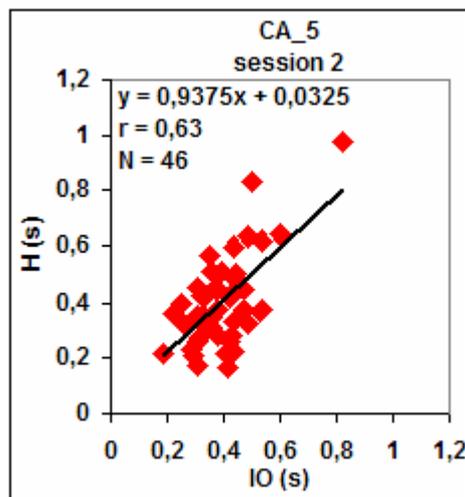
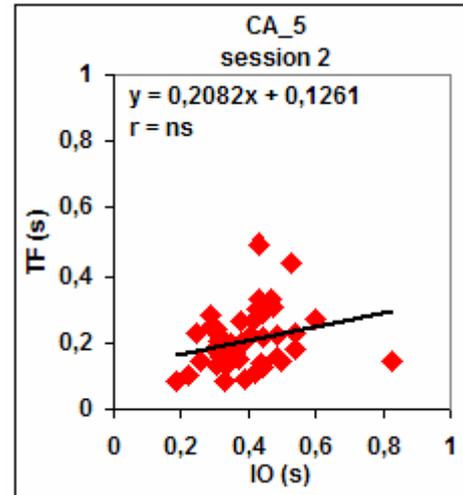
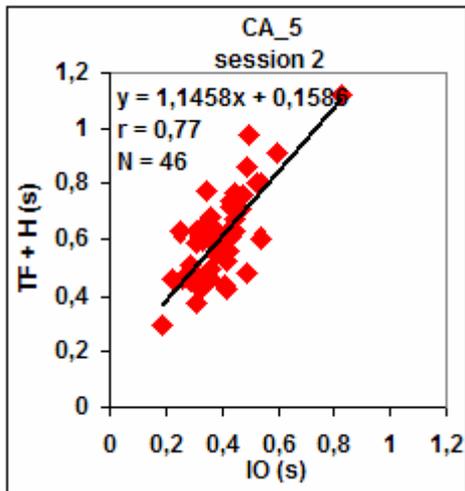


2. Locuteur AC

2.1. Première session : 4 ans et demi

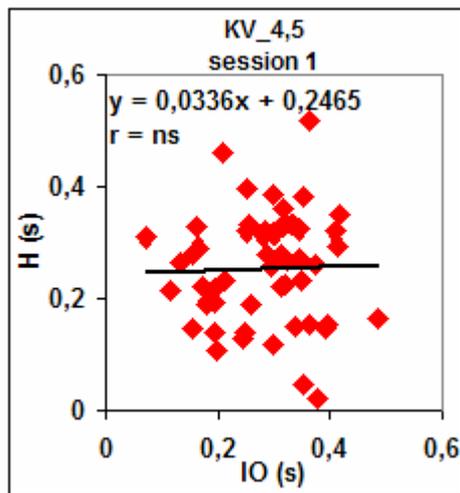
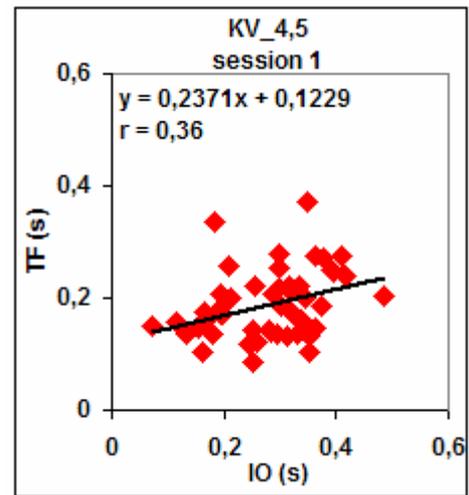
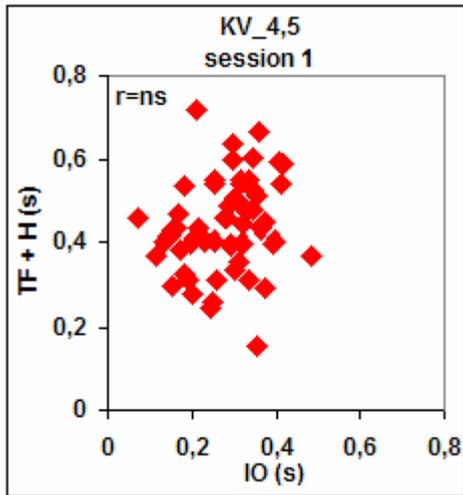


