



HAL
open science

De la substance à la forme : rôle des contraintes motrices orofaciales et brachiomanuelles de la parole dans l'émergence du langage

Amélie Rochet-Capellan

► **To cite this version:**

Amélie Rochet-Capellan. De la substance à la forme : rôle des contraintes motrices orofaciales et brachiomanuelles de la parole dans l'émergence du langage. Psychologie. Institut National Polytechnique de Grenoble - INPG, 2007. Français. NNT : . tel-00200933

HAL Id: tel-00200933

<https://theses.hal.science/tel-00200933v1>

Submitted on 22 Dec 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.

N° attribué par la bibliothèque

|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|

THESE

pour obtenir le grade de

DOCTEUR DE L'INP Grenoble

Spécialité : « *Sciences Cognitives* »

préparée au laboratoire Gipsa-lab, Département Parole et Cognition

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

présentée et soutenue publiquement

par

Rochet-Capellan Amélie

Le 23 Octobre 2007

***De la substance à la forme : rôle des contraintes motrices orofaciales et
brachiomanuelles de la parole dans l'émergence du langage***

Directeur de thèse : Jean-Luc Schwartz

Co-encadrant : Rafael Laboissière

JURY

M.	Pascal	PERRIER	, Président
M.	Noël	NGUYEN	, Rapporteur
M.	Rudolph	SOCK	, Rapporteur
M.	Jean-Luc	SCHWARTZ	, Directeur de thèse
M.	Rafael	LABOISSIERE	, Co-encadrant
M.	David	OSTRY	, Examineur

Table des matières

1	Introduction : problématique et cadre épistémologique	3
1.1	Problématique générale	3
1.2	Les sciences de la complexité, une révolution épistémologique . .	5
1.2.1	Du déterminisme au chaos	5
1.2.2	Les systèmes complexes	7
1.2.3	Les sciences complexes et le monde du vivant	8
1.3	L'encorporation des fonctions cognitives	9
1.4	Evolution des relations forme-substance en sciences du langage . .	10
1.4.1	La phonétique, discipline de la substance	11
1.4.2	La phonologie, discipline de la forme	11
1.4.3	Evolution des relations phonétique-phonologie et substance- forme	11
1.5	Encorporation de la parole et du langage	12
1.5.1	Trois axes pour encorporer la parole	13
1.5.2	Simuler l'émergence des formes de la parole	14
I	Comment la gestualité orofaciale contribuerait à l'émergence des formes de la parole... et du langage	17
2	De la motricité de la parole aux langues et au langage	21
2.1	L'importance de la motricité	22
2.1.1	Stetson et l'essence motrice de la parole	22
2.1.2	Motricité et perception de la parole	23
2.1.3	Des théories gestuelles aux théories unificatrices	24
2.1.4	Le contrôle moteur, une clé de l'unification forme-substance	25
2.2	La substance déterminerait la forme par l'intermédiaire du contrôle moteur	26
2.2.1	Paramètres du mouvement et quantification de l'effort . .	27
2.2.2	Structures coordinatives et phasage inter-articulateurs . . .	28
2.2.3	La mâchoire au centre du processus : la théorie Frame- then-Content	31
3	Coordination articulaire et effet LC	35
3.1	Introduction	35
3.2	L'effet LC : Des explications à l'expérimentation	36
3.2.1	L'hypothèse <i>simple first</i>	36

3.2.2	Limites de l'hypothèse <i>simple-first</i>	38
3.2.3	Une hypothèse alternative pour l'effet LC	39
3.2.4	Un scénario expérimental	40
3.3	Une étude comparative de la stabilité des structures L_aC_o et C_oL_a	41
3.3.1	Méthode	41
3.3.2	Résultats	44
3.4	Une étude articulatoire de la stabilité L_aC_o vs. C_oL_a	48
3.4.1	Méthode	48
3.4.2	Résultats	53
3.5	Discussion	62
3.5.1	Le système coordinatif mâchoire-langue-lèvre en action	62
3.5.2	Implications pour l'émergence des formes phonologiques	66
3.6	Conclusion	67
II Comment la gestualité brachiomanuelle entraînerait la gestualité orofaciale vers l'émergence de la parole... et du langage		69
4	Des liens entre la main, la bouche, la deixis et l'émergence du langage	73
4.1	Une origine vocale ou manuelle ?	73
4.1.1	Préférence pour la parole, des évidences pas si trompeuses ?	73
4.1.2	Au commencement était la main	74
4.1.3	Gestes des mains et parole, deux composantes indissociables	75
4.2	Arguments pour une base gestuelle main-bouche à l'émergence du langage	75
4.2.1	Recouvrement des aires cérébrales du langage et de l'action	76
4.2.2	Co-développements des gestualités orofaciale et brachiomanuelle et émergence du langage	76
4.2.3	Coordination main-bouche dans la communication et autres activités	79
4.3	Cas particulier du geste de pointé	83
4.3.1	Le déclic d'Helen Keller	83
4.3.2	Le geste de l'attention partagée	83
4.3.3	La coordination pointé-parole, comme base fondatrice du langage	85
5	Site de synchronisation de la parole et du geste de pointé	89
5.1	Introduction	89
5.1.1	Contexte et enjeux théoriques	89
5.1.2	Synchronisation main-voix dans la deixis	90
5.1.3	Synchronisation main-mâchoire dans la deixis	90
5.1.4	Hypothèses et expériences	91
5.2	Méthode	92
5.2.1	Langue et participants	92

5.2.2	Dispositif expérimental	94
5.2.3	Procédure	94
5.2.4	Post-traitements des signaux	96
5.2.5	Mesures et hypothèses	100
5.3	Résultats	100
5.3.1	Effet de la position de l'accent sur la coordination mâchoire- doigt	100
5.3.2	Analyse des mouvements du doigt	109
5.3.3	Analyse des cycles de la mâchoire	111
5.3.4	Petite synthèse des principaux résultats	115
5.4	Discussion	116
5.4.1	Pointer et dénommer...	116
5.4.2	... mais aussi : pointer en dénommant.	117
5.4.3	Profil de mâchoire et effet LC	120
5.5	Conclusion	122
6	Deux cycles de mâchoire pour un geste de pointé ?	123
6.1	Introduction	123
6.1.1	Contexte et enjeux théoriques	123
6.1.2	L'hypothèse 2 :1 : une hypothèse développementale	124
6.1.3	Evaluer l'hypothèse 2 :1 chez l'adulte	124
6.2	Méthode	125
6.2.1	Participants	125
6.2.2	Dispositif expérimental	125
6.2.3	Procédure	126
6.2.4	Enregistrement et post-traitement des données	127
6.2.5	Etiquetage et mesures	127
6.2.6	Hypothèses	129
6.3	Résultats	129
6.3.1	Un portrait $1_{syll.} = 2_{syll.} \leq 3_{syll.} = 4_{syll.} ?$	130
6.3.2	Etude de la trajectoire du doigt	133
6.3.3	Deux syllabes sur un geste de pointé ?	137
6.4	Discussion	138
6.4.1	Du 2 :1 ici...	138
6.4.2	... à confirmer ailleurs	138
6.4.3	Des éléments pour la coordination main-mâchoire	139
6.5	Conclusion	140
III	Des simulations, un cadre modélisateur et conclusion	141
7	Vers une définition d'une approche système complexe	145
7.1	La question de la compositionnalité et de la syntaxe	145
7.1.1	Démarche générale	146

7.1.2	Nowak et l'émergence de la compositionnalité	146
7.2	Dériver la forme de la substance	147
7.2.1	Optimisation de contraintes	147
7.2.2	De l'optimisation à la robotique	148
7.3	Des agents incorporés	153
7.3.1	Définition du système complexe étudié	153
8	Conclusion	157
8.1	Stabilité articulaire et effet LC	157
8.2	Coordination main-mâchoire dans la deixis	158
8.2.1	Une coordination autour de la partie accentuée de la parole	159
8.2.2	Deux syllabes sur un geste de pointé	159
8.3	Sur le chemin de la forme...	160
IV	Bibliographie	161
V	Annexes	177
A	Effet de la position de l'accent - sélection des données et synthèse des ANOVAs	179
A.1	Sélection des données : nombre d'essais rejetés selon les critères de sélection	180
A.2	Synthèse des ANOVAs	181
A.2.1	Occurrence des événements des trajectoires du doigt et de la mâchoire	182
A.2.2	Position des événements de la trajectoire du doigt par rapport à ceux de la mâchoire	184
A.2.3	Paramètres de la trajectoire du doigt	186
A.2.4	Paramètres de la trajectoire de la mâchoire	188
B	Fréquence d'oscillation de la mâchoire et du geste de pointé	191
B.1	ANOVAs pour les variables mesurées ne dépendant pas de l'ordre de désignation	192
B.2	ANOVAs pour les variables mesurées dépendant de l'ordre de désignation	192

Table des figures

1.1 Faculté du langage au sens strict et au sens large selon Hauser <i>et al.</i> (2002)	4
1.2 Le langage, un <i>spandrel</i> de l'évolution ?	4
1.3 Trois axes pour une incorporation de la parole	13
2.1 Progression d'une syllabe VC à une syllabe CV avec l'augmentation du débit, d'après Stetson (1951)	22
2.2 Durée limite de mouvement dans l'espace durée-amplitude	27
2.3 L'étude du mouvement dans l'espace des phases selon Kelso	30
2.4 Trois formes de syllabes CV pour limiter le mouvement	33
3.1 Développement sériel du contrôle moteur de la parole et explication <i>simple first</i> de l'effet LC.	37
3.2 Simulation de la fermeture du conduit vocal par la mâchoire dans trois modèles articulatoires, d'après Vilain <i>et al.</i> (1999).	38
3.3 Comparaison entre les profils de mouvements des lèvres dans la production de /babababa/ par le bébé et l'adulte d'après Munhall et Jones (1998)	39
3.4 Mouvements de la mâchoire, de la lèvre inférieure et de la pointe de la langue attendus à faible débit.	41
3.5 Procédure expérimentale : déroulement d'un essai de l'étude acoustique.	43
3.6 Etiquetage des courbes d'intensité et mesures.	44
3.7 Variation d'intensité entre la voyelle après la constriction labiale et la voyelle après la constriction coronale (ΔI) en fonction de la durée des productions CVCV ou CCV de tous les sujets.	45
3.8 Etude de la stabilité des attracteurs et des profils de locuteurs	47
3.9 Dispositif pour l'enregistrement des mouvements des articulateurs par articulographie électromagnétique (EMA)	49
3.10 Etiquetage des données articulatoires	51
3.11 Mesure du cycle de mâchoire, cycle de mâchoire normalisé et positions des constriction sur ce cycle.	52
3.12 Mesure articulatoire de l'attracteur, Δ_{Attr}	52
3.13 Variation d'intensité entre la voyelle après la consonne coronale et la voyelle après la consonne labiale (ΔI) en fonction de la durée pour les productions CVCV et CCV de tous les sujets.	54

3.14	Etude des durées des cycles de la mâchoire en fonction de la durée des productions CVCV et CCV pour tous les locuteurs confondus, pour les séquences dupliquées et pour les séquences L_aC_o et C_oL_a .	55
3.15	Amplitude des gestes de mâchoire (strokes) montant et descendant en fonction de leur durée en phase d'accélération pour les productions avec un cycle de mâchoire par constriction et pour chaque sujet.	57
3.16	Etude de la coordination entre la mâchoire et les constricteurs en phase de fusion : distribution des productions des locuteurs selon la position de la constriction labiale et de la constriction coronale sur le cycle de mâchoire normalisé.	60
3.17	Réorganisation de la coordination entre la mâchoire et les constricteurs dans le passage de deux cycles de mâchoire pour deux constriction à un seul cycle pour deux constriction (fusion)	64
4.1	Le langage dans le cerveau, du localisationnisme aux circuits distribués	77
4.2	Les rythmes du langage dans les mouvements des mains des bébés, d'après <i>Petitto et al. (2001)</i>	79
4.3	La redécouverte du langage par Helen Keller	84
4.4	Enchassement du cadre de la parole dans le cadre du signe via le pied <i>Ducey-Kaufmann et al. (2007)</i>	87
5.1	Hypothèse de travail, une synchronisation autour de la partie accentuée de la parole	91
5.2	Dispositif expérimental mis en place pour l'enregistrement des mouvements du doigt et de la mâchoire dans une tâche de pointage manuel avec dénomination de cible.	93
5.3	Succession des écrans durant la phase expérimentale.	96
5.4	Signaux, étiquetage et paramètres des mouvements de la mâchoire et du doigt pour un essai /papá/.	98
5.5	Moyenne et erreur standard moyenne du temps écoulé entre le signal-go et les principaux événements des mouvements de la mâchoire et du doigt selon les conditions expérimentales pour chaque expérience.	102
5.6	Normalisation des cycles de la mâchoire et calcul des positions des événements de la trajectoire du doigt sur les cycles normalisés. . .	103
5.7	Répartitions de la position des principaux événements de la trajectoire du doigt sur les cycles de mâchoire normalisés selon la position de l'accent pour tous les essais et pour les moyennes des 20 participants.	104
5.8	Répartition des délais entre les principaux points de rencontre entre les trajectoires du doigt et de la mâchoire selon la position de l'accent pour tous les essais et pour les moyennes des participants. . .	105

5.9	Profils des participants. Etude de l'effet de la position de l'accent sur la position de la phase de tenue du geste de pointé relativement aux cycles de mâchoire normalisés.	108
5.10	Analyse des mouvements du doigt. Moyenne et erreur standard moyenne des paramètres du geste de pointé, de la phase de tenue et du geste de retour du doigt selon les conditions expérimentales.	110
5.11	Analyse des mouvements de la mâchoire. Moyenne et erreur standard moyenne des paramètres des cycles de la mâchoire selon les conditions expérimentales.	112
5.12	Profils de mâchoire et effet LC. Etude de la hauteur de la mâchoire et de l'énergie cinétique selon la séquence pour les productions accentuées sur la deuxième syllabes.	121
6.1	Etude des relations entre la durée de la syllabe et la durée du geste de pointé chez l'enfant et l'adulte.	124
6.2	Signaux, étiquetage et paramètres des mouvements de la mâchoire et du doigt pour un exemple /papapá/.	128
6.3	Moyenne et erreur standard moyenne du délai entre les apex des deux gestes de pointé selon les conditions expérimentales.	130
6.4	Etude des temps de réponse et du délais entre les temps de réponse de la mâchoire et du doigt.	132
6.5	Etude des gestes de pointé. Moyenne et erreur standard moyenne de la durée, de l'amplitude et de l'amplitude du pic de vitesse des gestes de pointé selon les conditions expérimentales.	134
6.6	Etude de la trajectoire du doigt séparant les deux gestes de pointé. Moyenne et erreur standard moyenne des trois phases entre les deux gestes de pointé selon les conditions expérimentales.	135
6.7	Distributions de la différence entre la durée de la production du nom de la cible et la durée du geste de pointé qui l'accompagne selon les conditions expérimentales.	137
7.1	Têtes parlantes et jeux de langage, d'après Steels (2001)	151
7.2	Exploration-imitation chez un androïde bébé, d'après Serkhane (2005)	155

Liste des tableaux

3.1	Analyse des variations d'intensité (ΔI) entre la voyelle après la consonne coronale et la voyelle après la consonne labiale pour les productions CVCV et CCV de durées inférieures à 300 ms pour les 42 essais analysés et pour chaque séquence $L_a C_o$ et $C_o L_a$	46
3.2	Nombre de productions CVCV ou CCV correctes et nombre de productions gardées pour l'analyse articulaire pour chaque sujet et chaque séquence $L_a C_o$ et $C_o L_a$	53
3.3	Moyennes des variations d'intensité entre la voyelle après la consonne coronale et la voyelle après la consonne labiale pour les productions CVCV et CCV de durée inférieure a 300 ms par sujet et séquence.	54
3.4	Moyennes des amplitudes et des durées des cycles de la mâchoire portant la constriction labiale et la constriction coronale en phase d'accélération, pour chaque sujet et chaque paire $L_a C_o$ - $C_o L_a$	58
3.5	Moyennes de la durée entre la constriction labiale et la contraction coronale suivante normalisée par la durée entre les deux constrictions labiales consécutives (Δ_{Attr}), pour chaque sujet et chaque séquence.	61
4.1	Liens entre les gestualité orofaciale et manuelle au cours du développement, d'après Iverson et Thelen (1999)	78
6.1	Etude du délai entre les apex des deux gestes de pointé selon le nombre de syllabes pour chaque locutrice	133
A.1	Nombre d'essais rejetés pour cause d'erreur ou de masquage des marqueurs de l'Optotrak pour les deux expériences sur l'effet de la position de l'accent et des profils de mâchoire sur la coordination mâchoire-doigt.	180
A.2	Expérience 1 (position de l'accent et coordination doigt-mâchoire), étude des effets de la position de l'accent, de la cible et de la séquence sur l'occurrence des événements de la trajectoire du doigt et de la mâchoire.	182
A.3	Expérience 2 (profils de mâchoire et coordination doigt-mâchoire), étude des effets de la position de l'accent, de la cible et de la séquence sur l'occurrence des événements de la trajectoire du doigt et de la mâchoire.	183

A.4	Expérience 1 (position de l'accent et coordination doigt-mâchoire), étude des effets de la position de l'accent, de la cible et de la séquence sur la position des événements de la trajectoire du doigt relativement à ceux de la trajectoire de la mâchoire.	184
A.5	Expérience 2 (profils de mâchoire et coordination doigt-mâchoire), étude des effets de la position de l'accent, de la cible et de la séquence sur la position des événements de la trajectoire du doigt relativement à ceux de la trajectoire de la mâchoire.	185
A.6	Expérience 1 (position de l'accent et coordination doigt-mâchoire), étude des effets de la position de l'accent, de la cible et de la séquence sur les paramètres de la trajectoire du doigt.	186
A.7	Expérience 2 (profils de mâchoire et coordination doigt-mâchoire), étude des effets de la position de l'accent, de la cible et de la séquence sur les paramètres de la trajectoire du doigt.	187
A.8	Expérience 1 (position de l'accent et coordination doigt-mâchoire), étude des effets de la position de l'accent, de la cible et de la séquence sur les paramètres de la trajectoire de la mâchoire.	188
A.9	Expérience 2 (profils de mâchoire et coordination doigt-mâchoire), étude des effets de la position de l'accent, de la cible et de la séquence sur les paramètres de la trajectoire de la mâchoire.	189
B.1	Fréquence de la mâchoire et du geste de pointé, étude des effets du nombre de syllabes et de la position de la cible pour les mesures ne dépendant pas de l'ordre de désignation.	192
B.2	Fréquence de la mâchoire et du geste de pointé, étude des effets du nombre de syllabes et de la position de la cible sur les paramètres de la trajectoire du doigt et de la mâchoire, pour les deux désignations.	193

Remerciements

Merci :

A Jean-Luc, d'avoir encadré cette thèse et d'avoir toujours fait ce qu'il fallait quand il fallait, comme il le fallait.

A Rafael, pour m'avoir accueillie à Munich au Max Planck, pour avoir toléré mon impertinence, mes crises de panique et mes coups de gueule.

A Abry, parce que la vie est ainsi mal faite que l'imprévisibilité la rend belle et conne à la fois.

A Christophe, Coriandre et Arturo, dont l'ingénierie a rendu les manip possibles.

A David Ostry, pour m'avoir invitée à Montréal et pour m'y accueillir en post-doctorat.

A Razika et Nadine, pour s'être occupées de l'administratif qui fait appel à une connaissance et une patience qui me dépasse et m'angoisse.

A Christian Bulfone, Monique et Nino pour s'être occupés de mon ordinateur et pour avoir fait en sorte que la technique ne soit pas un problème.

A Jean-Michel Adam, Bernard Caylux, Alain Lecomte, Jérôme Gensel, pour m'avoir aidé à prolonger mes mains, ma mémoire et mon intelligence en m'apprenant l'informatique et pour m'avoir offert la chance d'enseigner dans leur département IMSS.

A Jean-Michel, Bernard et Christian, pour m'avoir facilité ma vie de monitrice et cette collaboration que j'ai vécue sans accro.

Aux thésards, post-docs et chercheurs de l'ICP ou d'ailleurs, Marion, Nico R. et Nico A., Lucie, Julie, Aude, Antoine, Pauline, Anahita, Marc, Xavier, Virginie D., Claire(s), Jihène, Nourhedine, Stefane, Julien, Annemie, Nathalie, Coriandre, Christophe, Anne, Hélène, Pierre, Pascal, Fred, Lucie Ménard,... ça a été vraiment super de travailler avec des gens comme vous tout autour !

A Chrystel Franzone, et rien de moins que ma reconnaissance *éternelle* pour m'avoir soutenue en ma difficile année de maîtrise et permis de finir ce que je voulais reporter.

Aux copains, pour les bouffes, les balades, les soirées, les rêves écolo,... Marion et Ben ; Nico et Virginie (qui, cela dit en passant n'a rien d'une boulette) ; Béran et Jer ; Fafa (attention, me voilà bientôt de nouveau dispo pour escalader les cimes) ; Arno ; Lucie et Stéphane RVB (aussi pour m'avoir tous deux supportée dans un bureau).

Spéciale dédicace à Marion et Lucie pour leur soutien existentiel qui rend les choses tellement plus claires.

A Alex, merci de m'avoir soutenue quand j'ai décidé de me lancer là-dedans et merci d'avoir accepté malgré tout. Merci aussi à Ninie, je garde égoïstement tous les supers moments qu'on a passé ensemble.

Aux B. extend Marie, Mireille, minus François et mon soutien verbal rétrospectif à regret d'avoir été physique et synchrone à Elisabeth et Alain pour cette difficile année 2007.

A ma fratrie, toujours là en cas de pépins. Si vous avez le courage de lire quelques-unes des pages du gros tas qui suit, vous lirez qu'un phonème n'a de sens dans l'émergence des lexiques des langues que par sa valeur distinctive. Dans notre systémique fratrie, je vous laisse le soin de deviner la suite... A ma Caro, qui ne sera pas la seule à ne rien comprendre de mon obscure prose, mais c'est surtout valable pour toi : je ne serais pas moi si tu n'avais pas été là, bébé rose presque comme les autres. A Patou, malgré nos différences et différents, tu as toujours été là pour recoller les pots cassés avec tes mains de magicienne du don de soi, et à JM qui a souvent tenu le pot de colle. A Kitou, qui partit un jour avec sa chaîne Hifi. A Fanny, parce que si on a pleuré, qu'est-ce qu'on a ri aussi ! Et à mon Timmy, dans notre meilleur des mondes possibles, c'est long de réapprendre à marcher.

A la bande des successeurs, Bilou, Lolo, Sara, Camilla, Emi, Baptiste, Nielsou, Eliote, Tom et La-Sans-Nom.

A mon père et ma maman parce que le suffisant à recevoir, c'est le mieux qu'on puisse donner et parce que c'est tellement plus facile de faire ce qu'on a envie quand on est assuré par un amour inconditionnel.

A mon Paul, qui n'est plus à moi depuis longtemps, mais qui était là avec moi, au premier rang des amphis de psycho, puis derrière la porte quand je rentrais tard et au téléphone quand j'étais loin, et maintenant assez grand pour te souvenir plus tard d'à quoi ça ressemble, faire une thèse, vu de l'intérieur...

RALALA, après tout ça, si toi tu ne sais pas, moi je sais pourquoi je dois te dire merci à toi, François !

A l'adolescente sous les bombes en 40, qui rêva d'être institutrice,

A la petite fugueuse avec sa valise rouge (?), qui rêva d'être médecin,

*A ma petite Crevette, si un jour tu as besoin d'un obscur conseil de ta vieille
Maman-sage, je te dirais que ta liberté est dans l'immatérialité de tes mains et
dans la bienveillance de ceux qui t'aiment.*

Résumé

Et si les propriétés sensori-motrices de la parole modelaient le langage ? Cette hypothèse a propulsé le langage dans le monde de la complexité et de la cognition en-corporée. Nous introduisons ici différents types d'arguments montrant le rôle de la motricité de la parole dans la genèse du langage. Motricité orofaciale, d'abord, avec l'idée que les propriétés de la coordination inter-articulateurs contraindraient la morphogenèse du langage. Motricité orofaciale et brachiomanuelle, ensuite, avec l'hypothèse que le langage émergerait de la coordination main-bouche portant l'acte de pointage par la voix et par la main. Dans ce cadre, nos expériences analysent les mouvements enregistrés chez des locuteurs du français dans différentes tâches afin d'établir les propriétés des coordinations mâchoire-langue-lèvres dans la parole, puis mâchoire-main dans le pointage. Ces recherches s'intègrent au cadre de recherche global et récent proposant d'étudier le langage comme un système-complexe.

And what if the sensori-motor properties of speech model language ? This hypothesis drove language in the field of complexity and embodied cognition. Here, we introduce different kind of evidences showing the part of the motricity of speech in the genesis of language. Orofacial motricity, first, with the assumption that the properties of inter-articulators coordination may constraint the morphogenesis of langage. Orofacial and brachiomanual motricity, then, with the hypothesis that langage may emerge from hand-mouth coordination that support the act of pointing by the voice and by the hand. Hence, our experiments analyze the recorded motions of speakers of french during differents tasks in order to establish the properties of jaw-tongue-lips coordinations in speech and of jaw-hand coordination in pointing. These studies lie whithin the global and recent research framework that propose to investigate language as a complex-system.

Boyard thèse...

Quand vous partez à l'aventure dans une thèse traitant d'un sujet en rapport avec le langage, vous entreprenez un voyage sans retour. Non pas parce que vous finirez enterré au pied d'une théorie géniale qui explique tout mais, bien au contraire, parce que vous avancerez vers un objet qui se définit et se redéfinit en même temps que votre progression. Ne trouvant pas LA théorie du langage, vous vous engagez dans des voies scientifiques multidisciplinaires et devrez-vous battre avec une multitude de concepts scientifiques de la physique, à la biologie en passant par la linguistique, la philosophie, la génétique, la psychologie, la neurologie, l'intelligence artificielle... Avec de la persévérance et de l'aide, vous viendrez à bout - ou presque - de certains de ces concepts qui, peut-être, vous aideront à l'ultime épreuve : la mise au clair, dans votre rapport de thèse, de ce que vous avez voulu faire au cours de cette aventure de trois ans. Bref, après avoir pataugé, mais quand même avancé dans cet océan de connaissances, et après avoir souvent eu des difficultés pour vous en extirper et retourner à la vie réelle, vous arriverez enfin devant la porte du trésor, des concepts plein les bras. Mais, le temps vous étant compté, vous serez contraint de vous en délester d'une partie pour passer la porte du sas qui se referme.

Le langage est un objet d'étude complexe, sensible - car faisant écho à la question de la nature humaine et à la pensée - multidisciplinaire, toujours mal défini, mais incroyablement passionnant. Les trois quarts de cette thèse étant un travail expérimental, le choix des concepts dans le quart théorique restant repose sur un souci d'ancrage de ces expériences dans un contexte épistémologique récent et sur une volonté de présenter l'origine des hypothèses et des méthodes utilisées. Face à la quantité et à la richesse des notions impliquées par cette démarche, ce quart théorique ne se veut en rien exhaustif : il relie des éléments entre eux et trace un chemin parmi tant d'autres possibles.

Dans cette première partie introductive, nous donnerons quelques éléments sur le langage en tant qu'objet d'étude. Nous verrons ensuite comment l'évolution des sciences au XXème siècle avec l'apparition de l'étude des systèmes complexes en sciences physiques et l'ouverture de frontière corps-cerveau a joué un rôle fondamental dans la conception des relations entre le langage et la parole.

Introduction : problématique et cadre épistémologique

1.1 PROBLÉMATIQUE GÉNÉRALE

Comme le remarquent souvent ceux qui l'étudient - linguistes, biologistes, informaticiens, mathématiciens... - le langage est un objet d'étude complexe. D'abord considéré comme (rien de moins) qu'un miracle, les chercheurs le découpent actuellement en sous-capacités comme l'imitation des expressions faciales, l'organisation des séquences gestuelles ou encore le partage de l'attention (Holden, 2004). Dans ce cadre, Hauser *et al.* (2002) subdivisent le langage en deux composantes : une faculté de langage au sens strict (FLN - *Faculty Language Narrow - sense*), structure cérébrale spécifique au langage et à l'espèce humaine, et une faculté de langage au sens large (FLB - *Faculty Language Broad - sense*), structure partagée avec d'autres fonctions cognitives et/ou d'autres espèces animales (voir figure 1.1). La FLN est une version évoluée de GU, la Grammaire Universelle de Chomsky (co-auteur de Hauser *et al.*, 2002), et les chercheurs de différents horizons s'accordent globalement sur l'existence d'une faculté cérébrale propre au langage. Les controverses concernent la nature de cette faculté, comme en témoignent les échanges entre Hauser-Fitch-Chomsky (le camp HFC) et Pinker-Jackendoff (le camp PJ). Pour HFC, la FLN ne serait pas le produit de la sélection naturelle mais plutôt d'une micro-mutation et comprendrait tout ce qui concerne le langage et qui n'est pas partagé avec les animaux ou avec d'autres fonctions cognitives (Hauser *et al.*, 2002; Hauser et Fitch, 2003; Fitch *et al.*, 2005). Ces auteurs insistent particulièrement sur la propriété récursive de la grammaire qui selon eux serait la composante minimale de la FLN. Ce point de vue s'oppose à celui de PJ pour qui la FLN serait le produit de la sélection naturelle, seul designeur de la nature capable de générer une telle complexité de formes (Pinker, 2003; Pinker et Jackendoff, 2005; Jackendoff et Pinker, 2005). Selon d'autres auteurs, le langage serait le fruit d'une évolution d'un ensemble de fonctions cognitives, une sorte de spandrel (voir figure 1.2) de l'évolution ou, en d'autres termes, la FLN serait vide. Enfin, selon Deacon (2003), une partie des aspects universaux du langage pourrait refléter la structure du monde ou plutôt des propriétés de l'interaction de l'individu avec cette structure, cette partie ne serait donc ni innée, ni acquise.

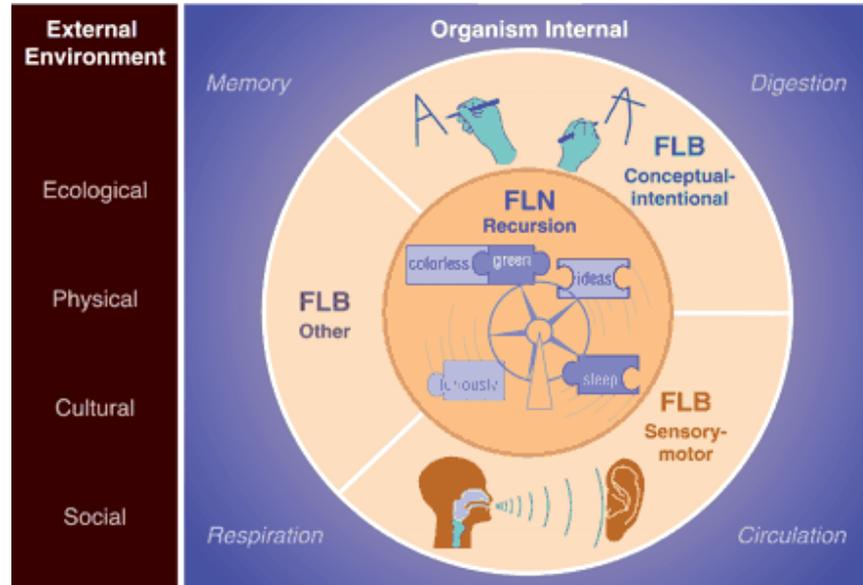


FIG. 1.1 – **Faculté du langage au sens strict et au sens large selon Hauser *et al.* (2002).** Dans leur article publié dans *Science* en 2002, Hauser, Fitch et Chomsky proposent de diviser la faculté de langage en deux grands modules : une faculté de langage au sens strict (la FLN - Faculty Language Narrow) et une faculté de langage au sens large (la FLB - Faculty Language Broad). La FLN regrouperait ce qui fait la spécificité du langage humain, qui n'a pas de précurseur dans le monde animal et qui n'est pas partagé avec les autres fonctions cognitives. Peut-être que la FLN se révélera être vide (voir Fitch *et al.*, 2005), mais le plus probable, selon ces auteurs, est qu'elle contienne au moins la fonction récurrente. Si l'hypothèse de l'existence d'une structure cérébrale spécifique au langage est globalement acceptée, la nature de cette structure et son origine sont discutées. L'idée d'un noyau récursif est particulièrement contestée, notamment par Jackendoff et Pinker (2005).

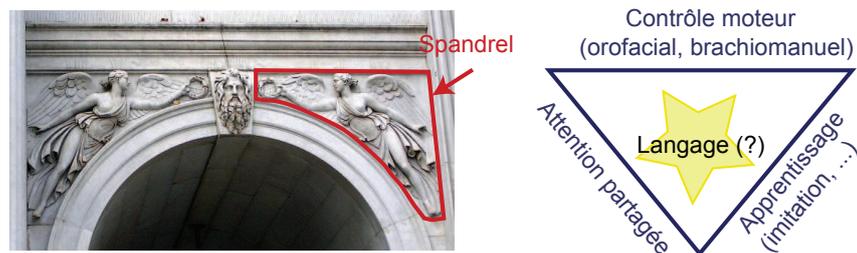


FIG. 1.2 – **Le langage, un *spandrel* de l'évolution ?** En architecture, un spandrel est l'espace triangulaire entre deux arches et la surface horizontale qui les surplombe. Il est souvent orné de dessins ou de sculptures sur les monuments historiques. Selon certains auteurs, le langage serait un spandrel de l'évolution, c'est-à-dire le produit d'un *espace* créé par la juxtaposition d'un ensemble de fonctions cognitives (tel que l'attention partagée, le contrôle des articulatoires, l'apprentissage, l'imitation et bien d'autres encore). Mais selon Pinker et Bloom (1990), en admettant que le langage soit un spandrel, ce n'est pas le spandrel en soi qui est intéressant mais la manière dont il est investi. Or, étant donné la complexité du langage, cet investissement n'a pu se faire que par un processus de sélection naturelle.

A l'heure actuelle, ces approches ne peuvent être départagées de manière certaine mais les résultats des recherches multidisciplinaires sur le langage vont dans le sens d'une circuiterie neuronale distribuée, connectant des aires cérébrales réalisant des traitements cognitifs génériques. La mise en place de cette circuiterie pourrait être le produit de la phylogenèse (héritage génétique) et/ou s'acquérir au cours de l'ontogenèse (développement de l'enfant). Nous verrons qu'une partie de ces aires cérébrales sont celles qui sont impliquées dans le contrôle moteur de la gestualité orofaciale et brachiomanuelle. D'autre part, le débat HFC-PJ est un débat qui se base essentiellement sur la composante syntaxique de la grammaire, or c'est aux formes de la phonologie que nous nous intéresserons ici.

A la suite de Liljencrants et Lindblom (1972), nous apporterons des éléments supplémentaires à une démarche scientifique visant à dériver la forme du langage (sa structuration comme un système), de la substance qui le véhicule, ses modalités d'expression et de perception. Nous montrerons en quoi la recherche actuelle impose de considérer le langage comme une faculté indissociable du contrôle moteur de la gestualité orofaciale mais aussi, de la gestualité brachiomanuelle. Plus précisément, nous verrons comment l'étude expérimentale de la gestualité dans la communication contribue à déterminer le rôle des contraintes motrices dans la morphogenèse du langage. Sur le plan épistémologique, cette démarche peut s'envisager comme un produit dérivé de l'explosion des sciences de la complexité en physique au cours du XX^{ème} siècle et de leur propagation à la biologie puis aux sciences humaines. Elle s'accompagne d'une abolition de la frontière entre le corps et le cerveau ou, en d'autres termes, d'une *encorporation* des fonctions cognitives, néologisme que nous utiliserons dans ce rapport pour *Inscription corporelle de l'esprit* traduisant le terme anglais *embodiment* (Varela, 1991, 1993).

1.2 LES SCIENCES DE LA COMPLEXITÉ, UNE RÉVOLUTION ÉPISTÉMOLOGIQUE

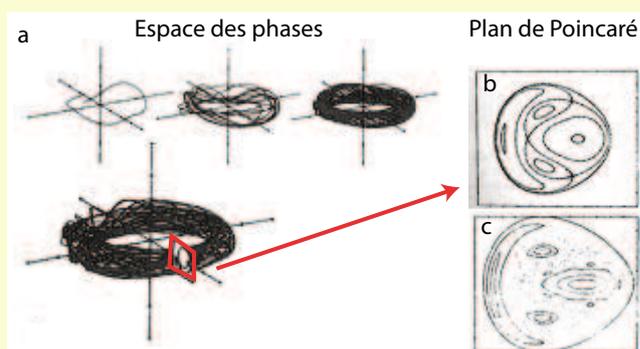
Bien que l'information ne soit pas encore totalement passée dans la culture collective, le XX^{ème} siècle a vu s'écrouler l'un après l'autre les murs des certitudes qui entouraient la forteresse de la physique newtonienne (Thuan, 1998, p.105). Le dernier mur en date à être tombé est celui du déterminisme Newtonien et Laplacien, supposant qu'un état t du monde physique est prédictible de l'état $t - 1$ et prédictible de l'état $t + 1$. Le responsable de cette remise en question d'une conception de la science ancrée dans les têtes depuis Aristote est la découverte du chaos, une des sources principales des sciences de la complexité.

1.2.1 Du déterminisme au chaos

Supposée par Henri Poincaré à la fin du XIX^{ème} siècle dans le cadre de la résolution du problème des trois corps (voir encadré, p. 6), la présence du chaos dans

LES CAPRICES DE LA LUNE ET LA DÉCOUVERTE DU CHAOS

Depuis le II^{ème} siècle, les astronomes-mathématiciens développèrent des modèles pour calculer la trajectoire de la Lune. Mais ces modèles étant faux, ils accumulaient les erreurs au cours du temps et la Lune déviait de plus en plus des trajectoires prédites. En introduisant un modèle où la trajectoire de la Lune est déterminée par la force gravitationnelle exercée par la Terre et le Soleil, Isaac Newton (1643-1727) parvint à approximer la trajectoire de la Lune avec une erreur de 1/6 de degré. Le problème des trois corps était né : si les équations de Newton fonctionnent pour deux corps, elles semblent ne pas fonctionner pour trois. La question de la Lune intéressa les plus grands mathématiciens des XVIII^{ème} et XIX^{ème} qui, en se penchant sur le problème, *découvrirent de nouvelles mathématiques, mais la Lune refusait obstinément de se plier à leurs calculs* (Thuan (1998), p.124). Après 40 années de recherche sur la question de la Lune, Leonhard Euler (1707-1783) conclut ses travaux par *Je ne vois pas comment cette recherche sur la Lune peut aboutir, ni comment elle peut être utilisée à des fins pratiques* (Thuan (1998), p.124).



Mais les scientifiques sont perspicaces. De plus, les déviations de la Lune posent la question de la stabilité du système solaire. Les orbites des planètes sont-elles stables ou bien les planètes vont-elles entrer en collision ou dériver vers l'infini ? A la fin de XVIII^{ème} siècle, Henri Poincaré (1854-1912) introduit une nouvelle méthode d'approche du problème des trois corps consistant à représenter le système dans un espace abstrait dit espace des phases. Si le mouvement d'un corps comprend 6 dimensions (position et vitesse en y , x et z), un système à trois corps peut-être représenté par un seul point dans un espace à 18 dimensions. Ce point va dessiner des trajectoires dans l'espace des phases au cours du temps. La figure a. donne un exemple simplifié d'espace des phases à 3 dimensions (position en x , en y et vitesse) et la trajectoire qu'y décrit une orbite stellaire. Poincaré observe ces trajectoires complexes dans le plan de phase (coupe verticale de l'espace des phases). Il remarque que les courbes qui se dessinent dans ce plan sont régulières tant que le mouvement ne dépasse pas une valeur critique d'énergie (en b.), au delà de cette valeur, les zones de stabilité sont entremêlées avec des zones de chaos (en c.). Il y a donc de l'ordre sous le désordre apparent mais ce nouvel ordre ne se décrit pas en termes déterministes mais probabilistes.

Il faudra attendre la découverte de l'informatique et notamment les travaux du météorologue Edward Lorenz dans les années 1960 pour que la simulation confirme l'existence du chaos et que l'étude des systèmes complexes explose (figures et historiques d'après Thuan, 1998).

certaines systèmes physiques se confirme au XX^{ème} siècle et va induire l'émergence d'une nouvelle science : celle de l'étude des systèmes complexes. La découverte de ces systèmes est une véritable révolution épistémologique en sciences physiques. En effet, elle fait basculer le monde scientifique de la certitude à l'incertitude, du déterminisme au probabilisme et du réductionnisme à l'émergentisme. On est passé au XX^{ème} siècle d'une science explicative qui cherche à redécouvrir les lois d'une Nature rationnelle et à les reproduire en laboratoire à une science entre explication et description, qui établit des scénarii sur le fonctionnement de systèmes complexes et qui les évalue par la simulation. Une hypothèse épistémologique qui revient fréquemment est que, peut-être, la science ne parviendra jamais à tout expliquer, car pour cela, il faudrait qu'elle soit capable de prendre en compte tout - ou presque - ce qui s'est produit depuis le début de l'univers et plus encore. Ainsi, Zwirn (2006) conclut son livre sur les systèmes complexes par les considérations épistémologiques suivantes :

Les notions d'explication et de compréhension devront être modifiées et affaiblies lorsqu'on s'intéresse aux systèmes complexes. Si, comme cela est possible, le seul moyen de connaître le comportement d'un système est de le simuler et d'observer l'évolution du modèle, alors, il faudra renoncer à l'espoir de comprendre au sens actuel du terme, pourquoi le système se comporte comme il le fait. (...) Cela ne veut bien sûr pas dire que nous ne pourrions pas progresser dans la connaissance de certains des mécanismes fondamentaux mis en œuvre. Notamment, un programme de recherche pourrait consister à découvrir les conditions nécessaires et suffisantes pour l'apparition du phénomène. Mais la compréhension fine et totale de l'ensemble du processus risque de rester inatteignable, (p.205).

1.2.2 Les systèmes complexes

Cette révolution est globale en science car, comme le constate Zwirn, une fois qu'on est capable d'identifier un système complexe, on en voit partout. Ainsi, la définition des systèmes complexes regroupe dans une même catégorie des objets scientifiques que rien ne semblait relier avant la découverte de la complexité, telle que la construction des termitières, l'embryogenèse, les mouvements des planètes, la météorologie, la bourse, l'économie, les embouteillages, l'organisation spatiale des villes, l'Internet,... et le langage. En fait, ce qui fait l'unité des systèmes complexes c'est la nature des mécanismes qui les sous-tendent. Zwirn insiste sur quatre caractéristiques :

Le holisme. Le système doit être appréhendé comme un tout, le tout n'étant pas la somme de ses parties (ce que permet de faire l'espace des phases de Poincaré pour le problème Terre-Lune-Soleil, voir encadré, p 6).

L'émergence par auto-organisation. Le comportement global du système émerge des interactions entre ses composants, interactions régies par des règles simples.

Par exemple, plutôt que d'expliquer la structure en V d'un vol coordonné de canards par le fait que chaque canard a dans la tête un plan de vol et/ou par la présence d'un chef d'escadrille, on peut montrer par simulation que la structure du vol émerge de la combinaison de règles simples comme (a) voler dans la même direction que ses voisins ; (b) rester proche de ses voisins ; (c) ne pas entrer en collision avec ses voisins et (d) éviter les obstacles. De plus, la manière dont le système évolue est très sensible aux conditions initiales (c'est le très célèbre effet papillon d'Edward Lorenz).

Adaptation. Le système peut se modifier pour s'adapter à l'environnement mais une modification progressive des paramètres du système n'induit pas une modification progressive de son organisation. Au contraire, l'adaptation se fait par transition de phases. Une petite variation dans les paramètres peut bouleverser complètement le système, comme l'eau passe de l'état liquide à l'état solide quand la température descend en dessous de 0 degré, sans transiter par un état hybride.

Evolution. Des générations successives de systèmes complexes de même type, qui se reproduisent et se modifient, évoluent par un mécanisme de sélection naturelle en privilégiant les systèmes porteurs des modifications les plus adaptées à l'environnement alors que les autres disparaissent.

La principale (pour ne pas dire l'unique) méthode d'approche des systèmes complexes est la simulation informatique. Elle utilise différents algorithmes : les algorithmes génétiques, les systèmes multi-agents, les réseaux de neurones etc. Ainsi, la simulation a aussi envahi les recherches sur le langage.

L'explication de l'organisation de la nature sans faire appel à un grand architecte ou organisateur est la voie que la science doit emprunter si elle ne veut pas tomber dans une circularité finissant par expliquer la forme par elle-même (la forme est comme ça parce que celui qui l'a générée en a décidé ainsi). La découverte du processus d'auto-organisation est une grande avancée en ce sens. Cependant, à l'heure actuelle, elle ne peut expliquer à elle seule la morphogenèse du vivant.

1.2.3 Les sciences complexes et le monde du vivant

Comme le constate Monod, les formes vivantes se distinguent des formes inanimées par le fait qu'elles résultent de forces intérieures alors que les forces extérieures interviennent comme obstacles, contraintes ou paramètres mais pas comme un moteur premier et direct (Lesne et Bourguine, 2006). Ainsi, le processus d'auto-organisation dans la morphogenèse des êtres vivants ne peut tenir actuellement la même part explicative que dans le monde physique. En effet, contrairement à la matière inanimée, le vivant évolue, et au fil de cette évolution se (s'auto)-redéfinit, il crée des fonctions biologiques de plus en plus complexes qui vont jusqu'à se redéfinir elles-mêmes, comme c'est le cas du langage. On peut retourner le problème dans tous les sens, et malgré tout l'attrait que présente l'auto-organisation,

il reste dans le vivant un autre designeur : la génétique qui fait qu'une cellule œuf de chien ne deviendra pas un chat et encore moins un arbre. Pour considérer le monde du vivant comme un système complexe, il faut donc non seulement le voir comme holistique, auto-organisé et adaptatif mais encore considérer l'adaptation sous l'angle de la sélection naturelle et de l'adaptation évolutive, le quatrième élément définissant les systèmes complexes selon Zwirn. Le monde de la complexité est donc pluriel.

Dans le cadre de la complexité, un des enjeux de la biologie actuelle est de trouver les facteurs externes qui agissent sur l'expression du code génétique car si un chien ne devient pas un chat, une cellule œuf de chien qui se développerait dans un environnement avec d'autres lois physiques que les nôtres deviendrait certainement un *drôle de chien*. De manière analogique, un des enjeux actuels des sciences cognitives est de délimiter les facteurs influant le développement de l'organisation d'un cerveau dont la structure de base est déterminée par la génétique, les petits hommes privés d'un environnement humain devenant de *drôles de petits hommes* (comme le montrent les cas répertoriés d'enfants sauvages, voir Malson, 1969). S'il y a de l'auto-organisation dans le vivant, dans l'interaction de l'organisme et de ses gènes avec l'environnement, et certainement dans l'évolution de ces gènes, le processus évolutif détermine l'individu. Chez l'homme, la complexité est décuplée par l'évolution culturelle et le phénomène de conscience, l'élément de *cinquième type* de la complexité, non considéré par Zwirn, qui s'intéresse essentiellement aux mathématiques et à la biologie.

1.3 L'INCORPORATION DES FONCTIONS COGNITIVES

Si la physique du XX^{ème} siècle bascule dans le monde du chaos et de la complexité sous l'impulsion de l'outil informatique, les progrès de l'intelligence artificielle (IA) - et surtout ses échecs, d'ailleurs - ont contribué à reconsidérer les relations entre le cerveau, le corps et l'environnement. À ses débuts, l'IA amène une conception des fonctions cognitives *du haut vers le bas (top-down)*. C'est la métaphore de l'ordinateur et de l'intelligence centralisée, héritée du dualisme cartésien corps/esprit : le cerveau programme, le corps exécute. D'abord adoptée en sciences cognitives pour décrire le fonctionnement du cerveau, cette métaphore s'efface progressivement et laisse de plus en plus de place à une conception du cerveau comme un système ouvert, qui se construit dans ses interactions avec le monde physique et social par l'intermédiaire des organes de perception et de production.

Un des grands fondateurs de cette conception, et principal responsable de son introduction en sciences cognitives, est Francisco Varela qui défend une *inscription corporelle de l'esprit* et développe le concept d'*enaction* (Varela, 1993). Ce concept réfère au fait que l'organisme doit être suffisamment incarné dans un en-

vironnement pour pouvoir se débrouiller dans cet environnement sans en posséder de représentation préalable : *Son monde émerge avec ses actions, c'est un monde "enacté"* (Kempf, 1998). Par conséquent, *la cognition n'est pas la représentation d'un monde prédéterminé par un esprit prédéterminé mais c'est plutôt la modification du monde et de l'esprit sur la base de l'histoire de toutes les actions qu'un être exerce sur le monde* (Varela, 1993).

La notion d'encorporation des fonctions cognitives a aussi été développée par le philosophe Andy Clark (Clark, 1997, 2003). Selon Clark, les fonctions cognitives ne sont pas des réalités distinctes du corps et du monde mais des éléments de contrôle d'activités incarnées. Ainsi, elles dépendent des propriétés physiques du corps et de l'environnement et de facteurs sociaux et culturels. D'après Clark, qui pousse à l'extrême la notion d'encorporation, l'être humain est un *cyborg-né*. Les outils ont prolongé ses mains (et plus généralement son corps) pour prolonger de plus en plus son intelligence. Les exemples les plus probants sont certainement le développement d'Internet et de la téléphonie mobile : nous laissons à nos outils le soin de mémoriser de plus en plus de choses pour nous. Notre mémoire est plus que jamais un système d'indexage. On se souvient où trouver l'information, et de moins en moins de l'information elle-même, et notre monde devient celui du tout accessible ici et maintenant.

L'ouverture de la frontière entre le cerveau, le corps et le monde dans les sciences cognitives est un des éléments justifiant l'approche du cerveau comme un système complexe. De la même manière, l'évolution des relations entre la substance et la forme ou entre la parole et la langue en sciences du langage va amener à développer de nouvelles méthodes d'approches dans le cadre de la complexité afin de dériver la forme de la substance.

1.4 ÉVOLUTION DES RELATIONS FORME-SUBSTANCE EN SCIENCES DU LANGAGE

La linguistique moderne a longtemps distingué l'étude de la forme (phonologie) de celle de la substance (phonétique). La distinction entre phonétique et phonologie émerge à la fin du XIX^{ème} siècle avec Baudouin de Courtenay. Elle se retrouve chez De Saussure au début du XX^{ème} siècle (Bally *et al.*, 1995) puis dans le cercle linguistique de Prague, dans les années 20 avec Troubetskoy (voir Boë, 1997; Jakobson, 1976). De nombreuses réflexions ont été consacrées à l'évolution des relations entre la phonétique et la phonologie. Le lecteur intéressé trouvera un échantillon représentatif de ces travaux dans l'historique critique de Boë (1997) et le manuel compilant les présentations d'une école thématique consacrée aux dichotomies forme-substance et phonétique-phonologie sous la direction de Nguyen *et al.* (2005), pour une littérature francophone ; et dans Pierrehumbert (2000) ou encore Davis *et al.* (2002), pour une littérature anglophone.

1.4.1 La phonétique, discipline de la substance

En linguistique, l'étude de la substance, c'est-à-dire des modalités de production et de perception de la parole, est allouée à la phonétique qui étudie les sons du langage dans leurs aspects acoustiques et articulatoires. Mais les sons constituent des objets instables, pouvant se diviser en caractéristiques indénombrables. Par exemple, un [a] ou un [i] en français différera selon le sexe, l'âge, la morphologie de la personne, le contexte phonétique mais aussi selon le contexte environnemental etc. Ainsi, au début du XX^{ème} siècle, en l'absence de cadre théorique global permettant d'organiser la multitude des observations, les travaux des phonéticiens se révèlent aux linguistes comme un amas de données chaotique. Dans les années 1940, Jakobson constatera : *La phonétique motrice et acoustique se sont révélées incapables de nous guider dans ce chaos [décomposition des sons en fractions diverses et indénombrables] et de dégager les caractères pertinents, les éléments constitutifs des sons et L'image rebutante de la multitude chaotique posa avec nécessité le principe antithétique, celui d'une unité organisatrice.* (Jakobson, 1976, p.36 à 38). Cette unité organisatrice, c'est le phonème, objet d'étude de la phonologie.

1.4.2 La phonologie, discipline de la forme

Cet échec de la phonétique à structurer ses résultats a amené les linguistes à proposer une seconde démarche étudiant non plus le son pour soi mais sous l'angle de sa fonction dans la langue. Le précurseur de cette démarche fut Baudouin de Courtenay qui mit l'accent sur le rôle des sons dans l'agencement des mots via leur caractère distinctif. Ainsi, pour De Saussure, la fonction du son est de permettre de différencier des chaînes sonores les unes des autres afin de distinguer des signifiants (les mots qui désignent) et donc, des signifiés (les objets désignés) (Bally *et al.*, 1995). L'unité distinctive, le phonème, n'existe que par sa valeur distinctive dans une langue donnée et donc, ne peut être étudiée que relativement au système phonologique de cette langue que constitue l'ensemble de ses phonèmes. Les langues sont donc des systèmes structurés. Par exemple, la distinction entre le e fermé "é" [e] et le e ouvert "è" [ɛ] est pertinente en français, le signifiant "thé" ayant un signifié différent de celui du signifiant "taie". [e] et [ɛ] sont donc deux phonèmes de la langue française qui se distinguent par le degré d'ouverture de la mâchoire. En revanche, en Russe, cette distinction n'étant pas pertinente, les signifiants "thé" et "taie" renverront à un même signifié pour un locuteur du Russe (Jakobson, 1976).

1.4.3 Evolution des relations phonétique-phonologie et substance-forme

Les linguistes structuralistes ont asservi l'étude de la parole (objet de la phonétique) à celle de la langue (objet de la phonologie). Pour De Saussure, l'étude de la parole est secondaire car la parole n'est pas essentielle au langage, elle n'est

qu'un média imposé par la nature : *les organes vocaux sont aussi extérieurs à la langue que les appareils électriques qui servent à retranscrire l'alphabet morse sont étrangers à cet alphabet, et la phonation, c'est-à-dire l'exécution des images acoustiques n'affecte en rien le système lui-même...* (Bally et al., 1995, p.36). L'objet de la linguistique, c'est la langue en tant que système conventionnel, acquis et organisé. Selon Jakobson avec l'étude de l'appareil vocal, *les linguistes se sont mis à tâtonner dans la physiologie* (Jakobson, 1976, p.26) et la phonétique acoustique n'est pas une meilleure candidate que la phonétique motrice pour classer les sons. Par conséquent, la description motrice, acoustique et auditive de la matière phonique doit se subordonner à son analyse structurale et donc, *la phonétique, discipline auxiliaire, doit être mise au service de la phonologie* (p.115). Dès lors, la recherche en phonologie va décrire les systèmes phonologiques des langues et mettre en évidence certains universaux et règles structurales (la forme). L'évolution de la recherche sur le cerveau et du courant cognitiviste va rapidement poser la question des bases cérébrales de ces universaux et de ces règles. Dans les années 1960, Chomsky introduira GU, la Grammaire Universelle, structure cérébrale propre de l'espèce humaine, génétiquement déterminée qui définit les principes de la langue que l'expérience paramétrise. La forme est donc dans le cerveau et la substance exécute. Cette approche s'accorde avec les conceptions top-down (du haut vers le bas) du cerveau qui dominent les sciences cognitives de l'époque.

Mais la fin du XXème siècle est, comme nous l'avons vu, marquée par un vent de renouveau épistémologique effaçant la frontière entre le corps et le cerveau et par conséquent entre la forme et la substance. Ces évolutions vont interpeller une partie de la communauté scientifique étudiant le langage en général, dont les linguistes et/ou les phonéticiens intéressés par la structure des systèmes phonologiques. Dans ce contexte, en 1972, Liljencrants et Lindblom vont affranchir la phonétique de la phonologie en introduisant une démarche visant à dériver la forme de la substance, ou encore l'organisation structurée des systèmes linguistiques, des modalités de perception et de productions de la parole. En effet, seule une explication des formes du langage par des facteurs externes permettra de les expliquer en rompant avec la circularité induite par la distinction forme-substance à savoir, *les formes du langage sont telles qu'elles sont car le langage est tel qu'il est...*

1.5 ENCORPORATION DE LA PAROLE ET DU LANGAGE

Dans une perspective d'unification de la forme et de la substance, les systèmes phonologiques des langues du monde sont supposés prendre forme sous l'influence d'un ensemble de contraintes de différents types et dans les activités de perception et de production de la parole. Ces contraintes étant partagées par tous les individus de l'espèce humaine, normalement constitués et vivant en groupe, leur action conjointe expliquerait les universaux observés dans les langues du monde. L'objectif de la recherche sur le langage adoptant cette perspective va être de déterminer

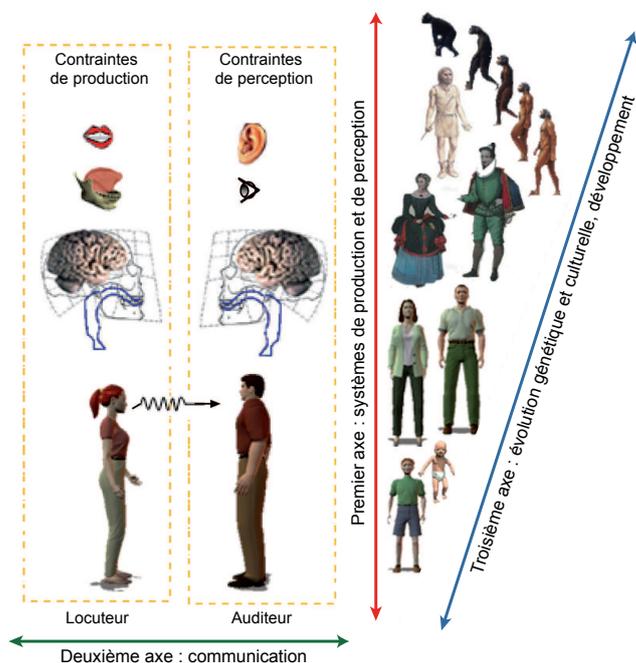


FIG. 1.3 – **Trois axes pour une encorporation de la parole.** Une théorie globale pour comprendre la parole et le langage devra considérer au moins trois axes de détermination substantielle. Le **premier axe** est centré sur l'organisme qui produit et perçoit la parole. Il étudie notamment les organes de production (exemple : lèvres, mâchoire et langue) et leur contrôle moteur ; la perception visuelle et auditive ainsi que les structures cérébrales qui les sous-tendent. Le **deuxième axe** s'intéresse à la parole dans le processus de communication. Son objectif est de comprendre comment les propriétés organiques peuvent être exploitées pour optimiser la communication par la parole. Le **troisième axe** étudie les origines génétiques et culturelles de la parole du point de vue phylogénétique et ontogénétique.

ces contraintes et d'élaborer des modèles pour tester leur validité via la simulation informatique.

1.5.1 Trois axes pour encorporer la parole

Globalement, les contraintes substantielles impliquées dans la détermination des formes de la parole peuvent être classées selon trois axes, représentés sur la figure 1.3.

Produire et percevoir. Le premier axe concerne les propriétés des organes de production et de perception, centraux et périphériques. Les recherches qui l'étudient vont décrire les configurations des articulateurs correspondant à la production des sons, élaborer des modèles biomécaniques des mouvements des articulateurs et de leur contrôle moteur et modéliser le système de perception.

Communiquer. Le deuxième axe comprend les facteurs liés à la situation de communication. Il étend l'objet d'étude corps-cerveau d'un organisme à au moins deux organismes et se centre sur la finalité communicative de la parole. La question simplifiée peut alors être posée en ces termes : parmi toutes les possibilités offertes par le corps et le cerveau, quelles sont celles qui optimisent la communication parlée ? Ainsi, selon Liljencrants et Lindblom, les systèmes sonores sont optimisés dans le cadre de l'efficacité communicative. Ils se développent notamment de manière à faciliter la production et la perception, les intérêts du locuteur et de l'auditeur.

Apprendre à utiliser un système de communication pré-existant. Le troisième axe considère l'évolution génétique et culturelle ainsi que le développement ontogénétique. Par exemple, selon Redford *et al.* (2001), les pressions sociales telles que l'identité d'un groupe et son histoire sont des forces stabilisatrices d'un système linguistique alors que d'autres pressions telles que l'impérialisme culturel ou économique peuvent induire des évolutions de ce même système.

La parole peut alors être considérée comme le résultat d'une interaction complexe entre les facteurs de ces différents axes, fortement imbriqués et qui interviennent sur des échelles de temps variées. Une telle complexité implique de se restreindre à certains facteurs et de les étudier d'abord de manière isolée afin de déterminer leur mode de fonctionnement et de délimiter leur action. La validité de ces facteurs et leur puissance prédictive concernant l'organisation des systèmes sonores des langues du monde peuvent ensuite être testées via l'expérimentation et la simulation informatique.

1.5.2 Simuler l'émergence des formes de la parole

Deux grands types de simulations ont été utilisées pour simuler l'émergence des formes phonologiques des langues du monde : les approches globales qui utilisent l'optimisation de contraintes et les approches locales qui utilisent l'auto-organisation (Berrah, 1998).

Approches globales. Les approches globales conçoivent les langues comme des solutions optimales à un même problème d'optimisation de contraintes, suffisamment complexe pour admettre plusieurs solutions équivalentes (Redford *et al.*, 2001). La formalisation des contraintes et de leurs interactions se fait via des fonctions d'optimisation inspirées des modèles de la physique.

Approches locales. Contrairement aux approches globales dont le but est de résoudre un problème d'optimisation, les approches locales définissent des règles locales d'interaction entre agents plus ou moins réalistes. L'organisation émerge de ces interactions via des cycles d'apprentissage itératif.

Nous reviendrons plus précisément sur la simulation au chapitre 7.

Le travail de recherche rapporté ici est une contribution à l'identification des contraintes substantielles pouvant déterminer les formes phonologiques des langues du monde¹. Dans la suite des travaux effectués en Master (dont est issue la figure 1.3), la première partie de ce rapport concerne la gestualité orofaciale dans la parole. Le chapitre 2 montre comment cette gestualité est revenue sur le devant

¹Comme nous l'avons vu précédemment, ces contraintes sont aussi bien productives que perceptives. La sous-représentation des travaux sur les contraintes perceptives dans la suite de ce rapport n'a pas pour but de défendre la suprématie des contraintes productives, ni d'alimenter une conception "motrice" de la perception. Elle résulte simplement d'un "zoom" sur la gestualité, puisque c'est sur cette dernière qu'ont porté nos travaux expérimentaux.

de la scène de la phonétique avec les travaux de Stetson et l'importance qu'elle a prise sur les plans théoriques et expérimentaux. Le chapitre 3 présente ensuite une recherche expérimentale mettant en relation les propriétés de la coordination inter-articulateurs avec un universel des lexiques phonologiques des langues du monde : la préférence pour les séquences lèvre-langue (cf. /pata/, consonne labiale puis coronale) plutôt que langue-lèvre (cf. /tapa/, consonne coronale puis labiale). La deuxième partie de cette thèse étend l'axe 1 du schéma de la figure 1.3 en y ajoutant la gestualité brachiomanuelle. Le chapitre 4 introduit une série d'arguments pour considérer la coordination entre la gestualité orofaciale et brachiomanuelle dans l'encorporation de la parole et plus généralement du langage. Les chapitres 5 et 6 présentent deux recherches expérimentales étudiant la coordination entre la mâchoire et la main dans une fonction capitale du langage : la deixis. Enfin, le chapitre 7 est consacré à la simulation et à la description du cadre général *système complexe* dans lequel s'intègrent nos travaux expérimentaux.

Première partie

Comment la gestualité orofaciale contribuerait à l'émergence des formes de la parole... et du langage



Résumé

Cette première partie s'intéresse au lien entre la motricité orofaciale exécutant la parole et les formes phonologiques des lexiques des langues du monde.

Le chapitre 2 introduit une démarche scientifique visant à déterminer le rôle du contrôle moteur des gestes orofaciaux dans l'émergence des formes phonologiques des langues. Cette démarche s'intéresse notamment aux bases motrices de la parole, et plus précisément aux contraintes motrices pouvant intervenir dans l'émergence des formes phonologiques et du langage. Couplés aux recherches neurologiques, les résultats des études qui en sont issues alimentent des théories ancrant l'origine du langage dans le contrôle moteur de la gestualité orofaciale, dont la plus élaborée en phonétique est certainement la théorie *Cadre puis Contenu* de Mac-Neilage (1998).

Dans le contexte de la théorie *Cadre puis Contenu*, le chapitre 3 propose et teste une nouvelle hypothèse pour expliquer la prévalence des séquences de type labiale-coronale (/pata/) par rapport aux séquences coronale-labiale (/tapa/). Cet *effet LC* pourrait être en partie le produit des modalités de la rencontre entre le contrôle moteur de la mâchoire (le *cadre*) et celui des articulateurs portés, la langue et les lèvres (le *contenu*). L'idée de base est que les propriétés motrices des différents articulateurs rendraient les formes labiale-coronale plus stables pour le système articuloire que les formes coronale-labiale. La tâche utilisée consiste en la répétition accélérée de séquences labiale-coronale et de leurs symétriques coronale-labiale. Les premières données acoustiques et articuloires enregistrées vont dans le sens d'une plus grande stabilité des formes labiale-coronale par rapport aux formes coronale-labiale.

De la motricité de la parole aux langues et au langage

L'être humain gesticule pour communiquer. Si on soumettait cette assertion à une population de culture française naïve quant aux recherches sur la parole, la première image qu'elle leur évoquerait serait probablement celle du locuteur de l'italien qui agite ses mains en parlant. Cependant, ils accepteraient aussi certainement l'idée que la parole est le premier média du langage. Mais, pour eux, comme pour de nombreux scientifiques dans l'histoire des sciences du langage, ce média serait d'abord sonore. Et pourtant, c'est la mise en mouvement du conduit vocal qui engendre le son et qui en détermine la forme, la précision et l'intensité. De nombreuses données montrent que la gestualité serait aussi utilisée pour la perception. Ainsi, après avoir été reléguée par la linguistique moderne à un rôle purement descriptif dans la caractérisation des systèmes phonologiques, l'étude des gestes de la parole est aujourd'hui un élément clé de la recherche visant à expliquer la genèse des formes des lexiques des langues. Cette progression doit beaucoup au développement de la connaissance sur le contrôle moteur. En effet, cette dernière a fourni un lien auparavant manquant entre l'organe susceptible d'héberger la forme, le cerveau, et les organes matérialisant cette forme, le conduit vocal. Ce lien bidirectionnel comble progressivement le fossé creusé par les linguistes entre forme/substance, langue/parole et phonétique/phonologie.

La première section de ce chapitre donne des éléments pour comprendre comment et pourquoi l'étude de la motricité de la parole est un des axes fondamentaux de la recherche actuelle sur le langage. La deuxième section introduit trois exemples d'approches étudiant le rôle des contraintes motrices dans la production de la parole et leur influence possible sur la composition des lexiques (axées respectivement sur l'estimation des coûts articulatoires, sur les structures coordinatives inter-articulateurs, et sur le rôle des gestualités de la mâchoire). Ces trois approches serviront de base à notre propre étude expérimentale, présentée au chapitre 3.

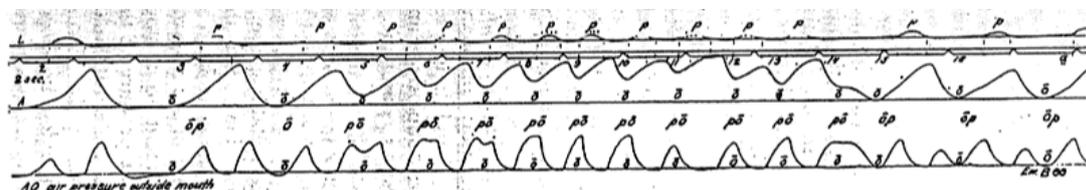


FIG. 2.1 – Progression d'une syllabe VC à une syllabe CV avec l'augmentation du débit. Quand la vitesse augmente, la répétition de « op-op » devient « po-po » pour retourner à « op-op » avec la décélération. La courbe du haut représente les mouvements de la lèvre et celle du milieu, la pression de acoustique dans la bouche. La distinction entre « op » et « po » apparaît sur la pression acoustique à l'extérieur de la bouche (en bas). Pour « op », la pression augmente pour le « p » puis redescend à zéro entre le « p » et le « o » suivant entre deux syllabes. Pour « po », la pression reste haute entre le « p » et le « o » suivant, d'après Stetson (1951).

2.1 L'IMPORTANCE DE LA MOTRICITÉ

L'opposition entre les aspects articulatoire et acoustique de la parole, qui oscille dans la littérature de la fin du XIX^{ème} et du début du XX^{ème} siècle entre une suprématie de l'articulatoire sur l'acoustique vs. de l'acoustique sur l'articulatoire, perd de sa puissance au cours du XX^{ème} siècle. En effet, si l'onde acoustique stimule l'oreille, les représentations qu'elle active ont une composante motrice. De plus, l'information gestuelle captée par le système visuel détermine le percept élaboré. Inversement, le paramétrage du contrôle de l'action s'acquiert grâce aux feedbacks auditif (Schwartz *et al.*, 2002) et somato-sensoriel (Tremblay *et al.*, 2003).