2.2. Deuxième session : 5 ans et demi

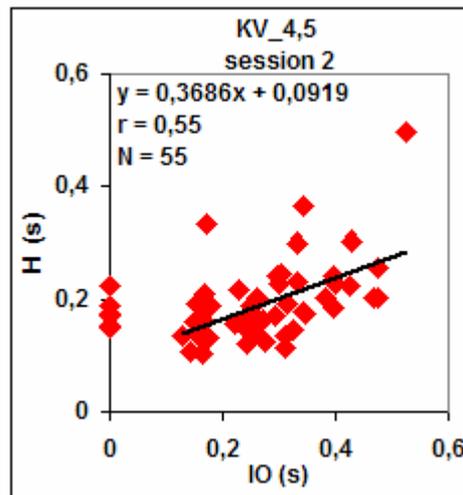
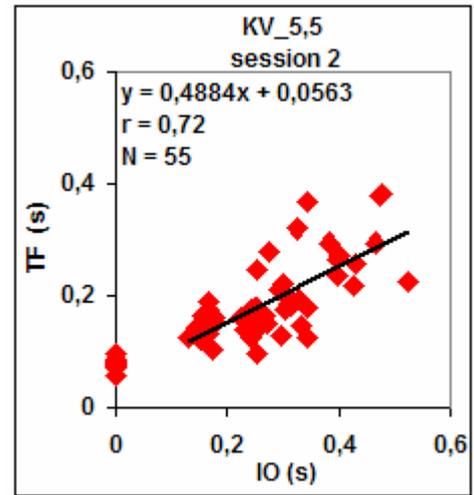
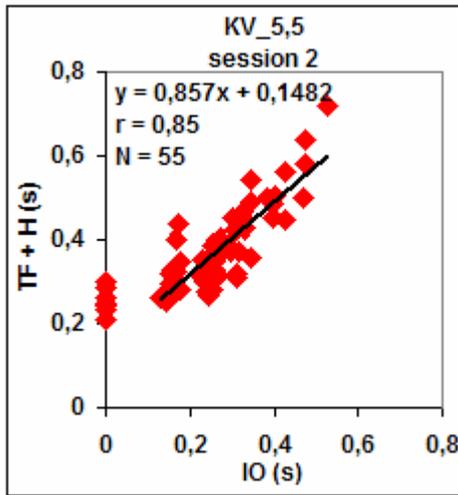


3. Locuteur KV

3.1. Première session : 4 ans et demi

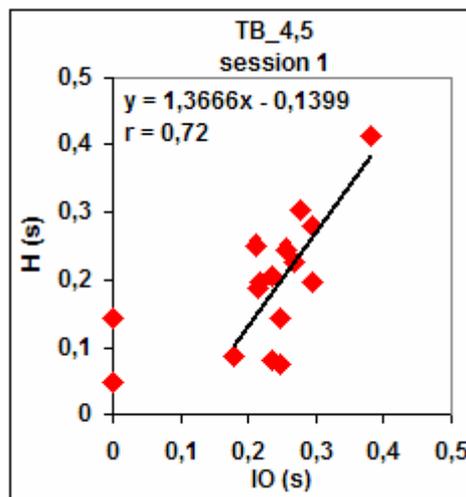
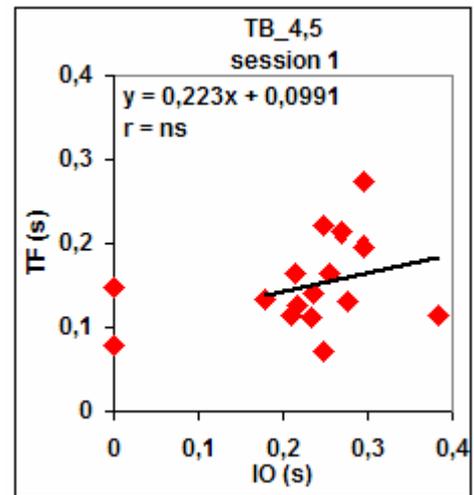
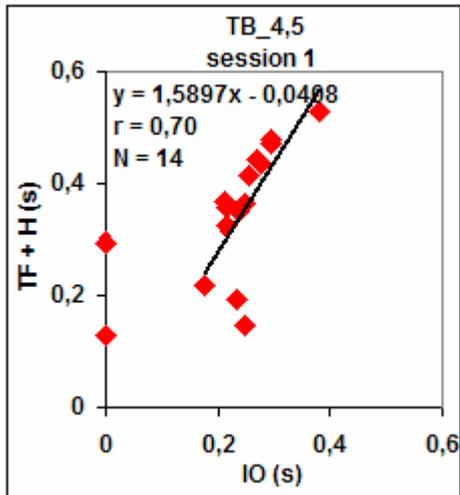