2.1.1 Stetson et l'essence motrice de la parole

Délaissée au début du XX^{ème} siècle avec le développement de l'acoustique, la phonétique articulatoire revient en force avec les nouvelles méthodes d'investigation permettant d'étudier le mouvement au cours de la production. Dès 1928, Stetson réintroduit la phonétique motrice et présente un ensemble de méthodes pour l'étude des mouvements de la parole. Il insiste sur l'essence gestuelle de la parole : *la parole est plus un ensemble de mouvements rendus audibles qu'un ensemble de sons produits par le mouvement*¹. La parole serait donc une branche de la gestualité humaine que l'oreille apprendrait à identifier. Selon Stetson, ce qui compte dans l'étude des gestes de la parole, c'est le mouvement en soi. Et le mouvement qui domine la parole est la succession d'ouverture et de fermeture de la mâchoire, liée à l'acte respiratoire et portant la syllabe CV.

Sur le plan expérimental, Stetson étudie l'évolution de la répétition accélérée de différents types de syllabes Consonne-Voyelle (CV) et Voyelle-Consonne (VC). Il constate que la répétition de « op » évolue vers « po », « eat », vers « tea » et

¹Speech is rather a set of movements made audible than a set of sound produced by movement, Stetson (1951), p.29

« up » vers « pu » (figure 2.1). D'après lui, pour une syllabe CV, le mouvement consonantique se coordonne avec le mouvement respiratoire alors que ce n'est pas le cas pour les syllabes VC. Cette différence expliquerait que la syllabe CV pourrait être produite plus rapidement que la syllabe VC. Toujours à l'aide du paradigme de répétition accélérée, Stetson va explorer la progression des répétitions de syllabes CVC en faisant aussi varier - ou non - les points d'articulation (cf. « pup » vs. « sam »). Il va montrer que, globalement, ces répétitions progressent en trois stades. Quand le débit est faible, la syllabe est articulée de manière régulière (cf. « pup-pup »). Avec l'augmentation du débit, la production évolue vers un stade de doublage de la consonne « puppup » pour aboutir à une fusion de la consonne finale avec la consonne initiale suivante (« pu'pu »). La phase de doublage de la consonne advient à une vitesse de 2 à 5 syllabes par seconde mais n'est pas toujours observée. La phase de fusion advient en moyenne vers 4 syllabes par seconde. Stetson note que la bascule en mode fusion se traduit par une augmentation nette du débit ; que le locuteur n'a pas conscience de la modification ; et qu'avec la décélération, il retourne à la forme initiale.

Ainsi, pour Stetson, la variation progressive du débit de production est une méthode d'investigation fondamentale en phonétique. Elle se justifie par le fait qu'en contexte écologique la parole tend systématiquement vers des vitesses importantes. De plus, la complexité de la parole impose à la phonétique d'utiliser un matériel contrôlé et simplifié afin d'extraire les caractéristiques importantes et d'alimenter des modèles simples et locaux. Ces modèles permettent de mieux comprendre les phénomènes de base impliqués dans la production de la parole. L'ensemble de ces remarques font de Stetson le précurseur de nombreux travaux visant à délimiter le rôle des contraintes motrices dans la constitution des lexiques, dont le travail expérimental présenté au chapitre suivant.

2.1.2 Motricité et perception de la parole

Avec Stetson, le mouvement revient en force dans l'étude de la production de la parole. Mais la découverte la plus importante ayant rendu l'étude de la motricité incontournable en parole est certainement celle du rôle de l'information articulaire dans la perception de la parole. En fait, plusieurs types d'arguments nouent les aspects sonores et gestuels de la parole avec, parmi les plus célèbres :

Le phénomène de lecture sur les lèvres qui concerne même les personnes sans trouble auditif. L'information visuelle complète et désambiguïse l'information acoustique ;

L'effet McGurk rapporté par McGurk et MacDonald (1976). Cet effet renvoie à l'illusion de percevoir /da/ au visionnement d'une vidéo du visage d'une personne de face prononçant /ga/ synchronisée avec une bande sonore diffusant le son /ba/ ;

La méthode Tadoma, utilisée par les personnes non-voyantes et non entendant.

Dans cette méthode, la parole est perçue via la récupération des gestes articulatoires en posant la main sur le conduit vocal ;

La découverte des neurones miroir chez le singe offre un lien substantiel manquant à la liaison entre la perception et l'action : il existe des neurones impliqués à la fois dans la perception et la réalisation d'une même action. Cette action peut être manuelle, comme un geste de saisie (Rizzolatti *et al.*, 1996) mais aussi orofaciale (Ferrari *et al.*, 2003). Un homologue du système miroir du singe existerait chez l'homme au niveau de l'aire de Broca, voir figure 4.1, p.77 ;

Le découverte du gène de la parole, FOXP2 sur le chromosome 7 (Lai *et al.*, 2001; Enard *et al.*, 2002). La mutation de ce gène chez la célèbre (malgré elle) famille KE se caractérise par des troubles de traitement du langage mais aussi par des problèmes dans le séquençage de la gestuelle de la parole. Les déficits verbaux (mais aussi non-verbaux) de la famille KE résulteraient d'une atteinte de la capacité à segmenter le mouvement ou de l'apprentissage procédural (Watkins *et al.*, 2002a,b).

2.1.3 Des théories gestuelles aux théories unificatrices

Théories accordant une primauté au geste et/ou à la production

Parmi les théories de la parole privilégiant la gestualité, la théorie motrice de la perception et la phonologie articulatoire sont souvent référées dans la littérature. Ces théories unifient les processus de perception et de production en leur supposant une seule et même unité : le geste ou la commande motrice qui l'exécute.

La théorie motrice de la perception suppose que la distinctivité perceptive ne serait pas dans le signal mais apprise par l'expérience linguistique consistant à lier des mouvements articulatoires avec le son produit. Le signal acoustique activerait la commande motrice requise à la production du son et percevoir ne serait ni plus ni moins que suivre en ligne des gestes du locuteur par un processus miroir (Lieberman et Mattingly, 1985). La découverte des neurones miroir constitue un argument supplémentaire en ce sens, selon les défenseurs de cette théorie.

La phonologie articulatoire de Browman et Goldstein (1989) introduit le geste articulatoire comme étant à la fois la primitive des représentations phonologiques et l'unité de production de la parole. En supposant une même unité pour les deux niveaux de description, cette théorie invalide la distinction phonologie-phonétique puisque décrire les unités distinctives de la langue revient à décrire les gestes de la parole. Cette approche fait l'hypothèse d'un invariant gestuel que le système perceptif serait capable d'extraire du signal. Or, l'idée selon laquelle percevoir serait comparer un invariant extrait du signal à une représentation mentale a été remise en cause au profit d'une incorporation plus large de la perception et de ses relations avec la production.

Théorie unificatrices de la perception et de la production

Dans leur démarche visant à dériver la substance de la forme, Liljencrants et Lindblom font l'hypothèse que les systèmes phonologiques des langues se constitueraient de façon à optimiser la communication en sélectionnant un système maximisant la distinctivité perceptive tout en minimisant le coût articulatoire. La production de la parole serait un compromis entre ces deux contraintes, aussi bien sur l'axe de l'évolution des systèmes phonologiques que sur l'axe individuel de la communication. La théorie de la perception pour le contrôle de l'action et la théorie Hyper et Hypo (H et H) sont deux exemples de théories élaborées dans ce cadre général.

Perception pour le contrôle de l'action. Si perception et action sont liées, cela ne veut pas dire qu'elles mettent en jeu les mêmes processus comme le suppose la théorie motrice. Pour Schwartz *et al.* (2002), le contrôle moteur et la perception de la parole seraient co-construits mais, contrairement à ce que défend la théorie motrice, cette co-construction ne consisterait pas à associer les mouvements articulatoires à des percepts mais à spécifier les paramètres du contrôle de l'action via la perception. Ainsi, la perception ne serait pas uniquement un miroir de l'action mais se développerait avec elle et les représentations qui la sous-tendent seraient multimodales et sensori-motrices.

Parole Hypo- et Hyper- articulée. Lindblom (1990a) rejette l'idée d'un invariant physique et privilégie une approche adaptative. Il étend les contraintes substantielles de la production de la parole à des données environnementales et contextuelles. Il distingue les contraintes orientées perception qui poussent la parole vers une hyper-articulation afin de maximiser la saillance perceptive et les contraintes orientées production qui poussent la parole vers une hypo-articulation afin d'économiser l'énergie. Les variations dans la production sont dues à ces phénomènes d'adaptation, à la plasticité du système et à des caractéristiques propres au locuteur. Mais la négociation entre contraintes perceptives et productives est une constante qui équilibre et structure le système.

2.1.4 Le contrôle moteur, une clé de l'unification forme-substance

Comme nous l'avons vu dans notre chapitre introductif, De Saussure et Jakobson insistent sur la nécessité de distinguer la langue de la parole et d'asservir l'étude de la deuxième à celle de la première. Mais si l'étude de la langue doit être indépendante de celle de la parole, De Saussure mentionne aussi l'interdépendance entre la parole et la langue, voire *la précédence* de la parole sur la langue. Pour lui, langue et parole se supposent l'une l'autre, tout en étant distinctes et devant être étudiées séparément. Quant à Jakobson, les traits distinctifs qu'il énonce pour décrire le système consonantique français sont tous des traits articulatoires². A la fin

²cf. nasales vs. non-nasales, voisée vs. non-voisée, occlusives vs. fricatives etc.

du XXème siècle, la découverte d'une influence bidirectionnelle entre le cerveau et le monde via les organes des sens et de l'action va donner un moyen à la substance d'agir sur la forme, et donc à la parole de déterminer la langue. Avant cette découverte, les scientifiques qui supposaient une influence de la substance sur la forme, aussi élaborées que pouvaient être leurs théories, se heurtaient au fait que les théories du cerveau ne fournissaient pas de moyen à la substance d'agir sur la forme, toutes occupées qu'elles étaient à établir la structure de notre ordinateur central de cerveau. Ainsi, si Stetson montre que la syllabe CV est plus *naturelle* pour le système de production que la syllabe VC, il faut attendre les théories du contrôle moteur, et la découverte de sa détermination par les propriétés en chair, en os et en muscles des organes qu'il déplace, pour utiliser ce résultat dans l'explication des systèmes phonologiques des langues du monde. L'influence de la substance sur la forme, jusqu'alors supposée, se concrétise. La phonétique motrice sort de sa condition de science de la description des formes pour devenir une science de l'explication des formes.

Ainsi, *la forme se fonderait dans la substance* et les phénomènes linguistiques, quelqu'ils soient, peuvent être appréhendés en termes d'interaction de facteurs. La primauté d'un facteur sur un autre ne devrait pas être présupposée mais démontrée via l'étude des interactions entre ces facteurs. Dans la suite, nous nous concentrerons sur l'étude des facteurs moteurs uniquement et plus spécifiquement sur la manière dont les propriétés biomécaniques du conduit vocal pourraient contribuer à déterminer la phonologie par l'intermédiaire du contrôle moteur.

2.2 LA SUBSTANCE DÉTERMINERAIT LA FORME PAR L'INTERMÉDIAIRE DU CONTRÔLE MOTEUR

Stetson ramène l'étude de la parole à un cas particulier de l'étude du mouvement. A partir de là, les recherches sur la production de la parole vont s'inspirer des résultats sur le mouvement en général et particulièrement sur ceux concernant les membres dans la locomotion. Dans ce cadre, Nelson *et al.* (1984) constate que la production de la parole est déterminée par le fait que la programmation motrice doit considérer des facteurs substantiels tels que les propriétés biomécaniques des articulateurs. De plus, l'influence de la biomécanique sur la commande motrice ne serait pas directe mais obéirait aux lois du contrôle moteur, comme le principe d'économie de l'effort. Une partie de la recherche consacrée aux mouvements des articulateurs s'est donc attelée à la quantification de l'effort articulaire via l'étude de la cinématique des mouvements. Mais comme l'avait introduit Stetson, le système de production de la parole est un système dynamique, les articulateurs peuvent se regrouper en structures coordinatives (ou synergies) qui évoluent sous la pression de facteurs externes. Ainsi, un autre pan de la recherche sur le mouvement des articulateurs va s'intéresser à l'aspect coordinatif de la production de la parole et fournir des descriptions du système en termes de relations et de transitions de

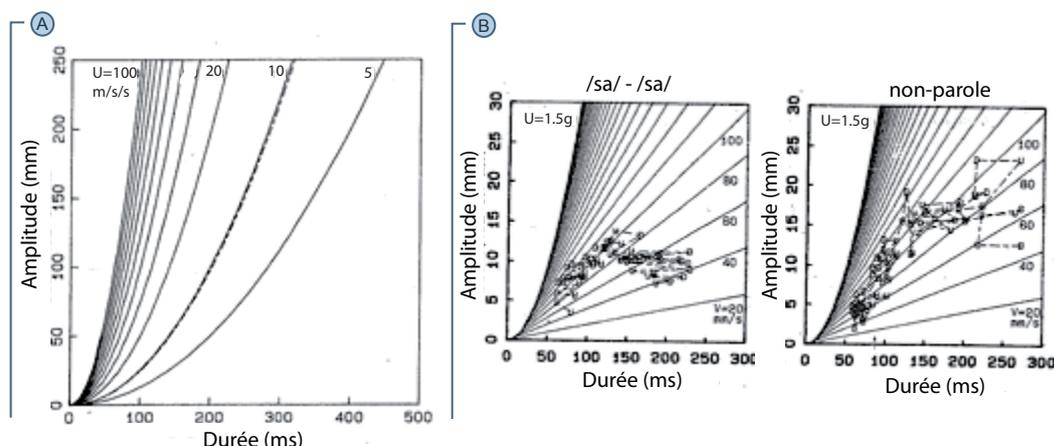


FIG. 2.2 – **Durée limite de mouvement dans l'espace durée-amplitude.** En A., les courbes représentent l'amplitude maximale du mouvement pour une accélération donnée (U) en fonction de la durée. La ligne en pointillés correspond à une valeur de U égale à l'accélération due à la gravité, d'après Nelson (1983). En B., mouvements d'ouverture et de fermeture de la mâchoire d'un locuteur répétant en accélérant la syllabe /sa/ (à gauche) et des mouvements de non-parole (oscillations verticales, à droite) représentés dans l'espace durée-amplitude. Quand le débit est lent, les mouvements de non-parole ont des amplitudes plus importantes que ceux des mouvements de parole. Quand la durée diminue, les deux distributions se rapprochent, l'amplitude diminue de façon à maintenir une accélération inférieure à 1.5g : la biomécanique de la mâchoire contraindrait les deux types de mouvement, d'après Nelson *et al.* (1984).

phase. Dans ce cadre, l'économie d'énergie pourra se ramener à la sélection d'un phasage préférentiel en réponse à des contraintes externes. Enfin, le contrôle moteur de la parole a été avancé comme une des clés des origines du langage.

2.2.1 Paramètres du mouvement et quantification de l'effort

Le mouvement des articulateurs s'étudie via la mesure de paramètres cinématiques tels que la durée, l'amplitude, la vitesse et l'accélération. A ce mouvement on peut associer un « coût » évaluant d'une manière ou d'une autre la consommation d'énergie requise à sa réalisation. Nelson (1983) introduit différentes mesures du coût du mouvement comme le pic de vitesse, l'accélération, ou la fréquence de changement de l'accélération (le jerk). L'optimalité d'un mouvement est un compromis entre la distance parcourue (amplitude), la durée, et l'accélération. Elle dépend de la tâche à effectuer. Dans l'espace durée-amplitude (ou temps-distance), et pour un mouvement sans frottement, la durée minimale (D) pour une accélération limite donnée (U) et une amplitude (A) s'obtient simplement en considérant le cas limite d'une accélération uniforme U pendant une première phase de durée $\frac{D}{2}$, puis une décélération uniforme -U pendant une seconde phase de même durée :

$$D = 2\sqrt{\left(\frac{A}{U}\right)}$$

Le graphique de gauche sur le figure 2.2, représente les courbes de relation entre durée limite et amplitude maximale selon différentes valeurs d'accélération. Pour

une accélération donnée, les mouvements réalisables sont à droite et en dessous de la courbe limite. Quand la durée du mouvement diminue, pour rester à une même valeur d'accélération, l'amplitude doit elle aussi diminuer. Si la tâche requiert de maintenir une amplitude importante malgré la diminution de la durée, l'accélération doit être augmentée. Cette représentation dans l'espace durée-amplitude a été utilisée par Nelson *et al.* (1984) pour comparer les mouvements de la mâchoire lors de la répétition accélérée de mouvements cycliques de non-parole (ouverture/fermeture de la mâchoire) aux mouvements de la mâchoire lors de la répétition accélérée de la syllabe /sa/. Les résultats obtenus pour un de leurs trois locuteurs sont représentés sur le graphique de droite de la figure 2.2, chaque point représentant un geste d'ouverture et un geste de fermeture. Pour ce locuteur, l'accélération limite est à 1.5 g (14.7 m/s^2) et l'amplitude des mouvements diminue avec la durée. Nelson et al. rapportent que les comportements des deux autres locuteurs sont différents avec une accélération limite à 0.5g (4.9 m/s^2) et de faibles amplitudes de mouvements. De plus, la différence observée entre les mouvements de la mâchoire dans la parole et la non-parole diminue avec la réduction de la durée des mouvements. L'amplitude des mouvements de la mâchoire varie moins que celle des mouvements de non-parole, ce qui peut s'expliquer par le fait que ces mouvements visent une cible phonétique (voir aussi Ostry et Flanagan, 1989, pour une étude comparant les mouvements de mâchoire en parole et non-parole).

L'évolution des paramètres cinématiques des mouvements avec l'accélération du débit de production a aussi été utilisée pour étudier la coordination entre articulateurs. Par exemple, Hertrich et Ackermann (2000) s'intéressent à la coordination lèvre-mâchoire et langue-mâchoire durant la répétition accélérée des syllabes CV (cf. /pa/, /pi/). Leurs résultats montrent que la variation de la vitesse n'affecte pas de la même manière l'amplitude des mouvements de la mâchoire et ceux de la lèvre inférieure. Selon les auteurs, la mâchoire et les constricteurs seraient donc contrôlés par des mécanismes distincts. De plus, leurs données rendent compte d'une grande variabilité entre leurs neuf locuteurs dans les effets de la vitesse mais aussi dans les effets de la hauteur de la voyelle (/i/ vs. /a/) et du lieu de l'articulation (/t/ vs. /p/).

La sensibilité des mouvements des articulateurs aux différents paramètres des mouvements et à la variabilité inter-individuelle est souvent utilisée comme contre-argument à l'idée d'un invariant gestuel. Cependant, la caractérisation cinématique des mouvements et ses échecs à mettre en évidence des régularités en ce qui concerne la manière dont les articulateurs se coordonnent motive une autre approche, l'approche en termes de phasage relatif entre articulateurs.

2.2.2 Structures coordinatives et phasage inter-articulateurs

La variabilité observée par les études caractérisant la coordination des articulateurs par des paramètres cinématiques rend l'extraction de principes de coordi-

nation difficile. C'est en partie ce constat qui a amené Kelso et ses collaborateurs à proposer une approche de la coordination des articulateurs dans la parole en termes de systèmes dynamiques.

De la dynamique partout

La démarche de Kelso est détaillée dans Kelso (1995), ce n'est pas une approche limitée à la parole mais à la description des comportements en général jusqu'au fonctionnement du cerveau. Kelso s'inscrit explicitement dans la révolution épistémologique des sciences de la complexité décrite dans le chapitre précédent. Sa démarche vise à combler les écarts entre les lois physiques décrivant les comportements des choses et celles décrivant les comportements des êtres vivants et plus précisément, de l'être humain. L'approche dynamique constitue selon lui une théorie unificatrice pour étudier l'émergence de la forme dans les systèmes complexes. Le contrôle moteur mais aussi la perception seraient gouvernés par les mêmes principes dynamiques de synergie, phasage, transition de phase etc. Un des axes des recherches de Kelso a visé à appliquer cette démarche à l'étude de la coordination des mouvements avec un intérêt particulier pour les articulateurs dans la parole.

Des mains qui galopent

Une étude souvent citée dans la littérature de la parole s'intéressant aux phénomènes d'économie d'effort et ayant inspiré les premiers travaux de Kelso sur la coordination du mouvement est celle de Hoyt et Taylor (1981) sur les allures du cheval. Hoyt et Taylor ont montré que la progression du pas au trot, puis du trot au galop avec l'augmentation de la vitesse s'accompagne d'une minimisation de la consommation d'oxygène, et donc d'énergie. D'autre part, l'augmentation de la vitesse fait basculer le système d'un mouvement alterné caractéristique du pas et du trot à un mouvement en phase caractéristique du galop. Haken *et al.* (1985) ont transposé cette progression à la coordination bi-manuelle. Dans leur étude, la tâche consiste à taper sur une table avec les index en alternant index gauche avec index droit. Ce mouvement est une analogie de celui du trot du cheval (« *Let your fingers do the walking* » Kelso, 1995, p.46)³. À partir d'une certaine vitesse, la répétition accélérée de ce mouvement alterné bascule vers un mouvement « en phase » dans lequel les deux index tapent simultanément sur la table, comme si les doigts « galopent ». Les auteurs ont proposé un modèle dynamique de ce phénomène en termes d'oscillateurs couplés connu sous nom de modèle HKB.

Phasage inter-articulateurs

Malgré le nombre important de composants neuromusculaires et de degrés de liberté impliqués dans la production de la parole, les formes qui sont produites sont correctes et distinctives. Selon Kelso *et al.* (1986), pour parvenir à gérer le nombre important de degrés de liberté, les structures motrices seraient contrôlées

³Littéralement : *Laissez vos doigts marcher*, il s'agit du titre de l'annonce qu'avait choisi Kelso pour le recrutement des participants à ses expériences.

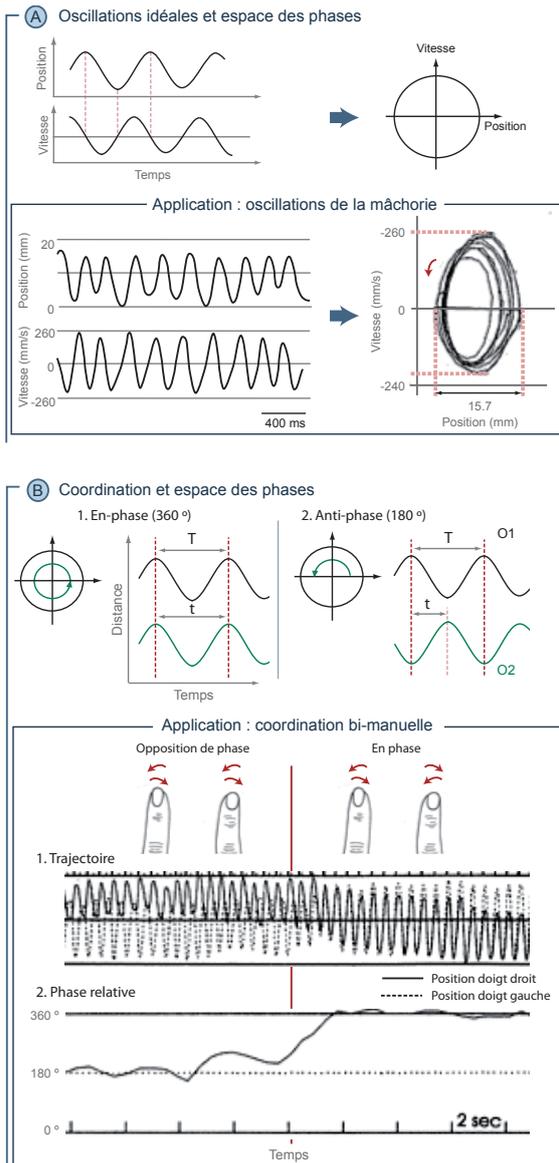


FIG. 2.3 – L'étude du mouvement dans l'espace des phases selon Kelso. En A., en haut : position et vitesse d'un oscillateur *idéal*. Son déplacement dans l'espace position-vitesse suit un cercle parfait. En bas, trajectoire et vitesse des mouvements de la mâchoire lors de la répétition de la syllabe /ba/ à un débit normal (Kelso *et al.*, 1985). Pour les mouvements réels, la trajectoire dans l'espace position-vitesse est un enchaînement de formes circulaires. En B., la méthode proposée par Kelso pour étudier la coordination entre deux systèmes oscillants (O_1 et O_2) consiste à choisir un oscillateur référant (ici, O_1) et à positionner le cycle de l'autre oscillateur (ici, O_2) relativement au cycle de l'oscillateur référant. Pour cela, un événement important du cycle de O_2 (ici, le pic de trajectoire) est positionné dans l'espace position-vitesse de O_1 . Cette mesure, ou *phase relative*, est une valeur angulaire. Soient T la période de O_1 et t la durée entre le début du cycle de O_1 et l'événement étudié alors, la phase relative se calcule par $t/T \times 360$, elle vaut 180° quand les deux systèmes oscillent en opposition de phase parfaite (O_2 est en position haute quand O_1 est en position basse) et 0 ou 360° quand ils oscillent en phase. L'étude de la phase relative au cours du temps permet de caractériser la coordination et de repérer des éventuels changements de phase. La figure du bas donne un exemple d'application à la coordination bi-manuelle, d'après Kelso (1995). Dans une tâche consistant à bouger les deux index de droite à gauche dans le même sens (mouvement en antiphase car les muscles homologues se contractent en alternance), avec l'accélération du mouvement, la coordination bascule vers une configuration où les index bougent en miroir (mouvement en phase). La phase relative (pic de trajectoire de l'index gauche par rapport au pic de trajectoire de l'index droit) passe de 180° à 360° .

comme des structures coordinatives ou *synergies* qui constitueraient l'unité de l'action. Les invariants dans la production de la parole pourraient donc être des stabilités coordinatives dont les formes résisteraient à des changements importants dans l'espace des paramètres, excepté dans les zones critiques où un faible changement entraînerait une transition de phase. Les études dynamiques de la production de la parole utilisent aussi le paradigme de répétition avec accélération. Cependant, elles ne s'intéressent pas aux paramètres cinématiques mais plutôt aux relations de phasage entre les articulateurs qui bougent en même temps, avec l'idée que le système évoluerait vers la coordination la plus stable avec l'augmentation de la vitesse. Ainsi, Kelso *et al.* (1985) réinterprètent les résultats de Stetson montrant la progression de la répétition d'une syllabe VC vers une syllabe CV en termes d'instabilité et de transition de phase. Leur explication est que la syllabe VC serait une forme instable alors que la syllabe CV serait un attracteur pour le système articulaire, d'où la bascule de VC vers CV plutôt que l'inverse. Kelso *et al.* (1986) vont aussi introduire une méthode pour étudier la coordination consistant à positionner les événements (début, fin de geste, pic de trajectoire ou de vitesse) d'un ou de plusieurs articulateurs sur le cycle d'un articulateur référent. Les mouvements ne sont donc plus représentés dans l'espace de leurs paramètres mais dans l'espace des phases relatives (voir figure 2.3). Comme Poincaré cherchait de l'ordre sous le désordre en introduisant l'espace des phases dans l'étude du problème des trois corps (voir encadré p. 6), Kelso et ses collaborateurs cherchent des régularités dans la coordination articulaire qui échapperaient aux approches cinématiques classiques. Ce paradigme a ensuite été repris lors de différentes recherches sur la parole (voir Sock et Löfqvist, 1995; Sock *et al.*, 1996)

L'idée selon laquelle la dynamique pourrait *tout* expliquer est en soit contestable (de la même manière que l'est celle du *tout* auto-organisé). La production de la parole ne peut être seulement considérée comme le produit d'un mouvement d'oscillateurs couplés, des facteurs cognitifs de haut niveau étant impliqués (comme l'attention, voir Treffner et Peter, 2002). Cependant, la méthode introduite par Kelso et la description de la coordination en termes de phasage relatif est une approche originale en parole et qui peut-être utilisée à des fins autres qu'une explication du *tout* dynamique. Elle fait partie des sources d'influence des études expérimentales présentées au chapitre suivant et dans la deuxième partie de ce rapport.

2.2.3 La mâchoire au centre du processus : la théorie Frame-then-Content

La théorie *Frame-then-Content* (*Cadre puis Contenu*) a été proposée par MacNeilage (1998) pour ancrer l'origine de la parole dans le contrôle moteur de la gestualité orofaciale. Cette théorie fournit un cadre pour expliquer les formes phonologiques par des contraintes substantielles et plus généralement, pour dériver le langage de la parole. Elle est issue des études développementales et exploite l'idée selon laquelle l'ontogenèse du contrôle moteur de la parole reproduirait sa phylo-

genèse.

Ontogenèse séquentielle du contrôle moteur de la parole

Les études du développement de la parole montrent que vers 3 mois, les bébés vocalisent (ou gazouillent) : c'est le stade des « arheu ». Vers 7 mois, ils commencent à babiller, ils répètent des séquences de type « bababa.. », « mamama... » : c'est le stade du babillage dupliqué. Vers 1 an, après avoir varié leur babillage, de « babababa » à « patapata », ils commencent à « parler », passant d'une activité rythmique à une activité sémantique. Ces stades vers l'acquisition de la parole sont universels. Selon MacNeilage et Davis (2000), cette séquence développementale témoigne d'une maturation nerveuse progressive et séquentielle du contrôle moteur des articulateurs. Ces auteurs font l'hypothèse que le bébé commencerait par contrôler les oscillations verticales de sa mâchoire. Du fait des dépendances anatomiques entre la mâchoire, la langue et les lèvres, ces oscillations porteraient la lèvre inférieure vers la lèvre supérieure, entraînant une fermeture labiale du conduit vocal. Associés à de la phonation, ces mouvements donneraient naissance à la protosyllabe CV (/ba/, /ma/,...). Ainsi, la mâchoire serait la porteuse de la parole et ses oscillations verticales constitueraient le *Cadre* (ou *Frame*) de la parole. L'acquisition du contrôle des articulateurs portés (langues, lèvres) et de leur coordination avec la mâchoire serait plus longue car plus complexe. Elle consisterait à ajouter un *Contenu* (ou *Content*) au cadre. Ce développement ontogénétique séquentiel du contrôle moteur de la parole (voir aussi figure 3.1, p. 37) proposé à partir d'observations des productions des bébés est appuyé par les études articulatoires montrant que les mouvements de la mâchoire de l'enfant de 1 an et de 2 ans sont plus proches de ceux de l'adulte que ne le sont ceux de la langue et des lèvres. Chez les jeunes enfants, les mouvements de la langues et des lèvres sont notamment plus variables et moins précis que ceux de l'adulte, Green *et al.* (2002, 2000).

Théorie darwinienne de la parole et du langage

D'après MacNeilage (1998) l'évolution de la parole aurait suivi les mêmes stades que son développement. D'abord vocalisant, nos ancêtres auraient progressivement développé le contrôle volontaire de leur mâchoire dans la communication puis, celui plus complexe des articulateurs portés et de leur coordination avec la mâchoire. Selon MacNeilage ce processus aurait suivi une évolution darwinienne, investissant le câblage neuronal de la gestualité orofaciale déjà en place pour l'alimentation (comme la sélection naturelle aurait détourné les panneaux solaires de la thermorégulation des oiseaux pour en faire des ailes pour voler). Ainsi, les aires cérébrales du contrôle moteur de l'alimentation se seraient progressivement spécialisées pour la parole. Au chapitre 4, nous verrons qu'une origine purement orofaciale de la parole et du langage est actuellement remise en cause par la découverte du rôle clé de la gestualité brachiomanuelle dans l'acquisition du langage et de la parole. Toujours est-il que la théorie *Frame-then-Content* est une source d'hypothèses et d'expérimentations riche et permet notamment d'expliquer certains universaux des

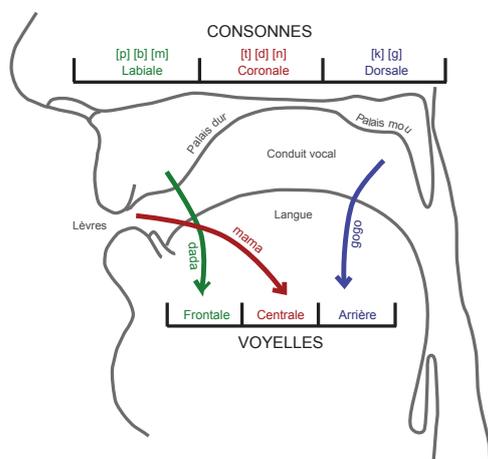


FIG. 2.4 – Trois formes de syllabes CV pour limiter le mouvement, d'après MacNeilage et Davis (2000). Le babillage et les langues du monde favorisent trois formes de syllabes CV (Consonne-Voyelle) : celles associant une consonne labiale à une voyelle centrale, une consonne coronale à une voyelle antérieure, et une consonne dorsale à une voyelle postérieure. Selon MacNeilage et Davis ces formes seraient favorisées car elles réduiraient les mouvements de la langue entre l'articulation de la consonne et de la voyelle.

systèmes syllabiques des langues du monde en termes de contraintes substantielles.

Frame-Content et les formes syllabiques

L'étude des systèmes syllabiques des langues du monde rend compte non seulement d'une préférence de la syllabe CV par rapport à VC mais aussi de préférences à l'intérieur des structures CV. Ainsi, trois co-occurrences consonne/voyelle prédominent dans les structures CV des langues du monde (Rousset, 2003; MacNeilage et Davis, 2000) :

1. Les consonnes bilabiales avec les voyelles centrales (cf. /ba/, /ma/);
2. Les consonnes coronales avec les voyelles antérieures (cf. /di/, /da/);
3. Les consonnes vélares avec les voyelles postérieures (cf. /gu/, /go/).

La dominance de ces trois formes caractériserait aussi le babillage et les premiers mots (Davis *et al.*, 2002; MacNeilage et Davis, 2000), amenant MacNeilage et Davis à élaborer une hypothèse explicative commune à l'ontogenèse et à la phylogénèse de ce phénomène. Cette hypothèse est que ces trois associations seraient privilégiées car elles minimiseraient les mouvements des articulateurs entre la production de la consonne et de la voyelle, ce qui simplifierait le contrôle (voir figure 2.4). Le même phénomène d'économie de mouvements et de simplicité motrice expliquerait certains universaux inter-syllabiques comme le fait que les formes associant une syllabe labiale (/pa/) à une syllabe coronale (/ta/), comme /pata/, sont plus fréquentes dans les lexiques des langues que leur symétrique, comme /tapa/. Ce phénomène, connu en tant qu'*effet LC* est l'objet du chapitre suivant.

Une étude expérimentale : rôle de la coordination articulatoire et effet Labial-Coronal

Ce chapitre reprend l'article « An articulatory basis for the Labial-to-Coronal effect : /pata/ seems a more stable articulatory pattern than /tapa/ »^a, Rochet-Capellan et Schwartz (2007). Les données analysées sont aussi celles présentées dans Rochet-Capellan et Schwartz (2005b, 2006). Des analyses de données supplémentaires sont disponibles dans Rochet-Capellan et Schwartz (2005a); Rochet-Capellan (2004). Ce travail n'aurait pu se faire sans l'aide de Christophe Savariaux et Coriandre Vilain et sans la participation de tous les locuteurs bénévoles.

^a« Une base articulatoire pour l'effet Labial-Coronal : /pata/ semble constituer une forme articulatoire plus stable que /tapa/ »

3.1 INTRODUCTION

Dans les inventaires phonologiques des langues du monde disponibles actuellement, les formes inter-syllabiques CV.CV (avec C=consonne et V=voyelle) de type « consonne labiale puis coronale » (lèvres puis langue, cf. /pata/) sont environ 2.5 fois plus fréquentes que leurs symétriques « coronale puis labiale » (langue puis lèvres, cf. /tapa/), voir Locke (1983); MacNeilage et Davis (2000); Rousset (2003). Cette asymétrie, connue sous le nom d'*effet Labial-Coronal* (ou *effet LC*), se retrouve dans les inventaires des premiers mots de l'enfant à l'âge d'environ 12-16 mois (Ingram, 1974; Locke, 2000; MacNeilage et Davis, 2000).

Dans une démarche de dérivation des formes de la substance, MacNeilage et Davis (2000) ont supposé que l'effet LC émergerait de l'auto-organisation de contraintes motrices et cognitives mises en présence au stade des premiers mots. Cependant, cette explication est fondée sur des prémisses contestables et n'a pas fait l'objet d'une évaluation expérimentale systématique. En gardant le principe

d'une démarche visant à établir les contraintes substantielles participant à l'émergence des formes, nous mettons ici en relation l'effet LC avec les propriétés du système coordinatif mâchoire-langue-lèvres. Notre hypothèse de base est que le séquençage « constriction labiale puis coronale » constituerait une forme de coordination plus stable pour le système articulatoire que l'inverse. Cette hypothèse peut être évaluée via une tâche de répétition accélérée habituellement utilisée pour étudier les propriétés des systèmes coordinatifs et fondée sur l'idée que le processus d'accélération amène le système en action vers sa forme de coordination la plus stable, dite *attracteur*.

Ici, nous proposons un scénario expérimental en deux temps afin de comparer la stabilité entre les formes labiales-coronales (L_aC_o) et coronales-labiales (C_oL_a)¹ : ① Détermination de l'attracteur de la répétition accélérée de bisyllabes CVCV L_aC_o (cf. /pata/) et C_oL_a (cf. /tapa/); et ② Etude des phénomènes articulatoires sous-tendant la progression vers l'attracteur. Ces deux temps se sont matérialisés à travers deux études expérimentales dont les résultats sont présentés puis discutés ici en tant qu'éléments pertinents pour la compréhension des phénomènes articulatoires impliqués dans l'effet LC et plus généralement dans l'organisation inter-syllabique de la parole.

3.2 L'EFFET LC : DES EXPLICATIONS À L'EXPÉRIMENTATION

3.2.1 L'hypothèse *simple first*

MacNeilage et Davis (2000) ancrent l'effet LC dans l'ontogenèse sérielle du contrôle moteur de la parole qui reproduirait l'évolution Frame-then-Content de la phylogenèse introduite par MacNeilage (1998), voir chapitre 2. Selon cette conception, illustrée de manière simplifiée sur la figure 3.1, au cours du développement ontogénétique, le contrôle moteur des oscillations verticales de la mâchoire, le *frame* ou *cadre*, précéderait le contrôle différencié des articulateurs portés (langue, lèvres et vélum), le *content* ou *contenu*. Cette précédenance du cadre sur le contenu induirait une dominance du cadre dans le babillage et les premiers mots se traduisant par une préférence pour la syllabe CV (Consonne-Voyelle, geste d'ouverture de la mâchoire) et une tendance à répéter la même syllabe (de « babababababa », vers 7 mois, à « papa », vers 12 mois). L'hypothèse forte de MacNeilage et Davis est, qu'en dehors de tout mouvement actif des articulateurs portés, les oscillations de la mâchoire fermeraient le conduit vocal par une mise en contact de la lèvre inférieure avec la lèvre supérieure, produisant des *proto-consonnes* ou *closants* labiaux (cf. /b/, /m/...). En revanche, les occlusions coronales et vélares nécessiteraient un geste actif de la langue, superposé à celui de la mâchoire.

¹Par la suite, L_aC_o sera employé comme adjectif pour labial(e)-coronal(e) et C_oL_a pour coronal(e)-labial(e)

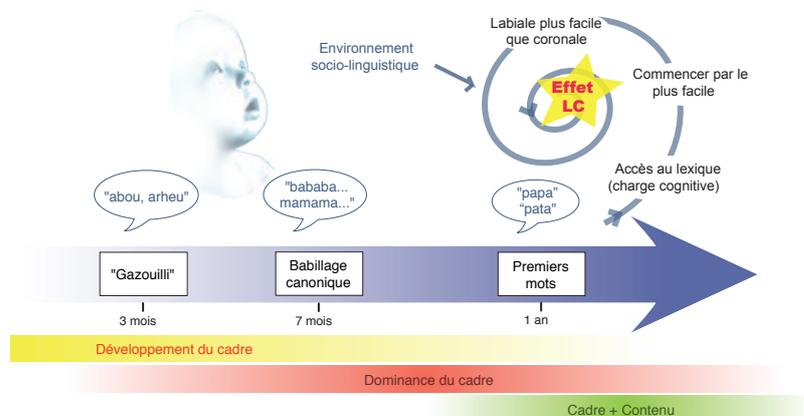


FIG. 3.1 – Développement sériel du contrôle moteur de la parole et explication *simple first* de l'effet LC. Selon MacNeilage et Davis (2000), au cours du développement, la maîtrise des oscillations verticales de la mâchoire (le *cadre* ou *frame*) précéderait le contrôle indépendant des articulateurs portés (langue, lèvres, dit *contenu* ou *content*). Ainsi, le babillage canonique serait sous la dominance du cadre. Dans ce contexte développemental, l'effet LC émergerait au stade des premiers mots de l'auto-organisation d'au moins trois contraintes : ① la charge cognitive nouvelle introduite par l'accès au lexique mental ; ② la tendance des systèmes moteurs à initialiser les séquences d'actions complexes par l'action la plus simple et ③ la supposition que les constrictions labiales seraient plus faciles à produire que les coronales. L'effet LC proviendrait donc d'une sorte de *régression* vers plus de simplicité motrice pour gérer la complexité cognitive.

Ainsi, les closants labiaux seraient des *cadre purs*, produits sur un simple geste de mâchoire et, de ce fait, seraient plus simples à réaliser que les closants coronaux ou vélaires. Sur la base de cette différence supposée entre les labiales et les autres formes de constrictions, MacNeilage et Davis considèrent l'effet LC comme un phénomène *simple first* (le plus *simple d'abord*) dont ils expliquent la raison d'être via un scénario d'émergence par auto-organisation impliquant au moins trois types de contraintes (voir l'extrémité droite de la figure 3.1) :

1. **Des contraintes cognitives** : l'accès au lexique mental requis à la production des premiers mots introduirait une charge cognitive nouvelle pour l'enfant ;
2. **Des contraintes motrices** : d'une manière générale, l'initialisation et la continuation des mouvements dépendent de systèmes de contrôle différents. Ce principe permet d'initialiser les mouvements de manière volontaire et de les poursuivre de manière automatique. En terme de charge cognitive, il serait donc moins coûteux de débiter une série d'actions complexes par l'action la plus simple.
3. **Des contraintes bio-mécaniques** : les consonnes labiales seraient plus simples à réaliser que les coronales. Les auteurs introduisent aussi l'existence possible de contraintes liées au changement de position de la langue, mais sans les définir ;

La logique de l'hypothèse *simple first* est donc que, pour faire face à la nouvelle charge cognitive (①), le système de la parole régresserait vers des formes phono-

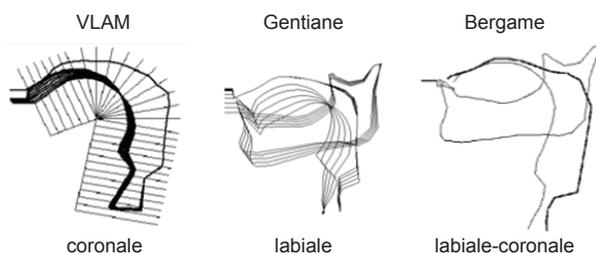


FIG. 3.2 – Simulation de la fermeture du conduit vocal par la mâchoire dans trois modèles articulatoires élaborés à partir de données cinéradiographiques de trois locuteurs différents. La fermeture obtenue est coronale pour le modèle VLAM, labiale pour le modèle Gentiane et labiale-coronale pour le modèle Bergame, d'après Vilain *et al.* (1999).

logiques plus simples à produire en termes de contrôle moteur (② et ③). Cette régression sélectionnerait les formes L_aC_o . MacNeilage et Davis supposent aussi une dimension culturelle à l'espace des contraintes, mais sans la définir. La préférence développementale pour les formes L_aC_o perdurerait ensuite dans la parole adulte et donc, dans les lexiques des langues du monde.

3.2.2 Limites de l'hypothèse *simple-first*

L'intérêt de l'hypothèse *simple-first* est d'abord de sortir l'effet LC de la catégorie des universaux innés du langage en tenant compte des contraintes substantielles. Cependant, elle dérive d'observations développementales et n'a pas été évaluée expérimentalement. Ainsi, l'assimilation des labiales à des *cadres purs* et l'ancrage de l'effet LC dans le développement se révèlent être deux présupposés contestables.

Problème de l'assimilation entre fermeture labiale et cadre pur

L'hypothèse *simple-first* repose sur l'idée que les consonnes labiales sont des *cadres purs* : langue et lèvres passives, les oscillations verticales de la mâchoire induiraient systématiquement une fermeture labiale du conduit vocal. Or, loin de rendre compte d'une dominance des formes consonantales labiales, les inventaires du babillage montrent que les formes coronales comme /da/ sont plus fréquentes que les labiales (Locke, 1983). De plus, le type de contact produit par la fermeture de la mâchoire pourrait dépendre d'abord de la morphologie du locuteur. Cette hypothèse a été testée par Vilain *et al.* (1999) via des simulations sur trois modèles articulatoires du conduit vocal correspondant à trois locuteurs adultes avec des morphologies différentes : les modèles VLAM (Maeda), Gentiane et Bergame. En introduisant une commande de fermeture de la mâchoire tout en maintenant les autres articulateurs passifs, le modèle VLAM a produit un contact labial, le modèle Gentiane, un contact coronal et le modèle Bergame, un contact labial et coronal, quasiment en synchronie (voir figure 3.2). Ainsi, les proto-syllabes /da/ et /bda/ pourraient être des *cadres purs* tout autant que /ba/.

Limites d'une origine développementale pour l'effet LC

Par ailleurs, l'hypothèse de MacNeilage et Davis ne considère pas les propriétés du système de production adulte. Or, Munhall et Jones (1998) ont montré que

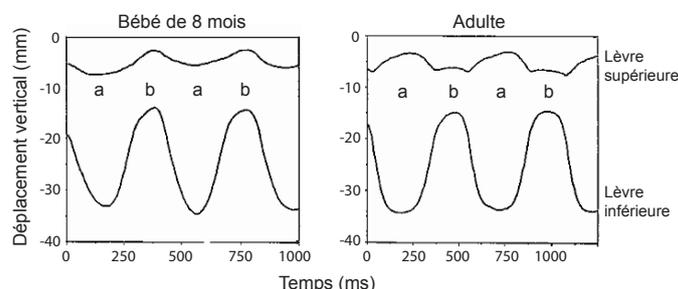


FIG. 3.3 – Déplacements verticaux des lèvres inférieure et supérieure lors de la production répétitive de « bababa » par un bébé de 8 mois (gauche) et un adulte (droite). Chez le bébé, les lèvres sont passives : la lèvre inférieure est portée par la mâchoire et fait *rebondir* la lèvre supérieure. Chez l'adulte, les lèvres sont actives et indépendantes, d'après Munhall et Jones (1998).

chez l'adulte, la consonne plosive labiale /b/ dans /babababa.../ est réalisée sur un mouvement actif des lèvres, superposé au mouvement de la mâchoire, comme on peut le voir sur la figure 3.3. Cette observation invalide la notion de *cadre pur* pour le système de production mature. D'une manière plus générale, l'hypothèse *simple first* néglige les phénomènes d'anticipation et de co-articulation propres à la parole adulte. Or, ces phénomènes encouragent à concevoir le contrôle moteur de la parole de manière globale et à considérer les propriétés des structures coordinatives (Kelso *et al.*, 1986; Nittrouer, 1991) impliquées dans la production des séquences $L_a C_o$ et $C_o L_a$ et pouvant contribuer à l'émergence de l'effet LC.

3.2.3 Une hypothèse alternative pour l'effet LC

Comme nous l'avons introduit au chapitre 2, pour une structure coordinative donnée, c'est-à-dire pour un ensemble d'effecteurs agissant de concert pour réaliser une séquence de gestes, certaines formes de coordination sont plus stables et économiques que d'autres. Ces formes constitueraient des attracteurs pour le système en action soumis à des contraintes externes telles que la vitesse. Ainsi, dans l'exemple de la figure 2.3. p.30, le mouvement en anti-phase des deux index évolue vers un mouvement en phase. Appliqué à l'étude de la parole, ces recherches ont engendré l'hypothèse selon laquelle, sur le plan de la production, les langues préféreraient les formes phonologiques intra- et inter-syllabiques les plus stables et économiques à produire et que le système articulatoire serait *attiré* vers ces formes sous l'effet de la vitesse². Ainsi, l'effet LC résulterait en partie de la plus grande stabilité de la coordination labiale puis coronale plutôt que coronale puis labiale pour le système mâchoire-langue-lèvres.

Sato *et al.* (2006) fournissent un premier argument allant en ce sens. Selon ces auteurs, la réalisation des séquences $L_a C_o$ dans les syllabes CCV (/psə/) ou

²On se limite ici aux aspects moteurs de la parole mais l'influence de la contrainte de stabilité et d'économie sur l'émergence des formes phonologiques interagit certainement avec d'autres facteurs comme la discriminabilité perceptive et/ou la stabilité perceptive (voir Sato *et al.*, 2007). Cependant, dans une tâche de répétition accélérée, le système est en grande partie *libéré* de la contrainte perceptive : le locuteur est son seul auditeur et la consigne encourage à aller le plus vite possible, sans égard pour la sortie.

CVC (/pəs/) se caractériserait par une anticipation de la constriction coronale dans la constriction labiale alors que l'inverse serait plus difficile. Cette asymétrie résulterait du fait que dans une séquence L_aC_o comme /psə/, le mouvement de la pointe de la langue pour le /s/ pourrait être préparé durant l'occlusion labiale alors que dans une séquence C_oL_a comme /spə/, les lèvres doivent rester ouvertes pour la production du /s/ avant de pouvoir se refermer pour le /p/. La synchronisation langue-lèvres serait donc plus difficile à réaliser que la synchronisation lèvres-langue (voir aussi Byrd, 1996; Zsiga, 1996; Surprenant et Goldstein, 1998; Chitoran *et al.*, 2002). Cette hypothèse en terme d'asymétrie de stabilité articulatoire présente au moins deux avantages : elle est compatible avec les propriétés du système de production mature et en maturation et elle est directement évaluable expérimentalement via le paradigme de répétition accélérée.

3.2.4 Un scénario expérimental

Le paradigme proposé ici pour évaluer l'hypothèse d'une plus grande stabilité articulatoire des formes L_aC_o relativement à leur symétriques C_oL_a comprend deux étapes. Une première étape d'étude comparative de la stabilité des formes L_aC_o et C_oL_a via le paradigme de répétition accélérée et une mesure permettant d'analyser les données d'un échantillon de locuteurs assez important. Une deuxième étape d'étude des phénomènes articulatoires impliqués dans le processus de répétition accélérée, restreinte à un échantillon de locuteurs de petite taille étant donné le coût en temps d'enregistrement et de traitement des données articulatoires.

Quel attracteur pour la répétition accélérée des formes L_aC_o et C_oL_a ?

L'utilisation du paradigme de répétition accélérée suit le raisonnement suivant. D'abord, la répétition lente de bisyllabes C_1VC_2V (cf. /pata/, /tapa/) devrait se caractériser par une alternance régulière de gestes de la mâchoire pour C_1 et pour C_2 . L'augmentation progressive de la vitesse de production devrait ensuite rapprocher les deux constriction et entraîner le système soit vers un attracteur C_1VC_2V (/pata/→/pta/) - correspondant à une réduction de la première voyelle - soit vers un attracteur C_2VC_1V (/pata/→/tpa/) - correspondant à une réduction de la deuxième voyelle. En partant du principe que la répétition accélérée est une tâche essentiellement motrice, la sélection d'un attracteur plutôt que l'autre devrait dépendre principalement de deux types de contraintes : les propriétés accentuelles de la langue des locuteurs (ici, le français) et la stabilité articulatoire de l'attracteur.

Propriétés accentuelles de la langue. La langue française se caractérise par une accentuation de la dernière syllabe. Cette propriété devrait tendre à préserver le séquençage initial en attirant le système vers une forme C_1VC_2V . Ainsi, l'attracteur devrait être de type L_aC_o pour les bisyllabes L_aC_o (/pata/→/patà/→/ptà/) et C_oL_a pour les bisyllabes C_oL_a (/tapa/→/tapà/→/tpà/).

Stabilité articulatoire. Si les formes L_aC_o sont effectivement plus stables que leurs symétriques C_oL_a , le système devrait être attiré vers la forme L_aC_o

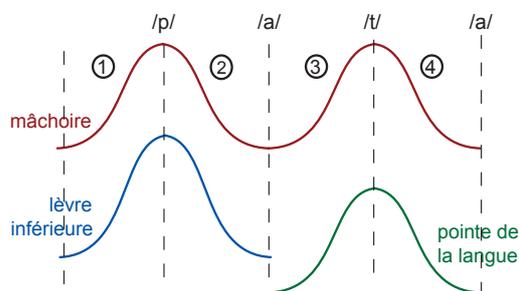


FIG. 3.4 – Mouvements de la mâchoire, de la lèvre inférieure et de la pointe de la langue attendus à faible débit. Chaque constriction devrait être portée par son propre cycle de mâchoire. En ① et en ② : mouvements de la mâchoire et de la lèvre inférieure pour la syllabe labiale (cf. /pa/). En ③ et ④ mouvements de la mâchoire et de la pointe de la langue pour la syllabe coronale (cf. /ta/).

pour les deux types de bisyllabes (cf. (/pata/-/tapa/) → /patà/ → /ptà/).

Cette étape d'évaluation de l'attracteur est l'objet de la première expérience rapportée ici, analysant les productions acoustiques de 28 locuteurs.

Quel substrat articulaire à l'évolution vers un attracteur ?

La phase de répétition lente d'une séquence /pata/, par exemple, devrait s'organiser de la manière suivante (voir figure 3.4) : ① Montée de la mâchoire et de la lèvre inférieure pour réaliser l'occlusion labiale (/p/); ② Descente de la mâchoire et de la lèvre inférieure pour réaliser la voyelle; ③ Montée de la mâchoire et de la pointe de la langue pour réaliser l'occlusion coronale (/t/); ④ Descente de la mâchoire et de la pointe de la langue pour réaliser la voyelle. Chaque constriction devrait donc être portée par son propre cycle de mâchoire et advenir sur la position haute de ce cycle. Le fait que les deux constriction soient réalisées par deux organes différents devrait permettre d'anticiper un geste dans l'autre et, avec l'accélération, faire évoluer le système vers une coordination d'un cycle de mâchoire pour deux constriction (cf. /ptá/). Cette progression pourrait se faire via une phase transitoire de deux syllabes sur un seul cycle de mâchoire (cf. /patá/). Cette possibilité a été négligée dans la littérature sur les mouvements des articulateurs étant donné que la plupart des études utilisent une tâche de répétition d'une même syllabe CV et donc, associent axiomatiquement la syllabe CV au cycle de mâchoire (Nelson *et al.*, 1984; Ostry et Flanagan, 1989; Perkell *et al.*, 2002; Perkell et Zandipour, 2002).

Cette deuxième étape est l'objet de la deuxième expérience rapportée ci-après, analysant les mouvements de la pointe de la langue, de la lèvre inférieure et de la mâchoire dans la répétition accélérée de bisyllabes $L_a C_o$ et $C_o L_a$ par cinq locuteurs.

3.3 UNE ÉTUDE COMPARATIVE DE LA STABILITÉ DES STRUCTURES L_aC_o ET C_oL_a

Cette première étude évalue l'hypothèse selon laquelle les formes L_aC_o seraient plus stables que les formes C_oL_a via le paradigme de répétition accélérée et sur la base d'une mesure acoustique de la réduction vocalique : la comparaison entre l'intensité des deux voyelles.

3.3.1 Méthode

Locuteurs

Les participants de l'étude étaient 15 locuteurs et 13 locutrices âgés de 19 à 34 ans avec le français pour langue maternelle et ne souffrant pas de trouble de la parole ni de problème auditif. Ils n'étaient pas informés des objectifs de l'expérience.

Matériel phonétique

Trois paires de bisyllabes L_aC_o et C_oL_a ont été sélectionnées : /pata/-/tapa/, /pasa/-/sapa/ et /fata/-/tafa/. Ce choix permet de contrôler un éventuel effet « plosive-fricative », c'est-à-dire la préférence pour les formes plosive-fricative (cf. /tfa/, /psa/) prédite par le principe de sonorité postulant une meilleure discriminabilité perceptive des formes plosive-fricative que fricative-plosive (Selkirk, 1984; Clements, 1990). La sélection de la voyelle /a/ repose sur le fait que sa réalisation nécessite une ouverture importante de la mâchoire sans implication significative de la langue ou des lèvres. Pour la passation de l'expérience, les six bisyllabes ont été arrangés de manière aléatoire en trois listes différentes.

Procédure

Les participants étaient assis face à un écran d'ordinateur relié à un micro. La consigne était de répéter en accélérant puis en décélérant les bisyllabes affichés à l'écran en suivant, autant que possible, le clignotement d'un carré noir-blanc-noir-blanc... et en produisant une syllabe sur le carré noir et une sur le carré blanc. L'objectif de ce métronome visuel n'était pas de contrôler précisément le débit des productions mais de donner des repères temporels globaux pour les phases d'accélération et de décélération afin d'éviter les énonciations *anarchiques* et d'inciter les participants à aller le plus vite possible. A cette fin, la fréquence du clignotement progressait de manière à devenir impossible à suivre en fin d'accélération. La consigne précisait qu'à ce stade, le débit maximal devrait être maintenu jusqu'au début du ralentissement du clignotement. Les participants étaient aussi encouragés à donner leur réponse sans s'interrompre, en dépit de la perception d'éventuelle(s) transformation(s) par rapport à la séquence initiale.

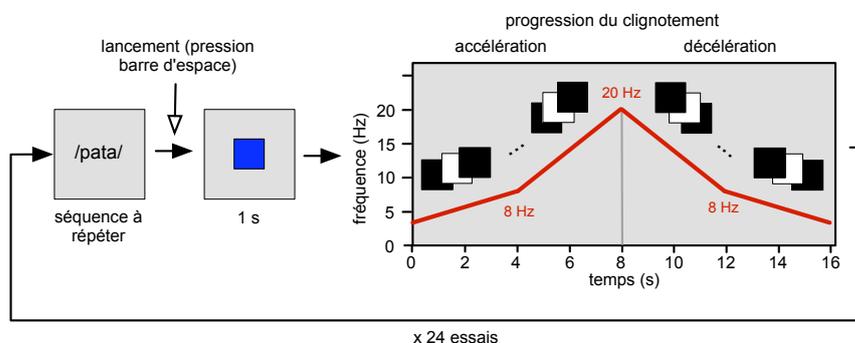


FIG. 3.5 – Déroulement d'un essai de l'étude acoustique. La séquence CVCV à répéter apparaît au centre de l'écran, le sujet démarre l'essai lorsqu'il est prêt en appuyant sur la barre d'espace. Après affichage d'un carré bleu pendant 1 s, le métronome visuel (clignotement d'un carré noir-blanc) démarre à 3.3 Hz, évolue à 8 Hz à mi-parcours de phase d'accélération (4 s) pour atteindre 20 Hz en fin de phase d'accélération (8 s). La progression est symétrique en décélération.

Après trois essais d'entraînement (*/vada/*, */daba/* et */dava/*), les listes expérimentales apparaissaient successivement. La figure 3.5 illustre le déroulement d'un essai. Le bisyllabe à répéter apparaissait au centre de l'écran et le lancement de l'essai était laissé à l'initiative du participant par pression de la barre d'espace. Cet événement induisait l'affichage d'un carré bleu pendant 1 s remplacé par le carré clignotant avec lequel la répétition devait débuter. La durée d'un essai était de 16 s réparties en 8 s d'accélération et 8 s de décélération (voir figure 3.5 pour la progression de la fréquence de clignotement). Certains locuteurs reprenaient parfois leur respiration en cours d'essai, puis poursuivaient leur énonciation avec approximativement le même rythme qu'avant cette interruption.

Mesures et hypothèses

Si les structures L_aC_o sont plus stables que les C_oL_a pour le système articulatoire, les répétitions devraient évoluer vers un séquençage de type L_aC_o jusqu'à atteindre une seule syllabe CCV de type L_aC_o , et ce, pour les bisyllabes L_aC_o et C_oL_a . Par exemple, si */pata/* (L_aC_o) est effectivement plus stable que */tapa/* (C_oL_a), le processus d'accélération devrait être : (*/pata/* ou */tapa/*) → */patá/* → */ptá/*. En revanche, s'il n'y a pas d'asymétrie de stabilité entre les deux formes, le fait que la langue française accentue la dernière syllabe devrait préserver le séquençage initial avec */pata/* → */patá/* → */ptá/* et */tapa/* → */tapá/* → */tpá/*. Finalement, les deux contraintes pourraient interagir et faire émerger les deux tendances avec une asymétrie articulatoire entre les attracteurs L_aC_o et C_oL_a superposée à un effet de séquençage initial. Les analyses décrites ci-après testent ces hypothèses via des mesures prosodiques basées sur la comparaison des intensités des deux voyelles. Les courbes d'intensité (énergie) du signal acoustique ont été obtenues pour chaque série de répétition en utilisant la fonction de calcul de l'énergie du logiciel Praat³. Les consonnes et les

³Praat est un logiciel libre développé par P. Boersma et D. Weenink de l'Institut des Sciences Phonétiques à Amsterdam. L'algorithme de la fonction de calcul de l'énergie acoustique utilise une

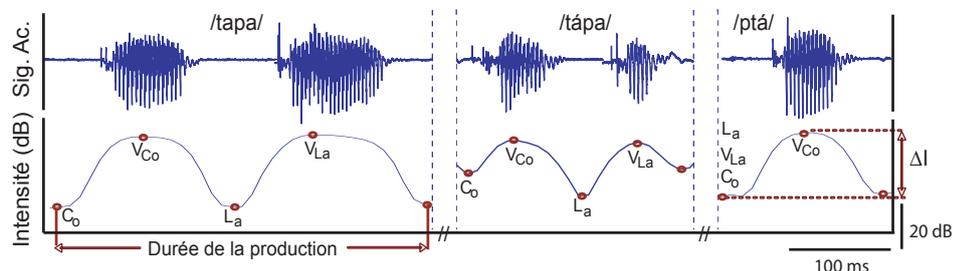


FIG. 3.6 – Etiquetage des courbes d'intensité et mesures. En haut : signal acoustique, en bas : courbe d'intensité en fonction du temps pour trois productions extraites d'une répétition accélérée de /tapa/. La répétition évolue progressivement de /tapa/ (à gauche) vers /tâpa/ (au centre) jusqu'à /ptá/ (à droite). L_a et C_o représentent le marquage des consonnes labiales et coronales et V_{La} et V_{Co} , le marquage des voyelles suivant ces consonnes. ΔI est la variation d'intensité entre V_{Co} et V_{La} .

voyelles ont été détectées automatiquement sur ces courbes, respectivement aux minima et maxima d'énergie, cet étiquetage a été ensuite validé manuellement (figure 3.6). En cas de réduction totale de la voyelle (production CCV) et pour maintenir la cohérence de l'étiquetage, la voyelle « disparue » et les deux consonnes ont été marquées au même minimum d'intensité (cf. figure 3.6, à droite). Les erreurs telles que la répétition d'une même syllabe ou les pauses ont été exclues de l'étiquetage et de l'analyse. Ces données ont permis de calculer deux paramètres pour chaque production CVCV ou CCV correcte : ① la durée de la production, temps écoulé entre la première consonne de la production et la première consonne de la production suivante, et ② la variation d'intensité entre les deux voyelles (ΔI), différence entre l'intensité de la voyelle après la consonne coronale (V_{Co}) et celle de la voyelle après la consonne labiale (V_{La}) :

$$\Delta I = I(V_{Co}) - I(V_{La})$$

Ainsi, dans les analyses ci-après, des valeurs positives de ΔI indiquent une réduction de la voyelle après la consonne labiale et donc, la présence de l'attracteur $L_a C_o$ (de type CVCV pour les valeurs de ΔI positives de faible amplitude et de type CCV pour les valeurs de ΔI positives de grande amplitude). Inversement, des ΔI négatives indiquent la présence de l'attracteur $C_o L_a$. L'hypothèse opérationnelle est donc que, si la forme $L_a C_o$ est plus stable que la forme $C_o L_a$, ΔI devrait devenir positive avec l'accélération quelque soit le type initial de la séquence.

3.3.2 Résultats

Les analyses concernent seulement les énonciations respectant l'alternance :

$$\ll \dots L_a V_{Co} V_{La} V_{Co} \dots \gg$$

avec L_a et C_o correspondant respectivement aux consonnes labiales et coronales et V à la voyelle /a/ ou à rien. Ce critère de sélection a exclu deux locuteurs ayant

fenêtre de Kaiser-20 de 42.6 ms (lobes latéraux inférieurs à -190 dB)

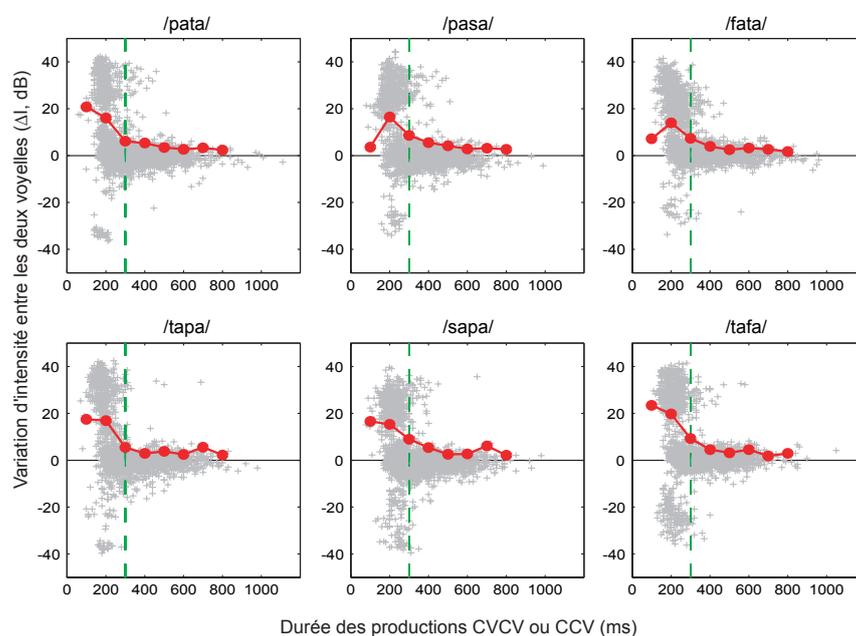


FIG. 3.7 – Variation d'intensité entre la voyelle après la constriction labiale et la voyelle après la constriction coronale (ΔI) en fonction de la durée des productions CVCV ou CCV de tous les sujets et par séquence. Un point du nuage représente une production CVCV ou CCV d'un sujet. La courbe superposée représente les écarts types de ΔI par intervalle de durée de 100 ms. En deçà de 300 ms, la dispersion augmente pour toutes les séquences.

tendance à répéter la même syllabe L_a ou C_o avec l'augmentation du débit (cf. /papapapa.../ pour /pata/), les analyses concernent donc 13 locuteurs et 13 locutrices.

Evolution de ΔI en fonction du débit : tendances globales

La figure 3.7 représente ΔI en fonction de la durée pour toutes les productions de tous les locuteurs. Globalement, les points se distribuent de manière similaire pour les trois paires L_aC_o et C_oL_a . Ces graphiques montrent que : ① les durées des productions s'étalent d'environ 100 ms à 700 ms ; ② pour les durées supérieures à 300 ms, les valeurs de ΔI sont proches de 0 dB ; ③ pour des durées inférieures à 300 ms, les valeurs de ΔI varient de -40 dB à +40 dB ; ④ les écarts types de ΔI (courbes rouges superposées aux nuages de points) augmentent nettement en deçà de 300 ms et ce, pour toutes les séquences étudiées. Pour la suite des analyses, 300 ms sera utilisée comme une durée seuil pour l'étude de l'attracteur.

Ces premières données montrent que, pour les débits lents, les locuteurs alternent les deux consonnes de manière régulière, sans réduction importante des voyelles. L'augmentation du débit induit ensuite une réduction de la voyelle après la consonne labiale (V_{L_a} , valeurs positives de ΔI) ou bien de la voyelle après la consonne coronale (V_{C_o} , valeurs négatives de ΔI). Par conséquent, les syllabes CCV L_aC_o (cf. /ptá/) et C_oL_a (cf. /tpá/) émergent toutes deux comme des attrac-

TAB. 3.1 – Analyse des variations d'intensité (ΔI) entre la voyelle après la consonne coronale et la voyelle après la consonne labiale pour les productions CVCV et CCV de durées inférieures à 300 ms pour les 42 essais analysés et pour chaque séquence. Les deux premières lignes correspondent aux moyennes et erreurs standards pour les 21 locuteurs. Les quatre dernières lignes présentent la répartition des moyennes positives et négatives avec, toutes les moyennes (lignes 5 et 6) et seulement les moyennes significativement différentes de 0 (lignes 6 et 8).

		pata	tapa	pasa	sapa	fata	tafa
Moy.		6.6 ^a	7.0 ^a	7.5 ^a	3.8	8.9 ^a	8.6 ^a
Err. std.		1.9	1.9	1.7	1.7	1.7	2.1
Nb. moy. positives	Toutes	30	30	27	21	37	33
	Signif.	19	15	18	13	26	21
Nb. moy. négative	Toutes	12	12	15	21	5	9
	Signif.	4	6	6	9	2	4

Comparaison avec 0, test de student bilatéral avec correction de Bonferroni :

^a $p \leq 0.01/6, ddl = 20$

teurs possibles pour le système. Cependant, les valeurs positives prédominent par rapport aux valeurs négatives pour les deux types de bisyllabes L_aC_o et C_oL_a . Cette tendance semble être globalement plus marquée pour les bisyllabes L_aC_o et pour les paires /fata/-/tafa/ et /pata/-/tapa/ par rapport à la paire /pasa/-/sapa/.

Analyse détaillée des attracteurs selon les bisyllabes et les locuteurs

Sélection des productions rapides. L'évaluation de la dominance de l'attracteur L_aC_o sur l'attracteur C_oL_a requiert une analyse détaillée des productions de durée inférieure à 300 ms. Or, le nombre de ces productions est très hétérogène d'un(e) participant(e) à un(e) autre. De plus, en dépit du guide visuel, une minorité de participants atteint rarement ce débit. Afin d'homogénéiser les données et de rendre les comparaisons inter- et intra-sujet(s) plus robustes, les analyses ont été restreintes aux participants ayant produit un minimum de cinq CVCV ou CCV de durée inférieure à 300 ms pour au moins deux essais sur les trois demandés et ce, pour chacun des six bisyllabes étudiés. Ce critère a sélectionné 11 locutrices et 10 locuteurs. Seuls les deux essais avec le plus grand nombre de productions de durée inférieure à 300 ms sont analysés ici. Les moyennes de ΔI sont calculées pour chacun des deux essais de chaque participant, ces deux moyennes permettent d'obtenir une moyenne par locuteur et par séquence puis une moyenne globale par séquence pour les 21 locuteurs. L'enjeu étant de déterminer s'il existe ou non une variation d'intensité entre les deux voyelles (moyenne significativement différente de 0) et si elle est positive (attracteur L_aC_o) ou négative (attracteur C_oL_a), l'analyse statistique utilise des t-tests bilatéraux avec correction de Bonferroni.

Dominance et répartition des attracteurs. Le tableau 3.1 donne les moyennes globales et la répartition des moyennes des 42 essais analysés. Ces données montrent que : ① Les moyennes globales sont positives pour tous les bisyllabes et diffèrent toutes significativement de 0 excepté pour /sapa/ (première ligne du tableau) ; ② Le ratio du nombre de moyennes positives (troisième ligne du tableau) sur le nombre de moyennes négatives (cinquième ligne du tableau) pour les 42 essais est de 2.5

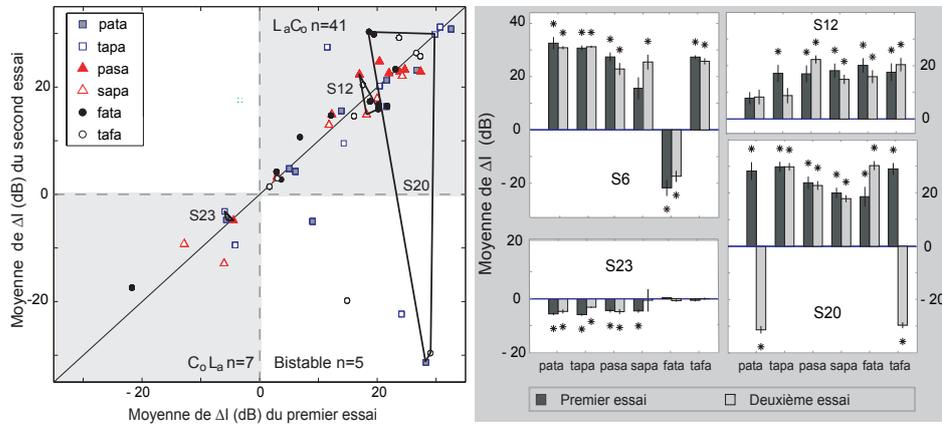


FIG. 3.8 – Stabilité de l'attracteur et profils de locuteurs. **A gauche** : moyenne de ΔI (variation d'intensité entre la voyelle après la consonne coronale et celle après la consonne labiale) du deuxième essai en fonction de celle du premier essai pour chaque séquence et pour les sujets dont les deux moyennes diffèrent significativement de 0 (test de Student bilatéral avec correction de Bonferroni, $p \leq 0.05/42$). Les points dans les zones en haut à gauche et en bas à droite correspondent aux cas où le locuteur a évolué vers des attracteurs différents pour ses deux essais. Les points dans les zones en haut à droite et en bas à gauche correspondent aux cas où le locuteur a évolué vers le même attracteur pour ses deux essais, respectivement L_aC_o et C_oL_a . **A droite** : moyennes de ΔI par séquence pour les deux essais de quatre locuteurs (les astérisques repèrent les moyennes significativement différentes de 0). S20 et S6 présentent un profil bistable, S12 un profil L_aC_o et S23, un profil C_oL_a . Les enveloppes convexes de S23, S20 et S12 sont représentées sur la figure de gauche.

(30/12) pour /pata/ et /tapa/, 7.4 (37/5) pour /fata/, 3.6 (33/9) pour /tafa/, 1.8 (27/15) pour /pasa/ et seulement de 1 (21/21) pour /sapa/. La différence entre la fréquence des moyennes positives et négatives est significative ($\chi^2(1) = 6.6$, $p < .001$); ③ La prévalence des moyennes positives sur les négatives est encore plus importante quand le calcul du ratio est restreint aux moyennes significativement différentes de 0 [(quatrième ligne) / (sixième ligne) du tableau]. Le rapport est alors de 4.75 pour /pata/, 2.5 pour /tapa/, 13 pour /fata/, 5.3 pour /tafa/, 3 pour /pasa/ et 1.4 pour /sapa/. Ici encore, la fréquence des moyennes positives est significativement différente de celle des moyennes négatives ($\chi^2(1) = 6.8$, $p < .001$). Ainsi, si les deux attracteurs L_aC_o (cf. /patá/, /ptá/) et C_oL_a (cf. /tapá/, /tpá/) sont possibles, l'attracteur L_aC_o émerge plus souvent que l'attracteur C_oL_a . Cette préférence est plus nette pour les bisyllabes L_aC_o que C_oL_a et pour les paires /pata-/tapa/ et /fata-/tafa/ que pour /pasa-/sapa/.

Variabilité intra-sujet. Il s'agit ici de délimiter le rôle éventuel du facteur sujet dans la détermination de l'attracteur, en d'autres termes : existe-t-il des profils de locuteurs et si oui, quels sont ces profils. La figure 3.8 représente les relations entre les moyennes de ΔI pour les deux essais sélectionnés, pour chaque locuteur et chaque séquence. L'analyse est restreinte aux cas pour lesquels ces deux moyennes diffèrent significativement de 0, soit 53 cas sur 21×6 . Les résultats montrent que les deux moyennes sont positives dans 41 cas, négatives dans 7 cas et ont des signes

différents dans 5 cas. La tendance dominante est donc d'évoluer vers un même attracteur pour les deux essais avec une prévalence nette de l'attracteur L_aC_o par rapport à l'attracteur C_oL_a . Cependant, dans certains cas, l'attracteur diffère d'un essai à l'autre. De plus, certains locuteurs ont un profil L_aC_o (cf. S12, figure 3.8, à droite) alors que d'autres, minoritaires, ont plutôt un profil C_oL_a (cf. S23). Enfin, certains locuteurs sont *bistables* : la catégorie de l'attracteur vers lequel ils progressent dépend de l'essai (S20) et du bisyllabe (S6).

En résumé, la répétition accélérée des bisyllabes CVCV L_aC_o et C_oL_a peut évoluer vers un attracteur L_aC_o ou C_oL_a . Cependant, et malgré la variabilité entre les séquences et les locuteurs, l'attracteur L_aC_o prévaut largement sur l'attracteur C_oL_a avec un attracteur [/*patá*/-/*ptá*/] pour /*pata*/ et /*tapa*/ ; [/*fatá*/-/*ftá*/] pour /*fata*/ et /*tafa*/ et [/*pasá*/-/*psá*/] pour /*pasa*/ et /*sapa*/.

3.4 UNE ÉTUDE ARTICULATOIRE DE LA STABILITÉ L_aC_o VS. C_oL_a

L'objectif est ici de comprendre comment le processus d'accélération entraîne le système articulatoire d'une forme CVCV L_aC_o ou C_oL_a vers un attracteur CCV L_aC_o ou C_oL_a sur la base de l'analyse d'enregistrements des mouvements de la mâchoire, de la langue et des lèvres lors de répétitions accélérées de séquences CVCV produites par cinq locuteurs.

3.4.1 Méthode

Locuteurs

Ces cinq locuteurs [trois femmes (S1, S3 et S4) et deux hommes (S2 et S5)] ont le français pour langue maternelle. Au moment de l'expérience, ils étaient âgés de 23 à 26 ans, ne présentaient pas de troubles auditifs ou de la parole et n'avaient pas participé à l'expérience précédente.

Matériel phonétique

Les bisyllabes CVCV L_aC_o et C_oL_a étaient les mêmes que ceux de l'expérience précédente (/pata/-/tapa/, /pasa/-/sapa/ et /fata/-/tafa/). Deux bisyllabes dupliqués (/papa/ et /tata/) supplémentaires ont été ajoutés comme items de contrôle afin de comparer la coordination entre les articulateurs dans le cas où les deux constrictions sont réalisées par le même organe avec celui où elles sont le fait de deux organes différents.

Procédure, enregistrement et traitement des données

Les séquences à répéter étaient énoncées par l'expérimentateur et le tempo n'était pas dirigé par un contrôleur externe du fait des conditions déjà difficiles de

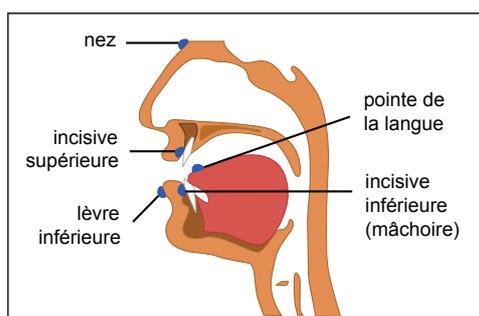
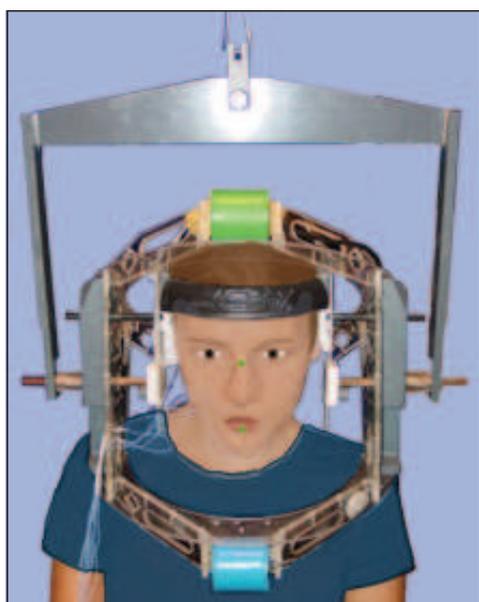


FIG. 3.9 – Dispositif pour l'enregistrement des mouvements des articulateurs par articulographie électromagnétique (EMA). L'EMA permet de suivre le déplacement de petites bobines réceptrices du champ électromagnétique calibré émis par trois grosses bobines de référence. Le système utilisé pour cette étude est l'EMA Cartens AG100, représenté sur la figure du haut. Il est composé d'un casque sur lequel sont fixées les trois bobines émettrices formant un triangle équilatéral et se positionnant respectivement au niveau du front, du menton et de la nuque de la locutrice. Ce casque est suspendu au plafond afin d'en alléger le poids pour le sujet. Sur la figure, on peut voir deux de ces bobines, celle du front (cylindre vert) et celle du menton (cylindre bleu). Un microphone fixé sur le casque permet l'enregistrement audio en parallèle de l'enregistrement du mouvement. Les bobines réceptrices sont collées dans le plan sagittal des articulateurs étudiés. Les signaux issus des bobines émettrices et réceptrices sont traités par une unité analogique connectée à un PC et le son est envoyé en parallèle sur l'entrée d'une carte acoustique. Les positions des bobines réceptrices sont calculées par rapport aux bobines émettrices ce qui permet de retrouver leurs coordonnées cartésiennes dans le casque.

Les positions des bobines réceptrices utilisées dans notre expérience sont illustrées sur la figure du bas. Pour la pointe de la langue, la bobine est positionnée à environ 1 cm du bout de la langue. Pour la lèvre inférieure, la bobine est positionnée à la limite entre la lèvre et la peau du menton. Enfin, pour la mâchoire, la bobine est positionnée entre les deux incisives inférieures. La bobine à la racine du nez et celle entre les incisives supérieures sont des points de référence pour la correction des mouvements de la tête lors du calcul post-expérimental des coordonnées des capteurs mobiles.

l'enregistrement des mouvements des articulateurs. Hormis ces différences dans le pilotage de l'expérience, la tâche des sujets était la même que dans l'expérience précédente. Il s'agissait de répéter de manière continue le bisyllabe CVCV en commençant à un débit lent et en l'augmentant progressivement jusqu'à une vitesse *aussi rapide que possible* puis de ralentir pour retourner au débit initial. La consigne précisait aussi d'essayer de ne pas s'arrêter au cours de l'essai et de continuer en dépit de toute impression de transformation. L'expérience comprenait trois essais (trois séries de répétition) pour chacun des huit bisyllabes soit 24 essais réalisés dans des ordres différents pour chacun des locuteurs.

Afin d'étudier la coordination entre la mâchoire, la pointe de la langue (réalisant la constriction coronale du /t/ et du /s/) et la lèvre inférieure (impliquée dans la constriction labiale du /p/ et labiodentale du /f/), les mouvements de ces articulateurs ont été enregistrés au moyen d'un Articulographe Electromagnétique (EMA) Cartens AG100 à une fréquence de 500 Hz. L'appareillage pour l'enregistrement est illustré et décrit sur la figure 3.9. L'enregistrement des mouvements de la mâchoire, de la pointe de la langue et de la lèvre inférieure se fait via trois bobines collées dans le plan sagittal du/de la participant(e) respectivement entre les incisives inférieures, à environ 1 cm de la pointe de la langue et sur le bas de la lèvre inférieure (figure 3.9, en bas). En plus de ces trois bobines, deux bobines de référence (à la racine du nez et entre les incisives supérieures) permettent de corriger les mouvements de la tête lors du post-traitement des données selon la procédure classique décrite par Hoole (1996). En bref, cette correction s'obtient en faisant coïncider, par translation et rotation, l'origine du repère sur la bobine des incisives supérieures. L'axe des ordonnées coïncide avec la droite joignant cette nouvelle origine à la bobine du nez et l'axe des abscisses est aligné avec le plan occlusal. Les coordonnées des trois bobines mobiles sont projetées dans ce nouveau repère. Ce sont les coordonnées ainsi obtenues que nous utilisons dans les analyses ci-après.

Enfin, le son était enregistré en parallèle via un microphone fixé sur le casque de l'articulographe, puis numérisé à 20 kHz.

Hypothèses

Les objectifs de cette étude sont d'évaluer l'hypothèse d'une réalisation de deux syllabes sur un seul cycle de mâchoire et d'étudier les phénomènes articulaires impliqués dans la progression vers l'attracteur.

1. La progression d'une coordination d'un geste de mâchoire par syllabe CV vers une coordination d'un geste de mâchoire par séquence CVCV devrait se produire avec l'augmentation de la vitesse. Elle résulterait de l'anticipation de l'une des constriction dans l'autre et devrait donc s'observer pour les six séquences L_aC_o et C_oL_a mais pas pour les deux séquences dupliquées (cf. /papa/ et /tata/).
2. D'après les résultats de l'étude acoustique, deux attracteurs sont attendus : L_aC_o et C_oL_a . Sur le plan articulaire, l'attracteur L_aC_o devrait se carac-

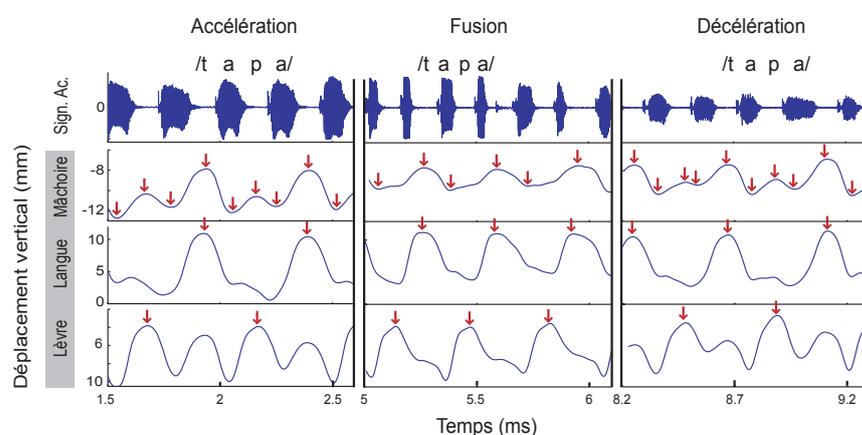


FIG. 3.10 – Etiquetage des données articulatoires. Signal acoustique, et déplacements verticaux de la mâchoire, de la pointe de la langue et de la lèvre inférieure au cours d'une répétition de /tapa/ par la locutrice S3. De gauche à droite : extrait de la phase d'accélération, de la phase avec un cycle de mâchoire pour deux syllabes (*fusion*) et de la phase de décélération. Les flèches sur les trajectoires repèrent les débuts et fins de strokes de la mâchoire et les constriction labiales et coronales. Les maxima secondaires non étiquetés sur les courbes de la langue et de la lèvre correspondent aux mouvements passifs induits par le geste de mâchoire accompagnant l'autre constriction.

tériser par un rapprochement entre la constriction labiale et la constriction coronale suivante. En revanche, l'attracteur $C_o L_a$ devrait se caractériser par un rapprochement entre la constriction coronale et la constriction labiale suivante (figure 3.12). Etant donné le petit nombre de locuteurs et la grande variabilité observée dans l'expérience acoustique, aucune attente n'est formulée concernant une éventuelle dominance de l'attracteur $L_a C_o$.

Mesures

La figure 3.10 présente un exemple de signaux acoustiques et articulatoires.

Traitement des données acoustiques. Le traitement des données acoustiques est le même que dans l'expérience précédente. La variation d'intensité ΔI entre les deux voyelles et la durée de chaque CVCV ou CCV correctement produit ont été calculées à partir de l'étiquetage des courbes d'intensité.

Traitement des données articulatoires. L'analyse articulatoire concerne les déplacements verticaux des articulateurs. Les maxima et les minima de ces trajectoires ont été repérés automatiquement aux valeurs 0 de la vitesse (dérivée première) de ces signaux, préalablement filtrés par un filtre passe-bas de Chebychev à une fréquence de coupure de 15 Hz.

Pour la mâchoire, cet étiquetage a permis de segmenter les trajectoires en mouvements (ou *strokes*) d'ouverture et de fermeture. Les signaux ont ensuite été visualisés afin de valider cet étiquetage et de repérer les périodes sans alternance régulière de gestes d'ouverture et de fermeture, propres à la production de parole sans

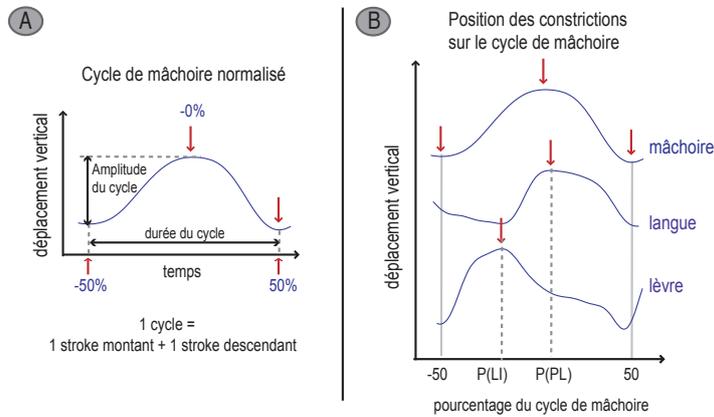


FIG. 3.11 – Mesure du cycle de mâchoire, cycle de mâchoire normalisé et positions des constriction sur ce cycle. **En A.** La durée et l'amplitude du cycle de mâchoire sont respectivement la somme de la durée et la moyenne de l'amplitude de ses gestes montants et descendants. **En B.** Position de la constriction de la pointe de la langue [P(PL)] et position de la lèvre inférieure [P(LI)] sur le cycle de mâchoire normalisé entre -50% (début du geste de fermeture) et +50% (fin du geste d'ouverture).

mâchoire (cf. syndrome du fumeur de pipe, Bognar, 1983). Seuls les strokes réalisés en dehors de ces périodes et avec une amplitude supérieure à 0.1 mm sont analysés. Ce seuil d'amplitude correspond à la précision théorique de l'EMA (Hoole, 1993). S'il est faible en comparaison aux mesures effectives de précision généralement obtenues (Hoole, 1996), il permet de pister les modalités de la disparition d'un cycle de mâchoire dans la progression de la forme CVCV à la forme CCV à l'intérieur des périodes de mouvement régulier. Le cycle de mâchoire est défini comme l'association d'un stroke de fermeture et d'un stroke d'ouverture adjacents. Sa durée et son amplitude sont, respectivement, la somme des durées et la moyenne des amplitudes de ses deux strokes [figure 3.11 (A)].

Pour la pointe de la langue et la lèvre inférieure, les constriction ont été repérées aux maxima de trajectoire sur la base de l'étiquetage acoustique des consonnes et des voyelles. La constriction labiale est la position la plus haute de la lèvre inférieure entre la voyelle après la consonne coronale (V_{Co}) et la voyelle après la consonne labiale (V_{La}). La constriction coronale est la position la plus haute de la pointe de la langue entre V_{La} et V_{Co} [figure 3.10 et 3.11 (B)]. Cette méthode s'inspire de celle utilisée par Hertrich et Ackermann (2000) et permet d'ignorer

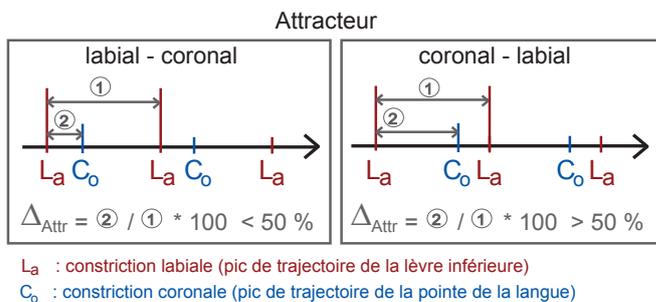


FIG. 3.12 – Mesure articulaire de l'attracteur (Δ_{Attr}). Une progression vers un attracteur labial-coronal (à gauche) se caractérise par un rapprochement entre la constriction labiale et la constriction coronale suivante et donc, par un Δ_{Attr} inférieur à 50%. En revanche, une progression vers un attracteur coronal-labial (à droite) se caractérise par un rapprochement entre la constriction coronale et la constriction labiale suivante et donc, par un Δ_{Attr} supérieur à 50%.

les maxima secondaires dus aux mouvements passifs induits par les gestes de mâchoire portant l'autre constriction.

Pour l'étude de la coordination entre les articulateurs, chaque constriction a été associée au cycle de mâchoire qui l'inclut temporellement et positionnée relativement à ce cycle [figure 3.11 (B)]. A cette fin, les cycles de mâchoire ont été normalisés de -50% (début du stroke montant) à +50% (fin du stroke descendant), 0% correspondant à la position haute du cycle (fin du stroke montant, début du stroke descendant), figure 3.11 (A) et (B). Les positions des constriction labiales et coronales sur ce cycle sont des indicateurs du phasage entre la mâchoire et la lèvre inférieure et la mâchoire et la pointe de la langue. Enfin, afin de déterminer si le rapprochement entre les deux constriction est de type L_aC_o ou C_oL_a , la durée de la constriction labiale à la constriction coronale suivante est comparée avec la durée entre deux constriction labiales successives (voir les modalités de calcul de Δ_{Attr} sur la figure 3.12).

Le tableau 3.2 résume le nombre total de productions correctes (données acoustiques) et parmi ces productions, le nombre de productions analysées sur le plan articulatoire.

3.4.2 Résultats

Analyse acoustique

La comparaison des valeurs de ΔI pour les séquences L_aC_o et C_oL_a des cinq locuteurs (figure 3.13, à droite) avec celles des 26 locuteurs de l'expérience précédente (figure 3.7, page 45) rend compte de comportements similaires. Dans les deux cas, ΔI est proche de 0 pour les durées supérieures à 300 ms alors qu'elle augmente en valeur absolue en deçà de 300 ms, avec des valeurs positives ou négatives. En revanche, pour les bisyllabes dupliqués L_aL_a (/papa/) et C_oC_o (/tata/) (à gauche), ΔI est toujours proche de 0, il y a donc peu ou pas de différence d'intensité entre les deux voyelles pour ces séquences.

D'autre part, les moyennes de ΔI pour les productions de durée inférieure à 300 ms (tableau 3.3) sont représentatives du panel de comportements observés

TAB. 3.2 – Nombre de productions CVCV ou CCV correctes (Tot.) et nombre de productions gardées pour l'analyse articulatoire (Ar.) pour chaque sujet et chaque séquence L_aC_o et C_oL_a .

Sujet	papa		tata		pata		tapa		pasa		sapa		fata		tafa	
	Tot.	Ar.														
S1	63	63	63	60	61	59	59	54	60	59	54	54	53	51	51	47
S2	51	49	53	52	58	47	52	48	62	43	59	43	65	59	58	49
S3	53	51	56	54	60	59	63	61	59	59	64	64	57	57	65	65
S4	66	63	68	60	76	52	73	55	74	73	62	60	69	64	71	53
S5	94	85	100	76	109	93	102	80	99	73	95	82	98	79	100	74

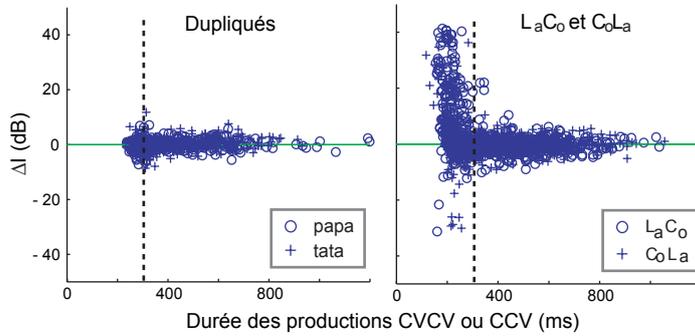


FIG. 3.13 – Variation d'intensité entre la voyelle après la consonne coronale et la voyelle après la consonne labiale (ΔI) en fonction de la durée pour les productions CVCV et CCV de tous les sujets. **A gauche** : pour les séquences dupliquées (/papa/ et /tata/), ΔI reste proche de 0 quelle que soit la durée. **A droite** : pour les séquences L_aC_o et C_oL_a , 300 ms apparaît comme une durée seuil avec une augmentation nette de la valeur absolue de ΔI en deçà de cette durée.

dans l'expérience précédente. Par exemple, peu des productions de la locutrice S3 atteignent des durées en deçà de 300 ms, montrant que cette locutrice évolue rarement vers une syllabe CCV. Ensuite, les locutrices S1 et S4 tendent à favoriser l'attracteur C_oL_a , avec plus de moyennes négatives que positives, alors que le contraire s'observe pour les locuteurs S4 et S5. Enfin, la tendance à évoluer vers un attracteur L_aC_o semble moins marquée pour la paire /pasa/-/sapa/ que pour les autres paires. Ces comportements reproduisent les profils de locuteurs et la variabilité entre les séquences caractérisant l'expérience précédente.

Forme de progression générale

L'exemple d'enregistrement sur la figure 3.10, page 51, illustre les trois étapes du processus d'accélération et de décélération et fournit un premier aperçu des phénomènes articulatoires impliqués dans la variation du débit de répétition. En début d'accélération (à gauche), chaque constriction est portée par son propre cycle de mâchoire. Cependant, l'amplitude et la durée des cycles de la mâchoire qui portent les constriction labiales (*cycles labiaux*) apparaissent plus faibles que celles des cycles qui portent les constriction coronales (*cycles coronaux*). Avec l'augmentation du débit, le cycle labial s'efface progressivement pour évoluer vers une pro-

TAB. 3.3 – Moyennes des variations d'intensité entre la voyelle après la consonne coronale et la voyelle après la consonne labiale pour les productions CVCV et CCV de durée inférieure à 300 ms pour les trois essais confondus, par sujet et séquence (n est le nombre d'observations pour le calcul de m).

Sujet	pata		tapa		pasa		sapa		fata		tafa	
	n	m	n	m	n	m	n	m	n	m	n	m
S1	31	0.8	29	-0.1	29	-1.4 ^b	27	-2.9 ^a	23	-0.2	24	-0.7
S2	25	23.2 ^a	11	9.2	25	11.7 ^b	20	5.5	30	20.3 ^a	20	13.9 ^a
S3	5	-0.8	9	-0.7	0	-	0	-	0	-	10	1.5
S4	43	-0.7	37	-4.7 ^c	37	-0.1	30	-2.4 ^a	28	0.3	29	-1.6 ^c
S5	67	6.9 ^a	59	4.7 ^a	56	1.7 ^a	45	-2.5 ^a	54	7.6 ^a	53	1.1

Comparaison avec 0, test de student bilatéral avec correction de Bonferroni :
^a $p \leq 0.01/30$, ^b $p \leq 0.05/30$, ^c p non significatif après correction de Bonferroni mais faible ($p \leq 0.01$)

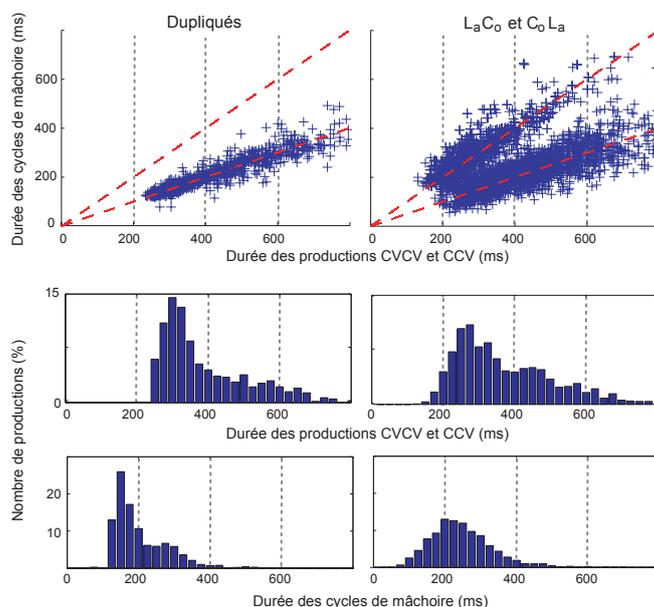


FIG. 3.14 – Étude des durées des cycles de la mâchoire en fonction de la durée des productions CVCV et CCV pour tous les locuteurs confondus, pour les séquences dupliquées (à gauche, /papa/, /tata/) et pour les séquences L_aC_o et C_oL_a (à droite). **En haut** : un point représente une production CVCV ou CCV pour un locuteur. La séparation en deux distributions pour les séquences L_aC_o et C_oL_a indique que ces séquences peuvent être réalisées de deux manières : avec un cycle de mâchoire par constriction ($y = x/2$) ou avec un cycle de mâchoire pour deux constriction ($y = x$). Les distributions des durées des productions (**au milieu**) et des durées de cycles de mâchoire (**en bas**) montrent que les durées des productions sont plus courtes pour les séquences L_aC_o et C_oL_a que pour les dupliquées alors que le contraire tend à s'observer pour les durées des cycles de mâchoire.

duction de deux syllabes sur un seul cycle (au centre). Cette phase de *fusion* se caractérise par une modification de la coordination entre les constricteurs et la mâchoire, avec la constriction coronale qui paraît advenir sur la position haute du cycle de mâchoire et la constriction labiale au début du geste de fermeture. Durant cette phase, les constriction labiales paraissent aussi plus proches des constriction coronales les suivant que de celles les précédant, ce qui indique un rapprochement de type L_aC_o (cf. /..tápatápa.../) plutôt que C_oL_a . Enfin, la décélération (à droite) ramène progressivement la coordination vers son état initial d'un cycle de mâchoire par constriction. La suite des analyses évaluée dans quelle mesure ces caractéristiques, compatibles avec nos hypothèses, sont généralisables à l'ensemble des sujets et des séquences CVCV.

Progression d'un cycle par constriction vers un cycle pour deux constriction

La première ligne de la figure 3.14 compare la durée des productions CVCV et CCV mesurée à partir de l'étiquetage des courbes d'intensité du signal acoustique (en x) à la durée des cycles de mâchoire sous-tendant ces productions (en y), pour toutes les productions de tous les locuteurs confondus. Pour les bisyllabes dupliqués (à gauche), les productions se distribuent autour d'une droite d'équation $y = x/2$; en d'autres termes, ces productions requièrent deux cycles de mâchoire. Si ce regroupement s'observe pour les bisyllabes L_aC_o et C_oL_a (à droite), ces séquences se caractérisent aussi par un regroupement autour d'une droite $y = x$ pour les durées de productions inférieures à environ 400-450 ms. Les productions

correspondant à ce second regroupement sont donc réalisées sur un seul cycle de mâchoire. Or, les analyses acoustiques de ΔI montrent que l'évolution vers une syllabe CCV advient pour des durées inférieures à 300 ms. Ainsi, la bascule vers un seul cycle pour deux constrictions se ferait avant la disparition totale de l'une des deux voyelles, ce qui confirme l'existence d'une phase avec deux syllabes pour un seul cycle de mâchoire. Cette phase constituerait une transition entre la phase de production de type CVCV sur deux cycles de mâchoire pour des débits lents et la phase de production de type CCV sur un seul cycle de mâchoire pour des débits rapides.

D'autre part, l'étude de la répartition des durées des productions CVCV et CCV (figure 3.14, au milieu) rend compte de valeurs équivalentes pour le groupe des dupliqués et pour le groupe $L_aC_o-C_oL_a$. Cependant, dans le premier groupe, les durées ne descendent pas en deçà de 225 ms et un pic apparaît autour de 300 ms alors que pour le deuxième groupe, environ 15% des productions ont des durées comprises entre 175 et 225 ms, avec un pic de distribution autour de 250-275 ms. En revanche, la répartition des durées des cycles de la mâchoire (figure 3.14, en bas) rend compte de valeurs plus faibles pour le groupe des dupliqués que pour le groupe $L_aC_o-C_oL_a$, avec, des pics de distributions respectivement autour de 150 ms et 225 ms. D'autre part, pour les deux types de séquences, la durée du cycle n'est jamais inférieure à 100 ms. Ce portrait est cohérent avec l'idée qu'à débit rapide, les séquences L_aC_o et C_oL_a peuvent être produites sur un seul cycle de mâchoire mais pas les séquences dupliquées. De plus, la différence des durées de cycles de mâchoire observée entre les deux groupes montre que la progression vers un seul geste de mâchoire pour deux constrictions permet d'atteindre des débits de parole plus rapides tout en gardant une durée de mouvement de mâchoire confortable, ce qui minimise probablement la consommation d'énergie.

Afin d'analyser plus précisément la bascule de deux cycles de mâchoire vers un seul, les productions des locuteurs ont été classées selon les trois phases précédemment décrites : la phase de fusion, qui inclut les productions avec deux constrictions sur un seul cycle et les phases d'accélération et de décélération incluant les productions avec un cycle de mâchoire par constriction, respectivement avant et après la phase de fusion.

Les mouvements de la mâchoire en phase d'accélération

La figure 3.15 étudie la relation entre l'amplitude et la durée des strokes de mâchoire selon les sujets et les séquences au cours du processus d'accélération. Pour les séquences dupliquées, deux types de comportements sont observés. Le premier est celui des locutrices S3 et S4 qui diminuent l'amplitude des mouvements de la mâchoire avec la durée et ce, comme le montrent les droites de régression, pour les mouvements du cycle labial (/papa/) comme pour ceux du cycle coronal (/tata/). De plus, pour ces locutrices, la distribution de l'amplitude en fonction de la durée tend à suivre une courbe théorique correspondant à une accélération de 0.25 g (voir

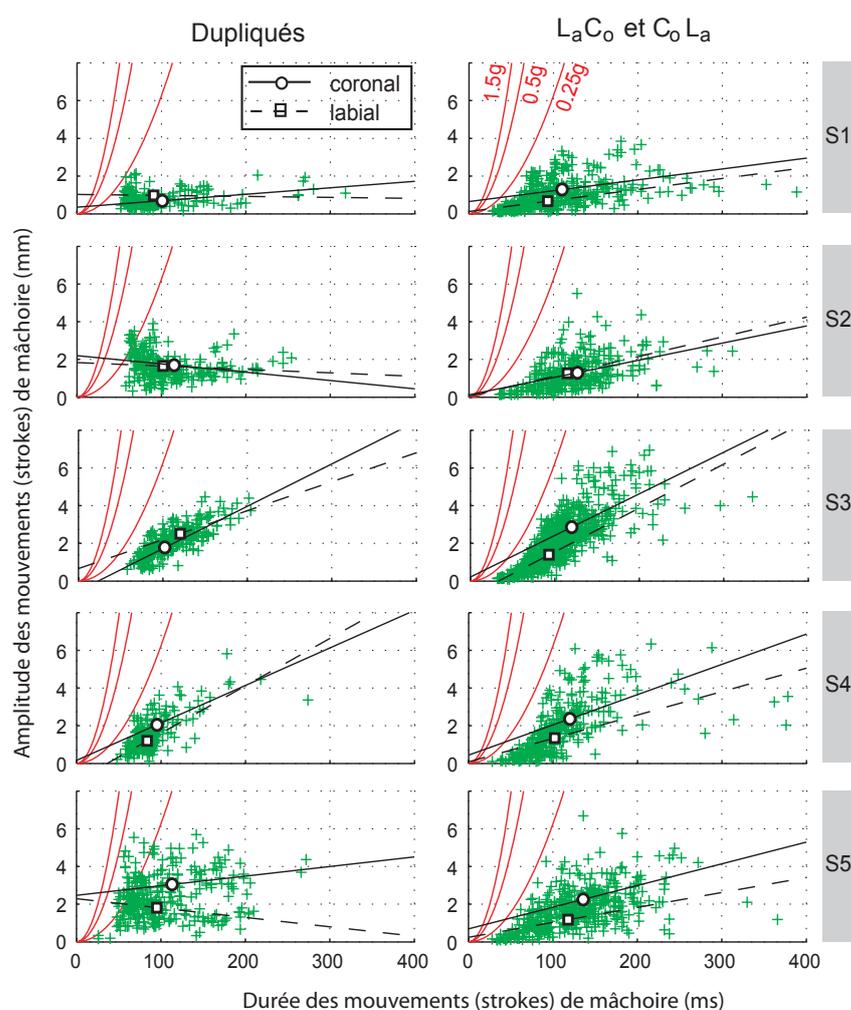


FIG. 3.15 – Amplitude des gestes de mâchoire (strokes) montants et descendants en fonction de leur durée en phase d'accélération pour les productions avec un cycle de mâchoire par constriction et pour chaque sujet. **A gauche** : séquences dupliquées (/papa/ et /tata/). **A droite** : séquences $L_a C_o$ et $C_o L_a$. Les trois courbes paraboliques sont des limites théoriques de durée minimum de mouvements de la mâchoire pour trois valeurs d'accélération, voir Nelson (1983). Les droites de régression et les moyennes des distributions (carrés et cercles) sont représentées séparément pour les gestes de mâchoire portant la constriction labiale (carrés) et pour ceux portant la constriction coronale (cercles).

TAB. 3.4 – Moyennes des amplitudes et des durées des cycles de la mâchoire portant la constriction labiale et la constriction coronale en phase d'accélération, pour chaque sujet et chaque paire $L_aC_o-C_oL_a$ (n est le nombre d'observations pour le calcul des moyennes). Les p correspondent aux résultats des tests de Student comparant les valeurs du cycle labial à celles du cycle coronal.

Sujet	Séquence	n	Amplitudes (mm)			Durées (ms)		
			labial	coronal	p	labial	coronal	p
s1	pata/tapa	47	0,58	0,91	S	163,6	198,6	S
	pasa/sapa	38	0,59	2,06	S	162,7	247,2	S
	fata/tafa	23	0,98	0,77	NS	272,6	222,1	NS
s2	pata/tapa	21	1,83	1,57	NS	231,9	244,0	NS
	pasa/sapa	28	1,07	1,48	S	209,1	285,2	S
	fata/tafa	23	1,00	0,83	NS	263,7	235,7	NS
s3	pata/tapa	59	1,17	2,75	S	178,3	229,7	S
	pasa/sapa	55	1,01	3,46	S	169,9	269,5	S
	fata/tafa	54	2,02	2,36	S	211,2	219,5	NS
s4	pata/tapa	23	1,16	2,67	S	174,7	247,0	S
	pasa/sapa	18	1,20	3,89	S	206,1	279,2	NS
	fata/tafa	33	1,57	1,34	NS	223,3	214,6	NS
s5	pata/tapa	35	0,85	2,53	S	204,7	268,0	S
	pasa/sapa	14	1,08	3,78	S	196,0	370,7	S
	fata/tafa	52	1,43	1,66	NS	267,8	247,0	NS

Test de Student avec correction de Bonferroni :
 S : $p \leq 0.01/15$, NS : $p \geq 0.05/15$

la description du modèle de Nelson, p. 2.2). Le second comportement caractérise les trois autres locuteurs et consiste à ne pas diminuer l'amplitude des mouvements avec la durée. Cependant, si S1 maintient une faible amplitude de mouvement pour toutes les durées, la diminution de la durée amène les distributions de S2 et S5 vers une courbe limite correspondant à une accélération de 0.5g. Ainsi, les locuteurs mettent en œuvre deux types de stratégies pour augmenter le débit de production : soit ils diminuent l'amplitude sans modifier l'accélération (cas de S3 et S4), soit ils augmentent l'accélération sans modifier l'amplitude (cas de S2 et S5). Ces deux grandes catégories de stratégies s'accordent avec les résultats de Sonoda et Nakakido (1986) montrant que l'accélération des mouvements de la mâchoire induit soit une augmentation de la vitesse (diminution de la durée) et peu de changement de l'amplitude soit, inversement, peu de modification de la vitesse mais une forte diminution de l'amplitude. Cette propriété a été reprise par Lindblom (1990a) pour expliquer les variations phonétiques et discutée par Perkell *et al.* (2002) à propos des différences inter-sujets.

Par opposition, la variabilité inter-individuelle est réduite pour le groupe $L_aC_o-C_oL_a$ pour lequel les cinq locuteurs diminuent l'amplitude des mouvements avec la durée, plus ou moins le long d'une courbe théorique de 0.25 g. D'autre part pour ces séquences, les droites de régression des strokes du cycle labial sont situées le plus souvent en dessous de celles des strokes du cycle coronal, ce qui montre que les amplitudes et les durées des mouvements de la mâchoire du cycle labial tendent à être plus faibles que celles du cycle coronal.

L'asymétrie entre le cycle labial et le cycle coronal

Afin d'évaluer le poids de l'asymétrie entre les cycles labial et coronal, les amplitudes et les durées des deux cycles ont été comparées pour chaque CVCV produit en phase d'accélération. Les résultats sont donnés par sujet et par couple de séquence dans le tableau 3.4. Les valeurs sont significativement plus faibles pour le cycle labial que pour le cycle coronal dans un certain nombre de cas alors que l'inverse n'est jamais observé. Cependant, la prévalence du cycle coronal sur le cycle labial dépend des consonnes labiale et coronale constituant la paire L_aC_o - C_oL_a . Elle est très nette pour /pasa/-/sapa/ - avec des différences d'amplitude et de durée en faveur du cycle coronal significatives pour tous les locuteurs - mais moins évidente pour /fata/-/tafa/ - avec des différences significatives seulement pour les amplitudes de la locutrice S3. D'autre part, l'asymétrie entre les deux cycles dépend aussi du locuteur. Par exemple, elle est systématique pour la locutrice S3, avec des différences d'amplitude en faveur du cycle coronal significatives pour les trois couples de bisyllabes L_aC_o - C_oL_a , alors que pour S2, elle apparaît seulement pour le couple /pasa/-/sapa/.

Ainsi, l'asymétrie entre le cycle labial et le cycle coronal est une tendance dominante du corpus analysé ici. Couplé au fait que l'amplitude et la durée des mouvements diminuent avec l'accélération, ce phénomène s'accorde avec l'idée d'un effacement progressif du cycle labial dans la progression d'un cycle par constriction vers un cycle pour deux constriction. Cette évolution ne peut se faire sans une réorganisation de la coordination entre les articulateurs.

Coordination de la mâchoire et des constricteurs au cours des trois phases

La figure 3.16 illustre la répartition des positions des constriction sur le cycle de mâchoire. Au cours des phases d'accélération et de décélération, les constriction labiale et coronale adviennent toutes deux autour de 0% du cycle de la mâchoire, soit sur la position la plus haute de la mâchoire. Ce résultat montre que la mâchoire accompagne alternativement la constriction labiale et la constriction coronale. Cependant, la concentration autour de 0% est moins homogène pour la pointe de la langue que pour la lèvre inférieure, que ce soit pour les bisyllabes dupliqués ou L_aC_o - C_oL_a . Cette différence pourrait résulter de la plus grande dépendance anatomique entre la lèvre inférieure et la mâchoire qu'entre la pointe de la langue et la mâchoire. D'autre part, pour le groupe L_aC_o - C_oL_a , la variabilité est plus importante en phase de décélération qu'en phase d'accélération. Ce phénomène peut s'interpréter comme une trace d'hystérésis : après la phase de fusion et donc, le passage à un mode de coordination de deux constriction par cycle de mâchoire, le retour à l'état initial d'un cycle de mâchoire par constriction se fait progressivement, avec une phase hybride entre les deux modes de coordination.

Pendant la phase de fusion, la coordination entre la mâchoire et les constricteurs se réorganise de façon à produire deux constriction sur un seul cycle de mâchoire. La figure 3.16 montre que la tendance générale est de réaliser la constric-

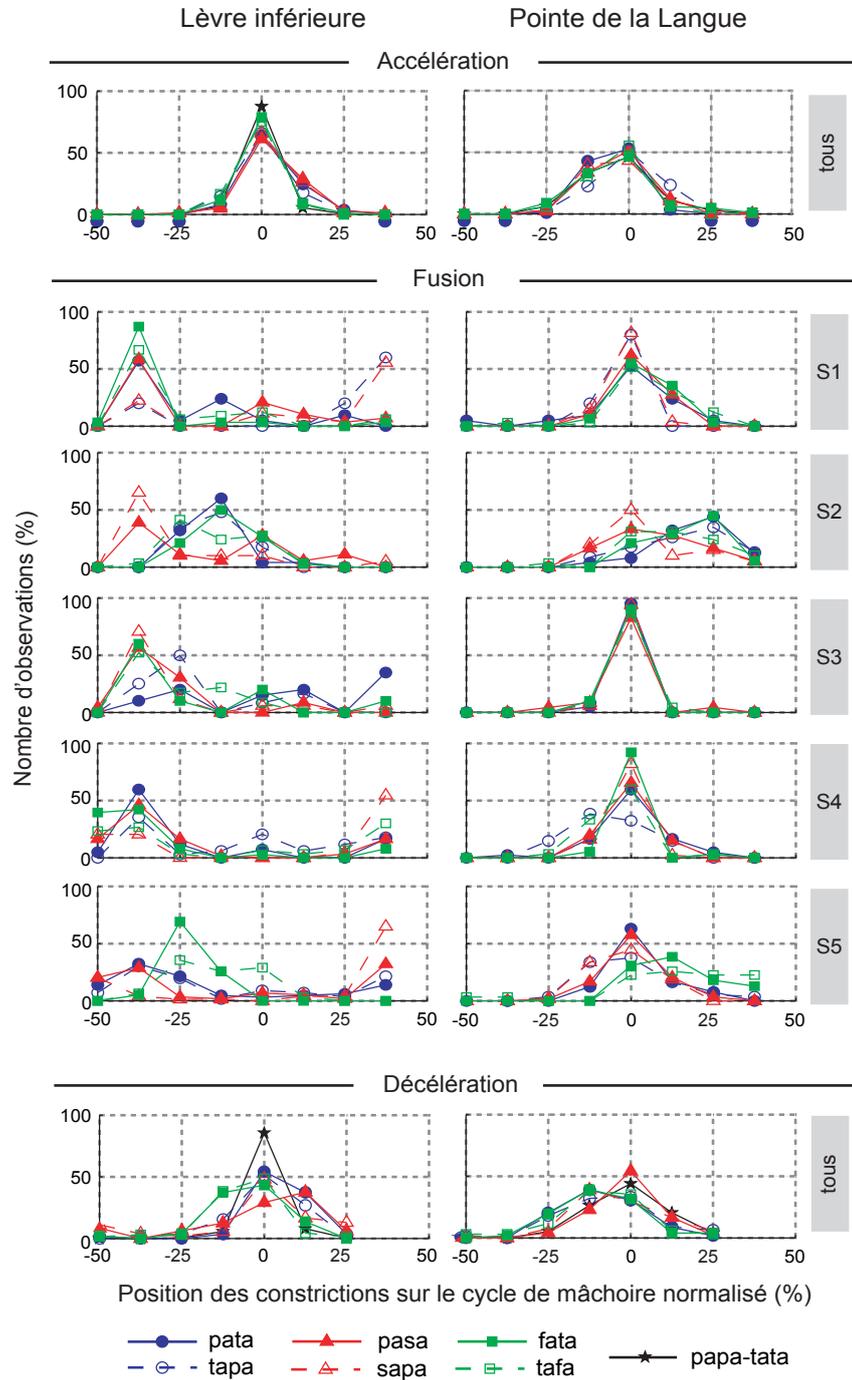


FIG. 3.16 – Distribution des productions des locuteurs selon la position de la constriction labiale (à gauche) et de la constriction coronale (à droite) sur le cycle de mâchoire normalisé. **En haut et en bas** : productions au cours de la phase d'accélération et de décélération, correspondant aux périodes de coordination d'un cycle de mâchoire par constriction, respectivement avant et après la fusion, pour tous les locuteurs confondus. **Au milieu** : productions au cours de la phase de fusion (un cycle de mâchoire pour deux constriction), analyse par locuteur.

TAB. 3.5 – Moyennes de la durée entre la constriction labiale et la constriction coronale suivante normalisée par la durée entre les deux constriction labiales consécutives (Δ_{Attr}), exprimée en pourcent, pour chaque sujet et chaque séquence (n est le nombre de productions CVCV ou CCV utilisées pour le calcul de la moyenne). Les valeurs inférieures à 50% indiquent une évolution vers l'attracteur L_aC_o et les valeurs supérieures à 50%, une évolution vers l'attracteur C_oL_a , (voir texte pour le détail).

Sujet	pata		tapa		pasa		sapa		fata		tafa	
	n	m	n	m	n	m	n	m	n	m	n	m
S1	20	49,3	5	53,0	27	50,8	26	51,0	31	53,7 ^a	32	52,0
S2	25	36,3 ^a	21	40,3 ^b	20	41,1	22	43,9	40	30,5 ^a	36	36,1 ^a
S3	19	44,3 ^b	11	38,6 ^a	22	41,5 ^a	17	42,5 ^a	10	44,8	23	43,7 ^a
S4	56	45,4 ^b	48	49,5	60	46,3 ^a	44	52,0	38	42,8 ^a	30	51,4
S5	67	42,1 ^a	58	40,8 ^a	67	51,8	71	53,5 ^b	35	37,9 ^a	41	44,9 ^b

Comparaison avec 0, test de Student bilatéral avec correction de Bonferroni :

^a $p \leq 0.01/30$, ^b $p \leq 0.05/30$

tion coronale sur la plus haute position du cycle de mâchoire (autour de 0%), et la constriction labiale sur une position basse, le plus souvent au début du geste de fermeture (juste après -50%), indiquant un rapprochement vers la constriction coronale suivante. Cependant, dans certains cas, une translation de 25% vers la droite du cycle de mâchoire apparaît, avec une constriction labiale autour de -25% (au cours du geste de fermeture) et une constriction coronale à +25% (au cours du geste d'ouverture). Comme pour les mesures précédentes, ces valeurs dépendent du locuteur et de la nature des consonnes labiale et coronale. La translation vers la droite s'observe surtout pour le locuteur S2 et la paire /fata/-/tafa/ alors qu'elle ne s'observe pas pour /pasa/-/sapa/. La différence entre /pasa/-/sapa/ et /fata/-/tafa/ fait écho à la différence entre les profils de cycles de mâchoire précédemment observée pour ces deux paires de bisyllabes. En effet, si le cycle labial est moins ample et plus court que le cycle coronal pour /pasa/-/sapa/, la différence est moins marquée pour /fata/-/tafa/. Or, la tendance à réaliser le /s/ sur un cycle de mâchoire plus ample et plus long que pour le /t/ pourrait résulter du fait que la coronale fricative /s/ requiert une position de mâchoire plus haute que la coronale plosive /t/, comme en témoignent diverses études (Keating, 1983; Keating *et al.*, 1994; Lindblad et Lundqvist, 1999; Rhardisse et Abry, 1994). De la même manière pour les consonnes labiales, la fricative /f/ nécessiterait une position de mâchoire plus haute que la plosive /p/, ce qui expliquerait à la fois l'absence d'asymétrie entre le cycle labial et le cycle coronal et le décalage vers la droite des constriction relativement au cycle de mâchoire, spécifiques à la paire à /fata/-/tafa/.

Caractérisation articulatoire des attracteurs L_aC_o et C_oL_a

Précédemment, les attracteurs ont été caractérisés de manière acoustique par la différence d'intensité entre la voyelle après la consonne coronale et celle après la consonne labiale (ΔI) avec des valeurs positives de ΔI pour l'attracteur L_aC_o et des valeurs négatives pour l'attracteur C_oL_a (tableau 3.3, p. 54). Les attracteurs peuvent

aussi se caractériser de manière articulatoire par la distance entre la constriction labiale à la constriction coronale suivante relative à la distance entre deux constrictions labiales successives (Δ_{Attr} , voir figure 3.12, page 52). Le tableau 3.5 donne les valeurs moyennes de Δ_{Attr} par sujet et par séquence. Les valeurs inférieures à 50% indiquent un rapprochement de type L_aC_o et les valeurs supérieures à 50%, un rapprochement de type C_oL_a . Bien que les analyses des ΔI et Δ_{Attr} concernent des productions répondant à des critères sélectifs différents (durées inférieures à 300 ms pour ΔI et productions réalisées sur un seul cycle de mâchoire pour Δ_{Attr}), les deux mesures donnent des profils de locuteurs cohérents.

Ainsi, les productions de la locutrice S1, caractérisées par des valeurs de ΔI négatives, présentent des valeurs de Δ_{Attr} majoritairement supérieures à 50%, ce qui confirme la tendance de cette locutrice à évoluer préférentiellement vers un attracteur de type C_oL_a . À l’opposé, les productions du locuteur S2 se caractérisent par des valeurs de ΔI positives et des valeurs de Δ_{Attr} inférieures à 50%, ce qui valide la tendance de ce locuteur à favoriser l’attracteur L_aC_o . Pour la locutrice S4, les valeurs de Δ_{Attr} sont inférieures à 50% pour les bisyllabes L_aC_o (/pata/, /pasa/ et /fata/) et supérieures à 50% pour les bisyllabes C_oL_a (/sapa/ et /tafa/). Ces résultats confirment la tendance de cette locutrice à évoluer vers l’attracteur préservant le séquençage initial. Pour le locuteur S5, les valeurs de Δ_{Attr} sont inférieures à 50% pour tous les bisyllabes sauf pour /pasa/-/sapa/, ce qui, ici encore, est cohérent avec les valeurs de ΔI . Enfin, si les données acoustiques de la locutrice S3 n’ont pu être analysées du fait du nombre insuffisant de productions avec des durées inférieures à 300 ms, les données articulatoires de cette locutrice montrent qu’elle produit des bisyllabes sur un seul geste de mâchoire pour toutes les séquences étudiées et qu’elle privilégie alors l’attracteur L_aC_o (moyennes de Δ_{Attr} inférieures à 50%).

L’attracteur L_aC_o correspond donc bien à une diminution de l’intervalle temporel entre la constriction labiale et la constriction coronale suivante et l’attracteur C_oL_a , à une réduction de la distance temporelle entre la constriction coronale et la constriction labiale suivante. De plus, les données des cinq locuteurs analysés ici reproduisent la préférence pour l’attracteur L_aC_o observée pour les 21 locuteurs et ce, en dépit de la variabilité entre les sujets et les séquences. Cette asymétrie en faveur de l’attracteur L_aC_o s’accorde avec l’hypothèse selon laquelle l’anticipation du geste de la langue pour la constriction coronale au cours du geste des lèvres pour la constriction labiale serait plus stable pour le système articulatoire que l’inverse.

3.5 DISCUSSION

L’objectif des deux expériences rapportées ici était de tester l’hypothèse selon laquelle le séquençage L_aC_o serait une forme articulatoire plus stable pour le système coordinatif mâchoire-langue-lèvre que le séquençage C_oL_a . L’analyse acoustique des répétitions accélérées des 28 locuteurs de la première expérience

apporte un premier argument en faveur de cette hypothèse pour des locuteurs ayant le français pour langue maternelle. L'analyse articulatoire des productions des cinq locuteurs de la deuxième expérience permet de mieux comprendre les phénomènes articulatoires impliqués dans cette asymétrie et dans la progression d'une séquence CVCV vers un attracteur CCV avec l'augmentation de la vitesse de répétition. Enfin, discutés relativement à la question des origines des universaux phonologiques des langues, ces résultats peuvent être considérés comme de nouveaux éléments pour expliquer la phonologie par la phonétique ou, en termes plus généraux, l'émergence de la forme à partir de la substance.

3.5.1 Le système coordinatif mâchoire-langue-lèvre en action

L'étude articulatoire montre qu'avec l'augmentation du débit, la répétition des bisyllabes L_aC_o et C_oL_a évolue vers une coordination d'un cycle de mâchoire par syllabe vers un cycle de mâchoire pour deux syllabes pour éventuellement aboutir à une syllabe CCV. Cette progression serait plus économique pour la mâchoire et l'attracteur L_aC_o vs. C_oL_a émergerait de l'interaction de plusieurs facteurs.

Changement de programme

Au début de la répétition, quand le débit est faible, chaque syllabe est produite sur un cycle de mâchoire et chaque constriction advient sur la position la plus haute de la mâchoire. Cette coordination entre la mâchoire et les constricteurs se réorganise ensuite dans la progression vers un seul cycle de mâchoire pour deux constricteurs (phase de fusion⁴). Pour les cinq locuteurs de l'expérience articulatoire, cette phase se caractérise par deux types de coordination (figure 3.17, à droite) en relation avec deux profils de mouvements de la mâchoire durant la phase précédant la fusion (figure 3.17, à gauche).

Le profil de coordination A, schématisé sur la première ligne de la figure 3.17, se caractérise par une asymétrie entre les cycles labial et coronal en phase d'accélération avec un cycle labial plus court et moins ample que le cycle coronal (en haut, à gauche). Or, le processus d'accélération reposant sur une diminution de la durée et de l'amplitude des cycles labial et coronal, l'asymétrie entre les deux cycles fait que les strokes du cycle labial atteignent la durée de saturation de la mâchoire (évaluée ici à 50 ms par stroke) avant les strokes du cycle coronal : le cycle labial *se fonde* dans le cycle coronal. Ce phénomène préserve le phasage entre le maximum de trajectoire de la mâchoire et celui de la pointe de la langue alors que le maximum de trajectoire de la lèvre inférieure arrive au cours du stroke de fermeture ou à la fin du stroke d'ouverture. Ce premier profil est typique de la locutrice S3 et des séquences /pasa/-/sapa/.

⁴Dans Rochet-Capellan et Schwartz (2007), nous employons le terme de *splitting* pour désigner cette phase en référence aux deux distributions observées sur la figure 3.14 pour les séquences L_aC_o et C_oL_a . Cependant, le terme de *fusion* semble plus pertinent pour décrire le phénomène *intrinsèque* aux mouvements de la mâchoire, ce qui justifie son utilisation ici.

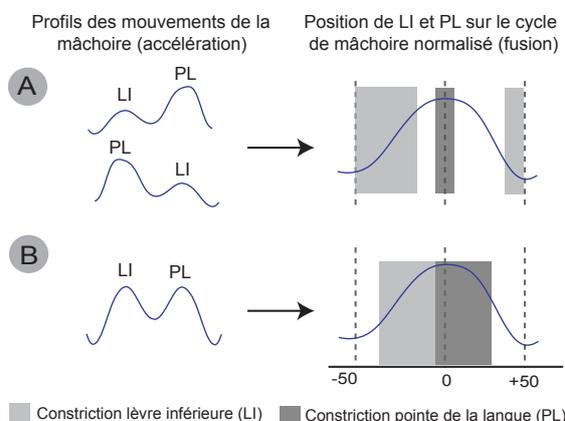


FIG. 3.17 – Réorganisation de la coordination entre la mâchoire et les constricteurs dans le passage de deux cycles de mâchoire pour deux constricteurs à un seul cycle pour deux constricteurs (fusion). Les comportements des sujets peuvent se classer en deux grandes catégories de coordination [(A) et (B), à droite], correspondant à deux profils de mouvements de la mâchoire observés avant la fusion [(A) et (B), à gauche].

Le profil de coordination B, schématisé sur la deuxième ligne de la figure 3.17 s’ancre dans l’absence d’asymétrie entre l’amplitude et la durée du cycle labial et celles du cycle coronal avant la fusion. Dans ce cas, une analyse supplémentaire des données montre que la progression vers un seul cycle se fait par une réduction du mouvement d’ouverture après la constriction labiale et du mouvement de fermeture vers la constriction coronale. Cette progression induit un déphasage entre les deux constricteurs et la mâchoire. La constriction labiale advient au cours du stroke de fermeture et la constriction coronale, au cours du stroke d’ouverture. Ce profil est typique du locuteur S2 et des séquences /fata/-/tafa/.

Ces résultats sont particulièrement intéressants dans le contexte des recherches qui associent majoritairement le cycle de mâchoire à la syllabe CV, c’est le cas des études articulatoires (Nelson *et al.*, 1984; Kelso *et al.*, 1985; Hertrich et Ackermann, 2000), des travaux de modélisation (Redford et Diehl, 1999; Redford *et al.*, 2001; Oudeyer, 2001) ou encore, des théories de production de la parole (MacNeilage, 1998). D’autre part, le comportement de la mâchoire pour soutenir l’accélération est proche de ceux décrits dans la littérature et s’accorde avec le principe général d’économie de l’effort.

Minimisation de l’effort

Dans la littérature, l’effort articulaire a été mesuré via différents indicateurs (voir Nelson, 1983 et chapitre 2). Nelson *et al.* (1984) ont étudié les mouvements de la mâchoire dans la répétition accélérée de la syllabe /sa/ et caractérisé ces mouvements par l’accélération maximale compatible avec la relation amplitude - durée (voir aussi figure 2.2, p. 27). D’après leurs données, les durées des gestes d’ouverture et de fermeture de la mâchoire ne sont jamais inférieures à 50 ms, seuil compatible avec nos données articulatoires. De plus, pour un de leur sujet, ils observent un maximum dans l’amplitude des mouvements autour d’une durée de 125 ms. Nelson *et al.* interprètent cet effet de « résonance » en termes d’économie d’effort et l’attribuent au fait que 125 ms serait une durée optimale pour les mouvements

d'ouverture et de fermeture de la mâchoire. Cette durée correspondrait au débit *naturel* de 3-4 cycles - et donc 3-4 syllabes CV - par seconde. Or, dans notre étude articulatoire, la répartition des durées des cycles de la mâchoire présente un pic autour de 225 ms, ce qui correspond à une durée moyenne de mouvement de fermeture et d'ouverture de 112.5 ms, proche des 125 ms observées par Nelson *et al.*. Nos données montrent aussi que pour les bisyllabes dupliqués /papa/ et /tata/, la durée des cycles de la mâchoire peut atteindre 150 ms, ce qui correspond à une durée de stroke de 75 ms, proche de la valeur de saturation de 50 ms. Enfin, l'estimation des valeurs maximales d'accélération indique des valeurs globalement plus faibles pour les séquences L_aC_o et C_oL_a que pour les dupliqués.

L'ensemble de ces résultats suggère qu'avec l'augmentation de la vitesse, la réalisation de deux constrictions sur un seul cycle de mâchoire serait plus économique pour la mâchoire que l'alternance rapide d'un cycle pour chaque constriction. En restant prudemment dans la métaphore, cette progression n'est pas sans rappeler celle des pattes du cheval qui, d'une alternance de plus en plus rapide dans le trot, basculent vers un mouvement en phase dans le galop.

Multi-déterminisme d'un système bistable

L'accélération de la répétition des séquences CVCV L_aC_o et C_oL_a amène le système mâchoire-langue-lèvre soit vers un attracteur L_aC_o (/pata/ → /patá/ → /ptá/), soit vers un attracteur C_oL_a (cf. /tapa/ → /tapá/ → /tpá/). Cependant, cette bistabilité est modulée par la préférence pour l'attracteur L_aC_o observée dans l'étude acoustique et qui s'accorde avec l'idée d'une plus grande stabilité articulatoire de la forme L_aC_o par rapport à la forme C_oL_a , modulo le fait que d'autres facteurs semblent intervenir dans l'émergence de l'attracteur.

Parmi ces facteurs, la coordination initiale L_aC_o vs. C_oL_a attire le système vers l'attracteur préservant l'ordre initial des constrictions. En effet, la préférence pour l'attracteur L_aC_o est plus nette pour les bisyllabes L_aC_o que C_oL_a , particulièrement pour la locutrice S4. Comme expliqué dans la description des hypothèses, ce comportement pourrait résulter du fait que la langue française accentue la dernière syllabe. Cependant, il semble aussi être le produit d'un non-respect de la consigne encourageant à poursuivre l'accélération même quand la production paraît différente de la séquence initiale : certains locuteurs semblent *s'interdire* d'accélérer pour ne pas basculer vers une forme différente de la forme CVCV initiale.

D'autre part, la nature fricative vs. plosive des consonnes labiale et coronale de la séquence CVCV à produire favorise l'un ou l'autre des attracteurs. Ainsi, l'attracteur C_oL_a est plus souvent observé pour /pasa/-/sapa/ que pour les deux autres paires de bisyllabes alors qu'inversement, la paire /fata/-/tafa/ favorise l'attracteur L_aC_o . Comme nous l'avons expliqué précédemment, cette asymétrie peut s'expliquer en termes d'amplitude de mouvement de la mâchoire avec des positions requises plus hautes pour /f/ que pour /p/, et pour /s/ que pour /t/.

Enfin, la détermination de l'attracteur paraît dépendre du facteur sujet, comme en témoignent les profils de locuteurs établis à l'issue de la première expérience avec une majorité de locuteurs plutôt L_aC_o une minorité plutôt C_oL_a et une autre minorité bistable.

D'autres facteurs non considérés ici pourraient aussi intervenir dans la sélection de l'attracteur. Par exemple, le rôle de la langue dans la préférence pour l'attracteur L_aC_o devra être délimité par des études supplémentaires utilisant des langues avec d'autres propriétés phonologiques et prosodiques. Des études complémentaires devront aussi évaluer la généralisation de la préférence de l'attracteur L_aC_o à d'autres modes et lieux d'articulation et à d'autres contextes vocaliques. Néanmoins, la démarche expérimentale présentée ici constitue un outil potentiel pour ces recherches futures. Elle fournit aussi de nouveaux éléments intéressants à discuter relativement aux universaux de l'organisation inter-syllabique des lexiques des langues du monde.

3.5.2 Implications pour l'émergence des formes phonologiques

Le sous-bassement théorique de ce travail expérimental est que les inventaires des langues du monde émergeraient de l'interaction de contraintes de différents types et notamment de contraintes liées aux propriétés physiques du système de production et à son contrôle moteur. Certaines associations de syllabes seraient favorisées car plus *faciles* à produire en termes d'économie d'énergie et de stabilité (plus simples à contrôler). Cette hypothèse a motivé de nombreux travaux expérimentaux et théoriques dont ceux de Liljencrants et Lindblom (1972); Lindblom *et al.* (1984); Kelso (1995); Davis *et al.* (2002) évoqués aux chapitres 2 et dont s'inspire largement la démarche expérimentale introduite ici. Dans cette démarche scientifique visant à expliquer les universaux des langues du monde, nos résultats apportent deux contributions : ① la répétition d'une même syllabe CV (cf. /papa/, /tata/) pourrait être plus coûteuse que l'alternance de deux syllabes CV impliquant des constricteurs différents pour les deux consonnes ; et ② l'ordre L_aC_o pourrait être plus stable que l'ordre C_oL_a .

Varier plutôt que dupliquer

L'analyse des lexiques des langues du monde révèle une dominance des formes CVCV alternant des constricteurs plutôt que celles répétant la même constriction (MacNeilage et Davis, 2000; Rousset, 2003). En d'autres termes, les langues préfèrent « pata » à « papa ». A priori, ce fait semble violer le principe de simplicité motrice, la duplication pouvant être supposée plus simple en terme de contrôle moteur, d'où sa dominance dans le babillage canonique et les premiers mots (MacNeilage et Davis, 2000). Cependant, cette hypothèse néglige le fait que la parole adulte est largement coarticulée, contrairement aux productions du bébé : elle implique des structures coordinatives, et non des organes contrôlés indépendamment les uns des autres. Ainsi, nos données montrent que les séquences L_aC_o et C_oL_a

peuvent être produites à des débits plus rapides que les séquences L_aL_a (/papa/) et C_oC_o (/tata/). Cette différence résiderait dans le phénomène d'anticipation d'une constriction dans l'autre rendu possible par l'implication de deux organes différents et qui permettrait la production de deux syllabes sur un seul cycle de mâchoire. Ce mode de coordination, incompatible avec les formes dupliquées, serait plus économique pour la mâchoire, autorisant des débits de production rapides sur des durées de cycles de mâchoire *optimales*, au-delà du seuil de saturation. Ainsi, pour le système de production mature de la parole, alterner serait plus économique que dupliquer pour atteindre des débits importants. L'analogie avec d'autres activités gestuelles comme jouer du piano ou taper à la machine vient ici *naturellement* : répéter un mouvement avec le même organe est plus lent que d'alterner deux organes. En d'autres termes, les constricteurs ne peuvent pas être à la fois et au four et au moulin ou encore, à l'anticipation et à la production. Mais, réitérons notre prudence interprétative en insistant sur le fait que toute généralisation ne pourra se faire que par la reproduction des résultats obtenus à des séquençages CVCV autres que L_aC_o et C_oL_a .