3.2. Deuxième session : 5 ans et demi

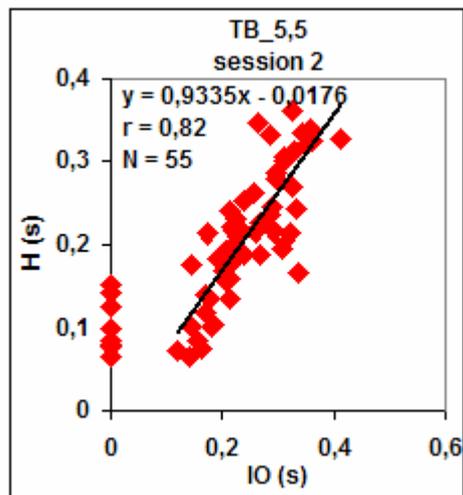
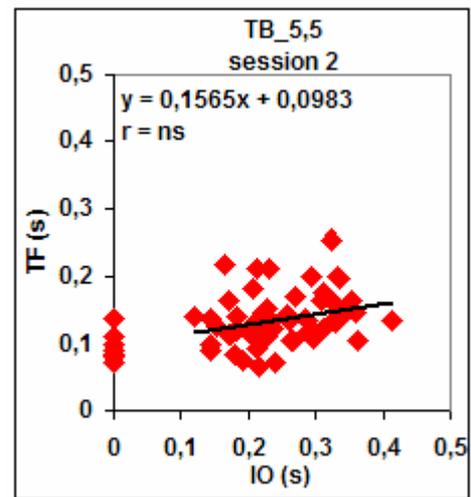
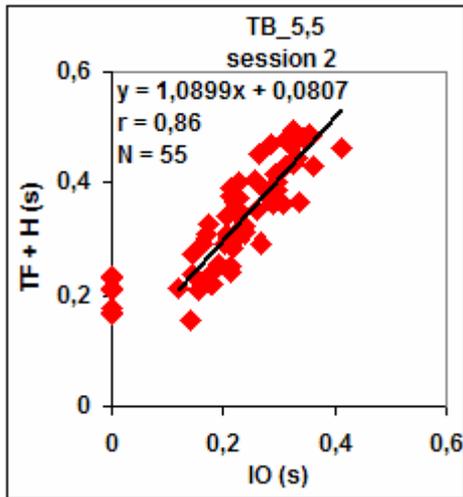


4. Locuteur TB

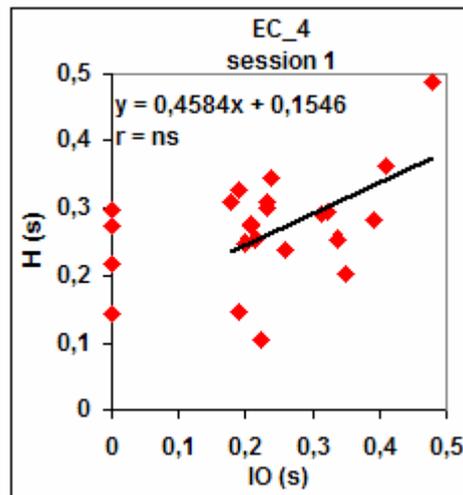
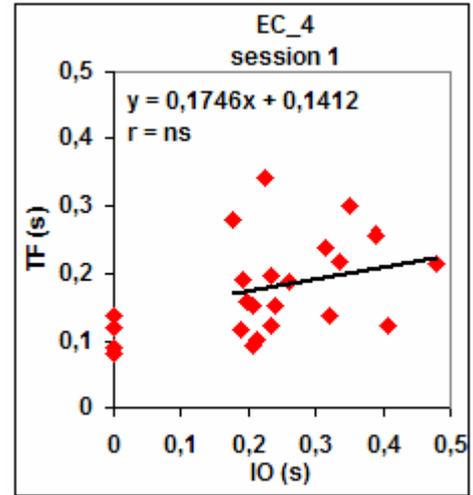
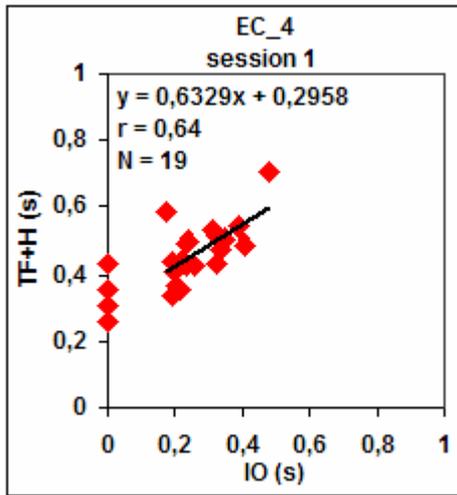
4.1. Première session : 4 ans et demi