Labiale-coronale plutôt que coronale-labiale

L'effet LC en tant que caractéristique des lexiques des langues du monde montre que parmi les formes variées, certaines sont préférées à d'autres. Les résultats de notre première expérience nous encouragent à défendre l'hypothèse selon laquelle les formes L_aC_o domineraient en partie du fait de leur plus grande cohérence et stabilité articuloire par rapport aux formes C_oL_a . Cette propriété expliquerait aussi la présence de l'effet LC dans les premiers mots de l'enfant, remettant en question l'hypothèse *simple first* de MacNeilage et Davis (2000). Il n'y a en effet aucune raison d'attribuer la dominance de l'attracteur L_aC_o dans la répétition accélérée des séquences CVCV L_aC_o et C_oL_a à un mécanisme du *plus simple d'abord* alors que la cohérence articuloire fournit une explication commune à nos résultats et à ceux observés chez l'enfant. De plus, elle peut-être interprétée comme une rencontre entre le cadre et le contenu de la parole, rencontre qui s'ancre au cours du développement dans l'acquisition de la coordination entre les deux systèmes et donc, compatible avec l'idée d'un développement cadre-puis-contenu. Le fait que la structure C_oL_a apparaisse aussi comme un attracteur possible n'est pas incompatible avec ce raisonnement. Il suggère simplement que la structure C_oL_a est aussi viable, bien que sous-optimale sur le plan articuloire et confirme la nécessité de concevoir les universaux des langues du monde comme le produit émergent d'interactions multifactorielles complexes.

Concernant cette question, Sato *et al.* (2007) ont notamment mis en évidence ce qu'ils ont appelé un *corrélât perceptif de l'effet LC*. Leur protocole expérimental consiste à soumettre des locuteurs de la langue française à des stimuli auditifs alternant une syllabe CV labiale (cf. /pa/, /pi/,...) et une syllabe CV coronale (cf. /ta/, /ti/,...). Ces syllabes ont été enregistrées individuellement par un phonéticien de langue maternelle française, à débit naturel de conversation et avec une même

intensité et intonation, puis composées en un flux ... $L_aC_oL_aC_oL_aC_o$... Soumis à un tel flux, les sujets doivent signaler tout changement auditif perçu. En mesurant la durée perceptive de chaque forme perçue (temps écoulé entre le moment où le sujet dit percevoir cette forme et celui où il dit en percevoir une autre), Sato *et al.* montrent que 70% des transformations sont des formes L_aC_o vs. C_oL_a et que les durées de perception des formes L_aC_o sont plus importantes que celles des formes C_oL_a (ratio de 1.4 pour les formes /pVtV/ par rapport aux formes /tVpV/). Ainsi, la plus grande stabilité des formes L_aC_o par rapport aux formes C_oL_a se retrouverait sur les plans moteur et perceptif.

3.6 CONCLUSION

Les expériences réalisées ici mettent en évidence une asymétrie articulatoire entre les séquences CVCV L_aC_o et C_oL_a : la répétition accélérée de ces deux types de séquences évolue plus souvent vers un attracteur L_aC_o que C_oL_a . Notre interprétation est que cette dominance de l'attracteur L_aC_o reposerait sur sa plus grande stabilité articulatoire pour le système coordinatif mâchoire-langue-lèvres par comparaison aux séquences C_oL_a . Cette asymétrie pourrait être une des causes possibles de l'effet LC observé dans les lexiques des langues.

De toute évidence, l'établissement d'une relation causale entre ces deux éléments ne peut se faire uniquement sur la base des données présentées ici : des expériences supplémentaires devront être réalisées afin de comparer la stabilité articulatoire des séquences L_aC_o et C_oL_a dans d'autres langues et, de manière plus directe, dans la production des premiers mots de l'enfant. Le paradigme et la méthode d'analyse acoustique proposés dans la première expérience devraient permettre de comparer la stabilité L_aC_o et C_oL_a dans les productions de locuteurs de différentes langues et de mettre à l'épreuve nos interprétations ainsi que de déterminer la part véritable du facteur articulatoire dans l'effet LC. Des études supplémentaires devraient aussi explorer les relations entre la stabilité perceptive et la stabilité articulatoire. Il serait notamment intéressant de voir si les différences observées dans nos expériences entre /sp/-/ps/ et /pt/-/tp/ ou /ft/-/tf/ ont un corrélat perceptif...

Enfin, l'étude plus systématique des processus d'anticipation de C_1 à C_2 dans la production de C_1VC_2V et leur lien avec les asymétries des cycles de la mâchoire pour C_1 et C_2 devraient faire avancer la compréhension des mécanismes de production de la parole en intégrant les phénomènes de coarticulation trop souvent négligés dans les études des mouvements de la mâchoire. Dans tous les cas, la détermination des facteurs impliqués dans l'émergence des formes phonologiques ne pourra se passer de l'expérimentation pour mettre en évidence ces facteurs et de simulations pour évaluer le rôle de ces facteurs et de leurs interactions. Nous reviendrons sur ce point au chapitre 7.

Deuxième partie

Comment la gestualité brachiomanuelle entraînerait la gestualité orofaciale vers l'émergence de la parole... et du langage



Résumé

Le chapitre précédent s'est intéressé à une première rencontre fondamentale dans la phylogénèse et l'ontogénèse de la parole : celle entre la mâchoire qui porte (le cadre) et les articulateurs portés, langue et lèvre inférieure (le contenu). Cette deuxième partie étudie une rencontre de deuxième type fondamental à l'émergence du langage : celle de la main et de la parole.

Le chapitre 4 rapporte les résultats d'études récentes établissant des liens forts entre les gestualités orofaciale et brachiomanuelle dans l'émergence de la parole et du langage. Il s'intéresse plus particulièrement à la coordination entre la parole et le geste de pointé, avancée comme une étape clé du développement et de l'évolution du langage. Il introduit notamment le cadre de recherche *Vocalize to Localize* (Vocaliser pour Localiser) qui propose de dériver la parole et le langage de la nécessité de localiser les choses dont nous avons besoin de parler. La rencontre entre la main et la parole dans l'acte de montrer (deixis) devient donc une équation clé des origines du langage qui pourrait intervenir tant sur le plan de l'acquisition que de la genèse des formes phonologiques, lexicales ou syntaxiques.

Dans ce cadre, les chapitres 5 et 6 présentent deux études expérimentales visant à fournir des éléments de résolution à cette équation en abordant deux grandes questions : ① Où les deux systèmes se rencontrent-ils ? Ou en termes plus opérationnels : quels sont les sites de synchronisation entre les deux systèmes ? et ② Comment cette rencontre entre deux systèmes moteurs aux propriétés si différentes est-elle portée par le cadre du signe (geste de pointé) et le cadre de la parole (oscillations verticales de la mâchoire) ? Ou en termes plus opérationnels : comment les oscillations de la mâchoire et le geste du bras se coordonnent-ils ?

Des liens entre la main, la bouche, la deixis et l'émergence du langage

Si l'être humain est un être de paroles, il se sert aussi de ses mains pour communiquer. Les gestes des mains complètent la parole, mais peuvent aussi la remplacer quand ses systèmes de perception et/ou de production sont altérés pour donner naissance aux langues signées. Le système brachio-manuel peut donc porter le langage tout aussi bien que le système orofacial. Ainsi, une des questions de la littérature sur les origines du langage est de déterminer pourquoi, si l'être humain peut *parler* avec ses mains, il parle avec sa bouche et si une communication manuelle aurait pu précéder la communication vocale au cours de l'évolution. Ces questionnements théoriques sur le rôle de la main dans les origines du langage et dans la communication sont alimentés par - et alimentent en retour - un ensemble de recherches contribuant à ancrer le langage dans la connexion des gestualités brachiomanuelles et orofaciales. Ces recherches fournissent différents arguments pour considérer la coordination main-bouche comme une composante clé de l'émergence du langage. Enfin, le geste de pointé brachiomanuel, de par son rôle dans l'attention partagée et dans le développement de la parole et du langage, fait de l'association pointé-parole une pierre fondatrice de l'émergence du langage.

4.1 UNE ORIGINE VOCALE OU MANUELLE ?

4.1.1 Préférence pour la parole, des évidences pas si trompeuses ?

Comment expliquer l'universalité du *choix* par les sociétés préhistoriques dispersées du signifiant vocal-auditif pour produire du sens alors que d'autres canaux étaient disponibles ? Pour continuer à paraphraser Hagège (1987, p. 22), relativement aux contraintes de notre monde physique, le canal vocal-auditif présente des avantages indiscutables par rapport au canal visuo-tactile : il fonctionne à distance et dans l'obscurité, sans qu'il soit nécessaire de toucher ou de voir l'interlocuteur. À ces arguments, Corballis (2003b) ajoute celui *tout aussi évident* de la libération de la main. En effet, la parole permet de communiquer tout en utilisant les mains pour manipuler des outils, porter des bébés, cueillir des fruits,... La question des avantages vs. inconvénients de la parole pour la survie de l'espèce est sujette à

controverse. Sans rentrer dans le détail, nous renverrons au débat autour de l'hypothèse de Lieberman selon laquelle l'anatomie du conduit vocal et notamment la descente du larynx aurait évolué pour la parole, augmentant le risque de mort par obstruction du larynx lors de l'alimentation (Lieberman, 2002; Dessalles, 2000). L'idée selon laquelle la descente du larynx serait l'élément clé de l'apparition de la parole est notamment remise en question par (Boe *et al.*, 2004; Boë *et al.*, 1999) qui proposent que c'est le contrôle moteur et non l'anatomie du conduit en soi qui compte, les simulations montrant qu'un conduit vocal de singe, de bébé ou d'homme de Néanderthal pourrait produire les sons de base de la parole. Indépendamment de ce débat, qui va plus loin que la simple question *pourquoi la parole*, sur le plan scientifique il paraît plus pertinent de se demander *comment la parole mais aussi la main ?* au lieu de *pourquoi la parole plutôt que la main ?*

4.1.2 Au commencement était la main

Par opposition à la théorie *Cadre puis Contenu* de MacNeilage (1998) dérivant le langage de la parole, et la parole du contrôle moteur des articulateurs dans la fonction alimentaire (voir chapitre 2), certains auteurs proposent une évolution du langage *De la main à la bouche*. Cette thèse est en vogue actuellement (Philippon, 2007) et trouve son plus célèbre défenseur en la personne de Michael Corballis (Corballis, 2003a,b) qui avance les arguments suivants :

Les primates non-humains ne parlent pas et leurs cris n'ont rien de commun avec la parole humaine : ils ne sont pas contrôlés volontairement (contrôle essentiellement sous-cortical), ils dépendent des états émotionnels et ils ne peuvent être ni segmentés, ni combinés.

Les primates non-humains gesticulent et les propriétés de ces gestes sont plus proches du langage humain que ne le sont celles de leurs cris. Ils présentent notamment un caractère dyadique et leur contrôle et compréhension impliquent des aires cérébrales proches de l'aire de Broca chez l'homme.

La bipédie aurait libéré la main et permis ainsi son utilisation pour la communication. Le contrôle moteur des articulateurs aurait été plus lent à se mettre en place du fait de sa plus grande complexité, notamment pour le contrôle de la langue.

Chez l'homme actuel, les gestes des mains accompagnent la parole et les personnes non-entendantes utilisent leurs mains pour communiquer.

Pourtant, à trop vouloir séparer l'évolution de la main et de la bouche et en écartant les vocalisations des origines premières du langage, Corballis se heurte au problème du transfert du langage de la main à la bouche et retombe dans la question *Pourquoi la parole et pas la main ?* Pour résoudre ce problème, il fait appel à une mutation du fameux *gène du langage*, FOXP2 (voir chapitre 2), qui aurait pu être l'ajustement final permettant à la parole de devenir autonome. Cette supposition fait, une fois de plus, de FOXP2, la solution *magic-bullet*¹ au problème de l'évolu-

¹Littéralement : *balle magique*, l'élément magique qui expliquerait tout.

tion du langage, pour reprendre les termes de Bickerton (2005). Mais pourquoi ne pas considérer que le langage aurait pu, dès le départ, se fonder sur une utilisation conjointe de la main et de la bouche ?

4.1.3 Gestes des mains et parole, deux composantes indissociables

Le scénario de plus en plus adopté pour une origine gestuelle du langage est que, ayant atteint un certain niveau cognitif, les individus auraient utilisé tous les moyens dont ils disposaient pour communiquer (voir Holden, 2004). La gestualité manuelle et la parole (geste orofaciaux et vocalisations) auraient co-évolué dès le départ pour porter la communication et seraient indissociables dans la construction du langage. La question *Qui a commencé avant ?* s'efface pour laisser la place à *Comment les deux systèmes ont co-évolué pour donner naissance au langage ?* Pour McNeill (1992) l'évolution aurait sélectionné la capacité de combiner la parole et les gestes par rapport à une signification. Arbib (2003), en réponse à Corballis (2003b) rejette l'idée d'un langage signé élaboré ayant précédé la parole. Selon lui, les deux systèmes se seraient influencés mutuellement dans une *spirale expansive*², bien que le proto-signe aurait pu, selon Arbib, précéder la proto-parole. Le développement de cette spirale expansive reposerait sur l'extension d'un système de neurones miroirs initialement en place pour les gestes de saisie et transféré ensuite à la parole (Arbib, 2005). Enfin, pour les auteurs s'intéressant au développement de la parole, la gestualité brachio-manuelle est indissociable de la gestualité orofaciale. Selon Iverson et Thelen (2003), le contrôle volontaire du conduit vocal serait, comme le défend Corballis (2003b), plus long à acquérir que le contrôle manuel et l'avance de la main sur la parole pourrait jouer un rôle dans le développement du contrôle vocal. Mais la relation entre l'activité manuelle et l'apparition du babillage au cours du développement ne signifie pas que le bébé élabore un système de communication avec ses mains avant de communiquer avec sa bouche. De plus, si le contrôle moteur volontaire de la main précède celui de la bouche, les deux systèmes coopèrent avant que l'enfant n'ait le temps de développer un langage intégral utilisant seulement des signes manuels. En revanche, sur le plan moteur, la main pourrait entraîner la parole ou la porter, hypothèse développée ci-après.

4.2 ARGUMENTS POUR UNE BASE GESTUELLE MAIN-BOUCHE À L'ÉMERGENCE DU LANGAGE

Trois types d'arguments peuvent être avancés pour défendre l'inextricabilité du langage des gestuelles brachiomanuelle et orofaciale qui le véhiculent et pour le rôle du développement de la coordination de la main et du conduit vocal dans l'émergence du langage :

Des arguments neurologiques, mettant en évidence des recouvrements entre les

²The doctrine of the Expanding Spiral, intitulé de la réponse d'Arbib à Corballis (2003b).

zones cérébrales du langage et de la gestualité ;

Des arguments développementaux, rendant compte d'une relation entre l'émergence du langage au cours du développement et l'utilisation de l'association main-parole dans la communication ;

Des arguments expérimentaux, mettant en évidence différents niveaux de coordination main-bouche en général et de coordination main-parole dans la communication en particulier.

4.2.1 Recouvrement des aires cérébrales du langage et de l'action

La conception traditionnelle Broca-Wernicke du langage est, comme l'annonce sans ambiguïté Lieberman (2007) dès le début de son article : incorrecte. Les corrélatés neuronaux du langage ne sont pas des modules spécialisés mais plutôt des circuits distribués impliquant des aires pouvant intervenir dans d'autres fonctions que le langage. D'après Pulvermuller (2005), langage, perception et action partageraient un même circuit temporo-pariéto-frontal avec, par exemple, une implication des zones motrices aussi bien que des aires classiques du langage dans le traitement et la représentation des mots d'action (voir figure 4.1, en bas). L'Aire Motrice Supplémentaire (SMA) serait aussi impliquée dans la production de la parole (Alario *et al.*, 2006) et dans le contrôle volontaire des mains (Lau *et al.*, 2004). De plus, l'aire de Broca générerait des fonctions plus génériques que la production de la parole, comme le séquençage des actions, la compréhension de l'action et l'imitation, requises aussi bien au langage parlé que signé (voir Nishitani *et al.*, 2005, pour une revue sur les fonctions de Broca). Broca serait notamment à l'homme ce que F5 est aux singes (figure 4.1, en haut), c'est-à-dire l'aire des neurones miroirs, activée quand les sujets observent, imaginent et imitent un geste de saisie d'objet ou de mouvements des lèvres (Rizzolatti et Arbib, 1998; Ferrari *et al.*, 2003). Broca est aussi une aire spécifique du langage signé, son altération pouvant engendrer des aphasies du signe (Jose-Robertson *et al.*, 2004; Emmorey *et al.*, 2002; Hickok *et al.*, 1998). Ces résultats conduisent à revoir la conception neurologique classique du langage, il n'y a plus une aire pour la compréhension (Wernicke) et une aire pour la production (Broca) spécifiques au langage parlé mais des circuits distribués mettant en relation différentes fonctions permettant de véhiculer le langage aussi bien par la main que par le conduit vocal.

4.2.2 Co-développements des gestualités orofaciale et brachiomanuelle et émergence du langage

Les études sur l'ontogenèse du langage fournissent un autre ensemble d'arguments pour une relation entre le langage, la parole et les gestes manuels. L'équipe de Goldin-Meadow a publié de nombreux travaux étudiant le rôle des gestes dans le développement cognitif et dans l'acquisition du langage. Les titres de certains des articles parlent d'eux mêmes : *Gesture paves the way for language develop-*

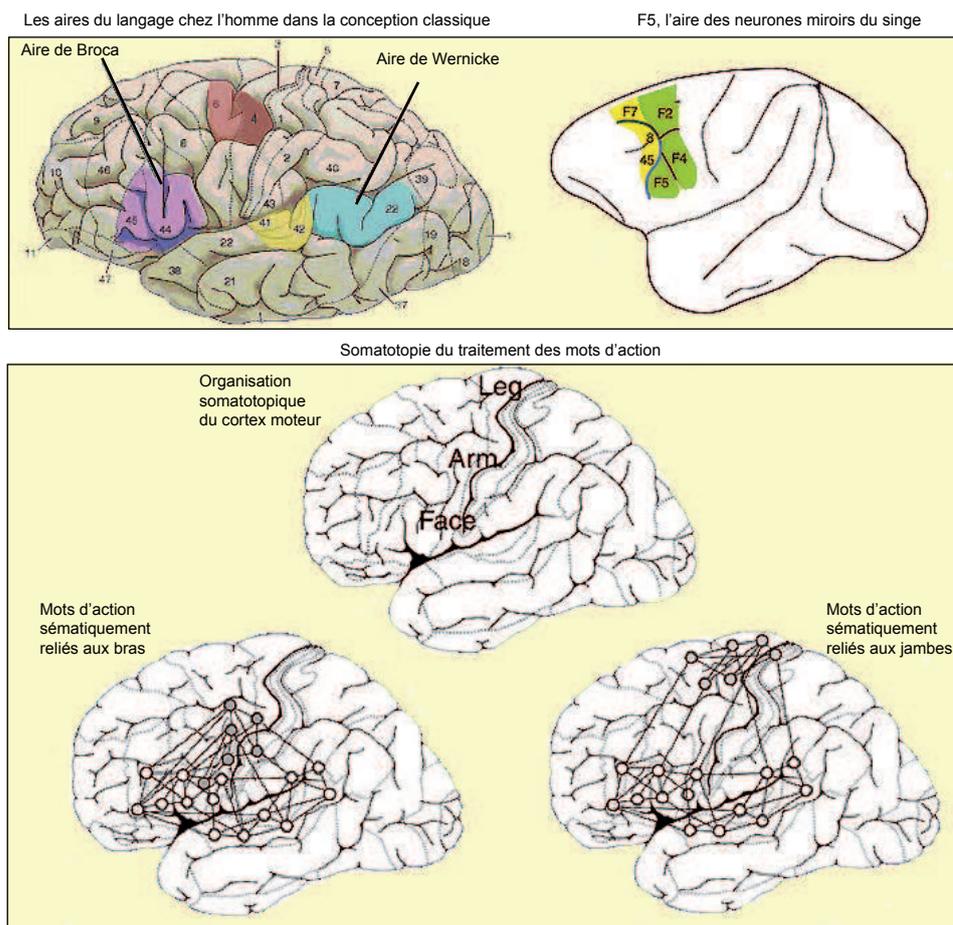


FIG. 4.1 – Le langage dans le cerveau, du localisationnisme aux circuits distribués. **En haut à gauche** : dans le localisationnisme classique, les aires du langage sont l'aire de Broca (production) et de Wernicke (compréhension), figure : <http://lecerveau.mcgill.ca>. **En haut à droite** : l'aire des neurones miroirs du singe, F5 serait homologue à l'aire 44 de l'homme, d'après Rizzolatti *et al.* (1996). **En bas** : le traitement des mots d'action implique les aires motrices aussi bien que les aires classiques du langage, avec une implication des aires motrices du bras (arm) pour le traitement des mots d'action en rapport avec le bras et une activation des aires motrices des jambes (leg) pour les mots d'action en rapport avec les jambes. Langage et action interagissent donc fortement, d'après Pulvermüller *et al.* (2005).

TAB. 4.1 – Liens entre les gestualités orofaciale et manuelle au cours du développement, d'après Iverson et Thelen (1999).

Période développementale	Manifestations	Liens oralité-manualité	
Naissance : Liens initiaux	Oral : succion, pleurs, sons végétatifs. Manuel : porter la main à la bouche, saisie réflexe, mouvements spontanés, incapacité à atteindre	Réflexe de Babkin (paume-bouche), coordination main-bouche spontanée.	La main et la bouche sont activées mutuellement
De 6 à 8 mois : Emergence du contrôle	Oral : « roucoulement », son volontaire, babillage reduplicatif. Manuel : début de la saisie, mouvement rythmique, babillage manuel	Débuts des mouvements rythmiques vocaux et manuels.	Entraînement mutuel des activités rythmiques des bras des mains et des articulateurs.
De 9 à 14 mois : Emergence des premiers gestes et des premiers mots	Oral : babillage varié, premiers mots. Manuel : début des premiers gestes et du contrôle précis des doigts	Les gestes communicatifs précèdent les premiers mots. Les gestes et la parole ont des référents différents. Quand la parole et les gestes cooccurrent, ils sont produits séquentiellement.	Le seuil d'activation des gestes est plus bas que celui de la parole. Pas d'activation simultanée de la parole et des gestes.
De 16 à 18 mois : Emergence de la synchronisation entre les gestes et la parole.	Plus de communication. Augmentation du vocabulaire. Poursuite de raffinement du contrôle moteur.	Début des mots avec une signification synchronisés aux gestes manuels co-produits.	La pratique de la communication diminue les seuils et augmente l'entraînement de l'activation jusqu'à la synchronie.

*ment*³ (Iverson et Goldin-Meadow, 2005), *Gesture is at the cutting edge of early language development*⁴ (Ozcaliskan et Goldin-Meadow, 2005), *How our hands help us learn*⁵ (Goldin-Meadow, 2005), *The importance of gesture in children's spatial reasoning*⁶ (Ehrlich *et al.*, 2006), etc. Parmi tant d'autres apports, ces travaux mettent en évidence une étape clé du langage : celle de l'association d'un mot et d'un geste véhiculant une information différente, comme pointer une tortue en disant « partir » pour « la tortue part », ou faire un geste signifiant « gros » en disant « monstre » pour « le monstre est gros ». Ces associations permettent à l'enfant d'exprimer une idée équivalente à celle qu'exprime une phrase de deux mots (Goldin-Meadow et Butcher, 2003, voir aussi la section 4.3.3). De plus, l'association des gestes à la parole semble motrice avant d'être perceptive, les gestes étant utilisés par les enfants non-voyants (Iverson et Goldin-Meadow, 1998) mais aussi par deux adultes non-voyants communiquant et même par des individus voyants quand ils ne se voient pas, lors d'une conversation téléphonique par exemple. Ces observations vont dans le sens d'un entraînement moteur mutuel des deux systèmes. D'autre part, les liens moteurs entre la main et la bouche sont présents dès la naissance et se développent progressivement pour aboutir à une synchronisation entre les deux types de productions. Cette progression est décrite dans Iverson et

³Le geste pave le chemin pour le développement du langage.

⁴Le geste est en proue du développement du langage précoce.

⁵Comment nos mains nous aident à apprendre

⁶L'importance du geste dans le raisonnement spatial des enfants

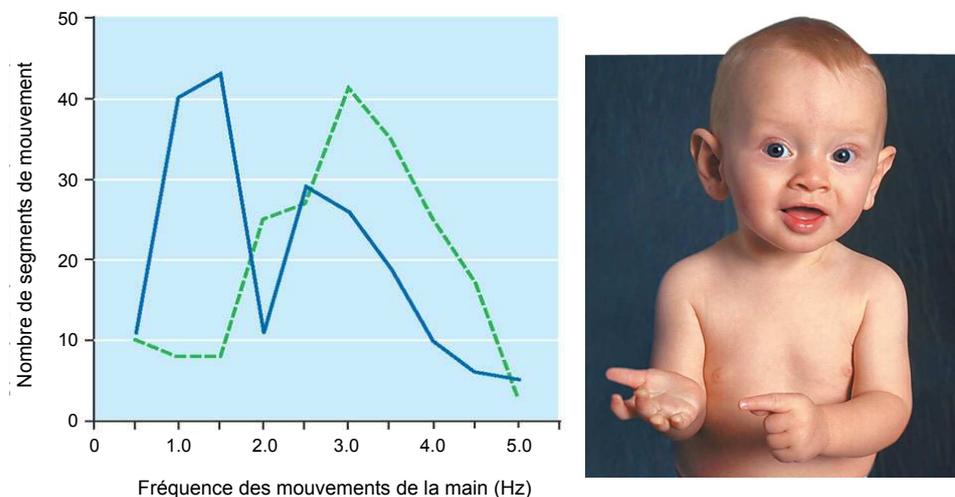


FIG. 4.2 – Les rythmes du langage dans les mouvements des mains des bébés, d'après Petitto *et al.* (2001). Petitto *et al.* ont comparé les fréquences des mouvements des mains produits par des bébés de 6, 10 et 12 mois entendants, exposés soit à la langue des signes, soit à la parole. Leur objectif était de contraster l'hypothèse non-linguistique du babillage, considérant celui-ci comme une activité purement motrice, et l'hypothèse linguistique, considérant le babillage comme le reflet de la sensibilité des bébés aux rythmes du langage qui les entourent et leur capacité à reproduire ces rythmes. Les distributions des fréquences des mouvements produits par chaque groupe de bébés sont représentées sur le graphique de gauche. La distribution des productions des bébés exposés au langage des signes présente deux pics, un entre 0.5 et 1.5 Hz et le second entre 2 et 3 Hz. Par opposition, la distribution des productions des bébés exposés à la parole présente un seul pic entre 2.5 et 3.5 Hz. De plus, les mouvements de basse fréquence des bébés exposés au langage des signes présentent la particularité d'être réalisés dans un espace restreint correspondant à l'espace *phonétique* des langues signées (photo. à droite). Selon les auteurs, ces mouvements linguistiques, qui se différencient des mouvements moteurs par leur fréquence rythmique, sont possibles car les bébés ont la capacité de capter et de reproduire les rythmes du langage.

Thelen (1999), voir tableau 4.2.2. Une des hypothèses fortes d'Iverson et Thelen, dans une perspective d'incorporation du langage dans la gestualité, est que l'activité manuelle rythmique capturerait et entraînerait celle du conduit vocal et que cette capture et cet entraînement seraient liés à l'émergence de la parole et du langage⁷. Enfin, Petitto *et al.* (2001) ont mis en évidence une rythmicité des gestes manuels spécifique au langage chez les bébés entendants élevés par des parents signants (voir figure 4.2).

4.2.3 Coordination main-bouche dans la communication et autres activités

Au moins depuis McNeill (1981), il est admis que la parole et les gestes de la main peuvent se coordonner dans les interactions face à face. Les études des

⁷We suggested that in the development of communication, rhythmic manual activity captured and entrained the coordination of the oral system and that both were linked with emerging speech and language., Iverson et Thelen, 1999, p.37

processus moteurs impliqués dans cette coordination se sont principalement développées autour de la question de la modularité vs. interactivité de la programmation et de l'exécution motrice des deux systèmes. Deux grands types d'études de la coordination entre la main et la bouche peuvent être distingués : des études analysant des gestes discrets produits en situation de laboratoire ou écologiques et des études analysant le couplage du système articulatoire et de la gestualité manuelle dans des tâches rythmiques. Les premières vont souvent dans le sens d'un contrôle modulaire et les secondes, d'un contrôle interactif, voir unifié.

Tâches non rythmiques

En laboratoire. La plupart des recherches de laboratoire utilisant des tâches non-rythmiques évaluent les niveaux d'interaction entre les deux systèmes à l'aide d'un paradigme de double-tâche. Ce paradigme compare les propriétés des réponses verbale et manuelle, dans une tâche consistant à produire simultanément une réponse verbale et manuelle, aux propriétés des réponses verbale et manuelle dans une tâche impliquant seulement une réponse verbale vs. seulement une réponse manuelle. Par exemple, dans Holender (1980), la double-tâche consiste à dénommer une lettre apparaissant sur un écran d'ordinateur tout en appuyant sur une touche du clavier avec le doigt. Elle est contrastée avec une tâche simple de dénomination vs. une tâche simple de réponse par pression d'une touche. Dans Castiello *et al.* (1991), il s'agit de prononcer /tah/ à l'illumination d'une cheville à saisir avec le pouce et l'index vs. seulement prononcer /tah/ vs. seulement saisir la cheville. Dans Levelt *et al.* (1985), puis dans Feyereisen (1997), la tâche consiste à pointer un objet avec l'index tout en le désignant par une expression démonstrative telle que « cette lampe » vs. seulement pointer l'objet vs. seulement désigner l'objet verbalement. Ces études aboutissent à des conclusions différentes quant aux niveaux d'interaction entre les deux systèmes, mais maintiennent globalement une idée d'indépendance de leur contrôle moteur.

- Castiello *et al.* montrent que les deux systèmes ont peu d'influence l'un sur l'autre : le temps de la réponse verbale, de l'illumination de la cheville au début de la voix, et le temps de la réponse manuelle, de l'illumination de la cheville au début du geste du doigt ne diffèrent pas en double-tâche et simple tâche. Autrement dit, la présence du geste a peu d'effet sur la parole et inversement.
- Holender montre que la présence de la réponse manuelle induit un report de la réponse vocale qui aurait pour objectif de synchroniser les deux systèmes. Ce ralentissement de la parole ne s'observe pas quand le sujet donne priorité à la parole et Holender conclut son article par l'idée d'un contrôle par deux processus indépendants en compétition pour un même ensemble de traitements. Le regroupement des réponses serait alors une conséquence de ces limitations.
- Levelt *et al.*, puis Feyereisen, montrent que la parole s'adapte au geste de pointé. Le report du début de la voix en présence du geste du doigt dépend

notamment de la distance à la cible et tend à rapprocher le début de la voix de l'extrémité du geste de pointé. Par opposition, le geste de pointé est peu, voire pas, affecté par la présence de la parole. De plus, une perturbation du geste du bras dans sa première phase d'exécution induit une adaptation de la réponse verbale. En revanche, si la perturbation est appliquée alors que l'exécution du geste est déjà avancée, la parole ne s'adapte plus à la perturbation. Pour Levelt *et al.* les deux systèmes interagiraient pendant la phase de planification et en début d'exécution du geste, la suite de l'exécution se ferait de manière modulaire. Nous reviendrons sur ces résultats au chapitre suivant.

Plus écologique. L'étude de Ruiter et Wilkins (1998) comparant les associations de gestes et de parole de locuteurs de l'Arrernte, une langue des aborigènes australiens, à ceux de locuteurs germanophones, lors de leur réponse à des questions sur des lieux et des événements, montre que la synchronisation entre les deux gestes n'est pas si évidente et pourrait dépendre de la culture. Kita (2003) va aussi dans ce sens. Concernant la relation entre le geste et la parole, il rapporte une étude basée sur des films de passants Japonais indiquant un itinéraire dans la rue. En ce qui concerne la relation pointage-parole, un des résultats de cette étude est que la production des mots « droite » et « gauche » serait facilitée par la réalisation du geste de pointé. Cependant, les conclusions des tâches écologiques sont parfois difficiles à interpréter, surtout quand elles se basent sur l'analyse de gestes différents comme dans Ruiter et Wilkins.

Tâches rythmiques

Les tâches rythmiques sont utilisées par les approches dynamiques du contrôle moteur pour étudier le couplage entre deux systèmes oscillatoires. Rappelons que ces approches trouvent leurs origines en physique et en biologie et supposent que le système nerveux central réduirait le nombre de degré de liberté en contrôlant les structures musculaires en mouvements comme une seule unité (Kelso, 1995). Pour le mouvement, ce type d'approche a concerné l'étude des membres dans la locomotion, puis a été étendu par l'équipe de Kelso au contrôle des articulateurs dans la parole (voir chapitre 2) mais aussi, à la coordination des gestes de la main et de la parole. En utilisant une tâche consistant à répéter le mot /stak/ tout en bougeant le doigt (flexion-extension), Kelso *et al.* (1981) ont montré une relation fréquentielle de 1 :1 entre la parole et le geste du doigt. De plus, la variation de l'intensité de la parole induit une variation de l'amplitude du geste du doigt et inversement, la variation de l'amplitude du geste du doigt induit des modifications de la parole analogue. Selon Kelso *et al.* (1981), le couplage entre la parole et le geste du doigt serait absolu.

Cependant les travaux ultérieurs de Smith *et al.* (1986), puis de Treffner et Peter (2002) vont plutôt dans le sens d'un couplage relatif. En reprenant la tâche de Kelso *et al.* (1981), Smith *et al.* montrent que la coordination absolue d'une syllabe pour

un cycle du doigt s'observe quand la vitesse de production n'est pas contrainte, avec une tendance à adopter une fréquence intermédiaire entre celle du geste du doigt effectué seul et celle de la répétition de la syllabe /stak/ réalisée seule. En revanche, elle disparaît quand les tâches sont réalisées à des débits différents. Selon les auteurs, la coordination entre les deux systèmes serait relative, avec une tendance à observer des relations harmoniques entre leurs fréquences respectives. Dans une perspective visant à expliquer les interactions entre la parole et les gestes manuels en termes de coordination plutôt que d'interférence et dans un cadre dynamique, Treffner et Peter (2002) proposent d'enrichir le modèle HKB (Haken *et al.*, 1985) des relations entre deux oscillateurs couplés en y incluant des processus de haut niveau tels que l'attention. Ce modèle, élaboré dans le cadre de l'étude de la coordination bimanuelle, rend compte de l'existence de phasages préférentiels correspondant à des modes de coordination stables. Avec l'augmentation de la vitesse de production, le système est attiré vers ces modes de coordinations préférentiels (voir chapitre 2). Afin de démontrer que la coordination entre les gestes manuels et la parole est compatible avec l'existence de modes cohérents d'un système dynamique plutôt qu'avec des interférences au niveau cognitif, Treffner et Peter utilisent une tâche consistant à répéter la syllabe /ba/ tout en tapant avec l'index sur la table. Ils comparent les mouvements de la mâchoire et ceux du doigt dans une condition *en phase* (mâchoire haute - doigt haut) à une condition *anti-phase* (mâchoire haute, doigt bas) en faisant varier le focus attentionnel (sur la parole vs. sur la main). Les résultats montrent que la coordination *anti-phase* induit des relations plus dispersées que la coordination *en phase*. Cette variabilité du phasage diminue avec l'augmentation du débit. D'autre part, les effets de l'attention sont plus importants en *anti-phase* qu'en *phase* : l'attention peut stabiliser ou modifier les états de coordination. Les auteurs proposent donc d'enrichir le modèle HKB en incluant des facteurs cognitifs comme l'attention.

Des relations entre les paramètres des gestes de la main et de la bouche ont aussi été mises en évidence dans des activités non-rythmiques par Gentilucci. Ces recherches montrent notamment une relation entre le contrôle moteur de la bouche et de la main dans les gestes de saisie. Par exemple, quand le participant doit ouvrir la bouche en saisissant un objet, l'ouverture de la bouche, comme celle de la main, s'adaptent à la taille de l'objet saisi (Gentilucci *et al.*, 2001). Cette influence de l'un des systèmes sur l'autre serait inter-modale, l'observation d'une action réalisée par un système (porter un fruit à la bouche) pouvant influencer les productions de l'autre (produire une syllabe) (Gentilucci *et al.*, 2004; Gentilucci, 2003).

Ces études montrent avant tout que la coordination entre la main et la bouche existe mais dépend du type de tâche. De plus, si les deux systèmes peuvent se coupler et former des synergies, avec une interaction entre les fréquences et les amplitudes des gestes de la parole et de la main, et des phasages préférentiels, les interactions entre les deux systèmes sont plus complexes que ne le prédisent les défenseurs d'une perspective dynamique stricte. La coordination dépend notam-

ment du type de geste manuel et surtout de la relation qu'il entretient avec la parole qui l'accompagne : quand les deux systèmes collaborent pour véhiculer du sens, comme c'est le cas du geste de pointé, la probabilité qu'ils se coordonnent est plus importante que lorsqu'ils visent des objectifs différents, ce qui pourrait expliquer les différences entre les résultats de Castiello *et al.* (1991) (où les réponses manuelle et verbale ne sont pas liées par les objectifs de la tâche) et ceux de Levelt *et al.* (1985) qui choisissent d'étudier le geste de pointé justement parce qu'il requiert une synchronisation entre les deux systèmes. Mais l'intérêt de l'étude du geste de pointé est qu'il est aussi un outil-clé de l'émergence du langage, notamment via le rôle que joue sa connexion avec la bouche dans le développement de la parole.

4.3 CAS PARTICULIER DU GESTE DE POINTÉ

La fonction première du langage est de communiquer des informations sur les objets de l'environnement. Si l'enfant est immergé dans le monde des mots dès sa vie intra-utérine, il doit apprendre à faire le lien entre les séquences sonores qu'il entend et les objets qu'elles désignent. Comme le rappelle si bien Helen Keller pour l'avoir redécouvert, la construction de ce lien est un déclic pour l'acquisition du langage. Pour mettre un nom sur un objet et véhiculer des informations sur cet objet, les interlocuteurs doivent avoir la capacité de diriger leur attention sur un espace délimité de leur environnement. Cette capacité, connue sous le terme *d'attention partagée* est liée au pointage brachiomanuel : le bébé commence par montrer avec le doigt, il associe ensuite de la parole à ses gestes de pointé pour enfin montrer avec le doigt et la parole. Pour de nombreux auteurs, le geste de pointé et sa coordination avec la parole sont des pierres fondatrices du développement et de l'évolution du langage.

4.3.1 Le déclic d'Helen Keller

Helen Keller (1880-1968) était une enfant normale jusqu'à devenir totalement sourde et aveugle à l'âge de 19 mois suite à une *fièvre cérébrale*. À partir de là, Helen vit dans ce qu'elle qualifiera elle-même plus tard un *non-monde*⁸. Privée de la capacité d'exprimer ce qu'elle ressent - les mouvements de sa bouche ne font plus de bruit, puis les quelques signes qu'elle utilise deviennent *de moins en moins adéquats*⁹ elle devient une enfant difficile avec des colères violentes. Son entourage pense que sa maladie l'a rendu idiot. En 1887, les parents d'Helen font appel aux services d'Anne Sullivan (1866-1936) pour l'instruction de leur fille. Dès son arrivée chez les Kellers, Anne épelle le nom des objets sur la paume des mains

⁸*Before my teacher came to me, I did not know that I am. I lived in a world that was no-world - Avant que mon professeur vienne à ma moi, je ne savais pas que j'étais. Je vivais dans un monde qui était un non-monde, Keller (1908)*

⁹*The few signs I used became less and less adequate (...). Keller (1902), p.11*



"We walked down the path to the well-house, attracted by the fragrance of the honey-suckle with which it was covered. Someone was drawing water and my teacher placed my hand under the spout. As the cool stream gushed over one hand she spelled into the other the word water, first slowly, then rapidly. I stood still, my whole attention fixed upon the motions of her fingers. Suddenly I felt a misty consciousness as of something forgotten, a thrill of returning thought, and somehow the mystery of language was revealed to me. I knew then that "w-a-t-e-r" meant the wonderful cool something that was flowing over my hand. (...) Everything had a name, and each name gave birth to a new thought." Keller (1902), p.15-16

"Nous marchions sur le chemin de la maison, attirées par le parfum du chevre-feuille dont il était recouvert. Quelqu'un puisait de l'eau et mon professeur mit mes mains en-dessous du tuyau. Tandis que l'eau fraîche coulait sur une de mes mains, elle épélaït sur l'autre le mot eau, d'abord doucement puis rapidement. Toute mon attention était rivée sur les mouvements de ses doigts. Soudainement, je pris vaguement conscience de quelque chose d'oublié, le tréssaïlement d'une pensée qui revenait, et c'est ainsi que le mystère du langage se révéla à moi. Je sus que « e-a-u » désignait la merveilleuse chose fraîche qui coulait sur mes mains. (...) Toute chose avait un nom, et chaque nom engendrait une nouvelle pensée."

FIG. 4.3 – La redécouverte du langage par Helen Keller. **A gauche**, Anne Sullivan lisant un livre à Helen Keller. **A droite**, extrait de l'autobiographie d'Helen Keller décrivant l'événement qui fit du 3 Mars 1887, le jour qu'elle considéra comme le plus important de sa vie.

d'Helen (voir la photographie d'Anne lisant un livre à Helen sur la figure 4.3.2). Helen devient rapidement capable de reproduire la *danse* des doigts d'Anne mais sans vraiment comprendre qu'elle désigne un objet précis. Et puis un jour, Helen fait le lien entre la séquence gestuelle d'Anne sur sa main et l'objet qu'elle désigne (voir l'émouvant et célèbre extrait de l'autobiographie d'Helen, figure 4.3.2). Elle sort alors de son *non-monde*. A partir de ce moment, son monde intérieur s'illumine et elle va n'avoir pas de cesse de poser la main de son instructrice sur les objets pour en demander le nom.

4.3.2 Le geste de l'attention partagée

Cette petite digression par l'histoire d'Helen Keller montre que le déclic du langage se fait chez elle quand elle comprend que les mouvements de son instructrice sur sa main ont un sens et que ce sens est de désigner un objet. Avant cette prise de conscience, Helen se focalise sur le doigt, au lieu de regarder le *ciel* qu'il montre... Chez l'enfant sans handicap sensoriel (ou avec prise en charge adéquate de son handicap), la capacité de partager l'attention sur un même espace de l'environnement s'établit progressivement au cours du développement et utilise le geste de pointé brachiomanuel (Tomasello *et al.*, 2007). Selon Butterworth (2003), la forme canonique du pointé émerge entre 9 et 12 mois. A 12 mois, les gestes de pointé constituent 60 % des gestes manuels communicatifs de l'enfant. A partir de 12 mois, le geste de pointé s'accompagne de vocalisations. Au départ, l'enfant pointe, puis, dans un second temps, cherche l'attention de l'adulte. A partir de 16 mois, il pointe seulement s'il a l'attention de l'adulte.

D'après Butterworth, le pointage manuel apparaîtrait spontanément, il ne serait pas appris. Cette universalité du geste de pointé est confirmée par les études interculturelles qui montrent cependant des variabilités dans la forme du geste, l'index n'étant pas toujours utilisé (Haviland, 2003; Wilkins, 2003). Pour certains auteurs, le geste de pointé impératif (*Donne-moi ça*, par opposition au pointé déclaratif, *Regarde ça*, selon la distinction de Baron-Cohen, 1995, cité par Ducey-Kaufmann, 2007) ferait suite à l'échec de gestes de saisie vers des objets hors de portée. Cette idée notamment défendue par Vygotski (1997) est remise en cause par différents arguments. Par exemple, pour Masataka (2003) le précurseur du geste de pointé serait le mouvement de flexion/extension de l'index, pour Butterworth (2003), le pointé servirait dès le départ des buts proto-déclaratifs, plutôt que proto-impératifs, et serait une fonction de la main différente de celle de la pince.

Toujours selon Butterworth (2003), le geste de pointé brachiomanuel est favorisé par rapport aux autres moyens d'attirer l'attention (comme le pointage par le regard ou par la tête) car il permet de désigner des objets au-delà de la périphérie corporelle. Le fait que les grands singes produisent des gestes de pointé en captivité avec apprentissage ou sans apprentissage, (Leavens *et al.*, 2004; Povinelli *et al.*, 2003) suscite de nombreuses questions notamment quant à l'origine biologique du geste de pointé (voir Leavens *et al.*, 2005; Leavens, 2004 pour un point de vue défendant des similitudes entre le pointage du grand singe et celui de l'homme ; et Povinelli *et al.*, 2003, 1999 pour une approche défendant des différences fondamentales entre les gestes des deux espèces). Cependant, si selon Leavens les grands singes vocalisent sur leur geste de pointé, montrant alors une préférence renforcée pour la main droite, seul l'être humain parle en pointant et qui plus est, apprend à le faire très tôt. Cette coordination des gestes orofaciaux dans la parole avec le geste du bras dans le pointage semble jouer un rôle fondamental dans le développement du langage.

4.3.3 La coordination pointé-parole, comme base fondatrice du langage

Selon Butterworth (2003), *le pointé est la voix royale vers le langage*¹⁰. Pour Goldin-Meadow (2007), l'âge auquel l'enfant utilise le geste de pointé détermine l'âge d'apprentissage et de la création du langage. Un des arguments allant en ce sens est la relation établie entre la coproduction d'un mot sur un geste de pointé et le langage propositionnel.

D'un mot sur un geste de pointé à deux mots

Le premier élément allant dans le sens du rôle de la coordination de la parole et du geste de pointé dans l'acquisition du langage est que l'association d'un mot et d'un geste de pointé véhiculant des informations complémentaires est une étape transitoire des énoncés d'un mot vers les énoncés de deux mots. Cette idée

¹⁰C'est d'ailleurs le titre de son article : *Pointing is the royal road toward language for babies*, (Butterworth, 2003)

évoquée précédemment est défendue par Goldin-Meadow et Butcher (2003). Ces auteurs rapportent une étude longitudinale réalisée à partir de films de 6 enfants en situation de communication, de l'étape de leurs premiers mots à celle de leurs productions à deux mots. Cette étude confirme que les premiers gestes manuels de l'enfant dans la communication sont déictiques (gestes qui montrent) et ne sont pas synchronisés avec des productions vocales. Selon Goldin-Meadow et Butcher, à cette étape, les deux systèmes ne sont pas unifiés. Dans une deuxième étape, l'enfant associe un mot et un geste de pointé (cf. montrer une bouteille en disant « bouteille »), les deux éléments fournissent des informations proches. L'apparition de ces associations s'accompagne d'une diminution des gestes isolés. Ensuite l'enfant associe un mot et un geste de pointé véhiculant des informations complémentaires mais non redondantes (cf. montrer une tortue en disant « partir »). La cohérence sémantique et la synchronisation temporelle des deux systèmes s'accordent selon Goldin-Meadow et Butcher avec l'idée d'un système de communication unifié, chaque élément faisant partie d'une même proposition. Enfin, vers 18 mois, l'enfant utilise deux mots. L'étape à laquelle l'enfant produit des combinaisons pointé-mot avec des informations complémentaires prédit l'étape des productions de deux mots. Volterra *et al.* (2005); Pizzuto *et al.* (2005) fournissent des résultats allant dans le même sens.

Vocaliser pour Localiser

Enfin, comme nous l'avons évoqué pour le rôle des gestes manuels en général, selon Iverson et Thelen (1999), la parole naîtrait d'une capture des fréquences rythmiques de la gestualité manuelle par le système moteur orofacial. Ainsi, les gestes *paveraient le chemin* pour le langage (Iverson et Goldin-Meadow, 2005), assertion que Tomasello *et al.* (2007) dérivent en : le pointé paverait le chemin de l'évolution et du développement pour l'émergence des symboles linguistiques. Dans ce contexte, Abry *et al.* (2004) ont défini un cadre de recherche pour les origines et le développement du langage autour de l'hypothèse *Vocalize to Localize* (*Vocaliser pour Localiser*). Selon cette hypothèse, la parole et le langage dériveraient de la nécessité de localiser les choses dont nous avons besoin de parler. Ainsi, le langage prendrait forme dans la rencontre entre l'outil primaire de vocalisation (les oscillations verticales de la mâchoire) avec l'outil primaire de localisation (le geste de pointé). Cette hypothèse fait de la rencontre entre la parole et le pointage brachio-manuel une équation clé des origines du langage qui jouerait un rôle essentiel aussi bien dans l'acquisition que dans la genèse des formes phonologiques, lexicales ou syntaxiques.

Le scénario introduit par Abry *et al.* pour une rencontre entre le cadre du signe et le cadre de la parole est représenté sur la figure 4.4. L'idée est que le cadre de la parole serait *enchâssé* dans le cadre du signe via un rapport entre les fréquences d'oscillations naturelles des deux systèmes de deux cycles de mâchoire pour un geste de pointé. Cette rencontre reposerait sur les propriétés des maturations nerveuses respectives des deux systèmes. Au chapitre 6 nous proposons un paradigme

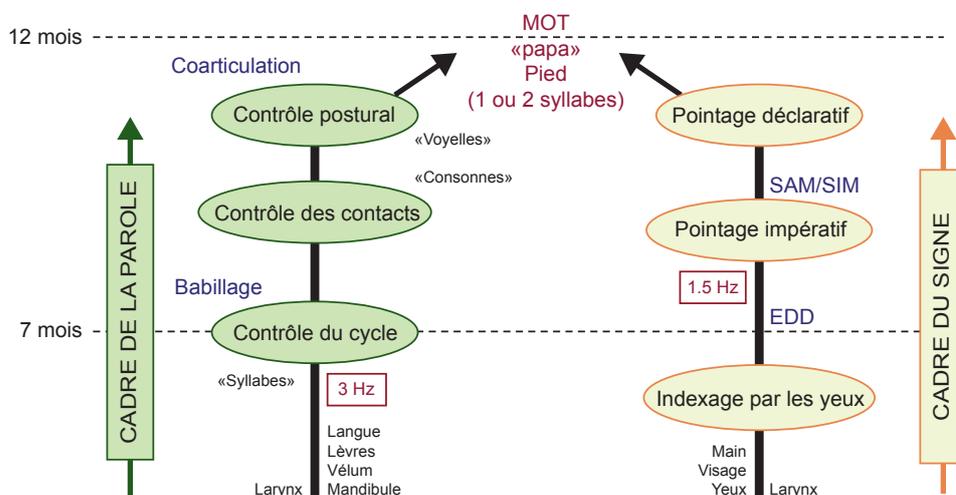


FIG. 4.4 – Enchâssement du cadre de la parole dans le cadre du signe via le pied (Ducey-Kaufmann *et al.*, 2007, version simplifiée). Le contrôle moteur du cadre de la parole et du cadre de signe se développent en parallèle dès la naissance jusqu'à se rencontrer à 12 mois. Le développement du cadre de la parole reprend celui proposé par MacNeilage et Davis (2000), voir figure 3.1, p.37, avec une préférence du contrôle de la mandibule donnant naissance à la syllabe puis une acquisition du contrôle indépendant des autres articulateurs pour aboutir au contrôle global sous-tendant la coarticulation. Le système de monstration débute par le regard puis, se poursuit par le pointage impératif puis le pointage déclaratif. Ce développement repose sur le développement du contrôle moteur du bras et sur le développement cognitif d'un système d'attention-intention partagé (SAM, Shared Attentional Mechanism, d'après Baron-Cohen (1995) ou SIM, Shared Intentional Mechanism, d'après Tomasello *et al.* (2005), cités par Ducey-Kaufmann *et al.*) développé après la détection de la direction du regard (EDD, Eyes Direction Detection). La fréquence de 3 Hz des oscillations de la mâchoire et celle de 1.5 Hz du geste de pointé aboutirait à la production de deux cycles de mâchoire sur un geste de pointé et favoriserait les mots d'une ou de deux syllabes, portés par l'unité accentuelle de la parole, le pied.

expérimental pour évaluer cette hypothèse d'un rapport 2 :1 chez l'adulte. Enfin, l'enchâssement de la parole dans le bras se ferait via l'unité rythmique de la parole, *le pied* constitué généralement de deux syllabes. Ainsi, selon Abry *et al.*, *le pied serait dans le bras*, c'est-à-dire que l'unité accentuelle du langage s'ancrerait sur le geste de pointé. Selon Abry *et al.*, cet ancrage pourrait expliquer la dominance des formes bisyllabiques dans les premiers mots (voir aussi Ducey-Kaufmann *et al.*, 2007; Ducey-Kaufmann, 2007).

Le précurseur de la deixis langagière

Le terme *deixis* vient du grec et signifie littéralement *force démonstrative, référence* (Oxford American Dictionaries). Ainsi, la fonction déictique du langage est sa fonction de montrer, désigner, référer aux objets de l'environnement physique ou conceptuel. Le pointage brachiomanuel est un système déictique puissant et universel. Pourtant, la *deixis* est aussi portée par la parole via la prosodie (« MADELEINE m'amena ») ou la grammaire (« C'est Madeleine qui m'amena », voir Lævenbruck *et al.*, 2005). Selon Diessel (1999), les démonstratifs seraient des objets linguistiques universaux et primaires, c'est-à-dire non dérivés lexicalement.

Au contraire, ils émergeraient dans une séquence développementale débutant par le geste de pointé manuel, se poursuivant par la production de parole sur ce geste de pointé pour aboutir au pointage par la voix. Dans ce cadre, le pointage a été considéré comme une forme canonique des démonstratifs et plus généralement, des indexicaux du langage (Haviland, 2000). Cependant, la co-production des deux formes de monstration perdure dans la communication entre adultes : on montre avec le doigt et avec la parole en même temps. Pour jouer son rôle dans la communication, cette association requiert que l'interlocuteur établisse un lien entre ce qui est dit et ce qui est pointé. Ce lien repose essentiellement sur le rapprochement événementiel de la partie de la parole qui désigne (le focus) avec la partie du geste qui montre (l'alignement avec la cible). Ce rapprochement pourrait se fonder sur la coordination du cadre du signe et du cadre de la parole, question que nous étudions au chapitre suivant.

Ces différents éléments incitent à développer des paradigmes expérimentaux étudiant les fondements d'une coordination main-parole relativement aux hypothèses théoriques avancées pour le rôle de la coordination main-bouche dans l'émergence des formes du langage. C'est dans cette voie que s'inscrivent nos travaux expérimentaux présentés dans les deux chapitres suivants. Leur objectif est de fournir des éléments de résolution à l'équation de la coordination main-parole dans la deixis en abordant deux grandes questions : ① Où les deux systèmes se rencontrent-ils ? Et plus précisément, la partie accentuée de la parole est-elle un site de synchronisation plausible pour la coordination pointé-parole ? et, ② Comment cette rencontre est-elle portée par le cadre du signe et le cadre de la parole ? Ou encore, le ratio de deux cycles de mâchoire pour un geste de pointé avancé par Abry *et al.* (2004) s'observe-t-il chez l'adulte et est-il un fondement plausible de la coordination de la mâchoire et du bras ?

La partie accentuée de la parole comme site de synchronisation pour la parole et le geste de pointé

Ce chapitre est le fruit d'une collaboration avec Rafael Laboissière et Arturo Galván à l'institut Max Plank des sciences cognitives et du cerveau de Munich. Des analyses plus succinctes des deux expériences rapportées sont aussi disponibles en anglais dans Rochet-Capellan et al. (2006, 2007a,b).

5.1 INTRODUCTION

5.1.1 Contexte et enjeux théoriques

Ce chapitre aborde la question des sites de synchronisation entre la parole et la main dans une tâche déictique (montrer une cible en la nommant). A l'issue du chapitre précédent, il ressort que la coordination entre la main et la bouche dans la communication serait une étape clé de l'évolution et de l'acquisition du langage. Le rôle de ce lien est particulièrement clair en ce qui concerne la fonction déictique. Ainsi, la deixis langagière pourrait prendre forme dans la coordination entre la parole et le pointage brachiomanuel. Les modalités de cette coordination, et par conséquent, de la synchronisation entre les deux systèmes contribueraient à façonner les mots du langage qui montrent : les démonstratifs (Diessel, 1999), mais aussi les premiers mots (Abry *et al.*, 2004). En deçà de ces objectifs théoriques liés à la question des origines du langage, l'étude des sites de synchronisation entre la main et la parole devrait enrichir la compréhension des processus de la coordination des deux systèmes. L'objectif est ici de contribuer à cet enrichissement en étudiant l'effet de la variation de la position de l'accentuation de la parole sur la coordination entre la parole et le geste de pointé brachiomanuel.

5.1.2 Synchronisation main-voix dans la deixis

D'après McNeill (1992), la coordination main-parole en général repose sur la nécessité de synchroniser le geste de la main avec le mot auquel il est relié de manière sémantique ou co-expressive. Dans le cas du geste de pointé, l'extrémité (ou l'apex) du geste de pointé se synchroniserait avec la partie de la parole qui montre (le focus). Si la question des sites de synchronisation n'a pas fait l'objet, à notre connaissance, d'étude systématique, l'étude princeps de Levelt *et al.* (1985) sur les processus de synchronisation entre la main et la voix dans la deixis rend compte d'une adaptation de la parole au geste de pointé. Cette adaptation tend à maintenir un rapprochement entre le début de la voix et l'apex du geste de pointé (voir chapitre précédent). Cependant, cette étude utilise un énoncé de type « cette lampe » dans lequel le focus est proche du début de la voix (« cette »). Les résultats sont donc aussi compatibles avec l'hypothèse d'un rapprochement entre le focus de la parole et l'apex du geste de pointé. Afin de départager ces deux interprétations des données, il est nécessaire d'étudier l'impact de la variation de la position du focus de la parole sur la coordination main-parole. De premiers travaux allant en ce sens et utilisant un focus grammatical sont actuellement en cours dans notre laboratoire et des résultats préliminaires disponibles dans Laurent (2007) avec des énoncés de type « CET homme-ci » vs. « Cet homme-CI ».

5.1.3 Synchronisation main-mâchoire dans la deixis

En parallèle de ces travaux, notre objectif est d'étudier l'effet de la position du focus sur la coordination parole-pointé en introduisant l'analyse des mouvements de la mâchoire sous-tendant cette coordination. Une des limites de l'étude de Levelt *et al.* (1985) (comme celles de Holender, 1980; Castiello *et al.*, 1991; Feyereisen, 1997) est qu'elle ne considère pas les mouvements des articulateurs. Les auteurs caractérisent la réponse verbale exclusivement par la voix et réservent le terme *geste* à la réponse manuelle en opposant *réponse gestuelle* à *réponse verbale*. Bien que plus terminologique qu'*idéologique* (et certainement matérielle : il est plus difficile d'enregistrer les mouvements des articulateurs que ceux de la main), cette opposition se heurte au fameux adage de Stetson (1951), selon lequel *la parole est plus un ensemble de mouvements rendus audibles qu'un ensemble de sons produits par le mouvement*, p. 29, (voir chapitre 2 pour plus de détails sur la conception de Stetson). D'autre part, comme nous l'avons exposé au chapitre précédent, les études développementales amènent à supposer que la connexion entre la parole et la main dans la deixis pourrait être une rencontre entre le cadre du signe (le geste de pointé) et le cadre de la parole (les oscillations verticales de la mâchoire), voir Ducey-Kaufmann *et al.* (2007).

Afin d'étudier l'effet de la position du focus sur la coordination entre la mâchoire et le geste de pointé de la main, nous proposons ici d'utiliser un matériel phonétique simple permettant d'interpréter la gestuelle de la mâchoire sans am-

bigüité, à savoir des séquences bi-syllabiques CVCV (C, pour consonne, V, pour voyelle), qui en condition de production normale, requièrent un cycle de mâchoire par syllabe. Le focus est opérationnalisé ici de manière simplifiée en variant la position de la syllabe accentuée : première ('CVCV) vs. deuxième syllabe (CV'CV). Il s'agit donc plus d'une accentuation lexicale que d'un focus proprement dit.

5.1.4 Hypothèses et expériences

L'hypothèse de base de ce travail est que, dans une tâche consistant à nommer une cible par un bisyllabe CVCV accentué sur la première ou sur la deuxième syllabe, l'apex du geste de pointé devrait se synchroniser avec la position la plus basse du premier cycle de mâchoire quand l'accent est sur la première syllabe (cf. /pápa/) et avec la position la plus basse du deuxième cycle de la mâchoire quand l'accent est sur la deuxième syllabe (cf. /papá/), voir figure 5.1. Cet alignement pourrait résulter d'une adaptation unidirectionnelle de la mâchoire, avec un geste de pointé identique pour les séquences 'CVCV et CV'CV, comme le suggèrent les résultats de Levelt *et al.* (1985), ou bien d'une adaptation bidirectionnelle des deux systèmes impliquant à la fois une modification des mouvements de la mâchoire et des mouvements de la main. Ces questions sont l'objet de la première expérience qui utilise les bisyllabes /papa/ et /tata/, accentués sur la première ou la deuxième syllabe comme noms de cibles. D'autre part, l'hypothèse d'une coordination mâchoire-geste de pointé, amène la question de l'influence des profils de la mâchoire sur la coordination et de son interaction avec l'effet de la position de l'accent. Or, au chapitre 2, nous avons mis en évidence des profils de mâchoire différents pour les bisyllabes labial-coronal (L_aC_o) et coronal-labial (C_oL_a) par rapport aux séquences dupliquées (/papa/, /tata/) d'une part, mais aussi d'une séquence L_aC_o (/pata/) à son symétrique C_oL_a (/tapa/). Cette question est abordée dans la deuxième expérience utilisant les bisyllabes /pata/ et /tapa/, accentués sur la première ou la deuxième syllabe comme noms de cibles.

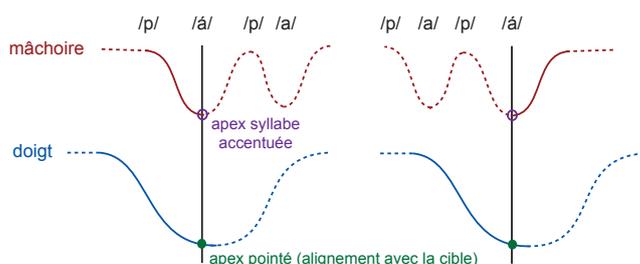


FIG. 5.1 – L'hypothèse de travail est que l'apex du geste de pointé devrait advenir, à proximité de la position basse du cycle de la mâchoire pour la voyelle accentuée. Aucune hypothèse spécifique n'est émise quant à la forme précise des gestes (d'où les pointillés).

NOTE MÉTHODOLOGIQUE

Le choix de traiter des différentes questions dans deux expériences séparées a été motivé par la volonté de limiter le nombre de facteurs expérimentaux et le nombre de niveaux de ces facteurs. Ce choix permet une analyse et une interprétation plus claires des résultats suivant la logique : la première expérience s'intéresse à l'effet de la position de l'accent et la deuxième expérience, à l'interaction éventuelle de cet effet avec les différences de profils de mâchoire entre les structures labiales-coronales et coronales-labiales. Chaque locut(rice)eur passait les deux expériences présentées ici dans des ordres contrebalancés. Bien que non idéale sur le plan méthodologique, cette solution nous a paru être le meilleur compromis entre nos objectifs et les contraintes matérielles rencontrées. L'effet de l'ordre de passation des expériences a été systématiquement contrôlé pour chacune des analyses statistiques.

D'autre part, la littérature ayant largement montré l'effet de la présence du geste de pointé sur l'initialisation de la parole, via la comparaison entre les performances en doubles tâches et tâches simples, les expériences n'incluent pas de condition parole-seule ou geste-seul. A posteriori, ce choix limite l'interprétation de certains résultats mais s'accorde, a priori, avec les objectifs des expériences visant à étudier l'effet de la position de l'accent et des profils de mâchoire sur la coordination doigt-mâchoire.

5.2 MÉTHODE

Exceptée la variation du type de bisyllabe avec lequel dénommer la cible, la méthode est la même pour les deux expériences. L'enregistrement synchrone des mouvements de la mâchoire et de la main a été réalisé via la mise au point d'un protocole expérimental utilisant un système Optotrak. La procédure expérimentale est inspirée de celle de Levelt *et al.* (1985). Une fois installé(e), le ou la participant(e) passait les deux expériences successivement, en commençant soit par la première, soit par la deuxième, avec une pause libre entre les deux expériences.

5.2.1 Langue et participants

Les participants étaient au nombre de vingt, quatre hommes et seize femmes, droitiers, âgés de 18 à 38 ans, et ne souffrant pas, à leur connaissance, de troubles de la parole ou de problème de vision. Ils avaient le portugais brésilien pour langue maternelle, langue qui autorise les paires de bisyllabes CVCV accentués sur la première ('CVCV, cf. /pápa/) ou sur la deuxième syllabe (CV'CV, cf. /papá/). Cette propriété de la langue (langue à accent lexical) est un critère de sélection important pour que la tâche d'accentuation soit *naturelle* pour les participants.

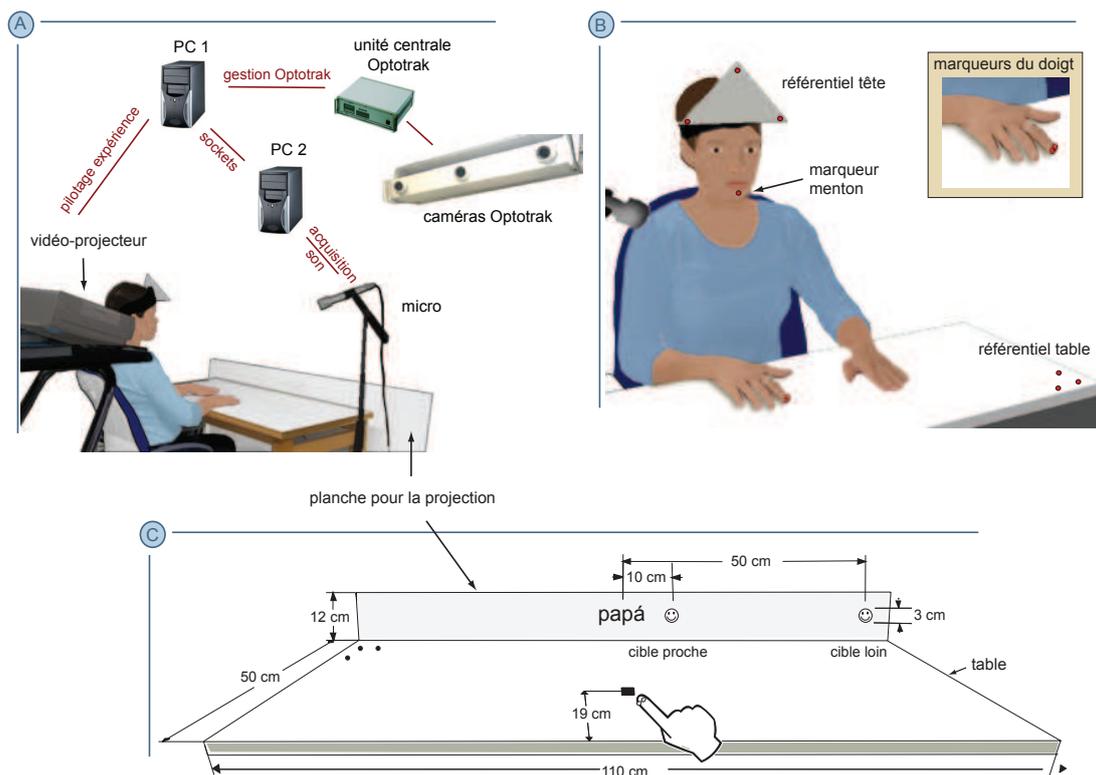


FIG. 5.2 – Dispositif expérimental. **En A** : Enregistrement des données et pilotage de l'expérience. La participante est assise à une table, la cible à pointer est projetée face à elle au moyen d'un vidéo-projecteur relié à la sortie vidéo d'un ordinateur (PC1) sur lequel s'exécute le programme de pilotage de l'expérience. PC1 gère aussi l'unité centrale de l'Optotrak (début et fin des enregistrements, récupération des données). Les caméras de l'Optotrak sont situées face à la locutrice à environ 2 m du sol. Les productions sonores sont enregistrées via un micro relié à la carte son d'un deuxième ordinateur (PC2). La synchronisation des signaux Optotrak et acoustiques se fait par communication entre les programmes des deux ordinateurs (sockets). **En B** : Positions des marqueurs Optotrak pour l'enregistrement des mouvements de la mâchoire (marqueur du menton) et du doigt (marqueur central et latéral). Les trois marqueurs sur la tête et sur la main sont des référentiels pour le calcul des coordonnées des marqueurs en mouvement lors du post-traitement des données. **En C** : Dimensions de la projection et de la distance à la cible. Le carré noir sur la table indique la position de départ et de retour du doigt pour le geste de pointé. La cible est une icône *smiley* projetée dans le champ visuel droit du sujet, à une position proche ou lointaine. Le nom par lequel elle doit être désignée est projeté au centre du champ visuel.

5.2.2 Dispositif expérimental

Installation

Le dispositif expérimental est représenté sur la figure 5.2, en A. La participante est assise à une table, dos à un vidéoprojecteur et face à une planche-écran sur laquelle est projetée la cible à pointer. Deux ordinateurs sont utilisés pour l'enregistrement du son et le pilotage de l'expérience et de l'Optotrak. Une communication entre les programmes s'exécutant sur chacun des ordinateurs permet de synchroniser les signaux sonores et Optotrak. Le vidéoprojecteur est installé de façon à respecter les dimensions indiquées sur la figure 5.2, en C.

Positionnement des marqueurs de l'Optotrak

Les positions des marqueurs de l'Optotrak (diodes émettrices infra-rouges) sont illustrées sur la figure 5.2, en B. Les pilotes expérimentaux ayant révélé que certain(e)s participant(e)s tendent à basculer leur index vers la droite au cours du geste de pointé, deux marqueurs sont utilisés pour le suivi du mouvement de l'index droit. Le premier, au centre de l'ongle, reste dans le champ des caméras de l'Optotrak quand la face de l'ongle est parallèle à la table. Le deuxième, sur la partie latérale gauche du bout du doigt, à la limite de l'ongle, est visible des caméras quand le doigt bascule vers la droite. Le troisième marqueur collé sur le menton permet de suivre les mouvements de la mâchoire. Dans l'absolu, ce marqueur dépend plus des mouvements d'un point de chair que de ceux de la mâchoire elle-même, la peau recouvrant le menton pouvant, dans certaines conditions, bouger indépendamment de la mâchoire. Cependant, relativement au matériel phonétique utilisé dans nos expériences, consistant en une consonne plosive (/p/ ou /t/) suivie d'une voyelle ouverte (/a/), la principale source de mouvement de ce point peut être attribuée à la mâchoire. Les trois marqueurs fixes collés à l'angle supérieur gauche de la table forment un référentiel pour le calcul des marqueurs mobiles (référentiel *table*). De plus, la tête pouvant bouger au cours de l'expérience, trois marqueurs supplémentaires sont disposés sur la tête de la participante à l'aide d'une équerre fixée sur un bandeau. Ce référentiel est indispensable pour corriger les mouvements de la tête lors du calcul des coordonnées de la mâchoire. Pour ces expériences, la fréquence d'échantillonnage des coordonnées tri-dimensionnelles des marqueurs était de 100 Hz.

5.2.3 Procédure

Tâche et plan d'expérience

La tâche était de pointer une cible avec l'index de la main droite tout en la nommant. Cette cible était une icône *smiley* pouvant apparaître à une position proche (10 cm du centre du champ visuel) vs. loin (50 cm du centre du champ visuel), toujours dans le champ visuel droit du/de la participant(e), voir figure 5.2, en C. Afin de varier le nom de la cible de manière simple et sans introduire de charge cogni-

tive supplémentaire comme la mémorisation de différentes associations cible-nom, la cible était toujours la même et son nom était affiché au début de chaque essai. Ce nom pouvait être /papa/ ou /tata/, dans la première expérience et /pata/ ou /tapa/, dans la deuxième expérience, accentué sur la première ou sur la deuxième syllabe. Chacune des deux expériences comportait donc trois facteurs à deux niveaux : position de la cible (*cible_{proche}* vs. *cible_{loin}*), position de l'accent (*accent₁* vs. *accent₂*) et séquence (*seq_{/pp/}* vs. *seq_{/tt/}*, pour la première expérience et *seq_{/pt/}* vs. *seq_{/tp/}* pour la deuxième), soit en tout, huit conditions expérimentales passées par tous les participants.

D'autre part, la tâche était de type *off-line* (voir Levelt *et al.*, 1985) : la réponse n'était pas donnée en réaction à l'apparition de la cible mais à l'émission d'un signal suivant cette apparition, ici, un changement de couleur de la cible. Ainsi, dans les analyses ci-après, les temps de réponse depuis le signal-go jusqu'à l'initialisation des mouvements n'incluent pas les processus d'identification (notamment le temps requis pour tourner les yeux et/ou la tête vers la cible) et de planification, comme c'est le cas en condition *on-line*, où la réponse est donnée en réaction à l'apparition de la cible.

Consigne et déroulement

La consigne était la même pour les deux expériences et était donnée par l'expérimentateur avant la mise en place des marqueurs Optotrak pour l'enregistrement :

« Une cible va apparaître en rouge sur la planche, elle représente un personnage dont le nom s'affichera au centre de la planche. Ce nom sera un mot de deux syllabes accentué sur la première ou sur la deuxième syllabe. Il s'agira de pointer cette cible avec l'index tout en la nommant avec le mot dès qu'elle deviendra verte, et ce, en respectant l'indication accentuelle ».

Afin d'aider le/la participant(e) à comprendre la tâche, l'expérimentateur l'invitait à pointer des objets dans la salle d'expérimentation tout en les nommant (cf. « table »+pointé, « chaise »+pointé etc.). Aucune contrainte n'était donnée sur la vitesse d'exécution du geste. La position de départ et de retour du doigt était indiquée par un carré noir collé sur la table, le/la participant(e) devait positionner l'extrémité de son index sur cette marque avant et après chaque essai.

Après la mise en place des marqueurs Optotrak, le/la participant(e) était invité(e) à lire la liste des mots utilisés pour dénommer la cible à haute voix afin de s'assurer qu'il/elle avait bien compris la consigne d'accentuation et d'effectuer les réglages requis à l'acquisition du son. Il/elle réalisait aussi quelques essais pour vérifier la visibilité des marqueurs par les caméras de l'Optotrak.

Le déroulement de chaque expérience est schématisé sur la figure 5.3. La passation se divisait en quatre blocs de 44 essais chacun, dont 4 essais d'entraînement suivis de 40 essais expérimentaux (5 essais par condition expérimentale). Les es-

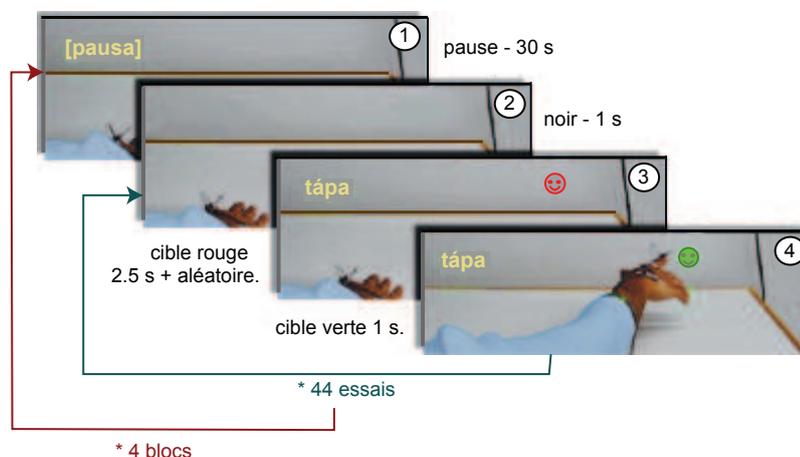


FIG. 5.3 – Succession des écrans durant la phase expérimentale. L'expérience est subdivisée en quatre blocs de 44 essais chacun. Les blocs sont passés successivement avec une pause de 30 s entre deux blocs indiquée par l'écran « pausa » (①). Un essai débute par l'affichage du nom de la cible au centre du champ visuel et de la cible en rouge dans le champ visuel droit du sujet, à une position proche ou loin (③). La durée d'affichage de cet écran est de 2.5 s \pm un temps aléatoire. La cible devient ensuite verte, c'est le signal auquel le/la participant(e) doit donner sa réponse (④). La cible verte reste à l'écran pendant 1 s. Deux essais sont séparés par un écran noir d'1 s (②).

sais expérimentaux étaient arrangés de manière aléatoire à l'intérieur de chaque bloc et différemment pour chaque participant(e). Les blocs étaient passés successivement, avec une pause de 30 s entre deux blocs, indiquée par un écran *pausa* suivi d'un écran noir pour une durée d'1 s. Un essai débutait par l'affichage de la cible en rouge et de son nom, au centre du champ visuel pour une durée de 2.5 s + un temps aléatoire tiré dans une distribution gaussienne de moyenne 0 s et d'écart type 0.15 s. La cible devenait ensuite verte pour une durée d'1 s, suivie de l'écran noir d'1 s, séparant deux essais consécutifs. La durée approximative de la phase expérimentale était de 15 min.

A la fin de la pause entre les deux expériences, quand le/la participant(e) s'estimait prêt(e) pour la deuxième expérience, l'expérimentateur indiquait qu'il s'agirait de la même tâche mais avec des mots différents pour les noms de cibles.

5.2.4 Post-traitements des signaux

Le post-traitement des données Optotrak comporte deux grandes étapes. ① Le calcul des coordonnées des marqueurs du doigt dans le référentiel table et de celle de la mâchoire dans le référentiel tête. ② L'évaluation de la direction principale du mouvement afin de passer d'une représentation tri-dimensionnelle à une représentation unidimensionnelle du mouvement.

Calcul des coordonnées des marqueurs du doigt et de la mâchoire

La détermination des coordonnées des marqueurs se fait en quatre étapes : ① Calcul du référentiel table à partir des coordonnées x-y-z des trois marqueurs collés sur la table ; ② Projection des coordonnées des marqueurs du doigt dans le référentiel table orthonormé, ③ Calcul du référentiel tête avec les coordonnées x-y-z des trois marqueurs de la tête ; ④ Projection des coordonnées du marqueur de la mâchoire dans le référentiel tête orthonormé. Les coordonnées ainsi obtenues sont ensuite utilisées pour calculer la direction principale du mouvement.

Calcul de la direction principale du mouvement

Le calcul de la direction principale du mouvement est effectué par une Analyse en Composantes Principales (ACP) des coordonnées tridimensionnelles des marqueurs. Ce choix se justifie par le fait que les mouvements des capteurs étudiés sont principalement linéaires et dans une seule dimension. Ainsi, sur l'ensemble des locuteurs, la part de variance expliquée par la première composante de l'ACP va de 95% (pour la mâchoire) à 99% (pour les marqueurs du doigt) et ce, pour les deux expériences. Cette première composante peut être assimilée aux mouvements verticaux pour le marqueur de la mâchoire - alternance d'ouverture et de fermeture - et au déplacement dans le plan parallèle à la table pour les marqueurs du doigt. L'ACP est précédée d'une interpolation linéaire des coordonnées visant à combler les échantillons perdus au cours de l'enregistrement suite à des masquages occasionnels des marqueurs. Les signaux ainsi obtenus sont ensuite filtrés par un filtre passe-bas de Butterworth à une fréquence de 15 Hz.

Étiquetage et contrôle des données

La figure 5.4 donne un exemple de signaux et d'étiquetage pour un essai /papá/. Les initialisations et les fins (apex) des mouvements du doigt et de la mâchoire sont repérées aux instants où le mouvement atteint 10% de son pic de vitesse, respectivement avant et après ce pic. Cette mesure est une mesure standard pour déterminer les débuts et les fins des gestes et permet d'automatiser l'étiquetage.

Segmentation de la trajectoire de la mâchoire. La trajectoire de la mâchoire pour la production d'un bisyllabe CVCV se compose d'un cycle par syllabe. Chaque cycle comporte un geste (ou stroke) d'ouverture (descendant) suivi d'un stroke de fermeture (montant). La fin d'un stroke peut être confondue avec le début du stroke suivant ou former une transition. Ce phénomène peut apparaître entre la fin du stroke de fermeture du premier cycle et le début du stroke d'ouverture du deuxième cycle (transition inter-cycles) ou bien entre le stroke d'ouverture et le stroke de fermeture d'un même cycle (transition intra-cycle). Pour chacune des deux expériences, les transitions inter-cycles s'observent dans environ 60% des essais avec une durée inférieure à 15 ms pour environ 75% des essais. Les transitions intra-cycles sont moins fréquentes et concernent seulement 40% des essais, pour le premier et le deuxième cycle. De plus, leur durée est inférieure à 15 ms dans 90 %

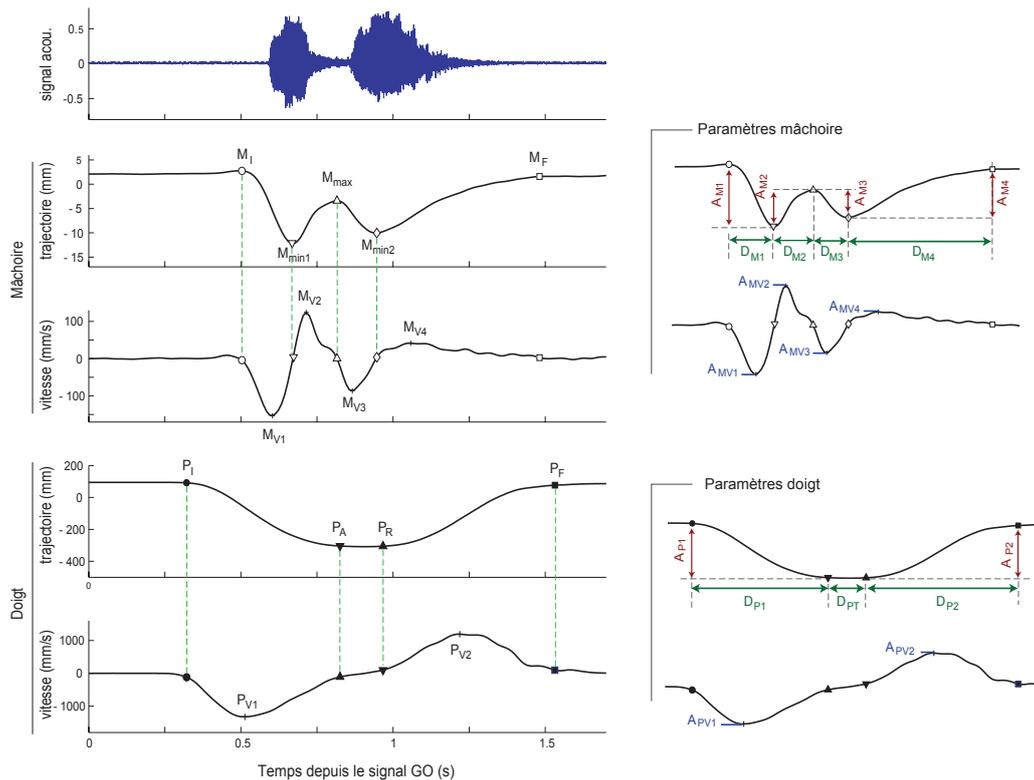


FIG. 5.4 – Signaux, étiquetages et paramètres des mouvements de la mâchoire et du doigt pour un essai /papá/. **A gauche**, de haut en bas : signal acoustique ; courbes de trajectoire et de vitesse de la mâchoire ; courbes de trajectoire et de vitesse du doigt. La trajectoire de la mâchoire comprend deux cycles, un pour chaque syllabe. Le premier cycle va de M_I , l'initialisation du premier stroke descendant (10% du pic de vitesse, M_{V1}) à M_{max} , la position de la mâchoire la plus haute entre les deux cycles (passage par zéro de la vitesse). Le deuxième cycle va de M_{max} à M_F , la fin du deuxième stroke montant (10% du pic de vitesse M_{V3}). M_{min1} et M_{min2} sont les positions les plus basses de la mâchoire pour le premier et le deuxième cycle. Pour le doigt, P_I , P_A sont l'initialisation et l'apex du geste de pointé et P_R et P_F sont l'initialisation et la fin du geste de retour. **A droite**, paramètres des mouvements. Les phases montantes et descendantes des cycles de la mâchoire sont caractérisées par leur amplitude, leur durée et l'amplitude de leur pic de vitesse, respectivement A_{Mn} , D_{Mn} et V_{Mn} . De la même manière, le geste de pointé et de retour du doigt sont caractérisés par leur amplitude (A_P et A_{PR}), leur durée (D_P et D_{PR}) et l'amplitude de leur pic de vitesse (V_P et V_{PR}). D_{PT} est la durée de la tenue du geste de pointé.

des essais. Ces observations globales rendent l'étude de ces transitions peu pertinente aux vus des objectifs de ce travail. Afin d'éviter ces délais entre les gestes dans la segmentation de la trajectoire, trois événements ont été repérés : la position la plus basse de la mâchoire entre la fin du stroke d'ouverture et le début du stroke de fermeture, M_{min1} pour le premier cycle et M_{min2} pour le deuxième, et la position la plus haute de la mâchoire entre la fin du stroke montant du premier cycle et le début du stroke descendant du deuxième cycle, M_{max} . Ainsi, pour le premier cycle, la phase descendante va du début du stroke ouvrant (M_I) à la position la plus basse (M_{min1}) et la phase montante, de M_{min1} à la position la plus haute entre les deux cycles (M_{max}). Pour le deuxième cycle, la phase descendante va de M_{max} à la position la plus basse, M_{min2} et la phase montante, de M_{min2} à la fin du stroke montant (M_F), cette dernière phase advient après la dénomination de la cible (dernière voyelle) et n'est pas contrôlée. Cependant nous avons choisi de l'inclure dans les analyses à titre indicatif.

Segmentation de la trajectoire du doigt. De son côté, la trajectoire du doigt peut être découpée en un geste de pointé, de l'initialisation, P_I , à l'apex, P_A et un geste de retour à la position indiquée sur la table, de P_R à P_F . Contrairement à la mâchoire, la fin du stroke d'aller n'est jamais confondue avec le début du stroke de retour. Au contraire, un plateau apparaît entre P_A et P_R avec une durée supérieure à 50 ms dans 75% des essais pour chacune des expériences. Cette phase de tenue représente la période pendant laquelle le doigt montre la cible.

Validation. Les trajectoires et les étiquetages ont été visualisés afin de vérifier systématiquement, pour chaque essai, les irrégularités de trajectoire dues au masquage trop étendu d'un marqueur et de valider l'étiquetage automatique. Les signaux acoustiques ont aussi été écoutés indépendamment afin de repérer les erreurs sur le nom de la cible (consonnes et/ou position de l'accent). L'analyse ci-après concerne uniquement les essais :

- sans erreur sur le nom de la cible ;
- et pour lesquels le début de la réponse de la mâchoire (M_I) et celle du geste de pointé (P_I) adviennent après le signal go ;
- et sans masquage important du marqueur de la mâchoire de M_I à M_F ;
- et avec au moins un des deux capteurs du doigt visible de P_I à P_F .

En ce qui concerne le mouvement du doigt, par défaut, le marqueur retenu est celui collé sur le milieu de l'ongle. Lorsque ce marqueur est masqué, c'est celui sur la partie latérale gauche de l'ongle qui est étudié (la corrélation moyenne entre les coordonnées des deux marqueurs au cours du temps est de .99 pour chaque coordonnée x, y et z). Des ANOVAs intra-sujets montrent que les facteurs « séquence », « position de la cible » et « position de l'accent » n'ont pas d'effet significatif sur les différents types d'erreurs (voir tableau A.1, p.180, en annexe).

D'autre part, certains essais accentués sur la première syllabe sont produits sur un seul cycle de mâchoire (cf. le/la participant(e) dit « tát' » pour /táta/), c'est le

cas pour 57 essais /táta/, répartis sur six locuteurs pour la première expérience et de 65 essais /páta/ répartis sur 8 locuteurs pour la deuxième. Pour ces essais, M_{max} , M_{min2} et M_F ont été marqués au même moment.

5.2.5 Mesures et hypothèses

L'hypothèse de travail est que la coordination entre la parole et le geste de pointé manuel serait contrainte par un rapprochement entre la partie accentuée de la parole et l'apex du geste de pointé. Par conséquent, en condition *accent*₁, l'occurrence de P_A devrait être proche de celle de M_{min1} alors qu'en condition *accent*₂, P_A devrait se rapprocher de M_{min2} . Cette hypothèse est évaluée ci-après via l'analyse des délais entre l'occurrence des principaux événements des gestes du doigt relativement à ceux de la mâchoire.

D'autre part, ce rapprochement entre l'apex du geste de pointé et la position basse de la mâchoire pour la voyelle accentuée pourrait résulter d'une adaptation unidirectionnelle de la mâchoire ou d'une adaptation bidirectionnelle de la mâchoire et du doigt. Ce dernier point est évalué ci-après via l'analyse de l'effet de la position de l'accent sur l'organisation temporelle et les paramètres des trajectoires de la mâchoire et du doigt. Ces différents paramètres (durée, amplitude et amplitude des pics de vitesse) sont représentés sur la figure 5.4, p.98, à droite. Dans le cas d'une adaptation unidirectionnelle de la mâchoire, seul le timing et les paramètres des mouvements de la mâchoire devraient être affectés par les variations expérimentales. Dans le cas d'une adaptation bidirectionnelle, les variations expérimentales devraient affecter les paramètres des mouvements de la mâchoire et du doigt.

5.3 RÉSULTATS

Les effets des trois facteurs sur les différentes mesures sont testés ci-après pour chaque expérience avec des ANOVAs à trois facteurs croisés intrasujets : *Position accent* × *Position cible* × *Séquence*. Quand les interactions entre les facteurs sont significatives, les comparaisons multiples indépendantes sont testées avec des t-tests appareillés avec correction de Bonferroni. Pour toutes les analyses (ANOVAs), le seuil de significativité est fixé à $p \geq 0.05$. Les ANOVAs ont été réalisées après contrôle systématique de l'effet de l'ordre de passation des deux expériences. Ce facteur contrôlé inter-sujet n'a pas d'effet notable sur les différentes mesures rapportées ici, il n'est donc pas mentionné par la suite.

5.3.1 Effet de la position de l'accent sur la coordination mâchoire-doigt

L'objectif est ici d'évaluer l'effet de la position de l'accent sur la coordination mâchoire-doigt et de voir s'il dépend de la position de la cible et de la séquence. La figure 5.5 représente la moyenne et l'erreur standard moyenne des délais entre le signal-go et l'occurrence des différents événements de la mâchoire et du doigt pour

la première expérience (en haut) et la deuxième (en bas). La comparaison entre les deux figures montre d'ores et déjà des organisations temporelles très proches entre les deux expériences. Cette similitude se retrouve pour toutes les analyses des mesures du positionnement des événements de la trajectoire du doigt relativement aux cycles de la mâchoire. Pour s'en convaincre, le lecteur pourra comparer les tables d'ANOVAs de l'expérience 1, tableaux A.2 et A.4, p.182 et p.184, à celles de l'expérience 2, tableaux A.3 et A.5, p.183 et p.185. Ainsi, seules les analyses des données de l'expérience 1 sont détaillées dans cette première partie des résultats, les conclusions étant en tout point applicables à l'analyse de l'expérience 2.

Points de rencontre entre la mâchoire et le doigt : tendances globales

La figure 5.5 met en évidence trois éléments importants en ce qui concerne l'effet de la position de l'accent sur la coordination mâchoire-doigt. D'abord, dans toutes les conditions expérimentales, l'initialisation du geste de pointé (P_I) advient avant celle du premier cycle de la mâchoire (M_I). Ensuite, en condition *accent*₁, l'apex du geste de pointé (P_A) est proche de la position basse du premier cycle de la mâchoire (M_{min1}) alors qu'en condition *accent*₂, P_A advient à proximité de la position haute de la mâchoire (M_{max}). Enfin, en condition *accent*₂, le début du geste de retour du doigt, P_R , est proche de la position basse du deuxième cycle de la mâchoire (M_{min2}) alors qu'il advient avant M_{max} en *accent*₁. D'autre part, la position de la cible et la séquence paraissent avoir peu d'effet sur ces différents points de rencontre. Ces premières données montrent que le rapprochement attendu entre l'apex du geste de pointé et la position basse de la mâchoire pour la voyelle accentuée pourrait être réalisé en condition *accent*₁ mais pas en condition *accent*₂. Elles invitent aussi à analyser plus précisément les différents rapprochements observés et plus globalement, la manière dont les événements de la trajectoire du doigt se positionnent par rapport aux cycles de la mâchoire selon la position de l'accent.

A cette fin, les principaux événements de la trajectoire du doigt, P_I , P_A , P_R et P_F , ont été positionnées relativement aux cycles de mâchoire normalisés de -100% (M_I) à +100% (M_F) avec M_{min1} à -50%, M_{max} à 0% et M_{min2} à 50%. La normalisation et le calcul des positions sont décrits sur la figure 5.6. Soit $P_{X/j}$ la position de l'événement P_X de la trajectoire du doigt sur les cycles de mâchoire normalisés, $P_{X/j}$ est :

- inférieure à -100% si P_X advient avant M_I ;
- entre -100% et -50% si P_X advient entre M_I et M_{min1} ;
- entre -50% et 0% si P_X advient entre M_{min1} et M_{max} ;
- entre 0% et 50% si P_X advient entre M_{max} et M_{min2}
- entre 50% et 100% si P_X advient entre M_{min2} et M_F ;
- supérieure à 100% si P_X advient après M_F .

Dans la suite, $P_{I/j}$ et $P_{A/j}$ réfèrent respectivement aux positions de l'initialisation et de l'apex du geste de pointé et $P_{R/j}$ et $P_{F/j}$, à celles du début et de la fin du geste de retour du doigt. La figure 5.7 représente la distribution de ces valeurs selon la positions de l'accent, pour les moyennes des 20 participants (barres) et pour les

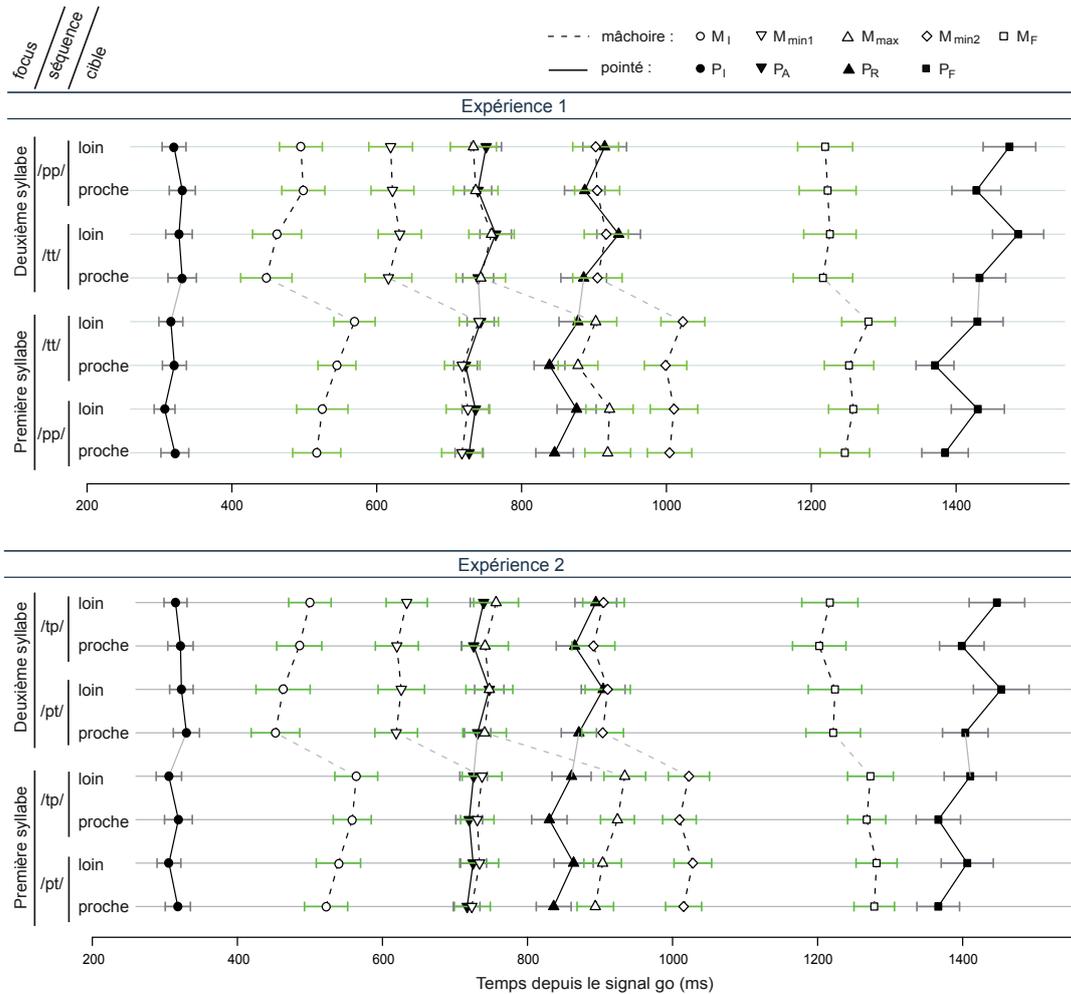


FIG. 5.5 – Moyenne et erreur standard moyenne du temps écoulé entre le signal-go et les principaux événements de la trajectoire de la mâchoire et du doigt selon la position de l'accent, la séquence et la position de la cible pour chaque expérience. Pour le doigt, P_I , P_A sont l'initialisation et l'apex du geste de pointé ; P_R et P_F sont l'initialisation et la fin du geste de retour. Pour la mâchoire, M_I et M_F sont l'initialisation du premier cycle et la fin du deuxième cycle, M_{min1} et M_{min2} sont les minima de trajectoire pour le premier et le deuxième cycle. M_{max} est le maximum de trajectoire entre les deux cycles. La comparaison entre les deux expériences rend compte de résultats très similaires. La comparaison des quatre lignes du bas aux quatre lignes du haut pour chaque expérience montre que la position de l'accent a un effet clair sur les relations entre les événements des deux systèmes et sur le timing de la mâchoire.

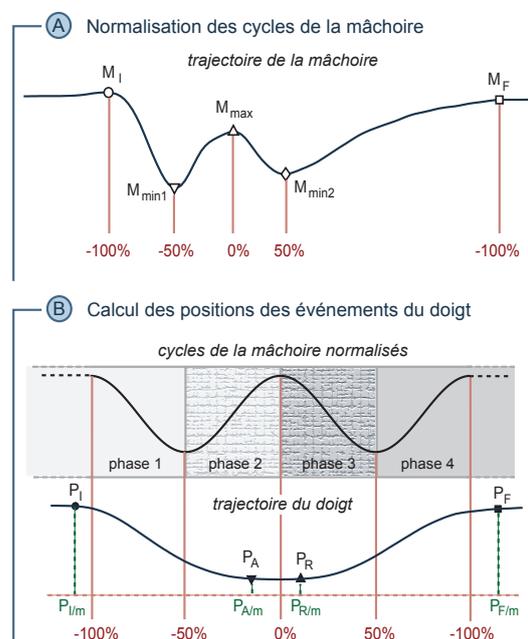


FIG. 5.6 – **En A.** normalisation des cycles de la mâchoire entre -100% (M_I , initialisation du premier cycle) et 100% (M_F , fin du deuxième cycle). La position la plus basse de la mâchoire est à -50% pour le premier cycle (M_{min1}) et à 50% pour le deuxième (M_{min2}). 0% correspond à M_{max} , la position de la mâchoire la plus haute entre les deux cycles. **En B.** calcul des positions des quatre points principaux de la trajectoire du doigt (P_I , P_A , P_R et P_F) sur les cycles de mâchoire normalisés. Soit P_X un de ces événements, le calcul de sa position $P_{X/j}$ consiste à déterminer dans quelle phase de la mâchoire est inclus P_X . On distingue quatre phases : de M_I à M_{min1} , de M_{min1} à M_{max} , de M_{max} à M_{min2} et de M_{min2} à M_F . Si P_X advient avant M_I ou après M_F , sa position est calculée relativement à la phase 1 dans le premier cas et à la phase 4 dans le deuxième. Soit M_d et M_f les occurrences du début et de la fin de la phase incluant P_X , et $M_{d,n}$ la valeur normalisée de M_d (par exemple, -100% pour la phase 1 ou -50% pour la phase 2) alors :

$$P_{X/j} = M_{d,n} + 50(P_X - M_d)/(M_f - M_d)$$

essais de tous les participants confondus (points). P_I et P_F advenant essentiellement en dehors des cycles de mâchoire, les valeurs de $P_{I/j}$ et $P_{F/j}$ sont mentionnées ci-après uniquement pour vérifier dans quelle mesure P_I advient avant M_I et P_F après M_F .

Les différents rapprochements ont aussi été caractérisés en termes de durées par la mesure des délais entre les événements qui tendent à se rapprocher. La répartition des valeurs de ces délais pour les 20 moyennes des participants et pour les essais de tous les participants confondus est représentée sur la figure 5.8. Les résultats détaillés des ANOVAs sur ces différents indicateurs de la coordination mâchoire-doigt sont disponibles en annexe dans le tableau A.4, p.184.

Position des événements du doigt relativement aux cycles de la mâchoire

Initialisations et fins des réponses. La figure 5.7 montre que l'initialisation du geste de pointé (P_I) débute avant celle du premier cycle de la mâchoire (M_I) pour toutes les moyennes des participants et pour plus de 80% des essais en condition *accent*₁ et pour 17 participants et environ 75% des essais en *accent*₂. En moyenne, le délai entre les deux réponses, $M_I - P_I$ est de 186 ms. Ce délai est 74 ms plus long en condition *accent*₁ qu'en *accent*₂ [$F(1, 19) = 14.1, p \leq 0.01$]; 20 ms plus long en condition *cible*_{loin} qu'en *cible*_{proche} [$F(1, 19) = 11.1, p \leq 0.01$] et 39 ms plus long en condition *seq*_{pp} qu'en *seq*_{tt} [$F(1, 19) = 13.1, p \leq 0.01$]. Cependant, l'effet de la position de l'accent sur $M_I - P_I$ n'interagit pas significativement avec les deux autres facteurs. D'autre part, d'après la figure 5.7 la fin du geste de retour du doigt (P_F) advient majoritairement après la fin du geste montant du deuxième

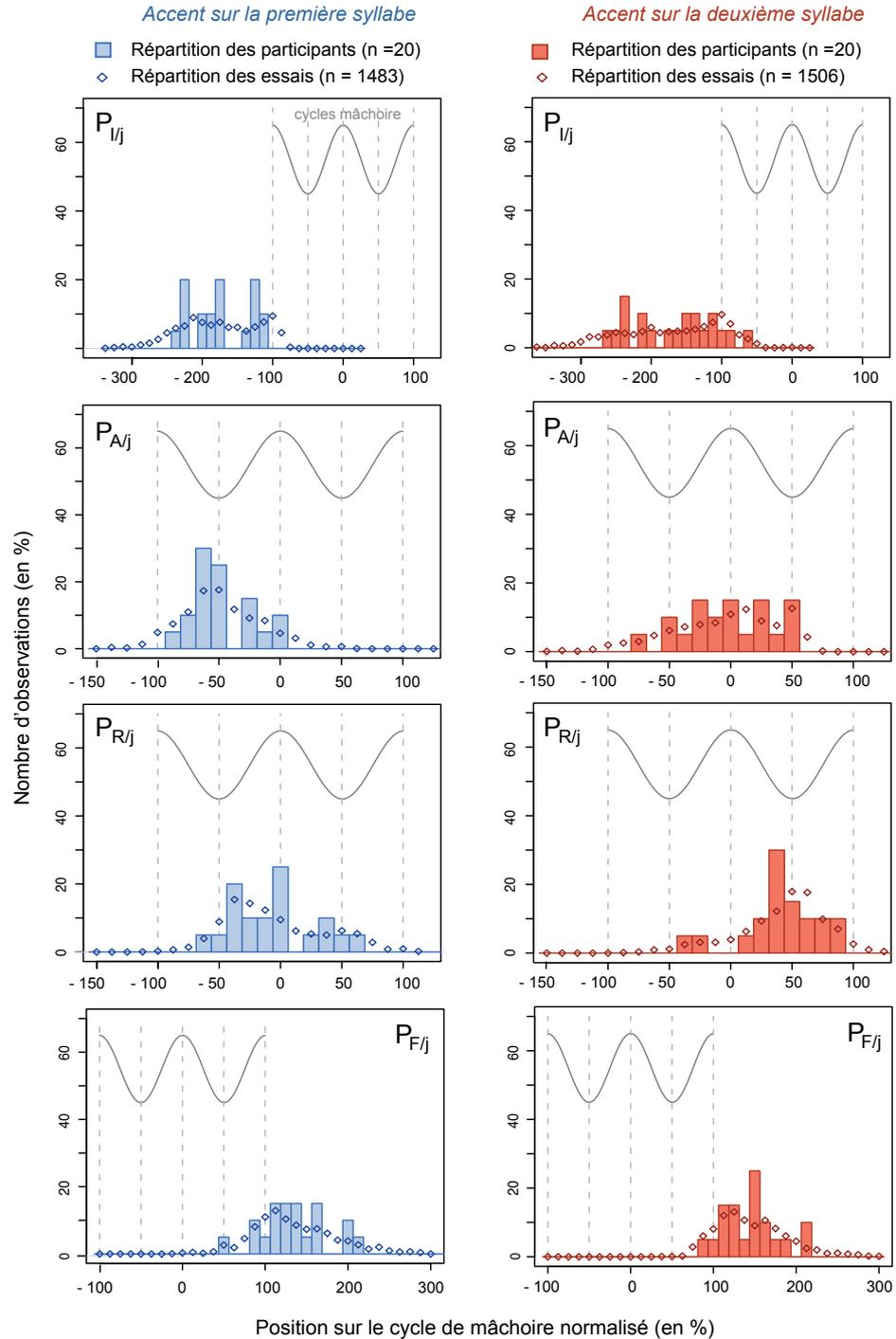
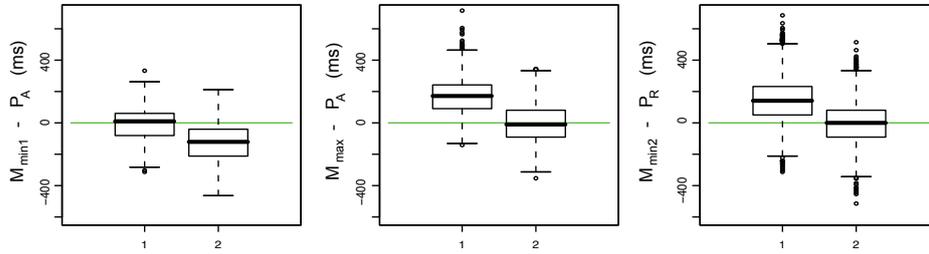


FIG. 5.7 – Répartitions de la position des principaux événements de la trajectoire du doigt sur les cycles de mâchoire normalisés, selon la position de l'accent, pour tous les essais et pour les moyennes des 20 participants. $P_{I/j}$ et $P_{A/j}$ sont les positions de l'initialisation et de l'apex du geste de pointé ; $P_{R/j}$ et $P_{F/j}$, sont les positions de l'initialisation et de la fin du geste de retour. Globalement, les différentes distributions suivent la même forme pour les moyennes des participants (barres colorées) et pour les essais (points superposés). Dans les deux conditions d'accentuation, les mouvements de la mâchoire sont compris entre l'initialisation du geste de pointé et la fin du geste de retour du doigt. La position de l'accent ($accent_1$ à gauche et $accent_2$ à droite) affecte clairement les distributions de $P_{A/j}$ et $P_{R/j}$.

— Répartitions des 1483 essais pour la condition accent 1 et 1506 essais pour la condition accent 2 —



Répartitions des moyennes des 20 participants

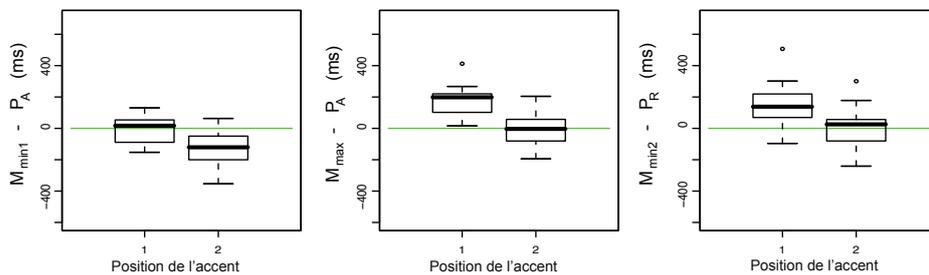


FIG. 5.8 — Répartition des délais entre les principaux points de rencontre entre les trajectoires du doigt et de la mâchoire selon la position de l'accent pour tous les essais (en haut) et pour les moyennes des participants (en bas). Pour le doigt, P_A et P_R sont le début et la fin de la tenue du geste de pointé (apex et début du geste de retour). Pour la mâchoire, M_{min1} et M_{min2} sont les positions basses pour la première et la deuxième voyelle et M_{max} est la position haute entre les deux cycles. Les limites inférieures et supérieures des rectangles correspondent au premier et troisième quartiles. La médiane est représentée par le trait horizontal plus épais à l'intérieur des rectangles. L'effet de la position de l'accent apparaît clairement pour les différents délais mais les synchronisations sont moins évidentes que ne le laissaient supposer les moyennes.

cycle de la mâchoire (M_F). Ainsi, la trajectoire du doigt, de l'initialisation du geste de pointé (P_I) au retour à la position de repos (P_F) englobe les deux cycles de la mâchoire et ce, dans les deux conditions d'accentuation. A l'intérieur de cet *enveloppement* de la mâchoire par le doigt, les choses se déroulent très différemment selon la position de l'accent.

Quand l'accent est sur la première syllabe. D'après la figure 5.5 deux points importants apparaissent en condition *accent*₁. D'abord, l'apex du geste de pointé P_A , qui correspond aussi au début de la tenue du geste, advient à proximité de M_{min1} , la position la plus basse de la mâchoire pour le premier cycle. Ensuite, la fin de la tenue du geste de pointé P_R advient avant M_{max} , le début du deuxième cycle. La deuxième syllabe tendrait donc à être réalisée sur le geste de retour du doigt.

P_A est-il proche de M_{min1} ? D'après la figure 5.7, en *accent*₁, les valeurs de $P_{A/j}$ sont toujours comprises entre -100% et 6.25%. P_A est donc inclus dans le premier cycle de la mâchoire. De plus, P_A advient sur le geste descendant du premier cycle, entre -100% et -43.75% pour 60% des essais et 14 participants. Enfin, 50% des essais et 11 moyennes se situent autour de -55.25% ($\pm 12.5\%$), soit autour de M_{min1} , la position basse du cycle. L'étude de la ré-

partition du délai de P_A à M_{min1} , $M_{min1} - P_A$, en *accent*₁ montre que ce délai est compris entre -80 ms et 60 ms pour 50% des essais avec une médiane à 10 ms. La répartition des moyennes des participants suit le même schéma. De plus, $M_{min1} - P_A$ dépend clairement de la position de l'accent avec une valeur moyenne de -7 ms en *accent*₁ et de -127 ms en *accent*₂ [$F(1, 19) = 53.2$, $p \leq 0.0001$]. En revanche, l'effet de la position de la cible et de la séquence ne sont pas significatifs. En *accent*₁, P_A advient donc quasiment toujours sur le premier cycle de mâchoire et tend à se rapprocher de M_{min1} .

P_R advient-il avant M_{max} ? En *accent*₁, P_R , le début du geste de retour du doigt, advient majoritairement pendant la phase montante du premier cycle de la mâchoire avec des valeurs de $P_{R/j}$ entre -50% et 0% (± 6.25) pour 14 participants et environ 70% des essais. Le délai entre P_R et M_{max} , $M_{max} - P_R$, dépend clairement de la position de l'accent avec une inversion de l'ordre d'occurrence des deux événements d'une condition à l'autre. En moyenne, P_R advient 46 ms avant M_{max} en condition *accent*₁ et 162 ms après M_{max} en condition *accent*₂ [$F(1, 19) = 127$, $p \leq 0.0001$]. Cette tendance reste la même pour les deux positions de la cible et les deux séquences (voir tableau A.4, p.184 en annexe pour le détail).

D'après cette analyse, en condition *accent*₁, la tenue du geste de pointé est majoritairement incluse dans le premier cycle de la mâchoire avec un apex du geste de pointé proche de la position basse du cycle et un début de geste de retour durant le geste de fermeture. Dans cette condition, la deuxième syllabe (non-accentuée) est donc réalisée en dehors de la tenue du geste pointé, sur le geste de retour, ce qui va dans le sens de sa *non-pertinence* pour la tâche de désignation.

Quand l'accent est sur la deuxième syllabe. D'après la figure 5.5, les deux points importants en condition *accent*₂ sont un rapprochement entre l'apex du geste de pointé (P_A) et le début du deuxième cycle de la mâchoire (M_{max}) d'une part et, d'autre part, entre le début du geste de retour P_R et la position la plus basse du deuxième cycle de la mâchoire (M_{min2}). En d'autres termes, la phase de tenue du geste de pointé semble incluse dans le geste d'ouverture du deuxième cycle.

P_A est-il proche de M_{max} ? En *accent*₂, P_A advient entre -50% et +50% des cycles de mâchoire normalisés, soit entre les deux minima de trajectoire de la mâchoire, pour 19 participants et environ 90% des essais. La distribution ne présente pas de pic autour de M_{max} comme le laissait attendre l'observation de la figure 5.5. Cependant, l'étude du délai entre les deux événements, $M_{max} - P_A$, montre qu'en condition *accent*₂, les valeurs et les moyennes de ce délai se distribuent autour de 0 ms avec une médiane à -10 ms et 50% des valeurs entre -90 ms et 80 ms. Comme pour les autres délais, la position de l'accent a un effet clair sur $M_{max} - P_A$, avec des valeurs moyennes de 173 ms en condition *accent*₁ et de -6 ms en *accent*₂ [$F(1, 19) = 98.6$, $p \leq 0.0001$]. L'interaction entre la position de l'accent et la séquence est significative mais ne change pas le sens de l'effet (+189 ms en *seq*_{ttj} et +168 ms en *seq*_{ppj})

$[F(1, 19) = 7.1, p \leq 0.05]$ (voir tableau A.4, p.184 en annexe pour le détail).

P_R est-il proche de M_{min2} ? En condition $accent_2$, P_R advient sur le deuxième cycle de mâchoire avec des valeurs de $P_{R/j}$ comprises entre 6.25% et 93.75% pour environ 90% des essais et 18 participants. De plus, la distribution des moyennes et des essais présente un pic autour de $50\% \pm 18.75\%$ correspondant à environ 50% des essais et 11 participants. Enfin, $P_{R/j}$ est inférieure à 31.25% pour moins de 20% des essais et seulement 5 participants. En $accent_2$, P_R advient donc majoritairement à la fin du geste d'ouverture (autour de M_{min2}) et sur le geste de fermeture du deuxième cycle de la mâchoire. D'autre part, la répartition de $M_{min2} - P_R$ en condition $accent_2$ (figure 5.8), présente, pour l'ensemble des essais, une médiane à 0 ms alors que 50% des essais se distribuent entre -90 ms et 80 ms. Pour les moyennes des participants, la médiane est à 26 ms et 50% des essais sont compris entre -80 ms et 55 ms. Le délai $M_{min2} - P_R$ dépend clairement de la position de l'accent avec une moyenne de 150 ms en $accent_1$ et de 2 ms en $accent_2$ [$F(1, 19) = 86.0, p \leq 0.0001$]. Ce délai dépend aussi de la position de la cible avec une moyenne 26 ms plus importante en $cible_{proche}$ qu'en $cible_{loin}$ [$F(1, 19) = 15.6, p \leq 0.001$]. Cependant, ni l'interaction entre les deux facteurs, ni l'effet de la séquence sont significatifs.

Ainsi, quand l'accent est sur la deuxième syllabe, la phase de tenue du geste de pointé se décale vers le deuxième cycle de la mâchoire et le geste de retour du doigt tend à advenir à proximité de la position basse du deuxième cycle, c'est-à-dire de la deuxième voyelle.

Si les résultats ne sont pas complètement en accord avec l'hypothèse d'une synchronisation entre l'apex du geste de pointé et la position basse de la mâchoire pour la voyelle accentuée, ils montrent des relations fortes entre les deux systèmes. En effet, malgré les variabilités, la partie accentuée de la parole (M_{min1} en $accent_1$ et M_{min2} en $accent_2$) est en général réalisée pendant la phase de tenue du geste de pointé (de P_A à P_R). La position de l'accent semble donc bien jouer un rôle déterminant dans la coordination des deux systèmes. Ce profil général amène une dernière question en ce qui concerne l'effet de la position de l'accent sur la coordination mâchoire-main, celle des profils des participants.

Profils des participants.

La question est ici de savoir si la tendance à réaliser la partie accentuée de la parole pendant la tenue du geste de pointé s'observe pour tous les participants ou bien si elle est sujette à des variabilités inter-individuelles fortes. La figure 5.9 représente les moyennes de la position du début ($P_{A/j}$) et de la fin ($P_{R/j}$) de la tenue du geste de pointé pour chaque participant et chaque condition d'accentuation. Le premier point important est que $P_{A/j}$ et $P_{R/j}$ se décalent vers la droite de la condition $accent_1$ à la condition $accent_2$ pour tous les participants. L'effet de la

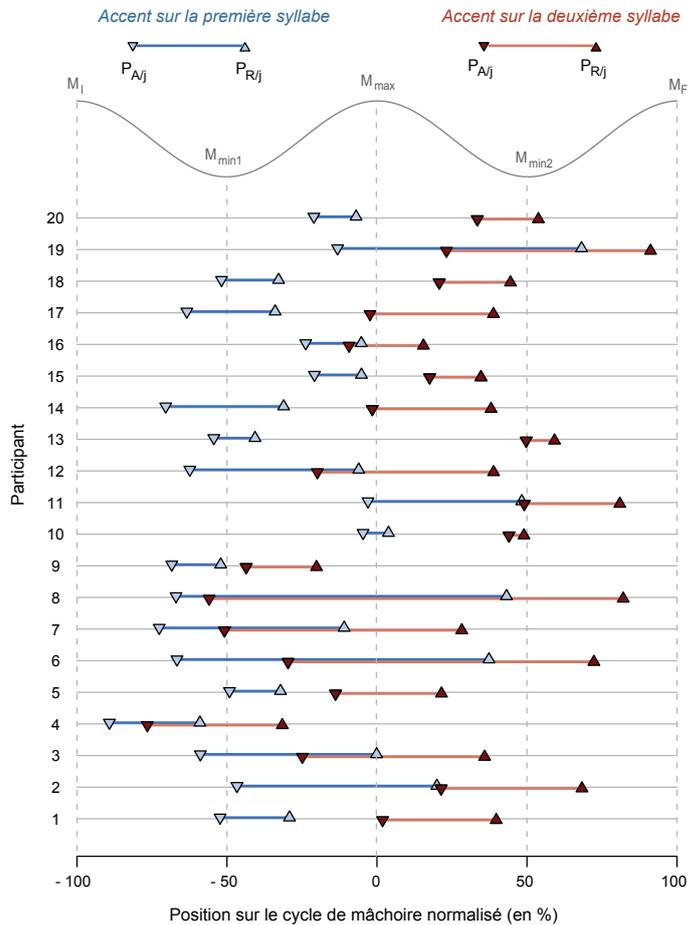


FIG. 5.9 – Profils des participants. Etude de l'effet de la position de l'accent sur la position de la tenue du geste de pointé (de l'apex du geste, P_A , au début du geste de retour, P_R) relativement aux cycles de mâchoire normalisés. Une ligne correspond aux résultats d'un(e) participant(e). Les moyennes de $P_{A/j}$ et $P_{R/j}$ en condition *accent₁* sont représentées en bleu clair et celles en conditions *accent₂*, en rouge foncé. Pour tous les participants, P_R et P_A sont décalés vers le deuxième cycle de mâchoire en *accent₂* par rapport à *accent₁*. De plus, la tenue du geste est principalement comprise entre les deux minima de la trajectoire de la mâchoire. Enfin, malgré la variabilité interindividuelle importante, des profils se dessinent (voir le texte pour le détail).

position de l'accent sur la coordination mâchoire-doigt est donc un effet fort, qui va dans le même sens pour tous les participants. Cependant, à l'intérieur de cet effet global, les comportements sont très variables. Ils peuvent néanmoins être classifiés en deux grandes catégories : ceux qui asservissent la fin de la tenue du geste de pointé à la position de l'accent et ceux qui basent la fin de la tenue du geste de pointé sur d'autres événements, comme la fin de la parole ou la première voyelle, quelle que soit la position de l'accent.

En résumé, à la question : qu'est-ce qu'il est important de dire pendant que le doigt montre la cible ? La réponse est ici claire, c'est avant tout la syllabe accentuée de la parole, et différents modes de coordination permettent d'y parvenir.

Ces différentes observations amènent une dernière question, celle de l'effet de l'accentuation sur les mouvements de la mâchoire et du doigt afin de déterminer si l'effet observé sur la coordination mâchoire-main est le produit d'une adaptation de la mâchoire et/ou du doigt. D'après la figure 5.5, p. 102, présentant l'organisation temporelle globale des mouvements dans les deux expériences, les cycles de la mâ-

choire sont avancés dans la condition *accent*₂ par rapport à *accent*₁. En revanche, l'organisation temporelle du geste de pointé du doigt ne semble pas affectée par la position de l'accent alors que le geste de retour paraît reporté en condition *accent*₂ par rapport à *accent*₁. Ces observations sont développées ci-après conjointement à l'analyse des paramètres des mouvements. Comme pour l'étude de la coordination, seules les analyses de la première expérience sont détaillées, les conclusions étant, ici encore, transférables à la deuxième expérience.

5.3.2 Analyse des mouvements du doigt

Timing

En moyenne, le début du geste de pointé, P_I , advient 322 ms après le signal-go et cette valeur ne dépend pas de la séquence. Une différence de +12 ms s'observe de la condition *accent*₁ à la condition *accent*₂ et de +8 ms de la condition *cible*_{loin} à la condition *cible*_{proche} mais les effets de la position de l'accent et de la cible ne sont pas significatifs. De la même manière, l'apex du geste de pointé, P_A , advient en moyenne 741 ms après le signal-go et ne dépend pas de la séquence. Une différence de +16 ms apparaît de la condition *accent*₁ à *accent*₂ et de *cible*_{loin} à *cible*_{proche}. Cependant, seul l'effet de la position de la cible est significatif [$F(1, 19) = 9.5$, $p \leq 0.01$]. Comme pour P_I et P_A , la séquence n'a pas d'effet significatif sur le début (P_R) et la fin (P_F) du geste de retour du doigt qui adviennent en moyenne 883 ms et 1429 ms après le signal-go. En revanche, ces deux mesures sont significativement affectées par la position de la cible avec, de *cible*_{proche} à *cible*_{loin}, une augmentation de 37 ms pour P_R [$F(1, 19) = 27.6$, $p \leq 0.0001$] et de 51 ms pour P_F [$F(1, 19) = 27.6$, $p \leq 0.001$]. Contrairement à P_I et P_A , les occurrences de P_R et P_F sont aussi significativement affectées par la position de l'accent avec, de *accent*₁ à *accent*₂, un report de P_R de 46 ms [$F(1, 19) = 16.0$, $p \leq 0.001$] et un report de P_F de 51 ms [$F(1, 19) = 22.1$, $p \leq 0.001$].

Paramètres des mouvements

Les valeurs des différents paramètres de la trajectoire du doigt sont représentées sur la figure 5.10. Le détail des ANOVAs sur ces différents paramètres est disponible en annexe dans le tableau A.6, p. 186. Globalement, les différents paramètres du geste de pointé et du geste de retour du doigt semblent dépendre uniquement de la position de la cible alors que la tenue du geste de pointé semble aussi dépendre de la position de l'accent.

Pour le geste de pointé, la durée (D_P), l'amplitude (A_P) et l'amplitude du pic de vitesse (V_P) sont en moyenne de 419 ms, 349 mm et 1394 mm/s. L'éloignement de la cible induit une augmentation significative de la durée [+25 ms, $F(1, 19) = 26.1$, $p \leq 0.0001$], de l'amplitude [+149 mm, $F(1, 19) = 2274$, $p \leq 0.0001$] et de l'amplitude du pic de vitesse [+538 mm/s, $F(1, 19) = 647$, $p \leq 0.0001$]. La position de l'accent n'a pas d'effet significatif sur la durée et l'amplitude du pic de vitesse du geste, mais tend à avoir un léger effet sur son amplitude avec une augmentation de

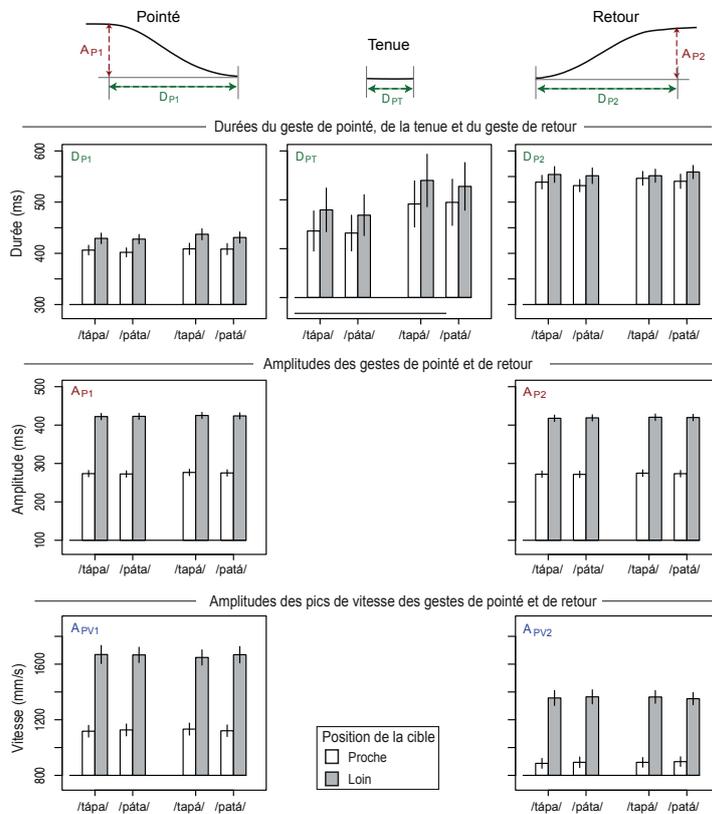


FIG. 5.10 – Analyse des mouvements du doigt. Moyenne et erreur standard moyenne des paramètres du geste de pointé, de la phase de tenue et du geste de retour du doigt selon la position de l'accent, la séquence et la position de la cible. D_P , A_P et V_P sont la durée, l'amplitude et l'amplitude du pic de vitesse du geste de pointé. D_{PT} est la durée de la tenue du geste de pointé. D_{PR} , A_{PR} et V_{PR} sont la durée, l'amplitude et l'amplitude du pic de vitesse du geste de retour. L'effet de la position de la cible apparaît clairement pour les différents paramètres alors que la position de l'accent semble affecter uniquement la durée de la tenue du geste. La variation de la séquence semble n'avoir aucun effet sur le mouvement du doigt.

2.4 mm en $accent_2$ par rapport à $accent_1$ [$F(1, 19) = 8.7, p \leq 0.01$]. De la même manière, l'effet de la séquence se limite à une augmentation de la durée de 3 ms en $seq_{/tt/}$ par rapport à $seq_{/pp/}$ [$F(1, 19) = 6.6, p \leq 0.05$].

Pour le geste de retour, la durée (D_{PR}), l'amplitude (A_{PR}) et l'amplitude du pic de vitesse (V_{PR}) sont en moyenne de 547 ms, 346 mm et 1126 mm/s, respectivement. Ce geste est donc plus lent que le geste de pointé. De plus, les différents facteurs n'ont pas d'effet significatif sur la durée du geste alors que, comme pour le geste de pointé, l'amplitude et l'amplitude du pic de vitesse augmentent significativement avec l'éloignement de la cible de +146 mm [$F(1, 19) = 1822, p \leq 0.0001$] et de +466 mm/s [$F(1, 19) = 734, p \leq 0.0001$], respectivement. La séquence n'a pas d'effet sur ces mesures. Comme pour le geste de pointé, l'amplitude du geste de retour augmente faiblement (+2 mm) mais significativement en $accent_2$ par rapport à $accent_1$ [$F(1, 19) = 4.5, p \leq 0.05$].

Enfin, la durée de la tenue du geste (D_{PT}) est en moyenne de 142 ms et est significativement plus longue en $cible_{loin}$ qu'en $cible_{proche}$ [+20 ms, $F(1, 19) = 13.6, p \leq 0.01$] et en $accent_2$ qu'en $accent_1$ [+30 ms, $F(1, 19) = 14.1, p \leq 0.01$]. Ni les effets d'interaction, ni l'effet de la séquence sont significatifs.

Logiquement, le geste de pointé est donc plus ample, plus long mais aussi plus rapide quand la cible est loin que quand elle est proche. L'effet de la position de la cible sur l'amplitude et la vitesse se retrouve sur le geste de retour. Cependant, celui-ci est beaucoup plus long que le geste de pointé, ce qui paraît cohérent étant donné qu'il est beaucoup moins contraint que le geste de pointé. D'autre part, de l'initialisation à l'apex, le geste de pointé n'est pas affecté de manière notable par la variation de la position de l'accent et de la séquence. Ces observations montrent que le geste de pointé vise avant tout à aligner le doigt avec la cible. Une fois cet objectif atteint, la tenue du geste est une phase plus perméable aux effets des autres facteurs et s'adapte notamment à la position de l'accent, avec une durée plus longue en $accent_2$ qu'en $accent_1$, due à un report du geste de retour. Ce report contribue à ce que la partie accentuée de la parole advienne pendant que le doigt pointe la cible.

5.3.3 Analyse des cycles de la mâchoire

Timing

Comme le montre la figure 5.5, p. 102, les débuts et fins de gestes de la mâchoire sont avancés en condition $accent_2$ par rapport à $accent_1$. La durée écoulée depuis le signal-go est en moyenne de 508 ms pour M_I , 674 ms pour M_{min1} , 824 ms pour M_{max} , 958 ms pour M_{min2} et 1253 ms pour M_F . Le détail des ANOVAs sur ces cinq facteurs sont disponibles en annexe dans le tableau A.2, p. 182.

L'avance des événements dans la condition $accent_2$ par rapport à $accent_1$ est de -63 ms pour M_I [$F(1, 19) = 15.9, p \leq 0.001$], -103 ms pour M_{min1} [$F(1, 19) = 57.0, p \leq 0.0001$], -162 ms pour M_{max} [$F(1, 19) = 105.6, p \leq 0.0001$], -102 ms pour M_{min2} [$F(1, 19) = 49.0, p \leq 0.0001$] et -40 ms pour M_F [$F(1, 19) = 5.7,$

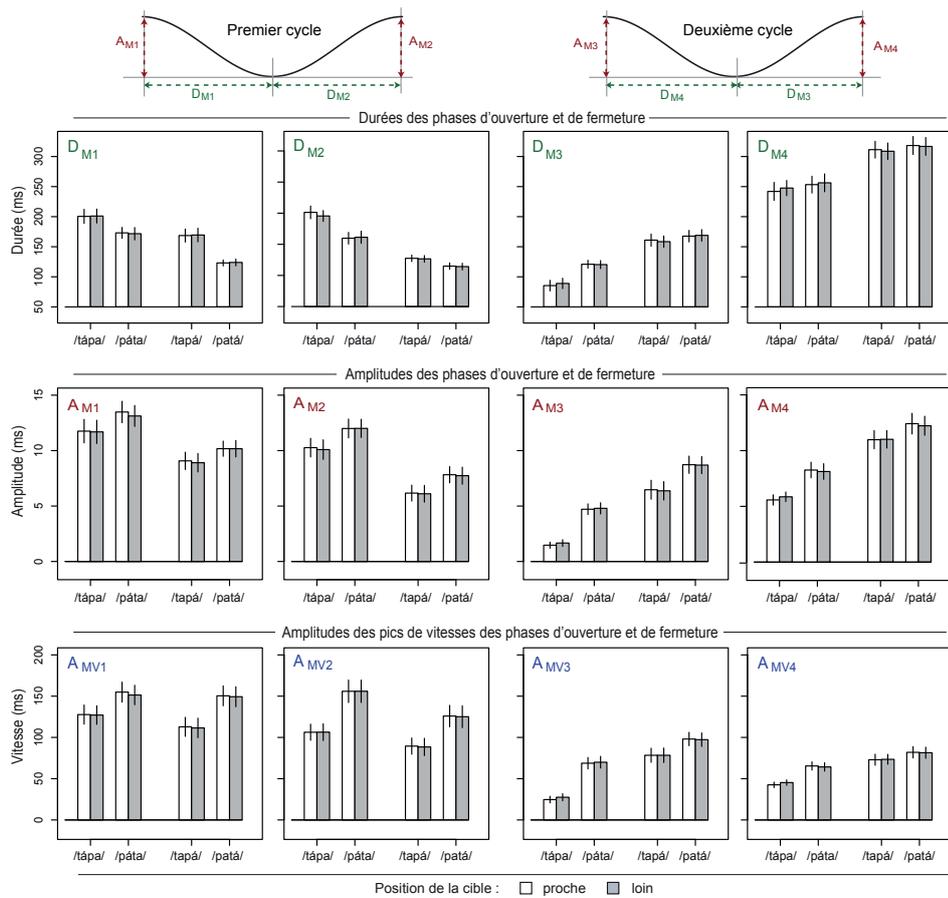


FIG. 5.11 – Analyse des mouvements de la mâchoire. Moyenne et erreur standard moyenne des paramètres des cycles de la mâchoire selon la position de l'accent, la séquence, la position de la cible. D_{Mn} , A_{Mn} et V_{Mn} sont la durée, l'amplitude et l'amplitude du pic de vitesse des quatre phases avec $n = 1$ et $n = 2$ pour la phase montante et descendante du premier cycle et $n = 3$ et $n = 4$ pour la phase montante et descendante du deuxième cycle. La valeur de ces différents paramètres dépend essentiellement de la position de l'accent et de la séquence.

$p \leq 0.05$]. L'occurrence de ces événements est aussi systématiquement reportée d'environ 10 ms en *cible_loin* par rapport à *cible_proche* avec une différence de 11 ms pour M_I [$F(1, 19) = 4.1$, non significative], 11 ms pour M_{min1} [$F(1, 19) = 6.8$, $p \leq 0.05$], 10 ms pour M_{max} [$F(1, 19) = 5.0$, $p \leq 0.05$], 10 ms pour M_{min2} [$F(1, 19) = 6.0$, $p \leq 0.05$] et 11 ms pour M_F [$F(1, 19) = 4.5$, $p \leq 0.05$]. Pour les cinq mesures, l'effet de la position de la cible n'interagit pas significativement avec celui de la position de l'accent. La séquence a un effet significatif sur l'occurrence de M_I , avec une avance de 39 ms en *seq_ut/* par rapport à *seq_pp/* [$F(1, 19) = 13.5$, $p \leq 0.01$] et un effet inverse pour M_{max} avec un report de 23 ms en *seq_ut/* par rapport à *seq_pp/* [$F(1, 19) = 9.2$, $p \leq 0.01$]. Cet effet n'interagit pas significativement avec celui de la position de l'accent ou de la cible. En résumé, en dépit de l'effet de la séquence sur les initialisations des gestes et de la position de la cible sur l'ensemble des événements, les positions basses des deux cycles de la mâchoire, M_{min1} et M_{min2} , sont avancées d'environ 100 ms en condition *accent2* par rapport à *accent1*. De son côté, l'éloignement de la cible induit une translation dans le temps d'environ 10 ms de tous les événements de la mâchoire.

Paramètres des mouvements.

La valeur des différents paramètres des mouvements de la mâchoire sont représentées sur la figure 5.11. Le détail des ANOVAs réalisées sur ces différents paramètres est disponible en annexe dans le tableau A.8, p. 188. Contrairement aux gestes du doigt, l'effet de la position de la cible sur les paramètres étudiés n'est jamais significatif. En revanche, ces paramètres dépendent clairement de la position de l'accent et de la séquence.

Pour le premier cycle, la durée, l'amplitude et l'amplitude du pic de vitesse sont respectivement de 166 ms (D_{M1}), 11 mm (A_{M1}) et 136 mm/s (V_{M1}) pour la phase descendante et de 150 ms (D_{M2}), 9 mm (A_{M2}) et 119 mm/s (V_{M2}) pour la phase montante. Pour la phase descendante, la durée, l'amplitude et l'amplitude du pic de vitesse sont plus importantes en condition *accent1* qu'en condition *accent2*. Cet effet de la position de l'accent est significatif pour la durée [D_{M1} , +40 ms, $F(1, 19) = 26.2$, $p \leq 0.0001$] et l'amplitude [A_{M1} , +2.9 mm, $F(1, 19) = 21.0$, $p \leq 0.001$] mais pas pour l'amplitude du pic de vitesse. Pour la phase montante, la durée, l'amplitude et l'amplitude du pic de vitesse sont significativement plus importantes en *accent1* qu'en *accent2* [D_{M2} , +59 ms, $F(1, 19) = 101.1$, $p \leq 0.0001$; A_{M2} , 4.1 mm, $F(1, 19) = 49.0$, $p \leq 0.0001$; V_{M2} , +24 mm/s, $F(1, 19) = 8.7$, $p \leq 0.01$]. D'autre part, les durées des deux phases sont plus importantes en condition *seq_ut/* que *seq_pp/* [D_{M1} +37 ms, $F(1, 19) = 14.3$, $p \leq 0.01$; D_{M2} , +25 ms, $F(1, 19) = 52.2$, $p \leq 0.0001$] alors que l'inverse s'observe pour les amplitudes [A_{M1} , -1.4 mm, $F(1, 19) = 9.8$, $p \leq 0.01$; A_{M2} , -1.7 mm, $F(1, 19) = 12.6$, $p \leq 0.01$] et les amplitudes des pics de vitesse [V_{M1} , -32 mm/s, $F(1, 19) = 21.3$, $p \leq 0.001$; V_{M2} , -43 mm/s, $F(1, 19) = 31.6$, $p \leq 0.0001$]. L'effet de la position de l'accent interagit significativement avec l'effet de la séquence pour la durée et l'amplitude des pics de vitesse des deux gestes. Ainsi l'augmentation

de D_{M1} en $accent_2$ est plus importante pour /p/ (+49 ms) que pour /t/ (+32 ms) [$F(1, 19) = 5.4, p \leq 0.05$], ce qui est cohérent avec le fait que l'augmentation du pic de vitesse V_{M1} est elle, plus importante pour /t/ (+15 mm/s) que pour /p/ (+3 mm/s) [$F(1, 19) = 12.5, p \leq 0.01$]. L'inverse s'observe pour la phase montante, avec une augmentation de la durée (D_{M2}) plus importante pour /t/ (+71 ms) que pour /p/ (+46 ms) [$F(1, 19) = 13.7, p \leq 0.05$] et une augmentation du pic de vitesse (V_{M2}) plus importante pour /p/ (+30 mm/s) que pour /t/ (+17 mm/s) [$F(1, 19) = 4.7, p \leq 0.05$].