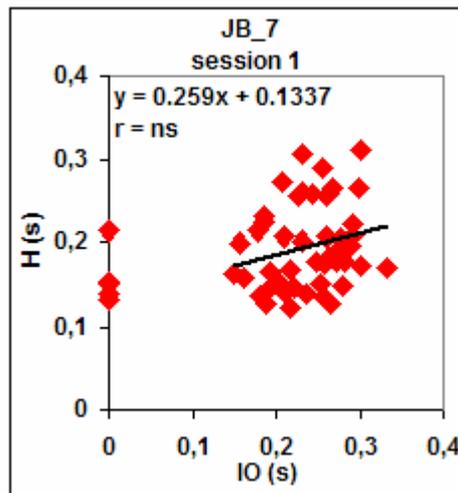
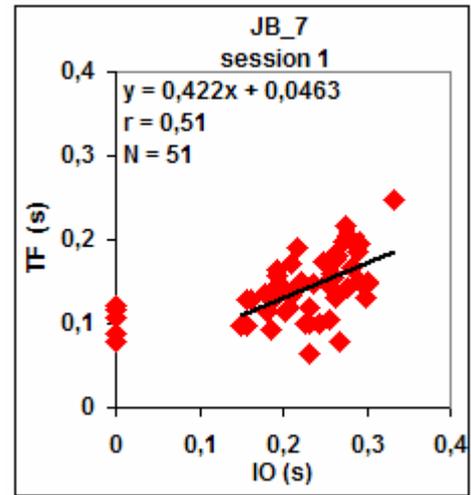
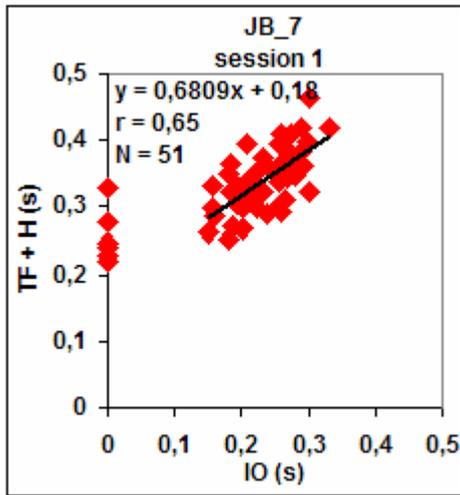
4.2. Deuxième session : 5 ans et demi



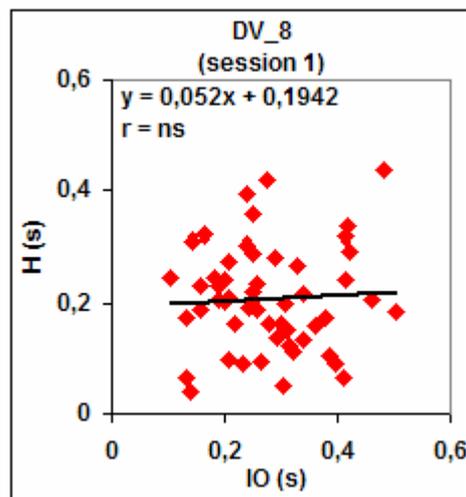
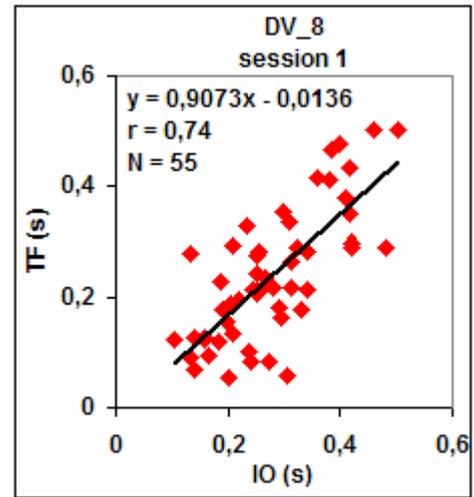
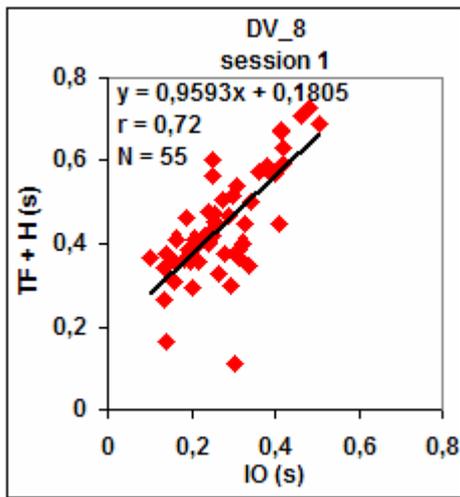
5. Locuteur EC : 4 ans