Logiquement, l'effet de la position de l'accent s'inverse pour le deuxième cycle de la mâchoire. Globalement, pour ce cycle, la durée, l'amplitude et l'amplitude du pic de vitesse sont respectivement de 134 ms (D_{M3}), 5.4 mm (A_{M3}) et 68 mm/s (V_{M3}) pour la phase descendante et de 282 ms (D_{M4}), 9.3 mm (A_{M4}) et 66 mm/s (V_{M4}) pour la phase montante. Pour les deux phases, la durée, l'amplitude et l'amplitude du pic de vitesse sont toujours significativement plus importantes en condition $accent_2$ qu'en $accent_1$ et en $seq/pp/$ qu'en $seq/tt/$. L'effet de la position de l'accent se traduit par une augmentation en $accent_2$ de 60 ms pour D_{M3} [$F(1, 19) = 42.3, p \leq 0.0001$] et 64 ms pour D_{M4} [$F(1, 19) = 61.5, p \leq 0.0001$]; 4.4 mm pour A_{M3} [$F(1, 19) = 40.2, p \leq 0.0001$] et 4.7 mm pour A_{M4} [$F(1, 19) = 22.4, p \leq 0.001$]; 40 mm/s pour V_{M3} [$F(1, 19) = 57.4, p \leq 0.0001$] et 23 mm/s pour V_{M4} [$F(1, 19) = 25.5, p \leq 0.0001$]. De son côté, la variation de la séquence induit une augmentation en $seq/pp/$ de 21 ms pour D_{M3} [$F(1, 19) = 29.1, p \leq 0.0001$] et 9 ms pour D_{M4} (non significatif); 2.7 mm pour A_{M3} [$F(1, 19) = 45.2, p \leq 0.0001$] et 1.9 mm pour A_{M4} [$F(1, 19) = 22.4, p \leq 0.001$]; 31 mm/s pour V_{M3} [$F(1, 19) = 32.5, p \leq 0.0001$] et 15 mm/s pour V_{M4} [$F(1, 19) = 29.1, p \leq 0.0001$]. Contrairement au premier cycle, l'effet de la position de l'accent est systématiquement plus important en $seq/tt/$ qu'en $seq/pp/$, que ce soit pour la durée, l'amplitude et l'amplitude du pic de vitesse des phases descendante et montante (voir en annexe le tableau A.8, p. 188 pour le détail).

De cette analyse détaillée des mouvements de la mâchoire, il ressort que ceux-ci dépendent avant tout des propriétés segmentales ($seq/tt/$ vs. $seq/pp/$) et supra-segmentales ($accent_1$ vs. $accent_2$) de la cible phonétique. En bref, si les mouvements de la mâchoire sont un ingrédient clé de la réalisation des deux syllabes, il y en a généralement *plus* quand la syllabe est accentuée que non accentuée et quand la consonne est /p/ que lorsque c'est /t/. L'augmentation de l'implication de la mâchoire globalement plus importante pour /tt/ que pour /pp/ quand la syllabe est accentuée par rapport à quand elle ne l'est pas pourrait alors montrer que l'accentuation requiert un *minimum de mâchoire*. D'autre part, si la position de la cible n'affecte pas les paramètres des mouvements de la mâchoire, l'adaptation de la mâchoire à la cible phonétique se superpose à un report d'environ 10 ms de tous les événements de la mâchoire en *cible_loin* par rapport à *cible_proche*. Ce report ne dépend pas de la position de l'accent et vient compenser en partie celui de 16 ms observé pour P_A , expliquant l'absence d'effet de la position de la cible sur le

positionnement de P_A relativement aux cycles de la mâchoire. Plus généralement, cet effet de la position de la cible sur l'organisation temporelle de la mâchoire va dans le sens d'une adaptation de la mâchoire au geste de pointé, en accord avec les résultats de Levelt *et al.* (1985).

5.3.4 Petite synthèse des principaux résultats

Ainsi, la synchronisation attendue selon l'hypothèse de travail entre la position basse de la mâchoire pour la voyelle accentuée et l'apex du geste de pointé tend à s'observer seulement quand l'accent est sur la première syllabe et correspond plus à un rapprochement qu'à une synchronisation forte entre les deux événements. Quand l'accent est sur la deuxième syllabe, le rapprochement se fait entre la position basse de la mâchoire pour la voyelle accentuée et la fin de la tenue du geste de pointé. Plus précisément, la coordination entre les deux systèmes apparaît guidée par la nécessité de réaliser l'accent de la parole pendant la tenue du geste de pointé. Ainsi, quand l'accent est sur la première syllabe, la tenue du geste de pointé est globalement incluse dans le premier cycle de mâchoire, avec un début autour de la position basse du cycle et une fin après cette position. En revanche, quand l'accent est sur la deuxième syllabe, la tenue du geste de pointé se décale vers le deuxième cycle de la mâchoire avec un début qui tend à advenir autour de la position haute de la mâchoire entre les deux cycles et une fin autour de la position basse du deuxième cycle. Ces profils de coordination s'observent quelles que soient la séquence et la position de la cible. D'autre part, l'étude des variabilités interindividuelles montre que le décalage de la tenue du geste de pointé du premier vers le deuxième cycle de mâchoire de la condition *accent*₁ à la condition *accent*₂ est partagé par tous les participants et que pour une grande majorité des participants, la coordination mâchoire-doigt réalise la partie accentuée de la parole dans la phase de tenue du geste de pointé et ce, en dépit de fortes variabilités dans la façon d'y parvenir.

Enfin, les analyses des deux systèmes montre que l'effet de la position de l'accent sur la coordination mâchoire-doigt serait principalement un compromis entre quatre types d'adaptations :

1. Une adaptation de la mâchoire au but phonétique, avec des durées, des amplitudes et des pics de vitesse plus importants quand le geste porte une syllabe accentuée que non accentuée et des valeurs spécifiques à la séquence (/pp/ vs. /tt/);
2. Une adaptation de l'initialisation de la réponse de la mâchoire à la tâche de coordination avec une réponse qui débute toujours après le début du geste de pointé et qui s'adapte à la position de l'accent de façon à rapprocher le cycle de mâchoire pour la voyelle accentuée de la tenue du geste de pointé ;
3. Une adaptation du geste de pointé à la cible spatiale avec une amplitude, une durée et un pic de vitesse plus importants quand la cible est loin que quand

elle est proche ;

4. Une adaptation de la durée de la tenue du geste de pointé, prolongée quand l'accent est sur la deuxième syllabe par un report du geste de retour jusqu'à la réalisation de la partie accentuée de la parole par la mâchoire.

5.4 DISCUSSION

La modification de la position de l'accent dans le mot désignant la cible affecte la coordination entre la mâchoire et le doigt d'une manière différente de celle prédite par l'hypothèse de travail. De plus, cette coordination implique au moins trois éléments : ① une adaptation *intrinsèque* de chacun des deux systèmes pour atteindre leurs cibles respectives ; ② une adaptation de la mâchoire au geste de pointé et ③ une adaptation de la tenue du geste de pointé à la parole. Après avoir détaillé chacun de ces points, nous reprendrons l'analyse des profils de la mâchoire dans les deux expériences afin de mesurer, à posteriori, l'apport potentiel des données de ces deux études à l'hypothèse d'une composante motrice pour l'effet LC étudié au chapitre 3.

5.4.1 Pointer et dénommer...

D'abord, chaque système fait le nécessaire pour atteindre sa cible : le doigt pointe et la parole dénomme.

Adaptation du doigt à la tâche de pointage de la cible

L'objectif du geste de pointé est d'aligner le doigt avec la cible. Pour ce faire, le geste s'adapte à l'éloignement de la cible : il parcourt une distance plus grande quand la cible est loin que quand elle est proche. Cette augmentation de l'amplitude du geste s'accompagne d'une augmentation de sa durée mais aussi de sa vitesse. La modification de la vitesse peut s'interpréter comme une adaptation à la tâche de pointé mais aussi comme une adaptation à la double-tâche, elle contribuerait alors à limiter le report de la parole pour maintenir les différents rapprochements observés. Levelt *et al.* (1985) n'étudient pas la vitesse du geste de pointé mais l'augmentation de la durée et de l'amplitude qu'ils observent avec l'éloignement de la cible est proche de celle obtenue dans notre étude et n'est pas affectée par la présence de la parole. L'augmentation de la vitesse du geste de pointé pourrait donc être une adaptation à la position de la cible plutôt qu'à la parole. Une comparaison plus précise de la vitesse du geste en tâche de pointage seul et de pointage avec parole serait néanmoins requise pour évaluer plus précisément l'effet de la parole sur la vitesse du geste de pointé.

Adaptation de la mâchoire à la cible phonétique

Le système moteur de la parole doit adapter les gestes des articulateurs pour réaliser la cible phonétique de façon à respecter les propriétés segmentales (les bonnes consonnes dans le bon ordre) et suprasegmentales (l'accent sur la bonne syllabe) de cette cible.

Globalement, dans les deux expériences, l'accentuation de la syllabe est portée par une augmentation de l'amplitude, de la durée et du pic de vitesse du cycle de la mâchoire qui porte cette syllabe. Cependant, l'effet de l'accentuation sur la durée et le pic de vitesse des gestes paraît plus important pour le deuxième que pour le premier cycle. L'accentuation aurait donc moins d'effet sur le premier que sur le deuxième cycle de la mâchoire. Cette différence peut s'expliquer par le fait que le premier cycle est directement contraint par la réalisation du deuxième qui borne sa durée.

Ensuite, le premier cycle de la mâchoire débute plus tôt pour la consonne /t/ que pour /p/ alors que l'inverse caractérise le deuxième cycle. Ce phénomène est commun aux deux expériences et induit une durée des phases du premier cycle plus longue pour /t/ que pour /p/ alors que l'inverse s'observe pour le deuxième cycle. D'autre part, l'amplitude et la valeur du pic de vitesse des gestes sont plus importantes pour /p/ que pour /t/, quel que soit le cycle. L'étude des positions de la mâchoire aux moments des constriction (voir dernière partie de la discussion) montre que la mâchoire serait plus haute pour le /p/ que pour le /t/. De plus, l'effet de l'accentuation sur les gestes de la mâchoire semble plus important pour /t/ que pour /p/. Ces deux éléments vont dans le sens d'une implication plus importante de la mâchoire dans la réalisation du /p/ que du /t/. Ces conclusions doivent cependant être pondérées par le fait que le capteur utilisé pour mesurer les mouvements de la mâchoire, s'il est porté par la mâchoire, est aussi *tiré* par les mouvements de la lèvre inférieure. Mais ce qui reste le plus important au regard des objectifs de cette étude est que les variations segmentales ne modifient pas l'occurrence des minima de trajectoire de la mâchoire et donc, leur relation avec les événements du geste de pointé, alors que la variation de la position de l'accent les affecte clairement.

5.4.2 ... mais aussi : pointer en dénommant.

L'adaptation propre à chaque système s'accompagne d'une adaptation à la tâche coordinative : la parole va à la rencontre de la main et la main attend la parole, de façon à ce que la partie accentuée de la parole soit réalisée pendant que le doigt montre la cible.

Quand la parole va à la rencontre de la main

L'hypothèse de travail prévoyait une rencontre entre la partie accentuée de la parole (la position basse de la mâchoire pour la syllabe accentuée) et l'apex du geste de pointé. Les résultats montrent que ce rendez-vous tend à advenir quand

l'accent est sur la première syllabe et qu'il peut être attribué au report de la réponse verbale. Deux éléments vont dans le sens de cette interprétation. D'abord, la littérature a largement montré que la réponse verbale commence plus tôt en tâche de parole-seule qu'en double-tâche (voir la section Introduction de ce chapitre et le chapitre précédent). Ce phénomène reposerait sur les latences de la mise en mouvement propres à chaque système, avec un système brachiomanuel plus lent que celui de la parole. Ensuite, l'initialisation du premier cycle de la mâchoire s'adapte à la distance de la cible afin de préserver le rapprochement entre la partie accentuée de la parole et l'apex du geste de pointé. Cette adaptation de la mâchoire à la position de la cible est aussi observée quand l'accent est sur la deuxième syllabe. Cependant, dans cette condition, la synchronisation attendue entre l'apex du geste de pointé et la partie accentuée de la parole n'est pas observée.

A priori, en considérant les résultats de Levelt *et al.* (1985), on pourrait attribuer l'échec de synchronisation dans la condition *accent*₂ au fait que la rencontre entre les deux systèmes ne se fait pas autour de la partie accentuée de la parole mais autour du début de la voix, qu'on peut évaluer ici par la position basse du premier cycle de la mâchoire. Les résultats invalident clairement cette hypothèse. D'abord, au niveau global, le rapprochement entre la position basse du premier cycle et l'apex du geste de pointé n'est pas observé quand l'accent est sur la deuxième syllabe. Dans cette condition, l'apex du geste de pointé advient entre les positions basses des deux cycles du fait d'une avance de la réponse de la mâchoire par rapport à la condition où l'accent est sur la première syllabe. La coordination entre les deux systèmes ne repose donc pas sur un rapprochement entre le début de la voix et l'apex du geste de pointé. Un point assez étonnant est que dans les deux expériences, quand l'accent est sur la deuxième syllabe, le début du premier geste de la mâchoire suit celui du geste de pointé d'environ 150 ms, soit un délai équivalent à celui entre l'apex du geste de pointé et le deuxième apex de la mâchoire. En d'autres termes, si le premier geste de la mâchoire débutait en même temps que le geste de pointé, le deuxième apex de la mâchoire pourrait se synchroniser avec l'apex du geste de pointé. Une explication possible du délai de la réponse de la mâchoire par rapport à celle du doigt serait que dans la tâche utilisée ici, l'initialisation de la parole pourrait s'ancrer sur le geste de pointé, via l'évaluation en ligne de la progression du doigt. L'adaptation de la parole au geste de pointé pourrait être un premier argument en ce sens. Cette hypothèse expliquerait le fait que lorsqu'ils introduisent une perturbation mécanique du geste de pointé, Levelt *et al.* (1985) observent une adaptation de la parole seulement si cette perturbation advient au début du geste. L'évaluation de cette hypothèse nécessitera de déterminer le type d'informations utilisées par le contrôle moteur de la parole pour démarrer sa réponse.

Mais, le rapprochement entre l'apex du geste de pointé et le début du deuxième stroke d'ouverture de la mâchoire (correspondant approximativement au moment de la constriction) quand l'accent est sur la deuxième syllabe rappelle aussi la syn-

chronie entre la main et la parole observée par Attina *et al.* (2006) dans le langage parlé-complété (LPC). Dans ce système de communication, la parole est complétée par des gestes de main mettant les doigts en contact avec différentes parties du visage afin de désambigüiser certains phonèmes et de fournir une information complémentaire aux personnes mal-entendantes. Attina *et al.* montrent que, chez des locuteurs du français, l'apex du geste de la main atteignant la position adéquate pour une syllabe donnée est synchronisé avec la constriction consonantale de cette syllabe dans la parole, ce qui correspond bien à une position haute de la mâchoire comme dans notre étude.

En résumé, la parole va à la rencontre de la main en adaptant l'initialisation de la mise en mouvement des articulateurs à la distance de la cible et à la position de l'accent. Cette adaptation suffit à synchroniser l'apex de la mâchoire pour la voyelle accentuée avec l'apex du geste de pointé quand l'accent est sur la première syllabe mais pas quand il est sur la deuxième syllabe. Dans ce dernier cas, l'apex de la mâchoire pour la voyelle accentuée est inclus dans la phase de tenue du geste de pointé, via une prolongation de cette phase.

Quand la main attend la parole

Contrairement à la mâchoire, le geste de pointé, de l'initialisation à l'apex, dépend principalement de la position de la cible : il est peu sensible à la variation de la position de l'accent. Cette observation montre que ce geste est principalement dirigé par la nécessité d'aligner le doigt avec la cible. Cependant, une fois cette tâche accomplie, la tenue du geste de pointé s'adapte à la position de l'accent. Ainsi, quand l'accent est sur la première syllabe, la tenue est réduite et prend fin avant la réalisation de la deuxième syllabe. En revanche, quand l'accent est sur la deuxième syllabe, la tenue est prolongée jusqu'à la fin de la phase descendante du deuxième cycle de la mâchoire. Cette adaptation advient après que le geste de pointé ait atteint sa cible et compense le délai d'initialisation de la réponse verbale.

Au vu de ces résultats, on peut s'interroger sur l'adaptation des deux systèmes avec des noms de cibles de plus de deux syllabes accentués sur la dernière syllabe. Est-ce que la mâchoire avancerait davantage sa réponse ? Est-ce que la tenue du geste de pointé prendrait fin avec la réalisation de la dernière voyelle ? Est-ce que l'initialisation du geste de pointé serait reportée ? Est-ce que l'apex du geste de pointé se synchroniserait avec la constriction de la consonne pour la syllabe accentuée ? Ici encore, quelles que soient les réponses à ces différentes questions, l'ensemble des résultats, s'il ne confirme pas totalement l'hypothèse de départ, met en évidence que les deux systèmes s'organisent de manière à ce que la partie accentuée de la parole advienne pendant que le doigt pointe la cible, organisation qui lie la partie du discours *qui montre* avec la partie du geste qui montre. Ainsi, l'accentuation de la parole constituerait un site déictique crédible pour la coordination main-bouche, qui pourrait être portée par une coordination main-mâchoire. La question est de savoir maintenant si le pattern de coordination observé ici avec un

matériel phonologique simple est reproduit dans la cas d'énoncés plus complexes avec une monstration par la voix plus réaliste, c'est-à-dire un *vrai* focus prosodique comme dans *MADELEINE m'amena*.

5.4.3 Profil de mâchoire et effet LC

Si la variation de la structure segmentale du nom de la cible n'a pas d'effet sur la coordination entre la parole et la main, les différents profils de mâchoire observés pour /t/ et /p/ appellent une étude plus détaillée de ces profils relativement aux questions évoquées au chapitre 3 traitant de l'effet LC. Rappelons que notre hypothèse pour expliquer le fait que les langues du monde préfèrent les structures labiale-coronale (L_aC_o , /pata/) aux structures coronale-labiale (C_oL_a , /tapa/) est que les structures L_aC_o seraient plus stables et plus économiques pour le système articulaire que les structures C_oL_a . Les structures L_aC_o pourraient notamment être produites avec des mouvements de mâchoire moins importants que les structures C_oL_a . En étudiant les productions accélérées de bisyllabes labiale-labiale (/papa/), coronale-coronale (/tata/), L_aC_o (/pata/) et C_oL_a (/tapa/) nous avons montré que les bisyllabes dupliqués étaient toujours réalisés sur deux cycles de mâchoire, un par syllabe alors que les bisyllabes variés pouvaient, avec l'augmentation du débit, être réalisés sur un seul cycle. La question est ici de déterminer si les profils de mâchoire des bisyllabes L_aC_o et C_oL_a produits à débit normal présentent les germes de cette progression vers un seul cycle. Nous avons étudié deux choses : la hauteur de la mâchoire pour la réalisation des deux constriction et l'énergie cinétique du mouvement de l'initialisation de M_I à M_{min2} , estimée par :

$$E_{cin} \propto \sum_{i=M_I}^{M_{min2}} v_i^2$$

avec v , vitesse du mouvement pour chaque échantillon i de M_I à M_{min2} . La moyenne et l'erreur standard de ces mesures sont représentés sur le figure 5.12 selon le bisyllabe. Seules les productions accentuées sur la deuxième syllabe sont étudiées, celles accentuées sur la première syllabe *négligeant* le deuxième cycle de la mâchoire, quel que soit le bisyllabe. Nos hypothèses sont les suivantes. D'abord, le /p/ serait réalisé sur une position de la mâchoire plus haute que le /t/, il serait donc plus avantageux pour la mâchoire de commencer par le /p/ puis d'effectuer le /t/ que l'inverse. L'énergie cinétique devrait donc être plus importante pour /papá/ que /tatá/ d'une part (le /t/ pouvant se faire sur une position plus ouverte que le /p/) et pour /tapá/ que pour /patá/, d'autre part.

D'après la figure 5.12, en A., la position de la mâchoire pour les constriction est plus haute pour /papá/ que pour /tatá/ avec une différence de +0.53 mm pour la première constriction (en M_I) et une différence de +1.0 mm pour la deuxième constriction (en M_{max}). Des t-tests appariés montrent que la différence n'est pas significative pour la première constriction mais l'est pour la deuxième [$t(19) = 2.1$,

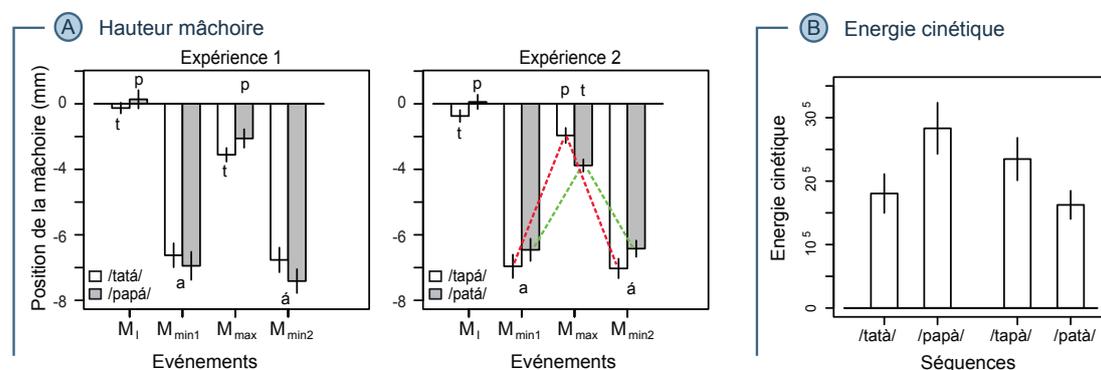


FIG. 5.12 – Profils de mâchoire et effet LC. En A, moyenne et erreur standard moyenne de la hauteur de la mâchoire (position du capteur estimant les mouvements de la mâchoire par rapport à l'origine du repère) au moment de la première constriction et de la deuxième constriction (M_1 , initialisation du premier geste d'ouverture de la mâchoire, M_{max} , position la plus haute entre les deux cycles de mâchoire) et des deux voyelles (M_{min1} et M_{min2} positions basses de la mâchoire) selon les séquences des deux expériences. Le capteur évaluant les mouvements de la mâchoire tend à être plus haut pour le /p/ que pour le /t/, quelle que soit la séquence, induisant une réduction du geste de fermeture entre les deux syllabes et du deuxième geste d'ouverture pour /patà/ par rapport à /tapà/. En B, moyenne et erreur standard moyenne de l'énergie cinétique de l'initialisation du premier geste de la mâchoire (M_1) à la fin de la phase d'ouverture du deuxième cycle (M_{min2}) selon les séquences des deux expériences. L'énergie cinétique est plus importante pour /papà/ que pour /tatà/ d'une part, et pour /tapà/ que pour /patà/, d'autre part.

$p \leq 0.05$]. Pour la deuxième expérience, la position de la mâchoire pour la première constriction est 0.86 mm plus haute pour /patà/ que pour /tapà/ [$t(19) = 2.5$, $p \leq 0.05$]. En revanche, pour la deuxième constriction (M_{max}), la mâchoire est 1.96 mm plus basse pour /patà/ que pour /tapà/ [$t(19) = 3.6$, $p \leq 0.01$]. Le /t/ serait donc réalisé sur une position de mâchoire plus basse que le /p/, et ce, qu'il précède ou suit un /t/ ou un /p/.

Ces différences dans les positions de la mâchoire s'accompagnent, dans la première expérience, de valeurs moyennes d'énergie cinétique plus importantes pour /papà/ que pour /tatà/ [+102717 (Unité normalisé, la masse étant absente) $t(19) = 4.3$, $p \leq 0.001$], figure 5.12, en B. Enfin, le point le plus intéressant au regard de notre hypothèse sur l'effet LC, est que dans la deuxième expérience, l'énergie cinétique est plus importante pour /tapà/ que pour /patà/ [+72284, $t(19) = 4.8$, $p \leq 0.001$].

Ces analyses vont dans le sens de dynamiques de mâchoire différentes selon les séquences. Pour les séquences dupliquées, l'implication de la mâchoire dans la réalisation de la constriction est plus importante pour la consonne /p/ que pour /t/. Ceci est dû au fait que la pointe de la langue peut fermer le conduit vocal sur une position basse de la mâchoire. En revanche, en condition de production normale, la fermeture du conduit vocal par mise en contact de la lèvre inférieure avec la lèvre supérieure requiert une position haute de la mâchoire. Ces propriétés font que pour /patà/, le geste de fermeture vers la deuxième constriction est réduit par rapport à /tapà/, ce qui expliquerait en partie que l'énergie cinétique soit plus faible pour

/patá/ que pour /tapá/. Ces résultats devront être confirmées par des études plus systématiques des cycles de la mâchoire, utilisant des méthodes d'enregistrement du mouvement de la mâchoire plus précises que l'Optotrak (comme l'EMA) afin de mieux distinguer la part du mouvement de la lèvre inférieure de celle de la mâchoire. Si ces propriétés s'avèrent reproductibles, elles constitueront un argument fort pour un facteur moteur à l'effet LC. Ces études pourront utiliser des séquences de types /atapa/ vs. /apata/, dans une langue accentuant la dernière syllabe comme le français. Ce type de séquence permettra de contrôler la position de départ (position ouverte pour le /a/) et de comparer deux cycles de mâchoire *pleins*, définis comme au chapitre 3 par l'association d'un geste de fermeture et d'un geste d'ouverture (un cycle pour /aC₁a/ et un cycle pour /aC₂a/ avec C₁ et C₂, première et deuxième consonne). L'utilisation de ces séquences permettra des comparaisons plus précises avec les résultats obtenus avec le paradigme de répétition accélérée.

5.5 CONCLUSION

L'objectif des deux études présentées ici était d'abord d'évaluer dans quelle mesure la coordination entre la parole et la main dans une tâche déictique peut s'organiser autour d'un rapprochement entre la partie accentuée de la parole et l'apex du geste de pointé. Dans l'ensemble, les résultats s'accordent avec l'idée d'un ancrage de la partie accentuée de la parole dans la phase de tenue du geste de pointé. Si l'interprétation de certains résultats appelle à des études supplémentaires c'est principalement le fait de la richesse des informations fournies par une étude détaillée des mouvements des deux systèmes, trop souvent négligée en ce qui concerne les gestes de la parole dans la littérature.

Enfin, s'il reste encore beaucoup de boîtes noires en ce qui concerne les processus de la coordination main-bouche dans la deixis, les données rapportées ici sont des arguments supplémentaires, dans la suite de Levelt *et al.* (1985), pour développer les études expérimentales de la coordination entre les deux systèmes. Ces études devront préciser les fondements permettant à la main et à la parole de se rencontrer et notamment de déterminer dans quelle mesure cette rencontre est effectivement une rencontre de cadres. Le chapitre suivant est un pas dans cette direction.

Deux cycles de mâchoire pour un geste de pointé ? Une étude préliminaire.

L'expérience rapportée ici a été réalisée dans le même contexte que les expériences du chapitre précédent, à l'institut Max Plank des sciences cognitives et du cerveau de Munich. Une présentation des données en anglais est disponible dans Rochet-Capellan et al. (2007c).

6.1 INTRODUCTION

6.1.1 Contexte et enjeux théoriques

Notre survol de la littérature sur la coordination main-bouche au chapitre 4 fait apparaître la co-production de la parole et du geste de pointé brachiomanuel comme une étape clé du développement du langage. Cependant, le développement de la maîtrise de cette coordination et son rôle possible dans la genèse des formes du langage est une question qui a été posée seulement récemment par Abry *et al.* (2004). Une hypothèse très forte de ces auteurs est que la rencontre entre la mâchoire dans la parole (*le cadre de la parole*) et le geste de pointé (*le cadre du signe*) serait portée par un rapport de deux cycles de mâchoire pour un geste de pointé. Selon Abry *et al.* ce rapport 2 :1 serait un élément pivot de l'émergence du langage. Ramené à des horizons théoriques moins vastes, la mise en évidence d'un tel rapport chez l'adulte constituerait un élément supplémentaire pour mieux comprendre les modalités de la coordination entre la main et la parole dans la deixis, notamment en ce qui concerne les modalités d'un couplage dynamique entre les deux systèmes. Dans ce cadre, ce chapitre propose et évalue un nouveau paradigme pour tester l'hypothèse d'un rapport de 2 :1 entre les oscillations de la mâchoire dans la parole et le geste de pointé chez l'adulte.

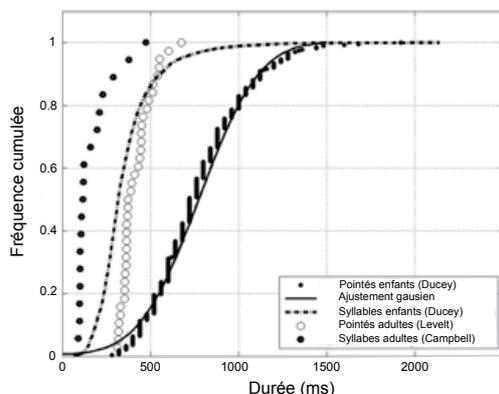


FIG. 6.1 – Etude des relations entre la durée de la syllabe et la durée du geste de pointé chez l'enfant et l'adulte (d'après Ducey-Kaufmann *et al.*, 2007; Ducey-Kaufmann, 2007). Les courbes représentent les distributions des fréquences cumulées (normalisées) des durées de 276 gestes de pointé et 4153 syllabes produits par des bébés ayant le français pour langue maternelle et âgés de 6 à 18 mois. Ces données sont comparées avec les données de Levelt *et al.* (1985) pour les gestes de pointé chez l'adulte et celles de Campbell (1992) pour les durées de la syllabe chez l'adulte.

6.1.2 L'hypothèse 2 :1 : une hypothèse développementale

Ducey-Kaufmann et Abry avancent une hypothèse qu'ils qualifient eux-mêmes d'*étonnante* pour répondre à une question non-élucidée à ce jour, à savoir le fait que les premiers mots de l'enfant comprennent généralement deux syllabes. Une formulation simplifiée et réduite de leur approche est que ces formes bisyllabiques émergeraient de l'*enchâssement* du *cadre de la parole* (au sens défini par MacNeillage et Davis, 2000, voir chapitre 4) dans le *cadre du signe*. Cet enchaînement se baserait sur un rapport entre les fréquences naturelles des deux systèmes de deux cycles de mâchoire pour un geste de pointé et reposerait sur les propriétés de leurs maturations nerveuses respectives. Cette hypothèse est largement décrite dans la thèse de Ducey-Kaufmann (2007). Pour ce qui nous intéresse, Ducey-Kaufmann fournit aussi deux résultats allant dans le sens du rapport 2 :1 chez l'adulte et chez l'enfant. Pour l'adulte, elle compare les durées moyennes de syllabes obtenues par Campbell (1992) aux durées moyennes de gestes de pointé obtenues par Levelt *et al.* (1985). Pour l'enfant, elle analyse un corpus de 4153 syllabes et 276 pointés produits par six enfants français de 6 à 18 mois. Les résultats sont résumés sur la figure 6.1 sous la forme de la distribution des fréquences cumulées des moyennes des différents mouvements. Si les productions de l'adulte paraissent globalement plus courtes que celles de l'enfant, que ce soit pour la parole ou pour la main, le rapport de 2 :1 tend à émerger dans les deux cas. Ces premiers résultats encouragent à évaluer plus directement l'hypothèse 2 :1 chez l'adulte.

6.1.3 Evaluer l'hypothèse 2 :1 chez l'adulte

Comme le constate Ducey-Kaufmann (2007), la plupart des études s'intéressant au couplage de la coordination main-bouche utilisent des tâches rythmiques. Ces travaux, généralement réalisés dans une perspective dynamique, étudient les relations entre les deux systèmes dans des tâches consistant à répéter à différentes vitesses un geste de la main ou du doigt tout en produisant de la parole, comme, par

exemple, taper avec le doigt en produisant une syllabe CV. Comme nous l'avons vu au chapitre 4, cette catégorie d'études a mis en évidence que la mâchoire et la main peuvent se coupler et que ce couplage présente des relations de phasages préférentielles. Cependant, aucune étude, à notre connaissance (et à celle de Ducey-Kaufmann), n'a été menée afin d'établir le rapport entre les fréquences préférentielles des oscillations mandibulaires dans la parole et du geste de la main dans le pointage. Ici, nous introduisons un paradigme expérimental original permettant d'évaluer l'hypothèse 2 :1 chez l'adulte. La logique à la base de ce paradigme est la suivante : si les deux systèmes sont liés par un rapport de deux cycles de mâchoire sur un geste de pointé alors, deux devrait être le nombre maximal de cycles de mâchoire pouvant être réalisés sur un geste de pointé. Si on considère que la syllabe CV (Consonne-Voyelle) est produite, en condition normale, sur un cycle de mâchoire, alors, le geste de pointé devrait rester le même quand il est associé à un énoncé de 1- et 2-syllabe(s) alors qu'il devrait augmenter de 2- à 3-syllabes pour être de nouveau stable de 3- à 4-syllabes. Afin de tester cette hypothèse, nous avons utilisé une tâche consistant à montrer une cible tout en la nommant avec une séquence de 1- vs. 2- vs. 3- vs. 4-syllabe(s) CV, deux fois de suite, sans arrêt entre les deux désignations. Selon l'hypothèse 2 :1, le délai entre les apex (extrema) des deux gestes de pointé devrait suivre un portrait :

$$1_{\text{syll.}} = 2_{\text{syll.}} < 3_{\text{syll.}} = 4_{\text{syll.}}$$

Ce paradigme a été évalué dans une première expérience passée par les neuf dernières participantes des expériences sur les sites de synchronisation présentées au chapitre précédent, toujours en fin de session expérimentale. Bien que l'expérience reste préliminaire, la richesse de ces premiers résultats nous encourage à les présenter ici.

6.2 MÉTHODE

6.2.1 Participants

Les participantes étaient donc neuf locutrices du Portugais brésilien, toutes droitières et sans trouble de la parole ou auditif répertorié. Contrairement aux études sur les sites de synchronisation, les propriétés accentuelles de la langue ne constituent pas un critère sélectif pour cette expérience.

6.2.2 Dispositif expérimental

L'expérience présentée ici utilisait le dispositif mis en place pour les expériences précédentes, voir figure 5.2, p.93. Sur le plan méthodologique, seul le matériel phonétique, la tâche, la consigne et la durée d'affichage des items étaient différents.

6.2.3 Procédure

Tâche et plan d'expérience

La tâche était de pointer une cible, toujours une icône smiley, avec l'index de la main droite tout en la nommant avec une séquence de 1- (/pá/), 2- (/papá/), 3- (/papapá/) ou 4- (/papapapá/) syllabes et ce, deux fois de suite. Par exemple, un essai /pá/ se décomposait en : [« pá » + geste de pointé] - [retour du doigt à la position de départ] - [de nouveau « pá » + geste de pointé] - [retour du doigt à sa position de départ]. Le premier facteur expérimental était donc le nombre de syllabes avec quatre niveaux. D'autre part, afin de varier la tâche et de tester la robustesse de l'effet du nombre de syllabes, la cible pouvait apparaître à une position proche (10 cm) ou loin (50 cm) du centre du champ visuel (voir figure 5.2, C., p. 93). L'expérience comportait donc deux facteurs intrasujets, le nombre de syllabes dans le nom de la cible (1_{syll.} vs. 2_{syll.} vs. 3_{syll.} vs. 4_{syll.}), et la position de la cible (*cible_{proche}* vs. *cible_{loin}*), soit huit conditions expérimentales passées par toutes les participantes.

Déroulement

Excepté l'augmentation de la durée d'affichage de la cible afin que celle-ci soit visible pendant la réalisation des deux désignations, le déroulement de l'expérience était identique à celui des expériences sur les sites de synchronisation (voir figure 5.3, p.96). La passation se subdivisait en quatre blocs de 44 essais chacun, dont 4 essais d'entraînement suivis de 40 essais expérimentaux (5 essais par condition expérimentale). Les essais expérimentaux étaient arrangés de manière aléatoire à l'intérieur de chaque bloc et différemment pour chaque participante. Les blocs étaient passés successivement, avec une pause de 30 s entre deux blocs, indiquée par un écran « pausa », pour une durée d'1 s, suivi d'un écran noir. Un essai débutait par l'affichage de la cible en rouge, à une position proche ou loin, et de son nom, au centre du champ visuel pour une durée de 2.5 s + un temps aléatoire, tiré dans une distribution gaussienne de moyenne 0 s et d'écart-type 0.15 s. La cible devenait ensuite verte pour une durée de 2.5 s, suivie de l'écran noir d'1 s, séparant deux essais consécutifs.

Consigne

Les participantes réalisaient cette dernière tâche à la suite des deux expériences sur les sites de synchronisation. Après une pause libre, l'expérimentateur introduisait la nouvelle tâche. La consigne était de pointer et de nommer la cible avec le mot affiché au centre de l'écran deux fois de suite pour chaque essai. La cible était toujours introduite comme la représentation symbolique d'une personne et la réponse devait être donnée dès le changement de couleur de la cible (du rouge au vert). La consigne précisait aussi que la longueur des noms de la cible pourrait varier de une à quatre syllabe(s) mais que ces noms seraient toujours accentués sur la dernière syllabe. Le geste du doigt devait partir de la position indiquée sur la

table et revenir à cette position, et ce, pour les deux gestes de pointé. Exceptée la contrainte de donner la réponse dès le changement de couleur de la cible, aucune contrainte n'était donnée sur la vitesse de l'exécution du geste.

Au cours des premiers essais, il est rapidement apparu que les participantes tendaient à confondre, à la lecture, les séquences /papapá/ et /papapapá/. Afin de contrer cette difficulté, l'expérimentateur lisait ces séquences à haute voix dès leur apparition à l'écran. Si cette aide diminuait les erreurs des participantes, elle constitue un problème méthodologique qui sera discuté ci-après.

6.2.4 Enregistrement et post-traitement des données

Toujours comme pour les expériences sur les sites de synchronisation, la position des marqueurs Optotrak était échantillonnée au cours du temps à une fréquence de 100 Hz et le son enregistré à une fréquence de 16 kHz. La procédure suivie pour le post-traitement des données est celle qui est décrite au chapitre précédent : après calcul des coordonnées des capteurs en mouvement dans le référentiel *table* puis du capteur de la mâchoire dans le référentiel *tête*, une Analyse en Composantes Principales permet d'obtenir une représentation unidimensionnelle du mouvement. Pour cette expérience, la première composante rend compte de 92% de la variance pour le marqueur de la mâchoire et de 98% de la variance pour chaque marqueur du doigt, elle constitue donc une représentation satisfaisante de la trajectoire du mouvement. Les procédures d'enregistrement et de traitement des données sont détaillées dans la section « Méthode » du chapitre précédent.

6.2.5 Etiquetage et mesures

La figure 6.2 donne un exemple de signaux pour une séquence /papapá/. Les débuts et les fins de gestes correspondent aux moments où le mouvement atteint 10% de son pic de vitesse respectivement avant et après ce pic. Dans la suite, l'indice n correspond à l'ordre de désignation, avec $n = 1$ pour la première et $n = 2$ pour la deuxième.

Etiquetage et mesures pour les mouvements de la mâchoire

Pour la mâchoire, M_{In} , est l'initialisation du geste d'ouverture pour la première syllabe et M_{An} , l'apex du geste d'ouverture pour la dernière syllabe. Ainsi, le délai de M_{In} à M_{An} (D_{Mn}) est la durée de production du nom de la cible, quel que soit le nombre de syllabes.

Etiquetage et mesures pour les mouvements du doigt

La trajectoire du doigt se compose de deux gestes de pointé successifs, chacun suivi d'un geste de retour à la position de repos. Les gestes de pointé vont de P_{In} (initialisation) à P_{An} (apex du geste), avec un pic de vitesse à P_{Vn} . Les gestes de retour vont de P_{Rn} (fin de l'alignement du doigt avec la cible) à P_{Fn} (le retour à la

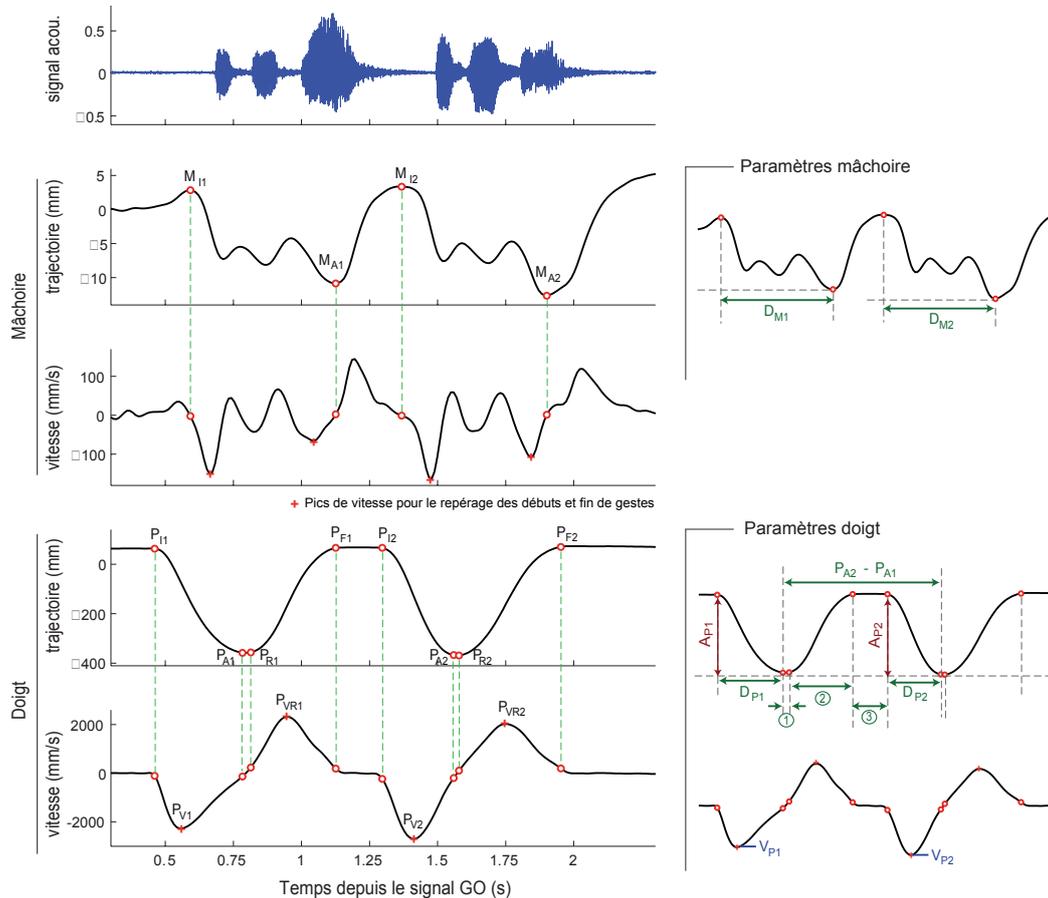


FIG. 6.2 – Signaux, étiquetages et paramètres des mouvements de la mâchoire et du doigt pour un exemple /papapá/. **A gauche**, de haut en bas : signal acoustique ; courbes de trajectoire et de vitesse de la mâchoire ; courbes de trajectoire et de vitesse du doigt. La trajectoire de la mâchoire se caractérise par un cycle par syllabe pour chacune des deux désignations et la trajectoire du doigt par un cycle par désignation (pointé + retour). Soit $n = 1$ la première désignation et $n = 2$ la deuxième, M_{In} est le début du geste d'ouverture de la mâchoire pour la première syllabe, repéré à 10% du pic de vitesse de ce geste, avant le pic. De la même manière, M_{An} est l'apex du geste d'ouverture de la mâchoire pour la dernière syllabe, repéré à 10% du pic de vitesse, après ce pic. Pour le doigt, P_{In} et P_{An} sont l'initialisation et l'apex des gestes de pointé et P_{Vn} , le pic de vitesse de ce geste. De la même manière, P_{Rn} et P_{Fn} sont le début et la fin du geste de retour et P_{VRn} , le pic de vitesse de ce geste. **A droite**, principales mesures étudiées. Pour la mâchoire, D_{Mn} est la durée de production du nom de la cible pour chaque désignation, de M_{In} à M_{An} . Pour le doigt, D_{Pn} , A_{Pn} et V_{Pn} sont la durée, l'amplitude et l'amplitude du pic de vitesse des deux gestes de pointé, de P_{In} à P_{An} . $P_{A2} - P_{A1}$ est le délai entre les apex des deux gestes. Enfin, trois phases séparent l'apex du premier geste de pointé de l'initialisation du deuxième geste de pointé. ① La tenue du premier geste, de P_{A1} à P_{R1} , ② le geste de retour, de P_{R1} à P_{F1} , et le plateau entre la fin du premier cycle (P_{F1}) et le début du deuxième cycle (P_{I2}).

position de repos) avec un pic de vitesse à P_{VRn} . Pour chacun des deux gestes de pointé, D_{Pn} , A_{Pn} et V_{Pn} sont la durée, l'amplitude et l'amplitude du pic de vitesse du geste. Comme expliqué dans la section introduction, l'hypothèse de travail se fonde sur l'analyse de la distance entre les deux cycles de pointé, mesurée ici par $P_{A2} - P_{A1}$, le délai entre les apex des deux gestes de pointé.

Seuls les essais sans erreur dans le nom de la cible et avec au moins un des deux marqueurs du doigt et le marqueur de la mâchoire visibles sont analysés ci-après. Le nombre d'essais analysés pour chaque locutrice et le nombre d'erreurs dans le nom de la cible sont donnés dans le tableau 6.1, p. 133.

6.2.6 Hypothèses

Cette expérience teste l'hypothèse d'un rapport de deux cycles de mâchoire pour un geste de pointé brachiomanuel. Comme expliqué en introduction, si un tel rapport lie les durées *naturelles* des deux systèmes, alors l'effet du nombre de syllabes sur le délai entre les deux apex du doigt $P_{A2} - P_{A1}$ devrait être significatif avec un portrait :

$$1_{syll.} = 2_{syll.} \leq 3_{syll.} = 4_{syll.}$$

Au moins trois questions demeurent concernant ce portrait.

1. D'abord, s'il est observé, résiste-t-il à l'augmentation de l'amplitude du geste de pointé induite par l'éloignement de la cible ? Si tel est le cas, la position de la cible devrait avoir un effet significatif sur $P_{A2} - P_{A1}$.
2. Ensuite, quelle phase de la trajectoire du doigt est modifiée pour sous-tendre ce portrait ? L'étude de Levelt *et al.* (1985) montre que le geste de pointé est relativement peu affecté par la présence de la parole. L'augmentation du nombre de syllabes de deux à trois augmente-t-elle la durée du geste ou bien seulement la phase de tenue et/ou de retour (voir figure 6.2) ?
3. Enfin, si comme le supposent Abry et Ducey-Kaufmann, la relation entre les fréquences préférentielles des deux systèmes permet de réaliser deux syllabes sur un geste de pointé, la durée de l'énonciation du nom de la cible (D_{Mn}) devrait être proche de la durée du geste de pointé (D_{Pn}) en condition $2_{syll.}$ alors qu'elle devrait être plus faible en $1_{syll.}$ et plus élevée en $3_{syll.}$.

6.3 RÉSULTATS

Les effets des facteurs expérimentaux sur les différentes mesures sont testés ci-après par des ANOVAs à deux facteurs intrasujets *nombre de syllabes* \times *position de la cible*. Le facteur *nombre de syllabes* ayant plus de deux niveaux, la sphéricité a été systématiquement testée avec le test de Mauchly. Quand le test est significatif, les valeurs des F de l'ANOVA sont données avec correction de

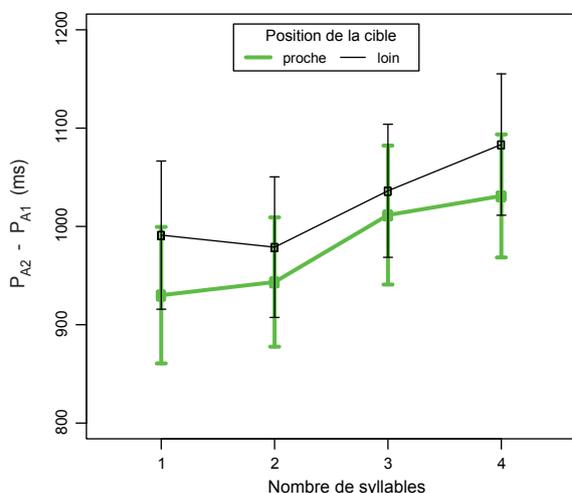


FIG. 6.3 – Moyenne et erreur standard moyenne du délai entre les apex des deux gestes de pointé ($P_{A2} - P_{A1}$) selon le nombre de syllabes et la position de la cible. Globalement, le portrait attendu $1_{syll.} = 2_{syll.} < 3_{syll.} = 4_{syll.}$ apparaît. La tendance à ce que la moyenne en $3_{syll.}$ soit inférieure à celle en $4_{syll.}$ n'est pas significative, voir le texte pour le détail.

Greenhouse-Geisser, d'où les degrés de liberté non-entiers. Pour les mesures impliquant l'ordre de désignation (premier vs. deuxième geste pour le doigt et première vs. deuxième énonciation pour la mâchoire) le facteur *ordre de désignation* est inclus comme troisième facteur intrasujet de l'ANOVA. Cette analyse permet de répondre à des questions telles que : la durée du premier geste de pointé diffère-t-elle de celle du deuxième geste ? Enfin, les comparaisons entre la condition $1_{syll.}$ et $2_{syll.}$ (C_1), entre $2_{syll.}$ et $3_{syll.}$ (C_2) et entre $3_{syll.}$ et $4_{syll.}$ (C_3) sont réalisées avec des t-tests appariés avec correction de Bonferroni ($p_{bf} = p/3$). Le détail de ces analyses pour l'ensemble des mesures est synthétisé en annexe dans les tableaux B.1 et B.2, p. 192 et 193.

6.3.1 Un portrait $1_{syll.} = 2_{syll.} \leq 3_{syll.} = 4_{syll.}$?

Etude du délai entre les apex des deux gestes de pointé, tendances globales

La figure 6.3 donne la moyenne et l'erreur standard moyenne du délai entre les apex des deux gestes de pointé ($P_{A2} - P_{A1}$) selon le nombre de syllabes et la position de la cible. Globalement, $P_{A2} - P_{A1}$ reste stable de $1_{syll.}$ à $2_{syll.}$ (961 ms dans les deux conditions) puis augmente de 63 ms de $2_{syll.}$ à $3_{syll.}$ (1024 ms) et de 33 ms de $3_{syll.}$ à $4_{syll.}$ (1057 ms). Le test de Mauchly est significatif (Greenhouse-Geisser : Epsilon = .47, $p \leq 0.05$), les données ne satisfont donc pas la contrainte de sphéricité. Cependant, l'effet du nombre de syllabes est significatif après correction [$F(1.4, 11.2) = 9.3$, $p \leq 0.01$]. Enfin, les comparaisons multiples montrent que la différence entre $2_{syll.}$ et $3_{syll.}$ est significative [$C_2 : t(8) = 3.1$, $p_{bf} \leq 0.05$] alors que la différence entre $3_{syll.}$ et $4_{syll.}$ n'atteint pas le seuil de significativité [$C_3 : t(8) = 2.5$, $p_{bf} = 0.034 \times 3$].

$P_{A2} - P_{A1}$ est aussi plus importante en condition *cible_{loin}* (1022 ms) qu'en *cible_{proche}* (979 ms). Cet effet est significatif [$F(1, 8) = 14.5$, $p \leq 0.01$]. D'autre

part, la différence entre $1_{\text{syll.}}$ et $2_{\text{syll.}}$ (C_1) est de -13 ms en condition *cible_proche* et de +12 ms en *cible_loin* ; celle entre $2_{\text{syll.}}$ et $3_{\text{syll.}}$ (C_2) est de -68 ms en condition *cible_proche* et de -57 ms en *cible_loin* ; enfin la différence entre $3_{\text{syll.}}$ et $4_{\text{syll.}}$ (C_3) est de -19 ms en condition *cible_proche* et de -47 ms en *cible_loin*. Malgré ces différences assez importantes pour les comparaisons C_1 et C_3 entre les deux conditions de position de la cible, l'effet d'interaction entre le nombre de syllabes et la position de la cible n'est pas significatif [$F(3, 24) = 1.7$]. Si la délimitation de cet effet d'interaction nécessiterait un échantillon de locuteurs plus large, particulièrement en ce qui concerne les comparaisons C_1 et C_3 , les tendances observées ici ne remettent pas en cause le saut entre $2_{\text{syll.}}$ et $3_{\text{syll.}}$.

Globalement, cette première analyse met en évidence le portrait attendu avec un saut net de $2_{\text{syll.}}$ à $3_{\text{syll.}}$ et un bémol pour l'absence de différence entre $3_{\text{syll.}}$ et $4_{\text{syll.}}$, particulièrement en condition *cible_loin*. Des conclusions plus précises quant à l'effet de la position de la cible sur ce portrait appellent un échantillon plus large de participants. Ces analyses vont donc dans le sens de l'hypothèse 2 : 1. Cependant, comme le montre le tableau 6.1, le nombre moyen d'erreurs sur le nom de la cible dépend du nombre de syllabes et est plus important en condition $3_{\text{syll.}}$ (2.3 erreurs) et $4_{\text{syll.}}$ (4.6 erreurs) qu'en condition $1_{\text{syll.}}$ (0.6 erreurs) et $2_{\text{syll.}}$ (0.9 erreurs). L'effet du nombre de syllabes est significatif [$F(3, 24) = 7.9, p \leq 0.001$] ainsi que le contraste entre les deux groupes [$t(8) = 3.4, p \leq 0.01$]. Ce problème, essentiellement dû à la confusion des séquences /papapá/ et /papapapá/ (voir section « Méthode ») amène la question : le portrait observé témoigne-t-il d'un rapport 2 : 1 entre les fréquences ou bien est-il le produit du fait que les conditions $3_{\text{syll.}}$ et $4_{\text{syll.}}$ sont plus difficiles que les conditions $1_{\text{syll.}}$ et $2_{\text{syll.}}$? Si les résultats s'expliquent par une asymétrie de difficulté alors, le portrait observé devrait aussi caractériser les temps de réponse, soit l'occurrence du début du premier geste d'ouverture de la mâchoire (M_{I1}) et celle du début du premier geste de pointé (P_{I1}).

Etude des temps de réponse

La figure 6.4, à gauche, représente les moyennes de P_{I1} et M_{I1} selon le nombre de syllabes et la position de la cible. Elle montre que M_{I1} tend à diminuer avec le nombre de syllabes avec des valeurs de 588 ms, 542 ms, 516 ms et 471 ms des 1- à 4-syllabes, respectivement. L'effet global du nombre de syllabes est significatif [$F(3, 24) = 10.6, p \leq 0.001$] mais pas les comparaisons multiples entre les conditions adjacentes (C_1, C_2 et C_3). L'effet de la position de la cible sur l'occurrence de M_{I1} n'est pas significatif. En ce qui concerne la réponse du doigt, le profil est très différent avec un report du geste dans les conditions $2_{\text{syll.}}$ (408 ms), $3_{\text{syll.}}$ (419 ms) et $4_{\text{syll.}}$ (400 ms) syllabes par rapport à la condition $1_{\text{syll.}}$ (348 ms). Cet effet du nombre de syllabes est significatif [$F(3, 24) = 7.2, p \leq 0.01$] ainsi que la comparaison C_1 , entre $1_{\text{syll.}}$ et $2_{\text{syll.}}$ [$t(8) = 3.1, p_{bf} \leq 0.05$]. En revanche, C_2 et C_3 ne sont pas significatives, ni l'effet de la position de la cible. Ainsi, le portrait caractérisant le délai entre les apex des deux gestes de pointé ne s'observe pas pour les temps de réponse de la mâchoire, ni pour ceux du doigt. Au contraire, ces

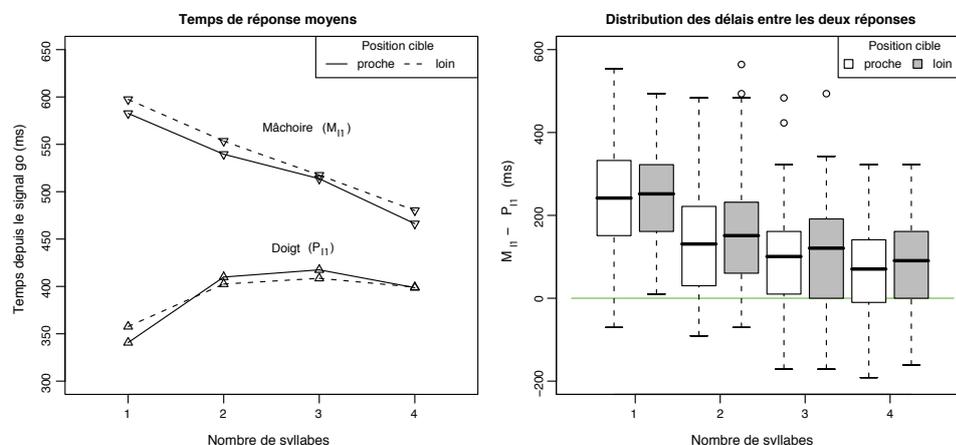


FIG. 6.4 – Étude des temps de réponse. **A gauche**, moyenne des temps de réponse de la mâchoire (M_{11} , initialisation du premier geste d'ouverture) et du doigt (P_{11} , initialisation du premier geste de pointé) selon le nombre de syllabes et la position de la cible. **A droite**, distribution du délai entre la réponse de la mâchoire et celle du doigt ($M_{11} - P_{11}$) pour tous les essais de toutes les locustres confondus et selon les conditions expérimentales. Les limites inférieures et supérieures des rectangles correspondent au premier et troisième quartiles. La médiane est représentée par le trait horizontal plus épais à l'intérieur des rectangles. Pour la majorité des essais, la réponse de la mâchoire advient après celle du doigt mais les deux réponses se rapprochent avec l'augmentation du nombre de syllabes. Le portrait $1_{syll.} < 2_{syll.} < 3_{syll.} = 4_{syll.}$ ne s'observe pas pour les temps de réponse de la mâchoire, ni pour ceux du doigt.

deux mesures sont affectées différemment par l'augmentation du nombre de syllabes, la tendance étant d'avancer la réponse de la mâchoire et de reporter celle du doigt, avec, pour cette dernière, un saut entre la condition $1_{syll.}$ et les trois autres conditions.

D'autre part, d'après le graphique de gauche sur la figure 6.4, la réponse de la mâchoire advient après celle du doigt dans les huit conditions expérimentales. Afin d'évaluer la force de cette tendance, le graphique de droite sur la figure 6.4, synthétise la répartition du délai entre les deux réponses ($M_{11} - P_{11}$) pour tous les essais des 9 participantes. Ces données montrent que $M_{11} - P_{11}$ est positif pour quasiment la totalité des essais en condition $1_{syll.}$ et $2_{syll.}$ et pour 75% des essais pour $3_{syll.}$ et $4_{syll.}$. De plus, $M_{11} - P_{11}$ tend à diminuer avec le nombre de syllabes (240 ms, 134 ms, 97 ms et 70 ms de 1- à 4-syllabes). Cet effet du nombre de syllabes est significatif [$F(3, 24) = 43.1, p \leq 0.0001$] alors que $M_{11} - P_{11}$ n'est pas significativement affecté par la variation de la position de la cible (voir tableau B.1, p. 192 pour le détail).

Ces résultats sur les temps d'initialisation vont à l'encontre d'une explication du portrait observé pour le délai entre les deux apex du doigt en terme d'asymétrie de difficultés. En revanche, ils pourraient témoigner d'une coordination entre les deux systèmes, avec la mâchoire qui avance sa réponse avec l'augmentation du nombre de syllabes alors que le doigt reporte la sienne.

TAB. 6.1 – Délai moyen (m) entre les apex des deux gestes de pointé ($P_{A2} - P_{A1}$) selon le nombre de syllabes pour chaque locutrice. Err. est le nombre d'erreurs sur le nom de la cible et n, le nombre d'essais analysés. Les trois dernières colonnes comparent les conditions 1_{syll.} à 2_{syll.} (C₁), 2_{syll.} à 3_{syll.} (C₂) et 3_{syll.} à 4_{syll.} (C₃). Le saut attendu entre la condition 1_{syll.} et 2_{syll.} s'observe pour les locutrices 1, 5 et 9 et tend à apparaître pour la locutrice 6.

Part.	Nombre de syllabe												Comparaisons		
	Err.	1		Err.	2		Err.	3		Err.	4		C1	C2	C3
		n	m		n	m		n	m		n	m			
1	0	40	667	0	40	669	2	36	813	4	32	884	-2 ^a	-144 ^d	-71 ^a
2	0	27	986	0	27	952	0	28	987	5	23	1045	33 ^a	-35 ^a	-58 ^b
3	0	40	781	1	39	769	1	39	782	0	40	806	12 ^a	-13 ^a	-24 ^a
4	3	36	936	2	36	946	6	34	996	11	29	1107	-10 ^a	-50 ^a	-111 ^b
5	1	39	943	2	38	1012	9	30	1124	6	31	1140	-69 ^a	-112 ^c	-16 ^a
6	2	37	880	3	36	874	1	39	965	8	30	948	6 ^a	-91 ^a	17 ^a
7	0	40	877	0	40	900	0	40	873	0	40	882	-23 ^a	27 ^a	-9 ^a
8	0	40	1405	0	40	1353	2	38	1364	6	34	1384	52 ^a	-11 ^a	-20 ^a
9	0	27	1171	0	30	1175	0	33	1312	1	30	1318	-4	-137 ^d	-6
Moy.	0.7	36	961	0.9	36	961	2.3	35	1024	4.6	32	1057	-1 ^a	-63 ^b	33 ^a

Comparaison avec 0, test de Student avec correction de Bonferroni :
^a $p \geq 0.05/3$, ^b $p \leq 0.05/3$, ^c $p \leq 0.01/3$, ^d $p \leq 0.0001/3$

Variabilité interindividuelle

La dernière question à propos de l'effet du nombre de syllabes sur le délai entre les apex des gestes de pointé concerne la variabilité interindividuelle. Le tableau 6.1 donne les moyennes de $P_{A2} - P_{A1}$ pour chaque participante selon le nombre de syllabes ainsi que les comparaisons C₁, C₂ et C₃ (ici, t-tests non appariés). Le portrait 1_{syll.} = 2_{syll.} < 3_{syll.} = 4_{syll.} apparaît clairement pour les locutrices 1, 5 et 9 et tend à s'observer pour la locutrice 6, soit pour environ la moitié des participantes. Les autres locutrices présentent des portraits variables avec des différences entre conditions adjacentes parfois élevées mais rarement significatives (cf. participante 2). Ces résultats accentuent la nécessité de reproduire l'expérience avec un échantillon de participants plus large et en corrigeant le problème méthodologique de la confusion entre /papapá/ et /papapapá/.

6.3.2 Etude de la trajectoire du doigt

Il s'agit ici de déterminer comment la trajectoire du doigt porte le portrait mis en évidence pour le délai entre les apex des deux gestes de pointé et notamment de vérifier si ce portrait s'observe aussi pour la durée des gestes de pointé, de P_{I1} à P_{A1} (D_{P1}) et de P_{I2} à P_{A2} (D_{P2}) (voir figure 6.2) et/ou s'il est le résultat d'une modification de la phase de tenue et/ou du mouvement de retour à la position de repos entre les deux gestes, de P_{R1} à P_{I2} .

Durée des gestes de pointé

Les durées moyennes des gestes de pointé selon le nombre de syllabes, la position de la cible et l'ordre de réalisation sont représentées sur la figure 6.5, à droite. En moyenne, D_{P1} est de 364 ms et D_{P2} de 356 ms, la durée du geste de pointé ne

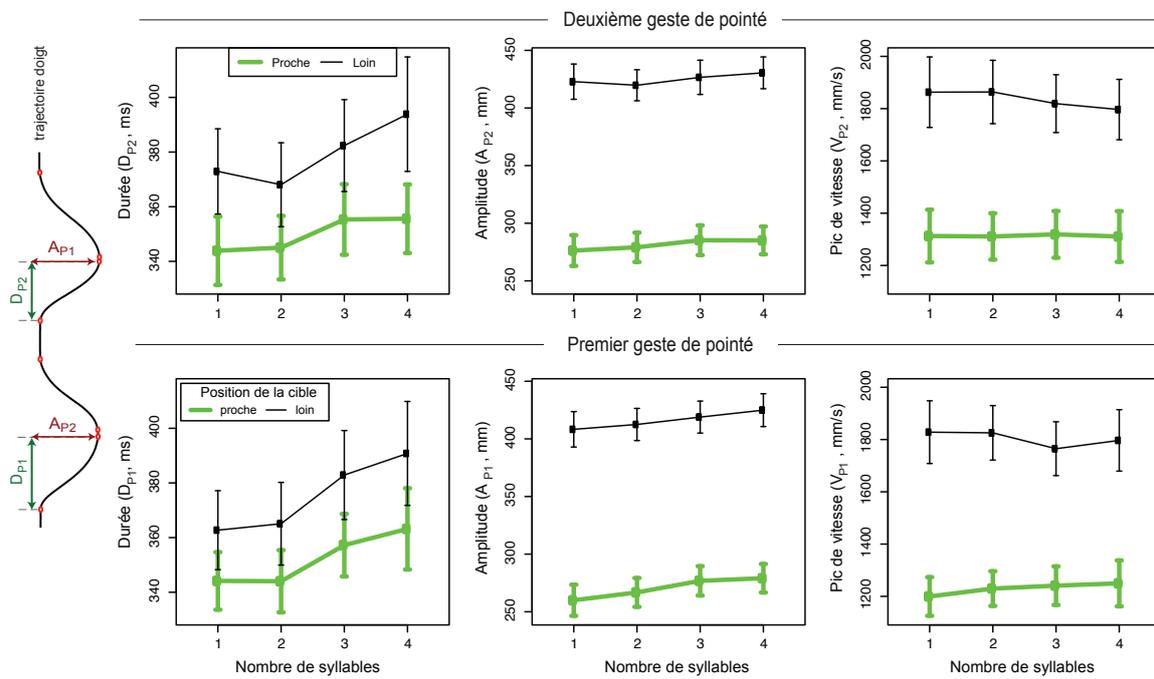


FIG. 6.5 – Étude des gestes de pointé. Moyenne et erreur standard moyenne de la durée (D_{Pn}), de l'amplitude (A_{Pn}) et de l'amplitude du pic de vitesse (V_{Pn}) du premier (en bas, $n = 1$) et du deuxième (en haut, $n = 2$) gestes de pointé selon le nombre de syllabes et la position de la cible. Le portrait $1_{syll.} = 2_{syll.} < 3_{syll.} = 4_{syll.}$ tend à apparaître pour la durée et l'amplitude des gestes. En revanche, l'amplitude du pic de vitesse augmente avec la distance de la cible et tend à diminuer avec l'augmentation du nombre de syllabes en condition *cible_{loin}*.

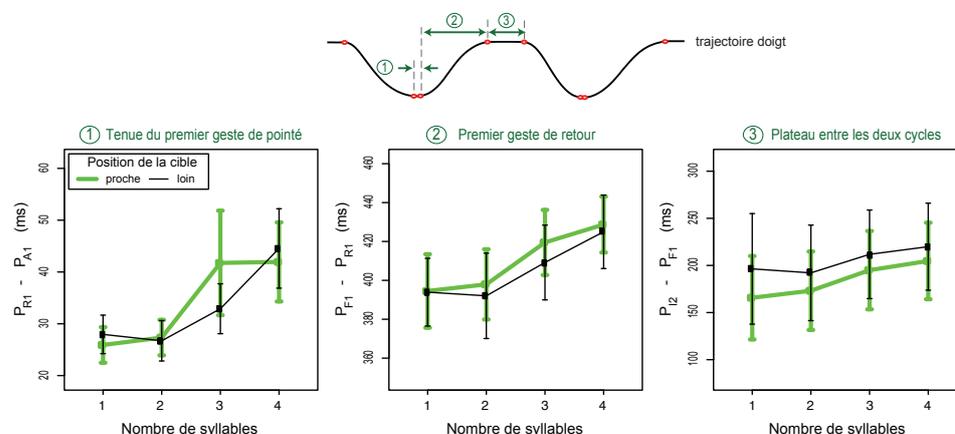


FIG. 6.6 – Étude de la trajectoire du doigt séparant les deux gestes de pointé. Moyenne et erreur standard moyenne des trois phases entre les deux gestes de pointé selon le nombre de syllabes et la position de la cible. ① Durée de la tenue du premier geste de pointé, de l’apex (P_{A1}) au début du geste de retour (P_{R1}). ② Durée du geste de retour, de P_{R1} à P_{F1} . ③ Durée du délai entre la fin du premier cycle et le début du deuxième, de P_{F1} à l’initialisation du deuxième geste de pointé (P_{I2}). Le portrait $1_{syll.} = 2_{syll.} < 3_{syll.} = 4_{syll.}$ tend à apparaître pour chacune des trois mesures.

dépend donc pas de l’ordre de réalisation. En revanche, l’effet de la séquence est significatif [Test de Mauchly : $Epsilon = 0.6, p \leq 0.05, F(1.7, 13.5) = 4.3, p \leq 0.05$] et va dans le sens du portrait observé pour le délai entre les deux apex avec une durée de geste de 356 ms en $1_{syll.}$ et $2_{syll.}$, de 370 ms en $3_{syll.}$ et de 376 ms en $4_{syll.}$. Si les comparaisons multiples C_2 et C_3 , ne sont pas significatives, le contraste 1- et 2-syllabes contre 3- et 4-syllabes est significatif [$t(8) = 2.4, p \leq 0.05$]. La durée des gestes de pointé est aussi significativement plus longue en condition *cible_loin* (377 ms) qu’en *cible_proche* (351 ms) [$F(1, 8) = 11.3, p \leq 0.01$]. Enfin, l’effet du nombre de syllabes interagit significativement avec l’effet de la position de la cible [$F(3, 24) = 3.1, p \leq 0.05$] mais les comparaisons multiples (C_1, C_2, C_3) et les contrastes (1- et 2-syllabes contre 3- et 4-syllabes) ne sont significatifs ni en *cible_loin*, ni en *cible_proche* (voir tableau B.2, p. 193 pour le détail).

De l’apex du premier geste de pointé au début du second

Comme le montre la figure 6.2, la trajectoire du doigt de l’apex du premier geste (P_{A1}) au début du deuxième geste (P_{I2}) se compose de trois phases : la tenue du geste de pointé, de P_{A1} à P_{R1} , le geste de retour, de P_{R1} à P_{F1} et la phase intermédiaire entre les deux gestes, de P_{F1} à P_{I2} . Les moyennes des durées de ces différentes phases sont représentées sur la figure 6.6. Le point le plus intéressant est que chacune des trois phases tend à reproduire le résultat observé pour la durée des gestes de pointé. D’abord, la durée de la tenue du geste de pointé ($P_{R1} - P_{A1}$) est de 27 ms en $1_{syll.}$ et $2_{syll.}$, de 37 ms en $3_{syll.}$ et de 43 ms en $4_{syll.}$, cet effet du nombre de syllabes est significatif [$F(3, 24) = 5.4, p \leq 0.01$]. La durée de la phase de retour ($P_{F1} - P_{R1}$) est de 394 ms, 395 ms, 414 ms et 427 ms de 1- à 4-syllabes, respectivement. Ici encore, l’effet du nombre de syllabes est significatif

$[F(3,24) = 9.7, p \leq 0.001]$. Enfin, la durée de la phase séparant la fin du premier cycle du doigt du début du deuxième ($P_{T2} - P_{F1}$) est de 181 ms, 183 ms, 204 ms et 212 ms de 1 à 4 syllabes, respectivement. L'effet de la séquence est significatif $[F(3,24) = 3.1, p \leq 0.05]$. Comme pour la durée du geste de pointé, les comparaisons multiples C_1 , C_2 et C_3 ne sont pas significatives. En revanche, le contraste entre le groupe 1- et 2-syllabe(s) et 3- et 4-syllabes est significatif pour la durée de la tenue $[t(8) = 2.4, p \leq 0.05]$ et celle du geste de retour $[t(8) = 3.4, p \leq 0.01]$ mais pas pour la phase intermédiaire $[t(8) = 2]$. Ainsi, tout comme la durée du geste de pointé, la durée de la tenue, du geste de retour et de l'intervalle entre les deux cycles du doigt dépendent du nombre de syllabes avec une tendance à observer des valeurs très proches pour 1- et 2-syllabe(s) et une augmentation de 2- à 3-syllabes. Il semble donc que le portrait observé pour le délai entre les apex des deux gestes de pointé et plus précisément, le saut de 2- à 3-syllabes soit le produit d'une modification de l'ensemble des différentes phases séparant les deux apex.

Amplitudes et pics de vitesse des gestes de pointé

Enfin, l'amplitude des gestes de pointé est en moyenne de 342 mm, 345 mm, 352 mm et 355 mm des conditions 1- à 4-syllabes, respectivement. L'effet du nombre de syllabes est significatif $[F(3,24) = 7.7, p \leq 0.001]$ et rend compte, ici encore, d'un regroupement 1- et 2-syllabes contre 3- et 4-syllabes. En effet, si les comparaisons multiples C_1 , C_2 et C_3 ne sont pas significatives, le contraste entre les deux groupes est significatif $[t(8) = 3, p \leq 0.05]$. Contrairement à la durée, l'amplitude des gestes de pointé dépend aussi de l'ordre de réalisation avec une amplitude plus importante pour le deuxième (353 mm) que pour le premier (343 mm) geste $[F(1,8) = 21.6, p \leq 0.01]$. L'interaction entre l'ordre de réalisation et le nombre de syllabes est significative $[F(3,24) = 3.9, p \leq 0.05]$ mais ne rend pas compte de différences notables (voir tableau B.2, p. 193 en annexe pour le détail). Enfin, l'amplitude des gestes dépend clairement de la position de la cible avec un geste plus ample en *cible_{loin}* (421 mm) qu'en *cible_{proche}* (276 mm) $[F(3,24) = 1563, p \leq 0.0001]$. De son côté, l'amplitude des pics de vitesse des gestes de pointé dépend essentiellement de la position de la cible avec des valeurs plus importantes en *cible_{loin}* (1820 mm/s) qu'en *cible_{proche}* (1272 mm/s) $[F(1,8) = 205, p \leq 0.0001]$. L'effet global du nombre de syllabes et de l'ordre de réalisation ne sont pas significatifs. Néanmoins, l'effet de la position de la cible interagit significativement avec le nombre de syllabes $[F(3,24) = 6.3, p \leq 0.01]$ avec une tendance à diminuer la vitesse avec l'augmentation du nombre de syllabes plus marquée en condition *cible_{loin}* que *cible_{proche}* et un saut entre les conditions 2- et 3-syllabes qui tendrait à apparaître en condition *cible_{loin}* (C_1 : 1 mm/s, C_2 : 53 mm/s, C_3 : -5 mm/s) mais pas en *cible_{proche}* (C_1 : -14 mm/s, C_2 : -10 mm/s, C_3 : 0 mm/s).

L'ensemble de ces données vont dans le sens d'un comportement du doigt globalement similaire pour 1- et 2-syllabes d'un côté et 3- et 4-syllabes de l'autre, avec, entre ces deux groupes : des durées et des amplitudes plus petites mais des

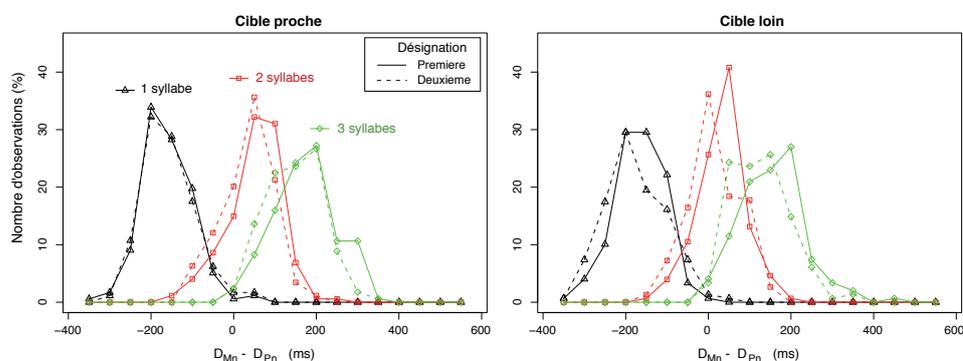


FIG. 6.7 – Distributions de la différence entre la durée de la production du nom de la cible (D_{Mn}) et la durée du geste de pointé qui l'accompagne (D_{Pn}) pour tous les essais des 9 locutrices confondus et selon la position de la cible, le nombre de syllabes et l'ordre de désignation ($n = 1$ pour la première désignation et $n = 2$ pour la deuxième). Globalement, la durée de production du nom de la cible est proche de celle du geste de pointé en condition $2_{syll.}$, alors qu'elle est nettement inférieure en condition $1_{syll.}$ et supérieure en condition $3_{syll.}$.

vitesses plus importantes pour 1- et 2-syllabes que pour 3- et 4-syllabes.

6.3.3 Deux syllabes sur un geste de pointé ?

Les analyses précédentes n'infirmant donc pas l'hypothèse selon laquelle les *fréquences naturelles* des deux systèmes permettrait de réaliser deux syllabes sur un geste de pointé (Abry *et al.*, 2004). Une des conséquences de cette supposition est que la durée de chaque dénomination de la cible (D_{Mn}), devrait être proche de la durée du geste de pointé qui l'accompagne (D_{Pn}) en condition $2_{syll.}$, alors qu'elle devrait lui être inférieure en $1_{syll.}$ et supérieure en $3_{syll.}$. Ainsi, $D_{Mn} - D_{Pn}$ devrait être inférieure à 0 en $1_{syll.}$, proche de 0 en $2_{syll.}$ et supérieure à 0 en $3_{syll.}$. La répartition de $D_{Mn} - D_{Pn}$ pour tous les essais de toutes les locutrices confondus et selon les conditions expérimentales est représentée sur la figure 6.7, le détail de l'ANOVA est disponible en annexe dans le tableau B.1, p. 192. En condition $1_{syll.}$, les valeurs de $D_{Mn} - D_{Pn}$ se répartissent de -300 ms à 0 ms avec un pic (environ 80% des essais) autour de -150 ms (± 50 ms). Elles se décalent ensuite vers la droite avec, en $2_{syll.}$, des valeurs de -150 ms à 200 ms et un pic (environ 90% des essais) autour de 50 ms (± 50 ms) et, en $3_{syll.}$, des valeurs de 0 à 400 ms avec un pic (environ 75% des essais) autour de 150 ms (± 50 ms). Ainsi, la durée de production du bisyllabe /papá/ est assez proche de celle du geste de pointé alors que celle du monosyllabe /pá/ est plus courte et celle du tri-syllabe /papapá/ plus longue. Cette relation entre la durée de production du nom de la cible et celle du geste de pointé évalue à deux le nombre de syllabes maximal pouvant, potentiellement, être réalisé sur un geste de pointé sans en affecter la durée.

6.4 DISCUSSION

Dans l'ensemble, ces résultats n'infirmen pas l'hypothèse d'une relation de deux syllabes sur un geste de pointé. De plus, s'ils doivent être pondérés du fait de problèmes méthodologiques et appellent une expérience plus contrôlée, ils suscitent aussi des questions intéressantes en ce qui concerne les fondements de la coordination entre le cadre du signe et le cadre de la parole.

6.4.1 Du 2 :1 ici...

Deux résultats principaux vont dans le sens de l'hypothèse 2 :1. D'abord, l'analyse des données révèle que l'effet du nombre de syllabes sur le délai entre les apex des deux gestes de pointé suit un portrait $1_{syll.} = 2_{syll.} < 3_{syll.} = 4_{syll.}$. Ce portrait s'observe aussi pour la durée des deux gestes de pointé et pour les phases séparant ces deux gestes. En revanche, il ne s'observe pas pour les temps de réponse de la mâchoire, ni pour ceux du doigt. Comme nous l'avons vu, ce dernier point est un élément allant à l'encontre d'une explication des résultats en termes d'asymétrie de difficulté. Il semble donc que le doigt tende à se comporter de manière équivalente pour 1- et 2-syllabes d'une part et pour 3- et 4-syllabes d'autre part. L'analyse de l'amplitude des gestes de pointé va aussi dans ce sens. Ensuite, l'étude de la relation entre la durée de la production du nom de la cible et celle du geste de pointé montre que ces durées sont proches quand le nom de la cible comporte 2-syllabes. Cette observation rejoint les résultats de Ducey-Kaufmann *et al.* (2007), voir figure 6.1, p. 124 ainsi que les résultats de nos expériences sur les sites de synchronisation.

D'autre part, l'influence de la position de la cible sur le rapport 2 :1 reste difficile à interpréter étant données les variations inter-individuelles assez importantes et la taille réduite de l'échantillon de participants. Il semble néanmoins que ce facteur ait peu d'effet sur le maintien du geste du doigt de 1- à 2-syllabe(s) et le saut de 2- à 3-syllabes alors qu'il pourrait avoir une influence sur la différence entre 3- et 4-syllabes. La tendance serait d'observer une différence entre les deux conditions quand la cible est loin mais pas quand elle est proche. Or, l'augmentation de la distance de la cible induisant une augmentation de la durée du geste on s'attendrait à observer le profil inverse. Ajouté à la tendance des locutrices à confondre les séquences de 3- et 4-syllabes, ce dernier point remet en question la pertinence d'une condition 4-syllabes pour tester l'hypothèse 2 :1.

6.4.2 ... à confirmer ailleurs

En effet, si l'absence de différence entre 3- et 4-syllabes renforce l'hypothèse 2 :1, la similitude entre 1- et 2-syllabe(s) et le saut de 2- à 3-syllabes sont deux observations qui suffisent à sa non-infirmation. Nous envisageons donc de reproduire l'expérience avec un plus grand échantillon de participants et un facteur *nombre*

de syllabes réduit à trois niveaux. Si les résultats observés ici sont reproduits, la question de l'effet de la fréquence des mots utilisés pourra ensuite être évaluée en utilisant des mots de la langue de 1-, 2- ou 3-syllabes fréquents vs. non-fréquents. D'autre part, une des hypothèses forte d' Abry et de ses collaborateurs est que l'*enchâssement* de la syllabe rythmique du babillage (le cadre de la parole) dans le geste de pointé (le cadre du signe) serait une histoire de *pied dans le bras*, c'est-à-dire, se ferait via l'unité rythmique (le pied) de la parole, voir figure 4.4, p. 87). Si tel est le cas, en utilisant notre paradigme, le saut de 2- à 3-syllabes ne devrait pas s'observer quand les mots ont le même nombre de pieds, comme c'est le cas en anglais pour « muse » et « music », isochronie qui devrait s'étendre à « musically ». Ce type d'expérience devrait fournir des résultats intéressants pour un *enchâssement* des unités de la parole dans la mâchoire et dans le bras.

La série d'expériences que nous envisageons devra aussi inclure des conditions de contrôle impliquant des tâches simples afin de délimiter la manière dont le doigt s'adapte à la parole et celle dont la parole s'adapte au doigt avec l'augmentation du nombre de syllabes. Ces études devront aussi investir l'effet de l'apprentissage sur les performances des participants, effet difficile à évaluer ici étant donné le nombre d'essais restreint par condition expérimentale, mais qui pourrait jouer un rôle puisque la tâche n'est pas une tâche habituelle et que les mesures concernent les durées des mouvements.

6.4.3 Des éléments pour la coordination main-mâchoire

Enfin, ces données fournissent de nouveaux éléments pour une meilleure compréhension de la coordination entre la mâchoire et le geste de pointé et réitère certains points mis en évidence par les expériences sur les sites de synchronisation avec :

- Une réponse de la mâchoire qui débute après celle du pointé. Bien que cette tendance soit moins forte en 3- et 4-syllabes qu'en 1- et 2-syllabes, elle domine toutes les conditions ;
- Une adaptation de la parole, qui avance sa réponse avec l'augmentation du nombre de syllabes. La comparaison des temps de réponse de la mâchoire obtenus dans la tâche principale à ceux dans une tâche de parole seule devrait permettre de mieux délimiter cette adaptation.

Contrairement à l'expérience sur les sites de synchronisation, qui utilisait toujours des bisyllabes, les données de cette expérience montrent que le mouvement du doigt n'est pas hermétique à l'augmentation du nombre de syllabes et que l'absence d'effet de la position de l'accent observée dans le chapitre précédent sur le geste de pointé pourrait résulter du fait que les énoncés comportaient toujours deux syllabes. Une augmentation du nombre de syllabes, au-delà de deux syllabes, pourraient donc modifier le geste de pointé (reporter son initialisation et augmenter sa durée et son amplitude), ce qui remettrait en question l'idée selon laquelle le geste du doigt ne serait pas ou peu affecté par la présence de la parole et par extension,

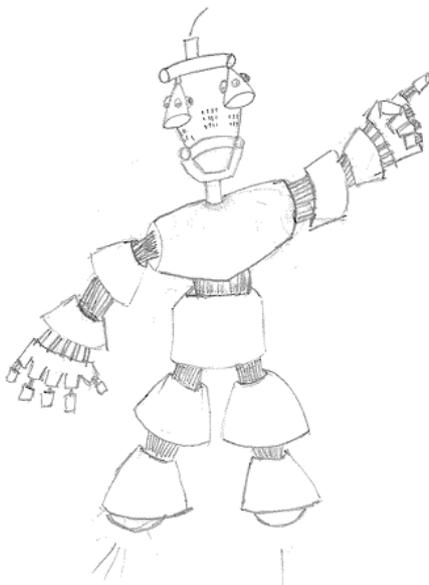
les niveaux de modularité vs. interactivité entre les deux systèmes établis par Levelt *et al.* (1985). Cette remarque renouvelle l'idée d'une tâche de désignation de cible faisant varier le nombre de syllabes de 1- à 3- ou 4-syllabes évoquée dans la discussion du chapitre précédent.

6.5 CONCLUSION

Les résultats obtenus dans cette première expérience vont dans le sens de l'hypothèse selon laquelle deux syllabes pourraient être réalisées sur un geste de pointé. Deux syllabes et non pas deux cycles de mâchoire, la durée des bisyllabes correspondant ici au temps écoulé entre, approximativement la première consonne et la dernière voyelle, soit, un cycle de mâchoire (ouverture-fermeture) suivi d'un geste d'ouverture et non deux cycles complets. Le rapport 2 :1 tel que présenté par Abry et Ducey-Kaufmann est bien un rapport de durée de strokes et non pas de cycles (et bien qu'ils en parlent en termes de fréquences, voir figure 4.4, p.87). En effet le geste de pointé n'est pas un mouvement cyclique, pas même dans notre expérience car entre le geste d'aller et de retour, il marque une pause pour désigner la cible. Toujours est-il que nos données montrent qu'il y a bien un rapport de 2 :1, même si ce n'est pas un rapport de fréquences à proprement parlé, ce qui réitère les questions du chapitre précédent : pourquoi la parole ne commence pas en même temps que la main pour faire les deux syllabes sur le geste de pointé quand l'accent est sur la deuxième syllabe ? Pourquoi le bras part avant la parole même quand le nom de la cible a plus de deux syllabes ? A supposer que le bras démarre *aussi vite que possible*, sur quoi se base la parole pour démarrer ?

Troisième partie

Des simulations, un cadre modélisateur et conclusion



Fin et suite

Ce dernier chapitre présente différents travaux utilisant la modélisation et la simulation computationnelle pour évaluer les théories sur les origines et l'évolution du langage et de ses formes. Ici encore, la richesse du domaine fait que cette présentation ne se veut pas exhaustive. Son but est essentiellement de rendre compte de la part donnée au sensori-moteur dans les différents modèles. Ce travail de bibliographie - bien qu'incomplet - a pour objectif de repositionner le travail expérimental réalisé au cours de cette thèse dans la problématique qui l'a motivé et qui consiste à aborder le langage comme un système complexe.

Cette présentation suit le gradient d'incorporation des modèles et des simulations évoqués : depuis les modèles qui ne considèrent pas les systèmes sensori-moteurs, en passant par l'optimisation de contraintes, jusqu'à la robotique de la parole qui, se heurtant au problème des systèmes devant apprendre par eux-mêmes afin de pouvoir gérer l'imprévisible et la flexibilité, ne peut que chercher les moyens d'inclure des modules de perception et de production réalistes.

Nous introduisons ensuite la démarche modélisatrice globale développée dans notre laboratoire ces 15 dernières années, puisque c'est de cette démarche que découlent nos travaux expérimentaux. Le système complexe qu'elle étudie est constitué de deux agents sensori-moteurs en interaction dans un monde d'objets. Cette démarche propose d'ancrer la compositionnalité dans le contrôle moteur sériel des articulateurs ; l'attention partagée dans la coordination main-bouche et de considérer la perception pour le contrôle de l'action. Nous conclurons sur les apports et les perspectives de ce travail de recherche.

Vers une définition d'une approche *système complexe* du langage

Le travail expérimental rapporté ici s'inscrit dans le projet de recherche « Pati papa ? Modélisation de l'émergence d'un langage articulé dans une société d'agents sensori-moteurs en interaction »¹. L'objectif global de ce projet est de faire progresser la compréhension du langage et plus particulièrement celle de ses origines en étudiant ce dernier comme un système complexe, au sens défini en introduction. Dans cette démarche, l'expérimentation et la simulation se complètent et s'alimentent mutuellement. Durant cette thèse, nous avons emprunté la voie de l'expérimentation. Nous consacrons cette dernière section à la simulation en suivant l'axe de l'incorporation (ou du réalisme sensori-moteur). Nous commencerons par présenter des exemples de modèles qui n'intègrent pas les systèmes sensori-moteurs et qui s'attachent à déterminer les contraintes de la communication ayant pu amener à la syntaxe. Nous introduirons ensuite des exemples de modèles cherchant à dériver la forme de la substance, ou le langage de la parole, et qui s'étendent, sur le plan méthodologique, de l'optimisation thermodynamique à la robotique de la parole. Nous verrons notamment comment nos résultats expérimentaux s'intègrent au cadre « Pati papa » qui *limite* la complexité à un système à deux agents sensori-moteurs en interaction dans un monde d'objets.

7.1 LA QUESTION DE LA COMPOSITIONNALITÉ ET DE LA SYNTAXE

Les modèles informatiques développés pour étudier le langage s'intéressent à différents aspects du langage. Certains de ces modèles utilisent la simulation informatique afin d'expliquer l'apparition de la compositionnalité et de la syntaxe par un processus de sélection naturelle. Bien que la syntaxe ne soit pas au centre

¹Projet soutenu par l'ACI Systèmes Complexes en Sciences Humaines et Sociales, regroupant l'Institut de la Communication Parlée de Grenoble (maintenant Département Parole-Cognition du nouveau laboratoire Gipsa), le Laboratoire Cognition et Développement de Paris, le Laboratoire Lidilem de Grenoble, le Laboratoire de Psychologie et NeuroCognition de Grenoble, l'équipe Laplace du laboratoire Gravir, maintenant LIG, de Grenoble, et l'Institut Max Planck, de Munich.

de cette thèse, ces travaux sont incontournables dans la littérature sur les origines du langage.

7.1.1 Démarche générale

Selon Christiansen et Kirby (2003), les approches génératives, que l'on pourrait appeler « chomskiennes », négligent le rôle de l'apprentissage dans le langage. Or, les comportements linguistiques ne sont pas déterminés, mais émergeraient du processus de l'utilisation répétée et de l'acquisition de génération en génération. De plus, la modélisation de la cognition située et des systèmes dynamiques dans la nature en utilisant les techniques de l'informatique permet de comprendre les relations entre les composants d'un système complexe et le produit émergent de leurs interactions. D'après Christiansen et Kirby, les modèles de simulation sont généralement constitués d' :

1. Une population d'agents ;
2. Un espace de signaux possibles (chaînes de caractères) ;
3. Un espace de significations possibles (représentation structurée avec des significations plus proches que d'autres) ;
4. Un modèle de production déterminant comment, en contact avec un concept, l'agent utilise sa connaissance pour produire un signal ;
5. Un modèle d'apprentissage définissant comment un agent individuel acquiert sa connaissance de la langue en observant les paires sens-signal produites par les autres agents.

7.1.2 Nowak et l'émergence de la compositionnalité

En utilisant ce type de modèle (mais en s'intéressant plus aux processus de l'évolution qu'à l'apprentissage), Nowak et ses collaborateurs ont décrit la progression des signaux aux mots puis des mots à la syntaxe par un mécanisme de sélection naturelle (Nowak, 2000; Nowak *et al.*, 2000, 2001; Komarova et Nowak, 2003). Ces travaux montrent notamment comment l'augmentation du nombre de concepts à communiquer rend, à partir d'un certain seuil, la communication compositionnelle (par combinaison de signaux) plus avantageuse que la communication holistique (un signal = une signification). La restriction du nombre de signaux au profit de leur combinaison limite en effet le nombre d'erreurs d'identification de ces signaux. De la même manière, une communication sans syntaxe impliquerait de mémoriser un mot pour chaque événement. Comme il existerait un seuil du nombre d'objets à désigner pour passer de l'utilisation de signaux holistiques associés à des objets à des signaux compositionnelles combinés pour désigner une infinité d'objets, il existerait un seuil syntaxique au-delà duquel laquelle la combinaison des mots en phrase deviendrait plus avantageuse qu'une communication mot-chose, mot-événement. De manière peut-être un peu simpliste, si les animaux en sont restés à une communication par signaux holistiques, associant au mieux,

comme les singes vervets trois cris à trois événements (alerte de la présence d'un prédateur volant, rampant ou marchant, voir Cheney et Seyfarth 1990), c'est parce qu'*ils n'ont pas assez de choses à dire*. Les travaux de Nowak permettent d'entrevoir l'apparition du langage dans la continuité du monde animal. Pour en revenir au débat Pinker-Jakobson/Hauser-Fitch-Chomsky autour de la question de l'évolution du langage (voir chapitre 1), on notera que Pinker (2003) utilise ces résultats comme des éléments supplémentaires pour défendre une évolution du langage par sélection naturelle. De plus, selon Pinker, l'approche de Nowak contribue aux *bonnes théories* de l'adaptation, celles qui expliquent en faisant appel à des résultats établis de manière indépendante, par les mathématiques et la simulation.

Mais à l'image des théories des origines du langage auxquelles ils se rapportent, ces travaux n'impliquent pas le système sensori-moteur dans les modèles utilisés. Dès le départ, les agents ont un ensemble de mots ou de sons. Le modèle n'explique pas d'où viennent ces signaux, ni comment le répertoire s'est formé. Si on reprend notre schéma définissant trois axes pour une étude incorporée du langage et de la parole (voir figure 1.3, p.13), on peut dire que ces modèles se positionnent essentiellement sur le troisième axe, celui de l'évolution et qu'ils se centrent sur la syntaxe. D'autres approches, auxquelles est liée plus directement notre démarche expérimentale, visent à dériver les formes phonologiques des langues sur la base de contraintes substantielles.

7.2 DÉRIVER LA FORME DE LA SUBSTANCE

Les approches expliquant l'émergence des formes phonologiques sur la base de contraintes sensori-motrices utilisent les données sur les analyses des lexiques des langues du monde (comme la base de données UPSID²) et les hypothèses explicatives en termes de contraintes afin d'élaborer des modèles informatiques permettant de confirmer ou d'infirmer ces explications. Nous présentons ici certaines de ces approches montrant la progression d'une simulation globale, basée sur l'optimisation d'un système en réponse à des contraintes, à des simulations locales dans lesquelles l'organisation du langage émerge via des interactions simples entre les unités du système.

7.2.1 Optimisation de contraintes et prédiction des systèmes vocaliques et syllabiques

L'optimisation de contraintes a été utilisée pour expliquer l'organisation des systèmes vocaliques et syllabiques des langues du monde.

² La base UPSID (UCLA Phonological Segment Inventory Database) élaborée par Maddiesson est un échantillon des systèmes phonologiques (consonnes, voyelles) actuels des langues du monde. Dix-sept familles plus trois langues isolées, réparties sur les cinq continents y sont représentées.

Prédiction des systèmes vocaliques

Selon Liljencrants et Lindblom (1972), la préférence pour les voyelles [i, a, u], caractéristique des systèmes vocaliques des langues du monde résulterait du fait que ces voyelles se situeraient aux extrémités de l'espace physiologique des voyelles possibles et donc, maximiseraient la saillance perceptive en diminuant la probabilité de confusion. Ainsi, les sons seraient sélectionnés pour un contraste perceptif maximal - ou au moins, suffisant (Lindblom, 1990b). Afin d'évaluer cette hypothèse Liljencrants et Lindblom (1972) quantifient le contraste perceptif par la distance entre deux voyelles dans l'espace des deux premiers formants. Ils modélisent les voyelles comme des particules qui se repoussent et migrent aux extrémités de l'espace acoustique. Leur simulation permet de prédire les systèmes vocaliques de 3 à 6 voyelles, au-delà de 6 voyelles, leur modèle fait des erreurs de prédictions, montrant que la dispersion ne serait pas le seul facteur en jeu dans la sélection des systèmes vocaliques. Au critère de dispersion qui définit la saillance perceptive de manière relative, par un éloignement des voyelles dans l'espace des deux premiers formants, Schwartz *et al.* (1997) ajoutent un critère de stabilité interne, la *focalisation*, correspondant à la préférence pour les voyelles avec deux formants qui convergent ($F1 = F2$ ou $F2 = F3$). Schwartz *et al.* évaluent leur théorie de la dispersion-focalisation en cherchant, pour un système de n voyelles parmi les N possibles, celles qui satisfont au mieux les deux principes (via la minimisation d'une fonction énergétique).

Prédiction des systèmes syllabiques

Un exemple de modèle de prédiction des systèmes syllabiques élaboré dans le cadre de l'optimisation de contraintes est le modèle ESS (Emergent Syllable Systems), par Redford *et al.* (2001). Ce modèle utilise un algorithme génétique pour sélectionner, au cours des générations, un répertoire de mots répondant au mieux à des contraintes perceptives et articulatoires. Les auteurs distinguent les contraintes qui se situent au niveau du vocabulaire, qui favorisent les mots courts, les mots différents et la réutilisation des syllabes, et les contraintes au niveau du mot, perceptives (saillance) et mécaniques (économie du coût articulatoire). Ces contraintes et leurs interactions sont formalisées afin de décrire le problème par une fonction dont l'optimisation simule l'émergence des formes syllabiques. Ce système permet notamment de prédire la dominance des syllabes CV et le fait que dans les systèmes syllabiques, les consonnes finales et les voyelles initiales soient défavorisées (cf. la syllabe VC est moins fréquente que la syllabe CV).

7.2.2 De l'optimisation à la robotique en passant par l'auto-organisation

Les deux exemples donnés précédemment font des hypothèses sur les contraintes perceptives et articulatoires sans impliquer directement les systèmes moteurs et sensoriels dans la simulation. S'ils sont globaux en termes de règles de résolution de problèmes, ils sont très locaux sur le plan des contraintes qu'ils considèrent.

Ces vingt dernières années, la recherche visant à dériver les formes phonologiques d'une interaction de contraintes substantielles a progressivement évolué vers la robotique de la parole, cherchant à faire émerger le langage dans des sociétés d'agents munis de systèmes sensori-moteurs de plus en plus évolués et interagissant selon des mécanismes communicatifs réalistes.

Premier pas vers une incorporation de la simulation : introduction des modèles de production et de perception

Lindblom *et al.* (1984) vont progresser de l'optimisation de contraintes à l'émergence, en élaborant un modèle pour la prédiction des systèmes syllabiques basé sur des contraintes purement physiologiques : les lois universelles gouvernant les relations entre la production et la perception des formants. Le modèle élaboré utilise un modèle articulatoire numérique pour définir les configurations de mouvements possibles disponibles pour la formation des systèmes phonétiques. L'espace articulatoire est ensuite traduit en espace acoustique : quel son correspond à une configuration articulatoire donnée ? Les types d'articulations traités sont ceux impliquant une transition d'un état fermé à un état ouvert (événement proche de la syllabe CV). Dans le modèle articulatoire, l'espace définissant l'ensemble des états fermés possibles est continu, de même que celui des états ouverts. Ainsi, l'espace des CV possibles, reliant un point de l'espace des états fermés (consonnes plosives) à un point de l'espace des états ouverts (voyelles) est lui aussi continu. Potentiellement, une infinité de trajectoires est possible et donc, une infinité de couples CV. La simulation va introduire des contraintes perceptives imposant une maximisation de la distance perceptive entre les couples CV et de la saillance perceptive entre la consonne et la voyelle. Ces contraintes perceptives vont interagir avec des contraintes articulatoires, favorisant les configurations articulatoires discriminantes et les articulations les moins extrêmes. Les simulations montrent que, sous l'influence de ces différentes contraintes, l'espace des formes CV possible se structure d'une manière comparable, selon Lindblom *et al.*, au phénomène d'auto-organisation des systèmes biologiques où la structuration émerge de l'interaction entre sous-systèmes, les sous-systèmes étant ici les formes CV.

Deuxième pas vers une incorporation de la simulation : utilisation des sociétés d'agents

Le niveau suivant dans l'implication des systèmes sensori-moteurs évalue dans quelle mesure les systèmes syllabiques qui émergent de l'interaction entre agents communicants peuvent ressembler à ceux caractérisant les langues du monde. Les modèles qui tentent de répondre à cette question n'ont plus l'optimisation de contraintes pour fil directeur mais définissent des règles locales d'interaction entre agents (ce qui leur vaut le nom d'*approches locales* par opposition aux *approches globales* exposées ci-dessus, voir Berrah, 1998). Parmi ces modèles, on peut distinguer ceux utilisant des agents avec des systèmes de production et de perception plus ou moins réalistes, que nous présenterons dans la section suivante et ceux qui

utilisent les mêmes mécanismes d'interaction entre agents mais des agents démunis d'*organes* de production et de perception. Par exemple de Boer (2000) cherche à prédire les systèmes syllabiques en utilisant des agents dotés d'une liste de mots, pointeurs sur un répertoire de phonèmes (une voyelle et huit consonnes). Chacun des agents a pour objectif d'imiter au mieux les productions de ses congénères. Le déroulement est le suivant : un *initiateur* produit un mot, un *imitateur* en analyse les phonèmes et cherche le mot de son répertoire le plus proche. Il produit alors le mot trouvé, l'*initiateur* analyse cette production. S'il trouve le même mot que celui initialement produit, le jeu est réussi et il communique cette réussite à l'imitateur par un retour non-linguistique. Les répertoires de mots sont évalués par attribution d'un score à chaque mot, fonction de la réussite et du nombre de jeux d'imitation dont il a fait l'objet. Si le jeu échoue, un agent peut ajouter un nouveau mot à son répertoire (s'il dispose des phonèmes le constituant) ou un nouveau phonème choisi en fonction du son perçu et du répertoire disponible. Les simulations aboutissent au partage des systèmes sonores entre agents. On peut assimiler ces formes de simulation à une optimisation indirecte ou *implicite*. Cependant, du fait des coûts de calcul importants des modèles computationnels de production et de perception réalistes, le modèle de de Boer se fonde sur des représentations abstraites en termes de traits distinctifs et n'inclut donc pas les contraintes physiologiques.

Le modèle de Glotin (1995) élaboré à l'ICP pour étudier l'émergence des systèmes vocaliques dans une société d'agents en interaction intègre une description acoustique et articulatoire des voyelles mais la transformation de l'acoustique en articulatoire étant très coûteuse en temps de calcul, les travaux associés de Berrah (1998), comme ceux de de Boer, ne garderont que la représentation acoustique.

Si les simulations décrites précédemment en sont indissociables, la robotique de la parole commence vraiment quand on se pose le problème de l'encorporation des agents dans un corps et dans le monde. Un premier ensemble de recherches réalisées en robotique de la parole est celui du laboratoire Sony Computer Science Laboratory à Paris sous la direction de Luc Steels. Ces recherches ont notamment défini des modalités d'interaction entre agents. Elles ont aussi proposé un modèle d'émergence de la parole purement auto-organisateur.

Troisième pas vers une encorporation de la simulation : la robotique de la parole

Jeux de langage et *grounding* sensori-moteur. Les jeux d'imitation utilisés par Glotin (1995); Berrah (1998); de Boer (2000) font partie de ce que Steels (2001) appelle des *jeux de langage*. La robotique de Steels est centrée sur le problème du dialogue entre humains et robots. Selon Steels (2001), les jeux de langage doivent intégrer tous les types d'activités requises pour le dialogue : la vision, le geste, la reconnaissance de formes, l'analyse de la parole, la synthèse et la conceptualisation. De plus, les agents doivent non seulement réussir le jeu mais aussi disposer de mécanismes d'apprentissage pour comprendre de nouveaux concepts. La figure 7.1



FIG. 7.1 – Têtes parlantes et jeux de langage.

Dans l'expérience des têtes parlantes, deux agents-robots munis d'une caméra orientable *pointent* (par orientation de la caméra) des formes géométriques de couleur sur un tableau blanc. Le *jeu de langage* consiste à élaborer un système conceptuel-commun, avec un mot pour chaque objet. Par exemple, le *locuteur* repère un objet (cf. carré bleu) et produit alors le mot associé à la représentation qui correspond le mieux à cet objet dans sa base conceptuelle. Si l'*auditeur* ne reconnaît pas le mot, il signale son incompréhension et le locuteur pointe alors l'objet. Quand les deux *locuteurs tombent d'accord* pour une association signal-référent, ils renforcent le score de cette association, d'après Steels 2001.

donne un exemple de jeu de langage et de mise en application avec les têtes parlantes. Mais Steels insiste sur le fait que le dialogue humain-robot requiert de relier les représentations symboliques aux traitements sensori-moteurs, ce qui recouvre le problème du *grounding* ou ancrage.

Le langage est inextricable des systèmes sensori-moteurs et il n'est pas fortuit que Steels, spécialiste en IA et en robotique, insiste sur ce point. Sans ce lien entre les systèmes moteurs et perceptifs, les robots resteront au stade d'un apprentissage et d'une adaptation très limitée. Nous rappellerons ici que par ses échecs, l'IA en général, et la robotique en particulier, ont beaucoup contribué à l'incorporation des fonctions cognitives. On pourrait dire que ce lien entre couche conceptuelle et sensori-motrice, au cœur des sciences cognitives actuelles, est aussi le point sensible de la robotique, exacerbé en robotique de la parole par les propriétés particulièrement complexes du langage humain. Cette problématique est aussi celle du Laboratoire d'Algorithmes et Systèmes d'Apprentissage de l'Ecole Polytechnique Fédérale de Lausanne et notamment des travaux d'Aude Billard sur l'apprentissage gestuel par imitation chez les robots (Calinon et Billard, 2007; Billard, 2002).

Auto-organisation. Un autre axe de recherche du laboratoire Sony est l'auto-organisation de la parole, au cœur de la thèse de Pierre-Yves Oudeyer (2003). Un premier modèle proposé par Oudeyer (2001) pour simuler l'émergence des systèmes syllabique utilise le jeu d'imitation de de Boer (2000) décrit ci-dessus mais avec des agents pourvus de systèmes de production et de perception plus réalistes. Ces agents sont munis d'un module de production et d'un module de perception. Le module de production comprend un modèle physique de conduit vocal et un système de contrôle qui spécifie les paramètres articulatoires pour donner forme au conduit. Le module de perception comprend un modèle de l'oreille interne. Enfin, les agents ont un module *cerveau* avec deux mémoires, une mémoire associative acoustique-articulatoire et une mémoire conceptuelle. Le modèle d'Oudeyer est

donc *plus incorporé* que celui de de Boer. Mais dans sa thèse, Oudeyer va plus loin. Il veut faire de l'*objet répertoire* un résultat de la simulation et non pas une donnée du modèle. Pour cela, il utilise des agents qui ne s'imitent pas, ne reproduisent pas les vocalisations qu'ils entendent et ne les stockent pas en mémoire pour les reproduire plus tard. Ces agents (différents de ceux de Oudeyer, 2001) sont dotés de cartes neuronales correspondants aux modalités perceptives et motrices connectées entre elles. Les connexions sont renforcées (ou inhibées) selon la loi de Hebb chaque fois qu'un son est perçu. L'activation des neurones articulatoires permet de produire les vocalisations. L'*interaction* entre agents est limitée à entendre les vocalisations des autres (et les siennes).

Les agents ne sont donc dotés d'aucune capacité sociale, à l'extrême même, ils ne font pas de différence entre les sons produits par les autres et leurs propres sons. Si on définit la communication comme étant l'émission d'un signal par un agent qui vise à modifier l'état interne ou le comportement d'un autre agent, alors évidemment les agents du système artificiel ne communiquent pas. Il n'y a pas de pression qui pousserait les agents à former des répertoires de sons contrastés les uns avec les autres.(...) Les simulations ont montré que si nous placions ces agents ensemble dans un environnement dans lequel ils soient à leur insu obligés à interagir les uns les autres (par l'intermédiaire de leurs vocalisations), alors l'ensemble de leurs vocalisations s'auto-organisent, Oudeyer, 2003, p.202

Ces travaux ne se situent plus vraiment dans le *grounding* de Steels. De plus, la part auto-organisatrice de la parole n'a pas de sens en dehors des situations de communication et des contraintes liées à ces situations. Les vocalisations dans le vide des agents d'Oudeyer sont inspirées du babillage moteur des bébés humains qui, selon Oudeyer, « *s'entraînent* » à activer leur cartes neuronales motrices et sensorielles. Mais que deviennent ces actions sans l'imitation ? Si le langage humain est complexe et qu'il ne peut être étudié que par des modèles simplifiés, faut-il en arriver à des niveaux de modélisation qui dénaturent complètement le problème pour le comprendre ? Faut-il faire table-rase pour comprendre l'émergence du langage ? Nous avons vu dans ce rapport que le langage avait dû se mettre en place au cours de l'évolution sur la base de capacités cognitives génériques, comme en témoignent aussi les découvertes récentes des circuits cérébraux du langage. Une réelle table-rase pour l'émergence du langage est donc inconcevable car parler - et même babiller - c'est bien plus que d'agiter sa mâchoire pour vocaliser dans le vide, il faudrait donc expliquer l'émergence de chacune des fonctions requises au langage... En d'autres termes, une simulation réaliste de l'émergence du langage ne peut, à l'heure actuelle, qu'inclure des hypothèses sur les fonctions cognitives pré-existantes au langage.

Une autre robotique de la parole, plus en accord avec le *grounding* de Steels est développée à l'ICP en collaboration avec l'équipe Laplace du laboratoire LIG

de Grenoble. Elle vise l'incorporation des agents dans le corps, dans l'interaction mais aussi dans le monde. Sur le plan méthodologique, elle utilise les algorithmes de l'apprentissage bayésien. C'est le cadre adopté par le projet « Pati papa », le *grounding* de cette thèse.

7.3 DES AGENTS INCORPORÉS EN INTERACTION DANS UN MONDE D'OBJETS EN ROUTE VERS « PATI-PAPA ? »

7.3.1 Définition du système complexe étudié

Le langage se développe chez l'être humain sur la base d'un système de communication préexistant : l'Homme ne s'est pas mis à articuler des sons et à construire un système phonologique *au cas où un jour il en aurait besoin pour communiquer*. Les bébés privés d'environnement humain ne babillent pas, et pire, en dehors de toute stimulation sociale (ou animale), ils dépérissent. Schwartz (2007) définit le système complexe étudié dans le cadre du projet « Pati papa », et plus généralement dans la robotique cognitive de la parole développée en collaboration avec l'équipe Laplace, par deux interlocuteurs en interaction. Cette interaction se fait via une boucle d'échanges perceptuo-moteurs impliquant les signaux vocaux de la communication et un système d'attention partagée impliquant la coordination des systèmes visuel, brachiomanuel et orofacial. Exposé de manière simplifiée, l'objectif est d'expliquer en termes de contraintes substantielles pourquoi le bébé dit «Pati papa ?» et pas «Tipa papapapa ?» ou « Titi papapapa » (et en n'oubliant pas que, *évidemment, les enfants disent bien autre chose que « Pati papa » (à supposer qu'ils le disent !)*, Schwartz 2007). Pour cela, le schéma conceptuel proposé ancre le langage dans trois éléments clés, qui deviennent de ce fait des enjeux de recherche, à savoir : ① la compositionnalité via le contrôle moteur sériel de l'action ; ② l'attention partagée via la coordination main-bouche dans la deixis et ③ les relations perception-action.

Compositionnalité. Le premier aspect à étudier, c'est le contrôle sériel de la mâchoire qui aurait permis le développement de la compositionnalité. En effet, si le modèle de Nowak vu précédemment fait émerger la compositionnalité de la nécessité d'avoir plus de choses à dire, il ne considère pas le *grounding* de cette compositionnalité qui ne peut être que dans le séquençage et l'ordonnement de la gestualité. D'ailleurs, comme nous l'avons vu au chapitres 2 et 4, le *gène du langage*, FOXP2 et l'*aire de production du langage*, Broca, seraient tous deux impliqués dans le contrôle sériel des articulateurs oro-faciaux ou brachiomanuels. C'est ici que se positionnent notre hypothèse et notre étude sur l'effet Labial-Coronal, dans le cadre de la théorie Frame-then-Content (Cadre-puis-Contenu). Nos résultats montrent que les propriétés coordinatives du système mâchoire-langue-lèvre favoriseraient les formes labiale-coronale (/pati/) plutôt coronale-labiale (/tipa/). De plus, une

fois le contrôle des articulateurs portés (le contenu) et de leur coordination en place (ou presque), les formes intersyllabiques variant les consonnes (/pati/) seraient favorisées par rapport aux formes répétant la même consonne qui caractérisent les premiers mots sous la dominance du frame (/papa/). Ainsi, les phénomènes d'anticipation et donc de coarticulation rendraient certaines formes intersyllabiques plus optimales que d'autres pour le système articulo-latoire. Est-ce que la stabilité articulo-latoire et l'économie des mouvements de la mâchoire permettront à la simulation de faire émerger /pati/ plutôt que /papa/ et que /tipa/ ?

Attention partagée. Le langage suppose une capacité de partage de l'attention vers un espace donné du monde physique ou abstrait. Comme pour la compositionnalité, pour que le système d'attention partagée se mette en place, il faut une gestuelle pour le porter et un système cognitif permettant de percevoir et de comprendre cette gestuelle. Si on ne pouvait pas bouger du tout, y compris le regard, on serait bien en mal de montrer ce dont on parle ou veut parler. Or, comme nous l'avons vu au chapitre 4, montrer à l'autre ce que je veux qu'il me donne (pointage impératif) ou ce dont je veux parler (pointage déclaratif) est une étape déterminante de l'acquisition du langage. De plus l'association de la parole au geste de pointé constitue une première forme de syntaxe qui va porter l'enfant vers la phrase de deux mots (cf. « Pati papa »). Or, le fait qu'on trouve chez les singes vervets des cris d'alerte qui réfèrent un prédateur et, chez les grands singes, des gestes de pointage avec parfois des vocalisations, pourrait constituer un ancrage phylogénétique de l'attention partagée via l'association « Vocalise-to-Localize ». Les travaux que nous avons présentés aux chapitres 5 et 6 étudiant la coordination entre la main et la parole dans la deixis, montrent que l'arrivée du geste de pointé tend à se rapprocher de la partie accentuée de la parole et que le geste de pointage et la parole pourraient être liés par une relation de deux syllabes sur un geste de pointé. Cette relation a été interprétée par Abry *et al.* (2004); Ducey-Kaufmann (2007) comme un fondement possible d'une progression d'un babillage, activité rythmique sans découpage, à une parole discrète de mots bisyllabiques. Est-ce que cette relation permettra à la simulation de faire émerger /papa/ plutôt que /papapapapa/ ?

Lien entre action et perception. Enfin, pour que signaux et communication il y ait, il faut un système perceptuo-moteur qui lie perception et action. Comme nous l'avons vu au chapitre 2, de plus en plus d'éléments lient l'action à la perception dans la parole et la théorie de la perception pour le contrôle de l'action permet de lier les deux systèmes tout en conservant leur nature respective. L'approche bayésienne développée par Pierre Bessière et ses collaborateurs est un outil particulièrement adapté pour que le système apprenne à lier perception et action en « expérimentant le monde par lui-même et par imitation » (Bessière *et al.*, 1998a,b; Lebeltel *et al.*, 2004). Cette méthode a été utilisée dans la thèse de Serkhane (2005) pour explorer les processus

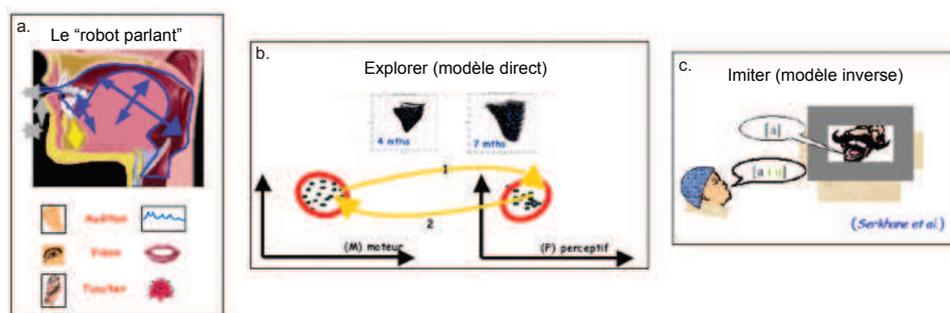


FIG. 7.2 – Vers une robotique de la parole, exploration-imitation chez un androïde bébé. A partir d'un modèle articulatoire du conduit vocal à 7 degrés de liberté, et disposant de capteurs auditif, visuel et tactile, simulé sur ordinateur (a), on peut définir un processus d'exploration vocale progressive (voir les triangles acoustiques produits par le bébé, s'élargissant entre 4 mois et 7 mois) qui permet d'apprendre des correspondances entre gestes articulatoires et sons et de construire un « modèle direct » (b). L'imitation consiste ensuite à tirer selon un « modèle inverse » des gestes moteurs à partir des percepts fournis (c), d'après Serkhane (2005); Schwartz (2007).

d'apprentissage par imitation et exploration chez un *androïde bébé* (voir figure 7.2).

Ainsi dans ce schéma conceptuel, le langage, « poussé » par des coordinations entre main, bouche, oreille et œil, apparaît potentiellement charpenté par des coordinations entre systèmes sensoriels et moteurs (Schwartz, 2007). L'étape suivante est de trouver le bon formalisme pour intégrer ces éléments à un modèle computationnel afin de voir dans quelle mesure, le langage tel qu'on le connaît peut émerger d'un tel *grounding* ou ancrage substantiel. Le problème n'est pas simple car les modèles réalistes du contrôle moteur des articulateurs considérant les propriétés biomécaniques des organes périphériques sont complexes et coûteux en temps de calcul (voir Buchaillard *et al.*, 2007; Payan et Perrier, 1997 pour le cas de la langue, Shiller *et al.*, 2002; Laboissière *et al.*, 1996 pour le cas de la mâchoire). Ces problèmes seront au centre de la thèse que Clément Moulin-Frier débute cette année au Département Parole et Cognition du Gispa-lab (voir Moulin-Frier, 2007, pour des résultats préliminaires).

Conclusion

Cette thèse s’ancre dans une démarche proposant de dériver le langage de la parole en évaluant comment les propriétés des systèmes de perception et d’action qui *matérialisent* le langage peuvent contribuer à mettre en forme les répertoires sonores des langues du monde.

L’idée d’un *grounding* (ancrage) sensori-moteur au langage est une conception qui s’accorde avec les sciences cognitives actuelles et plus particulièrement avec l’*encorporation* (ou l’*Inscription corporelle*) des fonctions cognitives dans les échanges bi-directionnels entre le cerveau et le corps et entre l’organisme [cerveau↔corps] et son environnement. Ainsi, le langage ne serait pas une fonction à laquelle l’évolution aurait, un beau jour, fait une place dans le cerveau mais plutôt une faculté qui se serait installée progressivement dans le cadre du développement de capacités cognitives génériques et de manière à satisfaire des contraintes imposées par ses interfaces¹. Or, ces interfaces incluent les systèmes de perception et d’action qui se développent de manière interdépendante dans l’interaction avec l’environnement. Si la structure du cerveau a un fondement génétique, les connexions inter-neuronales s’établissent dans l’interaction avec l’environnement via les organes d’action et de perception. Cette interaction *imprime* le corps et le monde dans le cerveau et rend ces trois acteurs (cerveau↔corps↔monde) indissociables dans l’émergence des fonctions cognitives.

Dans le but de contribuer à l’établissement du *grounding* sensori-moteur du langage, nous nous sommes intéressés dans cette thèse aux relations entre les propriétés motrices des systèmes orofacial et brachiomanuel et les formes du langage.

8.1 STABILITÉ ARTICULATOIRE ET EFFET LC

Dans une première série d’expériences, nous avons étudié l’*effet Labial-Coronal (LC)* qui réfère à la tendance forte qu’ont les langues du monde à utiliser plus

¹Et c’est finalement (et tardivement) Chomsky lui-même qui le dit : *The language faculty interfaces with other components of the mind/brain. (...) Recent work suggests that language is surprisingly ‘perfect’ in this sense, satisfying in a near-optimal way some rather general conditions imposed at the interface* (Chomsky, 1996, p.29-30).

souvent des séquences composées d'une consonne labiale suivie d'une consonne coronale (séquence L_aC_o , comme /pata/) que leur symétrique (séquence C_oL_a , comme /tapa/). Cette tendance pourrait être, au moins en partie, une réponse à des contraintes motrices. Nous avons proposé que la part motrice de l'effet LC ne serait pas dans le développement sériel du contrôle moteur comme le supposent MacNeilage et Davis (2000), mais dans les propriétés du système coordinatif mâchoire-langue-lèvre. Ces propriétés rendraient les séquences L_aC_o plus stables articulatoirement que leurs symétriques C_oL_a . Afin de tester cette hypothèse, nous avons développé un paradigme expérimental inspiré des travaux de l'équipe de Kelso sur les systèmes dynamiques. Par des mesures acoustiques et articulatoires, nous avons montré que, chez nos locuteurs français, la répétition continue et accélérée de séquences L_aC_o (/pata/) ou C_oL_a (/tapa/) produit un flux «...tapatapatapata...» pouvant évoluer vers un attracteur L_aC_o (/pta/) ou C_oL_a (/tapa/), avec une nette dominance de l'attracteur L_aC_o pour les deux types de séquences. De plus, la progression de la séquence CVCV sur deux cycles de mâchoire (un par syllabe) vers l'attracteur CCV sur un seul cycle, transite généralement par une phase CV'CV (/patá/) de deux syllabes sur un cycle de mâchoire. Cette progression prend naissance dans une asymétrie nette entre le cycle de mâchoire qui porte la consonne labiale et celui qui porte la consonne coronale. Ces deux derniers résultats réinterrogent l'association systématique de la syllabe CV au cycle de la mâchoire généralement admise dans la littérature. Enfin, la progression d'une forme CVCV à une forme CCV ne s'observe pas pour les séquences répétant la même syllabe (comme /papa/). Toujours réalisées sur deux cycles de mâchoire, ces séquences seraient donc plus coûteuses pour le système articulatoire que les séquences alternant des consonnes différentes. Nous avons interprété ce résultat comme un nouvel élément pour expliquer la dominance des formes alternant les constriction par rapport aux formes répétant la même constriction dans les lexiques des langues du monde.

8.2 COORDINATION MAIN-MÂCHOIRE DANS LA DEIXIS

Nous nous sommes ensuite intéressés à un autre point de contact majeur entre la substance motrice et la forme linguistique, à savoir, les relations entre la main et la bouche dans la communication. En effet, de plus en plus d'arguments font de la coordination main-bouche une étape clé de l'acquisition du langage, particulièrement dans les actes de pointage (deixis). Notre attention s'est donc portée sur la coordination entre la main et la parole dans la deixis autour de deux grandes questions : la question des sites de synchronisation des deux systèmes et la question des relations entre leurs fréquences naturelles, qui pourrait porter la synchronisation. Afin d'apporter de premiers éléments de réponse à ces questions, nous avons mis en place un paradigme de pointage de cible avec dénomination, utilisant un système Optotrak pour l'enregistrement des mouvements de la mâchoire et de la main. Ce paradigme permet de manipuler facilement les propriétés segmentales et suprasegmentales du nom de la cible et d'évaluer ainsi les effets de ces variations

sur la coordination mâchoire-main.

8.2.1 Une coordination autour de la partie accentuée de la parole

Dans nos deux premières expériences, nous avons étudié l'effet de la variation de la position de l'accent lexical dans le nom de la cible (mots bisyllabiques accentués sur la première, /pápa/, ou deuxième syllabe, /papá/) sur la coordination mâchoire-main. Nous avons mis en évidence une coordination temporelle forte entre les deux systèmes. Plus précisément, quand l'accent est sur la première syllabe, la parole démarre de façon à synchroniser la partie accentuée de la parole (apex du geste de la mâchoire pour la première syllabe) avec l'apex du geste de pointé. Quand l'accent est sur la deuxième syllabe, la parole avance sa réponse mais cette avance ne suffit pas à faire coïncider la partie accentuée de la parole (apex du geste de la mâchoire pour la deuxième syllabe) avec l'apex du geste de pointé. Cependant, le doigt tend à prolonger son alignement avec la cible pour attendre la réalisation de la composante accentuée de la parole.

8.2.2 Deux syllabes sur un geste de pointé

Dans une troisième étude, nous avons examiné une proposition émise récemment par Abry et al. selon laquelle il existerait une relation entre les fréquences propres des systèmes mandibulaire et brachio-manuel, se traduisant par la possibilité de produire deux cycles de mâchoire sur un geste de pointé. Cette relation a été mise en évidence chez de jeunes enfants par Ducey-Kaufmann (2007). Selon Abry *et al.* (2004) l'harmonie naturelle de deux cycles de mâchoire pour un geste de pointé aboutirait à l'*enchâssement* de deux syllabes pour un mot-référence, ce qui expliquerait la tendance des langues à favoriser les formes bi-syllabiques. Afin de tester ce qu'Abry et Ducey-Kaufmann appellent *leur étonnante hypothèse (astounding hypothesis)* nous avons évalué de manière directe le rapport 2 : 1 chez des locuteurs adultes. La tâche que nous avons utilisée consiste à pointer en dénommant une cible deux fois de suite, les noms de cibles pouvant contenir une, deux, trois ou quatre syllabes. Les résultats de cette expérience, encore préliminaires, montrent que le geste de pointé peut (potentiellement) englober une ou deux syllabe(s) sans adapter sa durée alors qu'un passage à trois ou quatre syllabes implique une augmentation de la durée du geste et donc, une adaptation du geste de pointé à la parole.

Cette série d'expériences originales a soulevé différentes questions. De plus, les relations fortes entre le geste de pointé et la parole encouragent à délimiter plus précisément leurs niveaux d'interaction. Dans la littérature, cette interaction est interprétée soit en terme de compétition pour l'accès à des ressources communes, soit comme le fait que les deux systèmes seraient imbriqués en une seule structure coordinative. La réponse à cette question passera par le développement d'études dans la lignée de Levelt *et al.* (1985) visant à délimiter dans quelles conditions

les deux systèmes s'influencent. Cette question est au cœur de notre projet de recherche postdoctoral que la Fondation Fyssens a choisi de financer.

8.3 SUR LE CHEMIN DE LA FORME...

Cette thèse se termine par ce qui est en réalité son point de départ, à savoir, la question de la modélisation et de la simulation de l'émergence d'un langage articulé dans une société d'agents sensori-moteurs en interaction. Nous avons vu que, si les simulations de l'émergence du langage ont évolué, via le développement de la robotique de la parole, vers des agents sensori-moteurs de plus en plus réalistes, de gros efforts restent à faire en ce sens. Mais dans une démarche ayant pour objectif avoué d'ancrer le langage dans les systèmes sensori-moteurs et de dériver la forme de la substance sensori-motrice, l'expérimentation et la simulation sont deux outils complémentaires qui s'alimentent mutuellement. En effet, la simulation doit formaliser les résultats de l'expérimentation afin d'évaluer le pouvoir explicatif des contraintes mises en évidence. Inversement, l'expérimentation doit valider les résultats de la simulation et découvrir de nouvelles contraintes.

Si le chemin partant de la substance pour rejoindre la forme est long et escarpé, il est aussi peut-être le seul qui permettra d'élaborer des *bonnes théories* de la morphogenèse du langage, celles qui font émerger la forme de facteurs extrinsèques et rompent avec la circularité consistant à expliquer la forme par elle-même (cf. *les formes du langage sont comme elles sont, car le langage est tel qu'il est*). Cette démarche devrait gommer progressivement le spectre du panglossisme voltérien² appliqué au langage et peut-être finir par révéler le vrai visage de notre : *meilleur des langages possibles*.

²Pangloss est un personnage de la nouvelle *Candide* de Voltaire qui répète sans arrêt que les choses ont une finalité, qu'il n'y a pas d'effet sans cause, dans notre *meilleur des mondes possibles*

Quatrième partie

Bibliographie



- ABRY, C., VILAIN, A. et SCHWARTZ, J.-L. (2004). Introduction : Vocalize to localize ? A call for better crosstalk between auditory and visual communication systems researchers : From meerkats to humans. *Interaction Studies*, 5:313–325.
- ALARIO, F.-X., CHAINAY, H., LEHERICY, S. et COHEN, L. (2006). The role of the supplementary motor area SMA in word production. *Brain Research*, 1076(1):129–136.
- ARBIB, M. A. (2003). The evolving mirror system : a neural basis for language readiness. In CHRISTIANSEN, M. et KIRBY, S., éditeurs : *Language Evolution : The States of the Art*, pages 182–200. Oxford University Press.
- ARBIB, M. A. (2005). From monkey-like action recognition to human language : An evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences*, 28(2):105–124.
- ATTINA, V., CATHIARD, M.-A. et BEAUTEMPS, D. (2006). Temporal measures of hand and speech coordination during french cued speech production. In GIBET, S., COURTY, N. et KAMP, J.-F., éditeurs : *Lecture Notes in Artificial Intelligence*, volume 3881, pages 13–24. Springer-Verlag.
- BALLY, C., SECHEHAYE, A. et MAURO, T. D. (1995). *Cours de linguistique générale par Ferdinand de Saussure*. Broché.
- BERRAH, A. (1998). *Evolution d'une Société Artificielle d'Agents de Parole : Un modèle pour l'Emergence des Structures Phonétiques*. Thèse de doctorat, INPG, Sciences Cognitives, ICP, Grenoble.
- BESSIÈRE, P., DEDIEU, E., LEBELTEL, O., MAZER, E. et MEKHNACHA, K. (1998a). Interprétation ou description (I) : Proposition pour une théorie probabiliste des systèmes cognitifs sensori-moteurs. *Intellectica*, 26-27:257–311.
- BESSIÈRE, P., DEDIEU, E., LEBELTEL, O., MAZER, E. et MEKHNACHA, K. (1998b). Interprétation ou description (II) : Fondements mathématiques de l'approche F+D. *Intellectica*, 26-27:313–336.
- BICKERTON, D. (2005). Beyond the mirror neuron - the smoke neuron ? *Brain and Behavioral Sciences*, 28(2):126.

- BILLARD, A. (2002). Imitation : a means to enhance learning of a synthetic proto-language in an autonomous robot. *In Imitation in Animals and Artifacts*, pages 281–311. MIT Press, K. Dautenhahn and C.L. Nehaniv édition.
- BOE, L.-J., HEIM, J.-L., ABRY, C. et BADIN, P. (2004). Neandertal vocal tract : Which potential for vowel acoustics ? *Interaction Studies*, 5(21):409–429.
- BOGNAR, E. (1983). Directions de recherches phonétiques consacrées à la mâchoire. *Bulletin de l'Institut de Phonétique de Grenoble*, 12:1–45.
- BOË, L.-J. (1997). Sciences phonétiques et relations forme/substance. *Histoire, épistémologie, langage*, 19:5–25.
- BOË, L.-J., MAEDA, S. et HEIM, J.-L. (1999). Neanderthal man was not morphologically handicapped for speech. *Evolution of Communication*, 77:3–49.
- BROWMAN, C. P. et GOLDSTEIN, L. (1989). Articulatory gestures as phonological units. *Phonology*, 6:201–251.
- BUCHAILLARD, S., BRIX, M. et and Y. PAYAN, P. P. (2007). Simulations of the consequences of tongue surgery on tongue mobility : Implications for speech production in post-surgery conditions. *International Journal of Medical Robotics and Computer Assisted Surgery*, 3:252–261.
- BUTTERWORTH, G. (2003). Pointing is the royal road to language for babies. *In KITA, S., éditeur : Pointing : Where Language, Culture, and Cognition meet*, pages 9–33. Lawrence Erlbaum Associates.
- BYRD, D. (1996). Influences on articulatory timing in consonant sequences. *Journal of Phonetic*, 24:209–244.
- CALINON, S. et BILLARD, A. (2007). Learning of gestures by imitation in a humanoid robot. *In Imitation and Social Learning in Robots, Humans and Animals : Behavioural, Social and Communicative Dimensions*, pages 153–177. Cambridge University Press, K. Dautenhahn and C.L. Nehaniv édition.
- CAMPBELL, N. (1992). *Multi-level timing in speech*. Thèse de doctorat, University of Sussex, U.K.
- CASTIELLO, U., PAULIGNAN, Y. et JEANNEROD, M. (1991). Temporal dissociation of motor responses and subjective awareness. A study in normal subjects. *Brain*, 114:2639–55.
- CHENEY, D. L. et SEYFARTH, R. M. (1990). *How Monkeys See the World : Inside the mind of Another Species*. Chicago : Chicago University Press.
- CHITORAN, I., GOLDSTEIN, L. et BYRD, D. (2002). Gestural overlap and recoverability : Articulatory evidence from Georgian. *In Laboratory Phonology VII*, pages 419–448. Berlin : Walter de Gruyter.

- CHOMSKY, N. (1996). *Powers and Prospects : Reflections on Human Nature and the Social Order*. Boston : South End Press.
- CHRISTIANSEN, M. H. et KIRBY, S. (2003). Language evolution : The hardest problem in science ? In CHRISTIANSEN, M. et KIRBY, S., éditeurs : *Language Evolution : The States of the Art*, pages 1–15. Oxford University Press.
- CLARK, A. (1997). *Being There : Putting Brain, Body and World Together Again*. MIT Press, Bradford Books.
- CLARK, A. (2003). *Natural-Born Cyborgs : Minds, Technologies and the Future of Human Intelligence*. Oxford University Press, NY.
- CLEMENTS, G. N. (1990). The role of the sonority cycle in core syllabication. In KINGSTON, J. et BECKMAN, M., éditeurs : *Papers in Laboratory Phonology I : Between the Grammar and the Physics of Speech*, pages 283–333. New-York : Cambridge University Press.
- CORBALLIS, M. C. (2003a). From hand to mouth : The gestural origins of language. In CHRISTIANSEN, M. et KIRBY, S., éditeurs : *Language Evolution : The States of the Art*, pages 201–218. Oxford University Press.
- CORBALLIS, M. C. (2003b). From mouth to hand : Gesture, speech, and the evolution of right-handedness. *Behavioral and Brain Sciences*, 26(2):199–260.
- DAVIS, B., MACNEILAGE, P. et MATYEAR, C. (2002). Acquisition of serial complexity in speech production : A comparison of phonetic and phonological approaches to first word production. *Phonetica*, 59(2-3):75–107.
- de BOER, B. (2000). Imitation games for complex utterances. In *Proceedings of BNAIC 2000*.
- DEACON, T. W. (2003). Universal grammar and semiotic constraints. In CHRISTIANSEN, M. et KIRBY, S., éditeurs : *Language Evolution : The States of the Art*, pages 111–139. Oxford University Press.
- DESSALLES, J.-L. (2000). *Aux Origines du langage, une histoire naturelle de la parole*. Hermès Science, Paris.
- DIESSEL, H. (1999). *Demonstratives : Form, Function, and Grammaticalization*. John Benjamins.
- DUCEY-KAUFMANN, V. (2007). *Le cadre de la parole et le cadre du signe : un rendez-vous développemental*. Thèse de doctorat, Sciences du langage, Université Stendhal, Grenoble.
- DUCEY-KAUFMANN, V., ABRY, C. et VILAIN, C. (2007). When the speech frame meets the sign frame in a developmental framework. A paraître dans la revue du congrès Emergence of Language Abilities (ELA) 2005.

- EHRlich, S. B., LEVINE, S. et GOLDIN-MEADOW, S. (2006). The importance of gesture in children's spatial reasoning. *Developmental Psychology*, 42:1259–1268.
- EMMOREY, K., DAMASIO, H., MCCULLOGH, S., GRABOWSKI, T., PONTO, L., HICHA, R. et BELLUGI, U. (2002). Neural systems underlying spatial language in american sign language. *Neuroimage*, 17:812–24.
- ENARD, W., PRZEWORSKI, M., FISHER, S. E., LAI, C. S. L., WIEBE, V., KITANO, T., MONACO, A. P. et PAABO, S. (2002). Molecular evolution of FOXP2, a gene involved in speech and language. *Nature*, 418:869–872.
- FERRARI, P., GALLESE, V., RIZZOLATTI, G. et FOGASSI, L. (2003). Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex. *European Journal of Neuroscience*, 17:1703–1714.
- FEYEREISEN, P. (1997). The competition between gesture and speech production in dual-task paradigms. *Journal of Memory and Language*, 36(1):13–33.
- FITCH, W. T., HAUSER, M. D. et CHOMSKY, N. (2005). The evolution of the language faculty : Clarifications and implications. *Cognition*, 97(2):179–210.
- GENTILUCCI, M. (2003). Grasp observation influences speech production. *European Journal of Neuroscience*, 17:179–184.
- GENTILUCCI, M., BENUZZI, F., GANGITANO, M. et GRIMALDI, S. (2001). Grasp with hand and mouth : a kinematic study on healthy subjects. *Journal of Neurophysiologie*, 86:1685–1699.
- GENTILUCCI, M., SANTUNIONE, P., ROY, A. et STEFANINI, S. (2004). Execution and observation of bringing a fruit to the mouth affect syllable pronunciation. *European Journal of Neuroscience*, 19:190–202.
- GLOTIN, H. (1995). Emergence du code phonétique dans une société de robots parlant. Mémoire de D.E.A., Sciences Cognitives, Institut National Polytechnique de Grenoble.
- GOLDIN-MEADOW (2005). How our hands help us learn. *Trends in Cognitive Science*, 9:234–241.
- GOLDIN-MEADOW, S. (2007). Pointing sets the stage for learning language - and creating language. *Child Development*, 78:741–745.
- GOLDIN-MEADOW, S. et BUTCHER, C. (2003). Pointing toward two-word speech in young children. In KITA, S., éditeur : *Pointing : Where language, culture, and cognition meet*, pages 85–107. Mahwah, N.J. : Erlbaum.