6. Locuteur JB : 7 ans



7. Locuteur DV : 8 ans



ANNEXE VII

Formulaire de consentement

Je consens à participer à l'étude suivante intitulée : la coarticulation anticipante chez l'adulte : Comparaison de données articulatoires entre sujets franco-qubécois et anglais, menée sous la direction de Lucie Ménard. Cette étude a pour principal objectif d'observer dans la parole les comportements anticipatoire parmi les sujets de différentes langues et de déterminer s'ils montrent des patrons similaires pouvant être décrits dans le cadre du modèle d'anticipation « *Mouvement Expansion Model* » (Abry & Lallouache 1991). **Ce projet de recherche est financé par le « Conseil de la Recherche en Sciences Humaines » (CRSH).**

Ma collaboration à cette étude consistera en la production aléatoire de courtes séquences qui feront l'objet d'un enregistrement audio-visuel et seront analysé dans le cadre de notre recherche. La procédure expérimentale se déroulera comme suit. Le signal acoustique sera enregistré par le biais d'un microphone. Les mouvements articulatoires des lèvres et mâchoire seront recueillies via une camera (face). Dans le but d'obtenir une détection des mouvements labiaux, les lèvres des locuteurs seront peintes en bleu avec un bleu à lèvre utilisé généralement pour les fêtes déguisées, facile à nettoyer. D'autre part, Les mouvements articulatoires des lèvres et mâchoire seront recueillis via un système de détection de mouvement (Optotrak). Cette procédure consiste en l'application de petits senseurs sur les lèvres, mâchoire et menton envoyant un signal lumineux infra rouge qui sera enregistré. Les deux images ci-dessous illustrent la procédure expérimentale utilisée.



Dispositif Optotrak



Dispositif lèvres bleues

L'enregistrement sera effectuée dans une chambre sourde au *Motor Control Lab* à l'Université Mc Gill le XX, XX et durera environ une heure.

Il va sans dire que des mesures précises seront prises par les chercheurs afin de s'assurer de l'utilisation sans risques pour le sujet du matériel (bâton coloré). Le maquillage utilisé sera hypo-allergène. Après chaque séance, le bâton sera soigneusement nettoyé à l'aide de mouchoirs. Tout le matériel ayant entré en contact avec la peau du sujet (mouchoir, pinceau, etc;) sera jeté.

Je reconnais que ma collaboration à cette étude sera apportée à titre gratuit et j'accepte d'y participer volontairement. Je comprends que je pourrai me retirer de l'étude à tout moment pour des motifs qui me concernent. Je reconnais que la responsable pourra interrompre ma participation en cours en tout temps après qu'elle aura débuté.

Il est convenu que les renseignements contenus dans mon dossier de recherche (mesures, données sociologiques) pourront être utilisés par la responsable de l'étude, à la condition cependant que les éléments qui pourraient être de nature confidentielle ne soient pas divulgués dans le public d'une façon telle que l'on puisse m'identifier. J'accepte également que les enregistrements soient conservés pour des recherches ultérieures.

Lucie Ménard
Professeure
Laboratoire de phonétique
Université du Québec à Montréal
Case postale 8888, succursale Centre-ville
Montréal (Québec), H3C 3P8
Tél. : (514) 987-3000, poste 3981
Télec. : (514) 987-4652

Signature :

Date :

N.B. : Prendre note que toute question, critique ou plainte peut être adressée à Lucie Ménard (987-3000 poste 3515). Si votre plainte ne peut être réglée directement avec la chercheuse principale, vous pouvez faire valoir vos droits par écrit auprès du Comité institutionnel d'éthique de la recherche avec des êtres humains de l'Université du Québec à Montréal (secrétariat : 987-3000 poste 773)