- GREEN, J., MOORE, C., HIGASHIKAWA, M. et STEEVE, R. W. (2000). The physiologic development of speech motor control : Lip and jaw coordination. *Journal of Speech, Language, and Hearing Research*, 43:239–255.
- GREEN, J., MOORE, C. et REILLY, K. (2002). The sequential development of jaw and lip control for speech. *Journal of Speech, Language, and Hearing Research*, 45:66–79.
- HAGÈGE, C. (1987). *L'homme de paroles. Contribution linguistique aux sciences humaines*.
- HAKEN, H., KELSO, J. A. S. et BUNZ, H. (1985). A theoretical model of phase transitions in human hand movements. *Biological Cybernetics*, 51(5):347–356.
- HAUSER, M., CHOMSKY, N. et FITCH, W. (2002). The faculty of language : What is it, Who has it, and How did it evolve ? *Science*, 298:1569–1579.
- HAUSER, M. D. et FITCH, W. T. (2003). What are the uniquely human components of the language faculty ? In CHRISTIANSEN, M. et KIRBY, S., éditeurs : *Language Evolution : The States of the Art*, pages 158–181. Oxford University Press.
- HAVILAND, J. (2000). Pointing, gesture spaces, and mental maps. In MCNEILL, D., éditeur : *Language and gesture*, pages 13–46. Cambridge University Press.
- HAVILAND, J. (2003). How to point in Zinacantán. In KITA, S., éditeur : *Pointing : Where Language, Culture, and Cognition meet*, pages 139–169. Lawrence Erlbaum Associates.
- HERTRICH, I. et ACKERMANN, H. (2000). Lip-jaw and tongue-jaw coordination during rate-controlled syllable repetitions. *Journal of Acoustique Society of America*, 107(4):2236–2247.
- HICKOK, G., BELLUGI, U. et KLIMA, E. (1998). The neural organization of language : Evidence from sign language aphasia. *Trends in Cognitive Sciences*, 2:129–136.
- HOLDEN, C. (2004). The origin of speech. *Science*, 303:1316–1319.
- HOLENDER, D. (1980). Interference between a vocal and a manual response to the same stimulus. In STELMACH, G. et REQUIN, J., éditeurs : *Tutorials in Motor Behavior*, pages 421–432. North Holland Publishing Company.
- HOOLE, P. (1993). Methodological considerations in the use of electromagnetic articulography in phonetic research. Rapport technique 31, Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München (FIPKM). pages 43–64.

- HOOLE, P. (1996). Issues in the acquisition, processing, reduction and parameterization of articulo-graphic data. Rapport technique 34, Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München (FIPKM). pages 158-173.
- HOYT, D. et TAYLOR, C. (1981). Gait and the energetics of locomotion in horses. *Nature*, 292:239–240.
- INGRAM, D. (1974). Fronting in child phonology. *Journal of Child Language*, 1:233–242.
- IVERSON, J. et GOLDIN-MEADOW, S. (1998). Why people gesture as they speak. *Nature*, 396:228.
- IVERSON, J. et GOLDIN-MEADOW, S. (2005). Gesture paves the way for language development. *Psychological Science*, 16:368–371.
- IVERSON, J. et THELEN, E. (1999). Hand, mouth, and brain : The dynamic emergence of speech and gesture. *Journal of Consciousness Studies*, 6:19–40.
- IVERSON, J. et THELEN, E. (2003). The hand leads the mouth in ontogenesis too. *Behavioral and brain sciences*, 26(2):225–226.
- JACKENDOFF, R. et PINKER, S. (2005). The nature of the language faculty and its implications for evolution of language (Reply to Fitch, Hauser, and Chomsky). *Cognition*, 97:211–225.
- JAKOBSON, R. (1976). *Six leçons sur le son et le sens*. Edition de minuit.
- JOSE-ROBERTSON, L. S., CORINA, D., ACKERMAN, D., GUILLEMIN, A. et BRAUN, A. (2004). Neural systems for sign language production : Mechanisms supporting lexical selection, phonological encoding, and articulation. *Human Brain Mapping*, 23(3):156–67.
- KEATING, P. (1983). Comments on the jaw and syllable structure. *Journal of Phonetic*, 11:401–406.
- KEATING, P., LINDBLOM, B., LUBKER, J. et KREIMAN, J. (1994). Variability in jaw height for segments in english and swedish VCVs. *Journal of Phonetic*, 22:407–422.
- KELLER, H. (1902). *The story of my life*. Bantam classic.
- KELLER, H. (1908). *The world I live in*. New-York : Review Book.
- KELSO, J. A. D., TULLER, B. et HARRIS, K. S. (1981). A "Dynamic Pattern" perspective on the control and coordination of movement. In MACNEILAGE, P., éditeur : *The production of speech*, pages 137–173. New-York : Springer-Verlag.

- KELSO, J. A. S. (1995). *Dynamic patterns : The self-organization of brain and behavior*. MIT Press, Cambridge, MA.
- KELSO, J. A. S., SALTZMAN, E. L. et TULLER, B. (1986). The dynamical perspective on speech production : Data and theory. *Journal of Phonetic*, 14:29–59.
- KELSO, J. A. S., VATIKIOTIS-BATESON, E., SALTZMAN, E. L. et KAY, B. (1985). A qualitative dynamic analysis of reiterant speech production : Phase portraits, kinematics, and dynamic modeling. *Journal of Acoustic Society of America*, 77(1):266–280.
- KEMPF, H. (1998). Entretiens avec Francisco Varela. *La Recherche*, 308:109–112.
- KITA, S. (2003). Interplay of gaze, hand, torso orientation, and language in pointing. In KITA, S., éditeur : *Pointing : Where language, culture, and cognition meet*, pages 307–328. Mawhaw, N.J. : Erlbaum.
- KOMAROVA, N. L. et NOWAK, M. A. (2003). Language, learning, and evolution. In CHRISTIANSEN, M. et KIRBY, S., éditeurs : *Language Evolution : The States of the Art*, pages 317–337. Oxford University Press.
- LABOISSIÈRE, R., OSTRY, D. et FELDMAN, A. (1996). The control of multi-muscle systems : human jaw and hyoid movements. *Biological cybernetics*, 74(4):373–384.
- LAI, C., FISHER, S., HURST, J., VARGHA-KHADEM, F. et MONACO, A. (2001). A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature*, 413:519–523.
- LAU, H., ROGERS, R., HAGGARD, P. et PASSINGHAM, R. (2004). Attention to intention. *Science*, 303:1208–1210.
- LAURENT, D. (2007). Pointage et démonstratifs : Exploration de la grammaire par la coordination geste-parole. Mémoire de D.E.A., Université Stendhal - Sciences du langage.
- LEAVENS, D. A. (2004). Manual deixis in apes and humans. *Interaction Studies*, 5:387–408.
- LEAVENS, D. A., HOPKINS, W. D. et BARD, K. A. (2005). Understanding the point of chimpanzee pointing : Epigenesis and ecological validity. *Current Directions in Psychological Science*, 14:185–189.
- LEAVENS, D. A., HOPKINS, W. D. et THOMAS, R. K. (2004). Referential communication by chimpanzees (Pan troglodytes). *Journal of Comparative Psychology*, 118:48–57.
- LEBELTEL, O., BESSIÈRE, P. et MAZER, E. (2004). Bayesian robot programming. *Advanced Robotics*, 16:49–79.

- LESNE, A. et BOURGUINE, P. (2006). *Morphogénèse, l'origines des formes*. Paris : Belin.
- LEVELT, W. J. M., RICHARDSON, G. et HEIJ, W. L. (1985). Pointing and voicing in deictic expressions. *Journal of Memory and Language*, 24:133–164.
- LIEBERMAN, A. et MATTINGLY, I. (1985). The motor theory of speech perception revised. *Cognition*, 21:1–36.
- LIEBERMAN, P. (2002). On the nature and evolution of the neural bases of human language. *American Journal of Physical Anthropology*, 119(S35):36–62.
- LIEBERMAN, P. (2007). The evolution of human speech : Its anatomical and neural bases. *Current Anthropology*, 48(1):39–66.
- LILJENCRANTS, J. et LINDBLOM, B. (1972). Numerical simulations of vowel quality systems : the role of perceptual contrast. *Language*, 48:839–862.
- LINDBLAD, P. et LUNDQVIST, S. (1999). How and why do the tongue gestures of [t], [d], [l], [n], [s], and [r] differ ? In *Proceedings of the 14th International Congress of Phonetic Sciences (ICPhS), San-Francisco*, pages 417–420.
- LINDBLOM, B. (1990a). *Explaining Phonetic Variation, A Sketch of the H and H Theory*, pages 403–439. Academic Publishers.
- LINDBLOM, B. (1990b). On the notion of "possible speech sound". *Journal of Phonetic*, 18:135–152.
- LINDBLOM, B., MACNEILAGE, P. et STUDDERT-KENNEDY, M. (1984). *Self-organizing Processes and the Explanation of Languages Universals*, pages 181–203. Berlin : Mouton.
- LOCKE, J. (2000). Movement patterns in spoken language. *Science*, 288:449–451.
- LOCKE, J. L. (1983). *Phonological Acquisition and Change*. Academic Press, New York.
- LØEVENBRUCK, H., BACIU, M., SEGEBARTH, C. et ABRY, C. (2005). The left inferior frontal gyrus under focus : an fMRI study of the production of deixis via syntactic extraction and prosodic focus. *Journal of Neurolinguistics*, 18:237–258.
- MACNEILAGE, P. (1998). The Frame/Content theory of evolution of speech production. *Behavioral and Brain Sciences*, 21:499–511.
- MACNEILAGE, P. et DAVIS, B. (2000). On the origins of internal structure of word forms. *Science*, 288:527–531.
- MALSON, L. (1969). *Les enfants sauvages*. 10/18.

- MASATAKA, N. (2003). From index-finger extension to index-finger pointing : Ontogenesis of pointing in preverbal infants. In KITA, S., éditeur : *Pointing : Where language, culture, and cognition meet*, pages 69–84. Mahwah, N.J. : Erlbaum.
- MCGURK, H. et MACDONALD, J. (1976). Hearing lips and seeing voices. *Nature*, 264:746–748.
- MCNEILL, D. (1981). Action, thought and language. *Cognition*, 10:201–208.
- MCNEILL, D. (1992). *Hand and Mind*. Chicago : University of Chicago Press.
- MOULIN-FRIER, C. (2007). Jeux déictiques au sein d'une colonie d'agents sensori-moteurs. Mémoire de D.E.A., Master ICA, Spécialité Sciences Cognitives, Institut National Polytechnique de Grenoble.
- MUNHALL, K. et JONES, J. (1998). Articulatory evidence for syllabic structure. *Behavioral and Brain Science*, 21:524–525.
- NELSON, W., PERKELL, J. et WESTBURY, J. (1984). Mandibule movements during increasingly rapid articulations of single syllables : Preliminary observations. *Journal of Acoustic Society of America*, 75(3):945–951.
- NELSON, W. L. (1983). Physical principles for economies of skilled movements. *Biological Cybernetics*, 46:135–147.
- NGUYEN, N., WAUQUIER-GRAVELINES, S. et DURAND, J. (2005). *Phonologie et phonétique : Forme et Substance*. Hermès, Paris.
- NISHITANI, N., SCHÜRMAN, M., AMUNTS, K. et HARI, R. (2005). Broca's region : From action to language. *Physiology*, 20(1):60–69.
- NITTROUER, S. (1991). Phase relations of jaw and tongue tip movements in the production of VCV utterances. *Journal of Acoustique Society of America*, 90(4 Pt 1):1806–1815.
- NOWAK, M. A. (2000). Homo grammaticus. *Natural History*, 109:36–44.
- NOWAK, M. A., KOMAROVA, N. L. et NIYOGI, P. (2001). Evolution of universal grammar. *Science*, 291:114–118.
- NOWAK, M. A., PLOTKIN, J. B. et JANSEN, V. A. A. (2000). The evolution of syntactic communication. *Nature*, 404:495–498.
- OSTRY, D. et FLANAGAN, J. (1989). Human jaw movement in mastication and speech. *Archives of Oral Biology*, 34:685–693.
- OUDEYER, P.-Y. (2001). The origins of syllable systems : An operational model. In MOORE, J. D. et STENNING, K., éditeurs : *Proceedings of the Twenty-third Annual Conference of the Cognitive Science Society*, pages 744–749. Laurence Erlbaum Associates.

- OUDEYER, P.-Y. (2003). *The self-organization of speech sounds*. Thèse de doctorat, Université Paris VI.
- OZCALISKAN, S. et GOLDIN-MEADOW, S. (2005). Gesture is at the cutting edge of early language development. *Cognition*, 96:101–113.
- PAYAN, Y. et PERRIER, P. (1997). Synthesis of V-V sequences with a 2D biomechanical tongue model controlled by the equilibrium point hypothesis. *Speech Communication*, 22(2-3):185–205.
- PERKELL, J. S. et ZANDIPOUR, M. (2002). Economy of effort in different speaking conditions. II. Kinematic performance spaces for cyclical and speech movements. *Journal of Acoustic Society of America*, 112(4):1642–1651.
- PERKELL, J. S., ZANDIPOUR, M., MATTHIES, M. L. et LANE, H. (2002). Economy of effort in different speaking conditions. I. A preliminary study of inter-subject differences and modeling issues. *Journal of Acoustic Society of America*, 112(4):1627–1641.
- PETITTO, L., HOLOWKA, S., SERGIO, L. et OSTRY, D. (2001). Language rhythms in baby hand movement. *Nature*, 413:35–36.
- PHILIPON, P. (2007). Parler avec les mains. *La Recherche*, 410:22.
- PIERREHUMBERT, J. (2000). The phonetic grounding of phonology. *Bulletin de la Communication Parlée*, 5:7–23.
- PINKER, S. (2003). Language as an adaptation to the cognitive niche. In CHRISTIANSEN, M. et KIRBY, S., éditeurs : *Language Evolution : The States of the Art*, pages 16–37. Oxford University Press.
- PINKER, S. et BLOOM, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences*, 13(4):707–784.
- PINKER, S. et JACKENDOFF, R. (2005). The faculty of language : What's special about it ? *Cognition*, 95(2):201–236.
- PIZZUTO, E., CAPOBIANCO, M. et DEVESCOVI, A. (2005). Gestural-vocal deixis and representational skills in early language development. *Interaction Studies. Social Behaviour and Communication in Biological and Artificial Systems*, 6(2): 223–252.
- POVINELLI, D., BIRSCHWALE, D. et CECH, C. (1999). Comprehension of seeing as a referential act in young children, but not juvenile chimpanzees. *British Journal of Developmental Psychology*, 17:37–60.
- POVINELLI, D. J., BERING, J. M. et GIAMBRONE, S. (2003). Chimpanzees "pointing" : Another error of argument by analogy ? In KITA, S., éditeur : *Pointing :*

- Where language, culture, and cognition meet*, pages 35–68. Mahwah, N.J. : Erlbaum.
- PULVERMULLER, F. (2005). Brain mechanisms linking language and action. *Natural Review of Neuroscience*, 6(7):576–82.
- PULVERMULLER, F., HAUKE, O., NIKULIN, V. et ILMONIEMI, R. (2005). Functional links between motor and language systems. *European Journal of Neuroscience*, 21:793–797.
- REDFORD, M., CHEN, C. et MIIKKULAINEN, R. (2001). Constrained emergence of universals and variation in syllable systems. *Language and Speech*, 44:27–56.
- REDFORD, M. et DIEHL, R. (1999). The relative perceptual distinctiveness of initial and final consonants in CVC syllables. *Journal of Acoustic Society of America*, 103(3):1555–1565.
- RHARDISSE, N. et ABRY, C. (1994). La coarticulation mandibulaire comme principe d'organisation syllabique. *In Proceedings des XXèmes Journées d'Etudes sur la Parole (JEP), Trégastel, France*, pages 421–426.
- RIZZOLATTI, G. et ARBIB, M. A. (1998). Language within our grasp. *Trends in Neurosciences*, 21(5):188–194.
- RIZZOLATTI, G., FADIGA, L., FOGASSI, L. et GALLESE, V. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, 3:131–141.
- ROCHET-CAPELLAN, A. (2004). L'économie articulatoire des mots du français : une autre explication de l'effet LC ? Mémoire de D.E.A., Master ICA, Sciences Cognitives, Institut National Polytechnique de Grenoble.
- ROCHET-CAPELLAN, A. et SCHWARTZ, J.-L. (2005a). The labial-coronal effect and CVCV stability during reiterant speech production : An acoustic analysis. *In Proceedings of the 9th International Conference on Speech Communication and Technology (InterSpeech 2005)*, pages 1009–1012, Lisbonne, Portugal.
- ROCHET-CAPELLAN, A. et SCHWARTZ, J.-L. (2005b). The labial-coronal effect and CVCV stability during reiterant speech production : An articulatory analysis. *In Proceedings of the 9th International Conference on Speech Communication and Technology (InterSpeech 2005)*, pages 1013–1016, Lisbonne, Portugal.
- ROCHET-CAPELLAN, A. et SCHWARTZ, J.-L. (2006). L'implication des contraintes motrices dans "l'effet Labial-Coronal". *In Proceedings des Journées d'Etudes sur la Parole, 2006 (JEP'2006)*, pages 567–570, Dinard, France.
- ROCHET-CAPELLAN, A. et SCHWARTZ, J.-L. (2007). An articulatory basis for the Labial-to-Coronal effect : /pata/ seems a more stable articulatory pattern than /tapa/. *Journal of the Acoustical Society of America*, 121(6):3740–3754.

- ROCHET-CAPELLAN, A., SCHWARTZ, J.-L., LABOISSIÈRE, R. et GALVÁN, A. (2006). Finger-jaw coordination during a deictic gesture with CVCV utterances : the effect of stress position. *In Proceedings of the International Seminar on Speech Production (ISSP'06)*, pages 193–200, Ubatuba, Brésil.
- ROCHET-CAPELLAN, A., SCHWARTZ, J.-L., LABOISSIÈRE, R. et GALVÁN, A. (2007a). Pointing to a target while naming it with labial-coronal or coronal-labial CVCV : the effect of consonants order and stress position on jaw-finger coordination. *In Proceedings of the 8th International Conference on Speech Communication and Technology (InterSpeech 2007)*, Antwerp, Belgique.
- ROCHET-CAPELLAN, A., SCHWARTZ, J.-L., LABOISSIÈRE, R. et GALVÁN, A. (2007b). The speech focus position effect on jaw-finger coordination in a pointing task. En révision pour le JSLHR.
- ROCHET-CAPELLAN, A., SCHWARTZ, J.-L., LABOISSIÈRE, R. et GALVÁN, A. (2007c). Two CV syllables for one pointing gesture as an optimal ratio for jaw-arm coordination in a deictic task : A preliminary study. *In Proceedings of the European Cognitive Science Conference (EuroCogSci07)*, pages 608–613, Delphes, Grèce.
- ROUSSET, I. (2003). From lexical to syllabic organization : favored and disfavored co-occurrences. *In In proceeding of the 15th International Congress of Phonetic Sciences (ICPhS)*, pages 715–718, Barcelona, Spain.
- RUITER, J. D. et WILKINS, D. (1998). The synchronization of gesture and speech in dutch and arrernte (an australian aboriginal language) : A cross-cultural comparison. *In SANTI, S., GUAITELLA, I., CAVÉ, C. et KONOPCZYNSKI, G., éditeurs : Oralité et Gestualité. Communication multimodale, Interaction. Actes du colloque Orange 98*, pages 603–607. Paris : L'Harmattan.
- SATO, M., SCHWARTZ, J.-L., ABRY, C., CATHIARD, M.-A. et LÆVENBRUCK, H. (2006). Multistable syllables as enacted percept : A source of an asymmetric bias in the verbal transformation effect. *Perception and Psychophysics*, 68(3):458–474.
- SATO, M., VALLÉE, N., SCHWARTZ, J.-L. et ROUSSET, I. (2007). A perceptual correlate of the Labial-Coronal effect. *Journal of Speech Language and Hearing Research*, 50:1466–1480.
- SCHWARTZ, J., ABRY, C., BOË, L.-J. et CATHIARD, M. (2002). Phonology in a theory of perception-for-action-control. *In DURAND, J. et LAKS, B., éditeurs : Phonology : from Phonetics to Cognition*, pages 255–280. Oxford : Oxford University Press.
- SCHWARTZ, J., BOË, L., VALLÉE, N. et ABRY, C. (1997). The dispersion-focalization theory of vowel systems. *Journal of Phonetics*, 25:255–286.

- SCHWARTZ, J.-L. (2007). *Eléments pour une morphogenèse des unités du langage*. Soumis.
- SELKIRK, E. (1984). *On the major class features and syllable theory*. MIT Press.
- SERKHANE, J. (2005). *Un Agent vocalisant : Etude et modélisation des mécanismes d'exploration vocale et d'imitation orofaciale dans le développement de la parole*. Thèse de doctorat, Sciences Cognitives, Institut National Polytechnique de Grenoble.
- SHILLER, D., LABOISSIÈRE, R. et OSTRY, D. (2002). Relationship between jaw stiffness and kinematic variability in speech. *Journal of Neurophysiology*, 88(5): 2329–2340.
- SMITH, A., MCFARLAND, D. H. et WEBER, C. M. (1986). Interactions between speech and finger movements : An exploration on the dynamic pattern perspective. *Journal of Speech and Hearing Research*, 29(4):471–80.
- SOCK, R. et LÖFQVIST, A. (1995). Some timing constraints in the production of bilabial stops. *Journal of Phonetics*, 23:129–138.
- SOCK, R., LÖFQVIST, A. et PERRIER, P. (1996). Kinematic and acoustic correlates of quantity in Swedish and Wolof : A cross-language study. *In Proceedings of the First European Speech Communication Association (ESCA) Tutorial and Research Workshop, Aufran-France*, pages 81–84.
- SONODA, Y. et NAKAKIDO, K. (1986). Effect of speaking rate on jaw movements in vowel sequence. *Journal of Acoustic Society of Japan*, 7:5–12.
- STEELS, L. (2001). Language games for autonomous robots. *IEEE Intelligent systems*, pages 16–22.
- STETSON, R. (1951). *Motor Phonetics : A Study of Movements in Action*. North Holland :Amsterdam.
- SURPRENANT, A. M. et GOLDSTEIN, L. (1998). The perception of speech gestures. *Journal of Acoustic Society of America*, 104(1):518–529.
- THUAN, T. X. (1998). *Le Chaos et l'Harmonie : La fabrication du réel*. Folio Essais.
- TOMASELLO, M., CARPENTER, M. et LISZKOWSKI, U. (2007). A new look at infant pointing. *Child Development*, 78(3):705–722.
- TREFFNER, P. et PETER, M. (2002). Intentional and attentional dynamics of speech-hand coordination. *Human Movement Science*, 21(5-6):641–97.
- TREMBLAY, S., SHILLER, D. et OSTRY, D. (2003). Somatosensory basis of speech production. *Nature*, 423:866–869.

- VARELA, F. (1991). *The embodied mind : cognitive science and human experience*. Cambridge : MIT press.
- VARELA, F. (1993). *L'Inscription corporelle de l'esprit. Sciences cognitives et expérience humaine*. Paris : Edition du Seuil.
- VILAIN, A., ABRY, C., BADIN, P. et BROSDA, S. (1999). From idiosyncratic pure frame to variegated babbling : Evidence from articulatory modelling. *In Proceedings of the International Congress of Phonetic Sciences (ICPhS'99)*, San-Fransisco, USA.
- VOLTERRA, V., CASELLI, M. C., CAPIRCI, O. et PIZZUTO, E. (2005). Gesture and the emergence and development of language. *In TOMASELLO, M. et SLOBIN, D., éditeurs : Beyond nature-nurture- Essays in honor of Elizabeth Bates*, pages 3–40. Mahwah, N. J. : Lawrence Erlbaum Associates.
- VYGOTSKI, L. (1997). *Pensée et langage*. Paris, La dispute. Ré-édition (Vygotski est mort en 1934).
- WATKINS, K., DRONKERS, N. et VARGHA-KHADEM, F. (2002a). Behavioural analysis of an inherited speech and language disorder : comparison with acquired aphasia. *Brain*, 125:452–464.
- WATKINS, K., VARGHA-KHADEM, F., ASHBURNER, J., PASSINGHAM, R., CONNELLY, A., FRISTON, K., FRACKOWIAK, R., MISHKIN, M. et GADIAN, D. (2002b). MRI analysis of an inherited speech and language disorder : structural brain abnormalities. *Brain*, 125:465–478.
- WILKINS, D. (2003). Why pointing with the index finger is not a universal (in sociocultural and semiotic terms). *In KITA, S., éditeur : Pointing : Where Language, Culture, and Cognition meet*, pages 171–215. Lawrence Erlbaum Associates.
- ZSIGA, E. C. (1996). Acoustic evidence for gestural overlap in consonant sequence. *Journal of Phonetic*, 22:121–140.
- ZWIRN, H. P. (2006). *Les systèmes complexes, mathématiques et biologie*. Odile Jacob.

Cinquième partie

Annexes

Effet de la position de l'accent - sélection
des données et synthèse des ANOVAs

A.1 SÉLECTION DES DONNÉES : NOMBRE D'ESSAIS REJETÉS SELON LES CRITÈRES DE SÉLECTION

TAB. A.1 – Nombre moyen (Moy.) d'essais rejetés pour cause d'erreur ou de masquage des marqueurs de l'Optotrak pour les deux expériences. « Dén. » correspond au nombre d'erreurs sur le nom de la cible ; « Init. », au nombre d'essais corrects mais pour lesquels la réponse de la mâchoire et/ou du doigt débute(nt) avant le signal-go ; « Doigt », aux essais corrects avec masquage des marqueurs du doigt et « Mâchoire », aux essais corrects avec masquage du marqueur de la mâchoire. Les colonnes Err. std., Min. et Max. donne l'erreur standard moyenne et le nombre minimum et maximum d'essais de chaque type sur les 20 participants. « Accent », « Cible » et « Séquence » donnent les résultats de l'ANOVA pour les effets principaux de la position de l'accent ($accent_1 - accent_2$), de la position de la cible (*proche - loin*) et de la séquence ($/tt/ - /pp/$ pour l'expérience 1 et $/tp/ - /pt/$ pour l'expérience 2). Seul l'effet de la séquence sur le nombre d'erreurs de dénomination est significatif pour l'expérience 2, il reste cependant de faible amplitude.

	Moy.	Err. std.	Min.	Max.	Accent	Cible	Séquence
Expérience 1							
Dén.	4.1	0.9	0	11	0.2 ^a	0.1 ^a	-0.1 ^a
Initialisation	2.4	1.9	0	39	0.3 ^a	-0.2 ^a	-0.2 ^a
Doigt	8.2	1.5	0	21	1.2 ^a	-2 ^a	0.3 ^a
Mâchoire	0.8	0.6	0	12	-0.3 ^a	-0.3 ^a	0.2 ^a
Expérience 2							
Dén.	3.4	1.14	0	18	-0.2 ^a	-0.2 ^a	0.8 ^b
Init.	3.5	8.5	0	38	0.6 ^a	-0.1 ^a	-0.2 ^a
Doigt	7.7	11.1	0	45	0.3 ^a	2.6 ^a	0.3 ^a
Mâchoire	1.3	1.5	0	20	-0.3 ^a	0.1 ^a	-0.5 ^a

Anova : ^{a)} $p > .05$; ^{b)} $p < .05$

A.2 SYNTHÈSE DES ANOVAS

NOTE POUR LA LECTURE DES TABLEAUX

Les tableaux ci-après synthétisent les différentes ANOVAs réalisées pour l'analyse de l'effet de la position de l'accent sur la coordination mâchoire-doigt présentées au chapitre 5. Ces ANOVAs comportent toujours trois facteurs intrasujets (*Position accent* \times *Position cible* \times *Sequence*). L'effet de l'ordre de passation des deux expériences 1-2 vs. 2-1 a été systématiquement contrôlé en tant que facteur intersujet de l'ANOVA. Etant très rarement significatif, il n'est pas intégré aux analyses ci-après. Chaque tableau donne, pour chaque variable mesurée, la moyenne globale et l'effet de :

- la position de l'accent, différence moyenne entre les conditions *accent*₁ et *accent*₂ ;
- la position de la cible, différence moyenne entre les conditions *cible*_{proche} et *cible*_{loin} ;
- la séquence, différence moyenne entre les conditions *seq.*_{/tt/} et *seq.*_{/pp/} pour l'expérience 1 et entre *seq.*_{/tp/} et *seq.*_{/pt/} pour l'expérience 2 ;
- la position de l'accent selon la séquence ;
- la position de l'accent selon la position de la cible ;
- la séquence selon la position de la cible ;
- la position de l'accent selon la séquence et la position de la cible ;

Lorsque l'effet d'interaction est significatif, l'effet de la position de l'accent dans chaque condition est testé avec des comparaisons multiples consistant en des t-tests appariés avec correction de Bonferroni dont le seuil de significativité est fixé à $.05/2$ pour les interactions doubles et $.05/4$ soit un p_{bf} indiqué de $p*2$ pour les interactions doubles et $p*4$ pour les interactions triples.

A.2.1 Occurrence des événements des trajectoires du doigt et de la mâchoire

Dans les tableaux A.2 et A.3, les mesures sont le temps écoulé entre le signal-go et : le début (P_I) et l'apex (P_A) du geste de pointé et le début (P_R) et la fin (P_F) du geste de retour du doigt ; le début (M_I , première consonne) et la fin (M_F) des cycles de la mâchoire ; la position de la mâchoire la plus basse pour la première (M_{min1}) et la deuxième (M_{min2}) voyelle et la position la plus haute pour la deuxième consonne (M_{max}), voir figure 5.4, p. 98 pour le repérage de ces événements.

TAB. A.2 – Expérience 1 (position de l'accent et coordination doigt-mâchoire), étude des effets de la position de l'accent, de la cible et de la séquence sur l'occurrence des événements de la trajectoire du doigt et de la mâchoire.

MESURES									
Doigt				Mâchoire					
P_I	P_A	P_R	P_F	M_I	M_{min1}	M_{max}	M_{min2}	M_F	
ms	ms	ms	ms	ms	ms	ms	ms	ms	ms
Moyennes globales									
322	741	883	1429	508	674	824	958	1240	
Effet de la position de l'accent ($accent_1 - accent_2$)									
-11 ^a	-16 ^a	-46 ^e	-51 ^e	63 ^e	103 ^f	162 ^f	102 ^f	38 ^c	
Effet de la position de la cible (<i>proche - loin</i>)									
9 ^b	-16 ^d	-36 ^f	-51 ^e	-11 ^b	-11 ^c	-10 ^c	-10 ^c	-11 ^c	
Effet de la séquence ($/tt/ - /pp/$)									
0	3 ^a	6 ^a	8 ^a	-39 ^d	-2 ^a	23 ^d	2 ^a	-7 ^a	
$(accent_1 - accent_2)$ selon la séquence									
$/tt/$	-15	-20	-49	-52	66	98	169	97	31
$/pp/$	-8	-12	-43	-51	60	109	155	107	44.7
P_{inter}	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>
$(accent_1 - accent_2)$ selon la position de la cible									
proche	-11	-15	-44	-52	58	99	158	97	30
loin	-12	-18	-48	-50	68	108	166	107	46
P_{inter}	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>
$(/t/ - /p/)$ selon la position de la cible									
proche	0	3	3	9	-40	-3	24	3	-6
loin	-1	3	9	6	-39	-1	22	1	-7
P_{inter}	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>
$(accent_1 - accent_2)$ selon la séquence et la position de la cible									
$/tt/.proche$	-10	-13 ^d	-40 ^{b'}	-48	70 ^{d'}	102 ^{f'}	175 ^{f'}	100 ^{f'}	30 ^{a'}
$/tt/.loin$	-20	-28 ^{b'}	-58 ^{e'}	-56	63 ^{e'}	94 ^{f'}	163 ^{f'}	94 ^{f'}	32 ^{a'}
$/pp/.proche$	-11	-17 ^{d'}	-49 ^{d'}	-57	46 ^{e'}	96 ^{f'}	141 ^{f'}	95 ^{e'}	30 ^{a'}
$/pp/.loin$	-4	-8 ^{a'}	-37 ^{e'}	-44	74 ^{e'}	122 ^{f'}	169 ^{f'}	120 ^{f'}	59 ^{d'}
P_{inter}	<i>a</i>	<i>c</i>	<i>c</i>	<i>a</i>	<i>c</i>	<i>d</i>	<i>d</i>	<i>c</i>	<i>c</i>

Anova : ^{a)} $p < 1$; ^{b)} $p < .1$; ^{c)} $p < .05$; ^{d)} $p < .01$; ^{e)} $p < .001$; ^{f)} $p < .0001$

Comparaisons multiples : ^{a')} $p_{bf} < 1$; ^{b')} $p_{bf} < .1$; ^{c')} $p_{bf} < .05$; ^{d')} $p_{bf} < .01$; ^{e')} $p_{bf} < .001$

TAB. A.3 – Expérience 2 (profils de mâchoire et coordination doigt-mâchoire), étude des effets de la position de l'accent, de la cible et de la séquence sur l'occurrence des événements de la trajectoire du doigt et de la mâchoire.

MESURES									
Doigt				Mâchoire					
P_I	P_A	P_R	P_F	M_I	M_{min1}	M_{max}	M_{min2}	M_F	
ms	ms	ms	ms	ms	ms	ms	ms	ms	ms
Moyennes globales									
317	729	866	1407	511	678	830	961	1246	
Effet de la position de l'accent ($accent_1 - accent_2$)									
-10 ^a	-14 ^a	-36 ^e	-38 ^e	71 ^e	107 ^f	167 ^f	116 ^f	59 ^d	
Effet de la position de la cible (<i>proche - loin</i>)									
10 ^c	-11 ^c	-30 ^e	-46 ^f	-12 ^c	-10 ^c	-11 ^c	-12 ^c	-6 ^a	
Effet de la séquence ($/tp/ - /pt/$)									
4 ^a	2 ^a	6 ^a	2 ^a	-33 ^e	-5 ^a	-18 ^d	7 ^a	11 ^a	
$(accent_1 - accent_2)$ selon la séquence									
/tp/	-15	-18	-38	-42	73	106	155 ^f	115	57
/pt/	-6	-10	-34	-35	69	107	180 ^f	118	61
p_{inter}	a	a	a	a	a	a	c	a	a
$(accent_1 - accent_2)$ selon la position de la cible									
proche	-7	-10	-35	-35	71	108	168	115	61
loin	-13	-18	-37	-42	70	106	167	118	57
p_{inter}	a	a	a	a	a	a	a	a	a
$(/tp/ - /pt/)$ selon la position de la cible									
proche	4	1	6	2	-35	-4	-16	9	15
loin	4	4	7	1	-31	-6	-20	6	8
p_{inter}	a	a	a	a	a	a	a	a	a
$(accent_1 - accent_2)$ selon la séquence et la position de la cible									
/tp/.proche	-12	-14	-35	-37	70	104	153	112	57
/tp/.loin	-17	-22	-41	-47	76	108	156	117	57
/pt/.proche	-3	-6	-35	-32	73	111	183	119	65
/pt/.loin	-9	-14	-34	-37	64	104	178	118	56
p_{inter}	a	a	a	a	a	a	a	a	a

Anova : ^a) $p < 1$; ^b) $p < .1$; ^c) $p < .05$; ^d) $p < .01$; ^e) $p < .001$; ^f) $p < .0001$

Comparaisons multiples : ^a) $p_{bf} < 1$; ^b) $p_{bf} < .1$; ^c) $p_{bf} < .05$; ^d) $p_{bf} < .01$; ^e) $p_{bf} < .001$

A.2.2 Position des événements de la trajectoire du doigt par rapport à ceux de la mâchoire

Dans les tableaux A.4 et A.5, les mesures étudiées sont les délais entre le début du premier cycle de la mâchoire (M_I) et le début du geste de pointé (P_I) ; entre la position basse de la mâchoire pour la première voyelle (M_{min1}) et l'apex du geste de pointé (P_A), entre la position haute de la mâchoire pour la deuxième consonne (M_{max}) et P_A et entre la position basse de la mâchoire pour la deuxième voyelle (M_{min2}) et le début du retour du geste du doigt (P_R). $P_{I/j}$, $P_{A/j}$, $P_{R/j}$ et $P_{F/j}$ sont les positions de P_I , P_A , P_R et de la fin du geste de retour du doigt (P_F) sur les cycles de mâchoire normalisés, voir figure 5.6, p.103.

TAB. A.4 – Expérience 1 (position de l'accent et coordination doigt-mâchoire), étude des effets de la position de l'accent, de la cible et de la séquence sur la position des événements de la trajectoire du doigt relativement à ceux de la trajectoire de la mâchoire.

MESURES								
Délais				Positions				
$M_I - P_I$	$M_{min1} - P_A$	$M_{max} - P_A$	$M_{min2} - P_R$	$P_{I/j}$	$P_{A/j}$	$P_{R/j}$	$P_{F/j}$	
ms	ms	ms	ms	%	%	%	%	
Moyennes globales								
186	-67	84	76	-171	-25	18	140	
Effet de la position de l'accent ($accent_1 - accent_2$)								
74 ^d	120 ^f	178 ^f	148 ^f	-5 ^a	-45 ^f	-48 ^f	-11 ^d	
Effet de la position de la cible (<i>proche - loin</i>)								
-20 ^d	5 ^a	7 ^a	26 ^e	6 ^c	-2 ^a	-7 ^d	-9 ^f	
Effet de la séquence (<i>/tt/ - /pp/</i>)								
-39 ^d	-5 ^a	20 ^d	-4 ^a	13 ^c	-0	0 ^a	-3 ^a	
<i>(accent₁ - accent₂) selon la consonne</i>								
<i>/t/</i>	81	118	189 ^f	146	-9 ^d	-45	-49	-15
<i>/p/</i>	68	122	168 ^f	150	-2 ^d	-45	-47	-8
P_{inter}	<i>a</i>	<i>a</i>	<i>c</i>	<i>a</i>	<i>c</i>	<i>a</i>	<i>a</i>	<i>a</i>
<i>(accent₁ - accent₂) selon la position de la cible</i>								
<i>proche</i>	69	114 ^f	173	142	-3	-43 ^f	-47	-11
<i>loin</i>	80	126 ^f	184	155	-7	-47 ^f	-50	-12
P_{inter}	<i>a</i>	<i>c</i>	<i>b</i>	<i>b</i>	<i>a</i>	<i>c</i>	<i>a</i>	<i>a</i>
<i>(/tt/ - /pp/) selon la position de la cible</i>								
<i>proche</i>	-40	-6	22	0	13	0	-2	-3
<i>loin</i>	-38	-4	19	-8	13	0	0	-3
P_{inter}	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>
<i>(accent₁ - accent₂) selon la séquence et la position de la cible</i>								
<i>/tt/.proche</i>	80	114	188	140	-9	-43	-48	-15
<i>/tt/.loin</i>	82	122	191	152	-9	-46	-51	-16
<i>/pp/.proche</i>	57	114	159	143	3	-42	-46	-8
<i>/pp/.loin</i>	78	129	177	157	-6	-47	-48	-8
P_{inter}	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>

Anova : ^{a)} $p < 1$; ^{b)} $p < .1$; ^{c)} $p < .05$; ^{d)} $p < .01$; ^{e)} $p < .001$; ^{f)} $p < .0001$

Comparaisons multiples : ^{a)} $p_{bf} < 1$; ^{b)} $p_{bf} < .1$; ^{c)} $p_{bf} < .05$; ^{d)} $p_{bf} < .01$; ^{e)} $p_{bf} < .001$

TAB. A.5 – Expérience 2 (profils de mâchoire et coordination doigt-mâchoire), étude des effets de la position de l'accent, de la cible et de la séquence sur la position des événements de la trajectoire du doigt relativement à ceux de la trajectoire de la mâchoire.

MESURES								
Délais				Positions				
$M_I - P_I$	$M_{min1} - P_A$	$M_{max} - P_A$	$M_{min2} - P_R$	P_I/j	P_A/j	P_R/j	P_F/j	
ms	ms	ms	ms	%	%	%	%	
Moyennes globales								
194	-51	101	95	-174	-31	12	133	
Effet de la position de l'accent ($accent_1 - accent_2$)								
81 ^e	121 ^f	181 ^f	152 ^f	-9 ^a	-46 ^f	-51 ^f	-18 ^e	
Effet de la position de la cible (<i>proche - loin</i>)								
-22 ^d	2 ^a	1 ^a	19 ^d	7 ^d	-1 ^a	-6 ^d	-7 ^e	
Effet de la séquence (<i>/tp/ - /pt/</i>)								
-37 ^e	-7 ^a	-20 ^d	1 ^a	13 ^e	2 ^b	4 ^c	4 ^b	
($accent_1 - accent_2$) selon la séquence								
/tp/	88	124	173	152	-12 ^{b'}	-46	-49	-13 ^{d'}
/pt/	75	117	190	152	-6 ^{a'}	-46	-535	-22 ^{f'}
p_{inter}	a	a	b	a	c	a	b	c
($accent_1 - accent_2$) selon la position de la cible								
proche	79	118	178	150	-9	-45	-51	-18
loin	83	124	185	155	-8	-47	-51	-18
p_{inter}	a	a	a	a	a	a	a	a
(<i>/tp/ - /pt/</i>) selon la position de la cible								
proche	-38	-6	-17	3	14	2	3	4
loin	-35	-9	-23	-1	12	3	4	4
p_{inter}	a	a	a	a	a	a	a	a
($accent_1 - accent_2$) selon la séquence et la position de la cible								
/tp/.proche	82	119	167	146	-12	-44	-48	-13
/tp/.loin	94	130	179	158	-12	-48	-50	-14
/pt/.proche	76	117	189	154	-7	-46	-54	-22
/pt/.loin	73	118	191	151	-5	-46	-51	-21
p_{inter}	a	a	a	a	a	a	b	a

Anova : ^{a)} $p < 1$; ^{b)} $p < .1$; ^{c)} $p < .05$; ^{d)} $p < .01$; ^{e)} $p < .001$; ^{f)} $p < .0001$

Comparaisons multiples : ^{a')} $p_{bf} < 1$; ^{b')} $p_{bf} < .1$; ^{c')} $p_{bf} < .05$; ^{d')} $p_{bf} < .01$; ^{e')} $p_{bf} < .001$

A.2.3 Paramètres de la trajectoire du doigt

Les tableaux A.6 et A.7, étudient l'effet des différents facteurs sur les paramètres de la trajectoire du doigt. D_P , A_P et V_P sont la durée, l'amplitude et l'amplitude du pic de vitesse du geste de pointé. D_{PR} , A_{PR} et V_{PR} sont la durée, l'amplitude et l'amplitude du pic de vitesse du geste de retour du doigt. D_{PT} est la durée de la tenue du geste, voir figure 5.4, p.103 pour le calcul de ces paramètres.

TAB. A.6 – Expérience 1 (position de l'accent et coordination doigt-mâchoire), étude des effets de la position de l'accent, de la cible et de la séquence sur les paramètres de la trajectoire du doigt.

MESURES							
Geste pointé			Tenue	Geste retour			
D_P	A_P	V_P	D_{PT}	D_P	A_P	V_{PR}	
ms	mm	mm/s	ms	ms	mm	mm/ms	
Moyennes globales							
419	349	1394	142	547	346.1	-1126	
Effet de la position de l'accent ($accent_1 - accent_2$)							
-5 ^a	-2.5 ^d	3 ^a	-30 ^d	-5 ^a	-2 ^c	1 ^a	
Effet de la position de la cible (<i>proche - loin</i>)							
-25 ^f	-149 ^f	-538 ^f	-20 ^d	-14 ^b	-146.2 ^f	466 ^f	
Effet de la séquence (/tt/ - /pp/)							
3 ^c	0.6 ^a	-4 ^a	3 ^a	2 ^a	0.3 ^a	2 ^a	
($accent_1 - accent_2$) selon la consonne							
/t/	-5	-3	3	-29	-3	-2.7	7
/p/	-5	-2	2	-30	-8	-1.3	-4
p_{inter}	a	a	a	a	b	a	a
($accent_1 - accent_2$) selon la position de la cible							
proche	-4	-2.9	-5	-30	-8	-2.2	6
loin	-6	-2	10	-30	-2	-1.7	-3
p_{inter}	a	a	a	a	a	a	a
(/tt/ - /pp/) selon la position de la cible							
proche	2	1.1	1	0	6 ^{c'}	0.9	6
loin	4	0.2	-9	6	-2 ^{d'}	-0.2	-2
p_{inter}	a	a	a	a	c	a	a
($accent_1 - accent_2$) selon la séquence et la position de la cible							
/tt/.proche	-2 ^{d'}	-3.1	-15 ^{d'}	-28	-7	-2.7	7
/tt/.loin	-8 ^{d'}	-2.9	21 ^{d'}	-30	2	-2.7	7
/pp/.proche	-6 ^{d'}	-2.8	6 ^{d'}	-31	-9	-1.8	5
/pp/.loin	-3 ^{d'}	-1.2	-2 ^{d'}	-29	-7	-0.7	-13
p_{inter}	c	a	c	a	a	a	a

Anova : ^{a)} $p < 1$; ^{b)} $p < .1$; ^{c)} $p < .05$; ^{d)} $p < .01$; ^{e)} $p < .001$; ^{f)} $p < .0001$

Comparaisons multiples : ^{d')} $p_{bf} < 1$; ^{b')} $p_{bf} < .1$; ^{c')} $p_{bf} < .05$; ^{d')} $p_{bf} < .01$; ^{e')} $p_{bf} < .001$

TAB. A.7 – Expérience 2 (profils de mâchoire et coordination doigt-mâchoire), étude des effets de la position de l'accent, de la cible et de la séquence sur les paramètres de la trajectoire du doigt.

MESURES						
D_P	A_P	V_P	D_{PT}	D_{PR}	A_{PR}	V_{PR}
ms	mm	mm/s	ms	ms	mm	mm/ms
Moyennes globales						
412	345.6	1425	137	541	342.9	-1129
Effet de la position de l'accent ($accent_1 - accent_2$)						
-4 ^a	-1.7 ^c	4 ^a	-22 ^e	-2 ^a	-1.5 ^b	3 ^a
Effet de la position de la cible (<i>proche - loin</i>)						
-21 ^f	-146.7 ^f	-552 ^f	-19 ^d	-15 ^c	-145.2 ^f	462 ^f
Effet de la séquence, ($/tp/ - /pt/$)						
-2 ^a	-0.6 ^a	-0.4 ^a	4 ^c	-5 ^a	-0.9 ^b	-8 ^a
$(accent_1 - accent_2)$ selon la séquence						
/tp/	-4	-1.6	-2	-19	-4	-1.4
/pt/	-4	-1.8	10	-24	0	-1.6
p_{inter}	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>
$(accent_1 - accent_2)$ selon la position de la cible						
proche	-2	-0.4	3	5	-4	-0.9
loin	-1	-0.8	-3	3	-6	-0.9
p_{inter}	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>
$(/tp/ - /pt/)$ selon la position de la cible						
proche	2	1.1	1	0	6 ^{c'}	0.9
loin	4	0.2	-9	6	-2 ^{a'}	-0.2
p_{inter}	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>c</i>	<i>a</i>
$(accent_1 - accent_2)$ selon la séquence et la position de la cible						
/t/.proche	-3	-1.3	-8	-20	-3	-1.2
/t/.loin	-5	-1.9	4	-19	-6	-1.6
/p/.proche	-3	-0.8	5	-29	3	-0.1
/p/.loin	-5	-2.9	15	-20	-3	-3
p_{inter}	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>

Anova : ^{a)} $p < 1$; ^{b)} $p < .1$; ^{c)} $p < .05$; ^{d)} $p < .01$; ^{e)} $p < .001$; ^{f)} $p < .0001$

Comparaisons multiples : ^{a')} $p_{bf} < 1$; ^{b')} $p_{bf} < .1$; ^{c')} $p_{bf} < .05$; ^{d')} $p_{bf} < .01$; ^{e')} $p_{bf} < .001$

A.2.4 Paramètres de la trajectoire de la mâchoire

Les tableaux A.8 et A.9, étudient l'effet des différents facteurs sur les paramètres de la trajectoire de la mâchoire. D_{Mn} , A_{Mn} et V_{Mn} sont la durée, l'amplitude et l'amplitude du pic de vitesse de la phase d'ouverture ($n = 1$) et de fermeture ($n = 2$) du premier cycle et de la phase d'ouverture ($n = 3$) et de fermeture ($n = 4$) du deuxième cycle, voir figure 5.4, p.103 pour le calcul de ces paramètres.

TAB. A.8 – Expérience 1 (position de l'accent et coordination doigt-mâchoire), étude des effets de la position de l'accent, de la cible et de la séquence sur les paramètres de la trajectoire de la mâchoire.

MESURES												
PREMIER CYCLE						DEUXIÈME CYCLE						
OUVERTURE			FERMETURE			OUVERTURE			FERMETURE			
D_{M1}	A_{M1}	V_{M1}	D_{M2}	A_{M2}	V_{M2}	D_{M3}	A_{M3}	V_{M3}	D_{M4}	A_{M4}	V_{M4}	
ms	mm	mm/s	ms	mm	mm/s	ms	mm	mm/s	ms	mm	mm/s	
Moyennes globales												
166	11	136	150	9	119	134	5.4	68	282	9.3	66	
Effet de la position de l'accent (1 – 2)												
40 ^f	2.9 ^e	9 ^a	59 ^f	4.1 ^f	24 ^d	-60 ^f	-4.4 ^f	-40 ^f	-64 ^f	-4.7 ^f	-23 ^f	
Effet de la position de la cible (<i>proche – loin</i>)												
0	0.1 ^a	2 ^a	2 ^a	0.1 ^a	0	0	0	-1 ^a	-1 ^a	0	0	
Effet de la séquence (/tt/ – /pp/)												
37 ^d	-1.4 ^d	-32 ^e	25 ^f	-1.7 ^d	-43 ^f	-21 ^f	-2.7 ^f	-31 ^f	-9 ^a	-1.9 ^e	15 ^f	
<i>(accent₁ – accent₂) selon la séquence</i>												
/tt/	32 ^e	2.7	15 ^d	71 ^f	4	17 ^c	-73 ^f	-4.9 ^f	-52 ^f	-65	-5.3 ^f	-29 ^f
/pp/	49 ^f	3.1	3 ^d	46 ^f	4.2	30 ^f	-48 ^f	-4 ^f	-28 ^f	-63	-4.1 ^f	-17 ^f
<i>Pinter</i>	<i>c</i>	<i>a</i>	<i>d</i>	<i>d</i>	<i>a</i>	<i>c</i>	<i>d</i>	<i>c</i>	<i>e</i>	<i>a</i>	<i>d</i>	<i>e</i>
<i>(accent₁ – accent₂) selon la position de la cible</i>												
proche	41	3	10	59	4.1	23	-61	-4.5	-42	-67	-4.8	-23
loin	40	2.9	9	58	4.1	24	-59	-4.3	-39	-61	-4.6	-23
<i>Pinter</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>b</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>
<i>(/tt/ – /pp/) selon la position de la cible</i>												
proche	37	-1.4	-32	27	-1.7	-43	-21	-2.7	-32	-9	-2.1 ^f	-16
loin	38	-1.3	-31	23	-1.8	-43	-21	-2.7	-31	-8	-1.7 ^f	-14
<i>Pinter</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>b</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>c</i>	<i>b</i>
<i>(accent₁ – accent₂) selon la séquence et la position de la cible</i>												
/tt/.proche	32	2.7	15	74	4.1	17	-76	-5	-54	-69	-5.4	-30
/tt/.loin	32	2.8	16	69	4	18	-70	-4.7	-51	-61	-5.2	-28
/pp/.proche	50	3.3	5	45	4.2	30	-47	-4	-29	-65	-4.2	-17
/pp/.loin	48	3	2	47	4.3	31	-49	-3.9	-27	-61	-4.1	-17
<i>Pinter</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>b</i>	<i>a</i>	<i>a</i>	<i>b</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>

Anova : ^{a)} $p < .1$; ^{b)} $p < .1$; ^{c)} $p < .05$; ^{d)} $p < .01$; ^{e)} $p < .001$; ^{f)} $p < .0001$

Comparaisons multiples : ^{a')} $p_{bf} < .1$; ^{b')} $p_{bf} < .1$; ^{c')} $p_{bf} < .05$; ^{d')} $p_{bf} < .01$; ^{e')} $p_{bf} < .001$

TAB. A.9 – Expérience 2 (profils de mâchoire et coordination doigt-mâchoire), étude des effets de la position de l'accent, de la cible et de la séquence sur les paramètres de la trajectoire de la mâchoire.

MESURES												
PREMIER CYCLE						DEUXIÈME CYCLE						
OUVERTURE			FERMETURE			OUVERTURE			FERMETURE			
D_{M1}	A_{M1}	V_{M1}	D_{M2}	A_{M2}	V_{M2}	D_{M3}	A_{M3}	V_{M3}	D_{M4}	A_{M4}	V_{M4}	
ms	mm	mm/s	ms	mm	mm/s	ms	mm	mm/s	ms	mm	mm/s	
Moyennes globales												
167	10.7	131	152	8.7	113	130	4.7	61	285	8.9	62	
Effet de la position de l'accent (1 – 2)												
36 ^f	3.1 ^f	13 ^d	61 ^f	4.3 ^f	24 ^e	-51 ^f	-3.8 ^f	-37 ^f	-58 ^e	-4.1 ^f	-20 ^f	
Effet de la position de la cible (<i>proche – loin</i>)												
2 ^a	0.3 ^c	2 ^a	-1 ^a	0	0	-1 ^a	0	0	5 ^b	0.1 ^a	0	
Effet de la séquence (<i>/tp/ – /pt/</i>)												
28 ^d	-0.1	-10 ^a	-13 ^d	2.3 ^e	44 ^f	25 ^f	3.1 ^f	37 ^f	4 ^a	1.6 ^f	14 ^f	
<i>(accent₁ – accent₂)</i> selon les consonnes												
/tp/	33	2.8	11	48 ^f	3.8 ^f	20	-40 ^f	-3.7	-31 ^f	-58	-3.6 ^f	-14 ^f
/pt/	39	3.3	14	73 ^f	4.9 ^f	28	-62 ^f	-3.8	-43 ^f	-58	-4.6 ^f	-26 ^f
<i>pinter</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>d</i>	<i>d</i>	<i>b</i>	<i>d</i>	<i>a</i>	<i>d</i>	<i>a</i>	<i>f</i>	<i>f</i>
<i>(accent₁ – accent₂)</i> selon la position de la cible												
proche	36	3.2 ^f	15 ^e	60	4.5	26 ^f	-53 ^f	-3.8	-38	-54	-4.1	-20
loin	36	2.9 ^f	11 ^d	61	4.2	22 ^f	-49 ^f	-3.7	-37	-61	-4.1	-20
<i>pinter</i>	<i>a</i>	<i>c</i>	<i>c</i>	<i>a</i>	<i>b</i>	<i>c</i>	<i>c</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>
<i>(/tp/ – /pt/)</i> selon la position de la cible												
proche	30	0	-10	-11	2.3	44	25	3.2	38	6	1.8	14
loin	25	-0.2	-9	-14	2.3	44	26	3	36	2	1.5	13
<i>pinter</i>	<i>a</i>	<i>b</i>	<i>a</i>	<i>a</i>	<i>b</i>	<i>a</i>						
<i>(accent₁ – accent₂)</i> selon la séquence et la position de la cible												
/tp/.proche	35	2.9	13	49	3.9	21	-41	-3.7	-32	-55	-3.5	-14
/tp/.loin	32	2.7	10	48	3.7	19	-39	-3.7	-31	-60	-3.6	-14
/pt/.proche	38	3.5	17	72	5.1	30	-64	-3.8	-43	-53	-4.6	-26
/pt/.loin	40	3.1	12	74	4.7	25	-60	-3.8	-42	-62	-4.7	-26
<i>pinter</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>						

Anova : ^{a)} $p < 1$; ^{b)} $p < .1$; ^{c)} $p < .05$; ^{d)} $p < .01$; ^{e)} $p < .001$; ^{f)} $p < .0001$

Comparaisons multiples : ^{a')} $p_{bf} < 1$; ^{b')} $p_{bf} < .1$; ^{c')} $p_{bf} < .05$; ^{d')} $p_{bf} < .01$; ^{e')} $p_{bf} < .001$

Fréquence d'oscillation de la mâchoire et du geste de pointé brachiomanuel - synthèse des ANOVAs

NOTE POUR LA LECTURE DES TABLEAUX

Les tableaux ci-après synthétisent les différentes ANOVAs réalisées pour l'étude testant l'hypothèse d'un rapport 2 :1 entre la fréquence d'oscillation de la mâchoire dans la parole et celle du doigt dans le pointé brachiomanuel présentées au chapitre 6. La première table résume les ANOVAs pour les variables dépendantes n'impliquant pas l'ordre de désignation, à deux facteurs intrasujets *Nombre de syllabes* × *Position de la cible*. La deuxième table résume les ANOVAs pour les variables impliquant l'ordre de désignation et donc, comportant trois facteurs intrasujets *Nombre de syllabes* × *Position de la cible* × *Ordre de désignation*. Les deux tableaux étudient, pour chaque mesure :

- L'effet du nombre de syllabes, avec les différences $1_{\text{syll.}} - 2_{\text{syll.}}$, $2_{\text{syll.}} - 3_{\text{syll.}}$, $3_{\text{syll.}} - 4_{\text{syll.}}$. Quand l'effet global du nombre de syllabe est significatif. Ces différences sont testées avec des t-tests avec correction de Bonferroni.
- L'effet de la position de la cible, différence moyenne entre les conditions *cible_{proche}* et *cible_{loin}*.
- L'effet du nombre de syllabes selon la position de la cible.

Le deuxième tableau étudie, en plus, l'effet de l'ordre de désignation et son interaction avec les deux autres facteurs. Les interactions triples n'étant jamais significatives, elles ne sont pas étudiées ici. Certaines des variables analysées ici ne sont pas mentionnées dans les analyses présentées au chapitre 6. Elles sont données ici comme informations supplémentaires.

B.1 ANOVAS POUR LES VARIABLES MESURÉES NE DÉPENDANT PAS DE L'ORDRE DE DÉSIGNATION

TAB. B.1 – Etude des effets du nombre de syllabes et de la position de la cible pour les mesures ne dépendant pas de l'ordre de désignation. P_{I1} et M_{I1} sont les temps de réponse du geste de pointé et de la mâchoire ; P_{A1} et P_{A2} sont les occurrences de l'apex du premier et du deuxième geste de pointé ; M_{A1} et M_{A2} sont les occurrences de l'apex du dernier geste d'ouverture de la mâchoire pour la première et la deuxième énonciation du nom de la cible ; P_{R1} et P_{F1} sont le début et la fin du premier geste de retour et P_{I2} est le début du deuxième geste de pointé. Ainsi, $M_{I1} - P_{I1}$ mesure le délai entre la réponse de la mâchoire et celle du doigt, $P_{A2} - P_{A1}$ mesure le délai entre les deux apex des gestes de pointé, $P_{R1} - P_{A1}$ mesure la durée de la tenue du premier geste de pointé, $P_{F1} - P_{R1}$ mesure la durée du premier geste de retour et $P_{I2} - P_{F1}$ mesure le délai entre les deux cycles du doigt.

MESURES								
	P_{I1} ms	M_{I1} ms	$M_{I1} - P_{I1}$ ms	$P_{A2} - P_{A1}$ ms	$M_{A2} - M_{A1}$ ms	$P_{R1} - P_{A1}$ ms	$P_{F1} - P_{R1}$ ms	$P_{I2} - P_{F1}$ ms
Moyennes globales								
	393.7	529	135.3	1000.7	965.6	33.6	407.6	194.9
Effet du nombre de syllabes								
1 _{syll.} - 2 _{syll.}	-59.4 ^{c'}	46.7 ^{d'}	106.1 ^{d'}	-0.6 ^{a'}	7.7 ^{d'}	-0.1 ^{a'}	-0.8 ^{a'}	-1.6 ^{d'}
2 _{syll.} - 3 _{syll.}	-11.2 ^{a'}	25.7 ^{a'}	36.9 ^{c'}	-62.7 ^{c'}	-65.6 ^{b'}	-10.3 ^{a'}	-19.4 ^{b'}	-20.8 ^{a'}
3 _{syll.} - 4 _{syll.}	18.4 ^{a'}	45.2 ^{b'}	26.8 ^{b'}	-33.2 ^{a'}	-39.4 ^{a'}	-5.9 ^{a'}	-12.5 ^{a'}	-8.9 ^{a'}
	p	d	e	f	e	d	d	e
Effet de la position de la cible (<i>proche - loin</i>)								
	-1 ^a	-6.1 ^a	-5.2 ^a	-43.4 ^d	-28.2 ^c	1.2 ^a	5.1 ^a	-20.4 ^a
Effet du nombre de syllabes selon la position de la cible								
1 _{syll.} - 1 _{syll.}								
Proche	-69.9	44.7	114.6	-13.4	-5.3	-1.4	-3.4	-7.4
Loin	-49	48.7	97.7	12.2	20.8	1.2	1.9	4.2
2 _{syll.} - 3 _{syll.}								
Proche	-9.9	24.4	34.3	-68.1	-70.3	-14.4	-21.6	-21.8
Loin	-12.5	27	39.5	-57.3	-60.9	-6.2	-17.1	-19.7
3 _{syll.} - 4 _{syll.}								
Proche	19.4	49.4	30	-19.4	-34.2	-0.2	-9.2	-9.8
Loin	17.4	40.9	23.5	-47	-44.6	-11.6	-15.8	-8.1
<i>P_{inter}</i>	a	a	a	a	a	a	a	a

Anova : ^{a)} $p < 1$; ^{b)} $p < .1$; ^{c)} $p < .05$; ^{d)} $p < .01$; ^{e)} $p < .001$; ^{f)} $p < .0001$
 Comparaisons multiples ($p_{bf} = p * 3$) : ^{a')} $p_{bf} < 1$; ^{b')} $p_{bf} < .1$; ^{c')} $p_{bf} < .05$; ^{d')} $p_{bf} < .01$; ^{e')} $p_{bf} < .001$

B.2 ANOVAS POUR LES VARIABLES MESURÉES DÉPENDANT DE L'ORDRE DE DÉSIGNATION

TAB. B.2 – Etude des effets du nombre de syllabes et de la position de la cible sur les paramètres de la trajectoire du doigt et de la mâchoire, pour les deux désignations. D_{Pn} , A_{Pn} et V_{Pn} sont la durée, l'amplitude et l'amplitude du pic de vitesse du premier ($n = 1$) et du deuxième ($n = 2$) gestes de pointé. De la même manière, D_{Tn} est la durée de la tenue de chacun des deux gestes et D_{PRn} , A_{PRn} et V_{PRn} sont la durée, l'amplitude et le pic de vitesse des deux gestes de retour. Pour la mâchoire, D_{Mn} et A_{Mn} sont la durée et l'amplitude pour la première et la deuxième énonciation. $D_{Mn} - D_{Pn}$ est la différence entre la durée de l'énonciation et celle du geste de pointé.

	MESURES									
	DOIGT							MÂCH.		
	D_{Pn} ms	A_{Pn} mm	V_{Pn} mm/s	D_{Tn} ms	D_{PRn} ms	A_{PRn} mm	V_{PRn} mm/s	D_{Mn} ms	A_{Mn} mm	$D_{Mn} - D_{Pn}$ ms
	Moyennes globales									
	364.2	348.3	1545.7	34.5	435.2	347.1	1371.6	433.9	9.8	69.7
	Effet du nombre de syllabes									
1 _{syll.} - 2 _{syll.}	0.4 ^{a'}	-2.7 ^{a'}	-6.7	-0.1 ^{a'}	0.7 ^{a'}	-2.9 ^{a'}	-19.1	-202.4 ^{f'}	1.2 ^{a'}	-202.8 ^{f'}
2 _{syll.} - 3 _{syll.}	-13.9 ^{a'}	-7.4 ^{b'}	21.5	-8.4 ^{a'}	-12.9 ^{b'}	-6.7 ^{a'}	7.4	-132.5 ^{f'}	0.3 ^{a'}	-118.6 ^{f'}
3 _{syll.} - 4 _{syll.}	-6.4 ^{a'}	-3 ^{a'}	-2.5	-1.7 ^{a'}	-9.7 ^{a'}	-2.8 ^{a'}	22.8	-114.5 ^{f'}	0.7 ^{a'}	-108.1 ^{f'}
p	c	e	a	c	d	d	a	f	d	f
	Effet de la position de la cible (<i>proche - loin</i>)									
	-26.3 ^d	-144.5 ^f	-548.1 ^f	2.5 ^a	6.5 ^a	-143.5 ^f	-605.9 ^f	-4.4 ^b	0 ^a	21.9 ^c
	Effet de la désignation (<i>dsignation₁ - dsignation₂</i>)									
	-0.9 ^a	-9.8 ^d	-57.5 ^b	-1.7 ^a	-55.1 ^d	-4.1 ^a	144.3 ^d	14.3 ^c	1.1 ^b	15.1 ^c
	Effet du nombre de syllabes selon la position de la cible									
	1 _{syll.} - 2 _{syll.}									
Proche	-0.5 ^{a'}	-4.8	-14.3 ^{a'}	-1.1	-1.1	-4.5	-10.9	-204.4	1.1	-203.9
Loin	1.3 ^{a'}	-0.6	0.9 ^{a'}	0.9	2.5	-1.3	-27.3	-200.4	1.3	-201.7
	2 _{syll.} - 3 _{syll.}									
Proche	-11.8 ^{a'}	-8.1	-9.5 ^{a'}	-11.1	-13.7	-7.9	-13.3	-130.6	0.5	-118.8
Loin	-16.1 ^{a'}	-6.6	52.6 ^{a'}	-5.7	-12.2	-5.4	28	-134.5	0.1	-118.4
	3 _{syll.} - 4 _{syll.}									
Proche	-3.1 ^{a'}	-1	-0.4 ^a	0.4	-5.9	-1.2	15.8	-115.4	0.4	-112.3
Loin	-9.7 ^{a'}	-5	-4.6 ^{a'}	-3.9	-13.5	-4.4	29.7	-113.6	0.9	-103.9
p_{inter}	c	a	d	a	a	a	a	a	a	a
	Effet du nombre de syllabes selon la désignation									
	1 _{syll.} - 2 _{syll.}									
Proche	-0.5	-4.8 ^{a'}	-14.3	-1.1 ^{a'}	-1.1 ^{a'}	-4.5 ^{a'}	-10.9	-204.4 ^{f'}	1.1	-203.9 ^{f'}
Loin	1.3	-0.6 ^{a'}	0.9	0.9 ^{a'}	2.5 ^{a'}	-1.3 ^{a'}	-27.3	-200.4 ^{f'}	1.3	-201.7 ^{f'}
	2 _{syll.} - 3 _{syll.}									
Proche	-11.8	-8.1 ^{b'}	-9.5	-11.1 ^{a'}	-13.7 ^{b'}	-7.9 ^{b'}	-13.3	-130.6 ^{f'}	0.5	-118.8 ^{f'}
Loin	-16.1	-6.6 ^{b'}	52.6	-5.7 ^{a'}	-12.2 ^{a'}	-5.4 ^{a'}	28	-134.5 ^{f'}	0.1	-118.4 ^{f'}
	3 _{syll.} - 4 _{syll.}									
Proche	-3.1	-1 ^{b'}	-0.4	0.4 ^{a'}	-5.9 ^{a'}	-1.2 ^{a'}	15.8	-115.4 ^{f'}	0.4	-112.3 ^{f'}
Loin	-9.7	-5 ^{a'}	-4.6	-3.9 ^{a'}	-13.5 ^{a'}	-4.4 ^{a'}	29.7	-113.6 ^{f'}	0.9	-103.9 ^{f'}
p_{inter}	a	c	a	e	c	d	a	e	a	d
	Effet de la position de la cible selon la désignation (<i>proche - loin</i>)									
<i>des.1</i>	2.2	-10.8	-83.3	-3	-56.5	-2.8	101.3 ^{f'}	14.2	1.1	12
<i>des.2</i>	-3.9	-8.8	-31.7	-0.3	-53.8	-5.4	187.2 ^{f'}	14.4	1	18.3
p_{inter}	a	a	b	a	a	a	d	a	a	a

Anova : ^{a)} $p < 1$; ^{b)} $p < .1$; ^{c)} $p < .05$; ^{d)} $p < .01$; ^{e)} $p < .001$; ^{f)} $p < .0001$

Comparaisons multiples ($p_{bf} = p * 3$) : ^{a')} $p_{bf} < 1$; ^{b')} $p_{bf} < .1$; ^{c')} $p_{bf} < .05$; ^{d')} $p_{bf} < .01$; ^{e')} $p_{bf} < .001$