



Multimodalité de la communication langagière humaine : interaction geste/parole et encodage de distance dans le pointage

Chloe Gonseth

► To cite this version:

Chloe Gonseth. Multimodalité de la communication langagière humaine : interaction geste/parole et encodage de distance dans le pointage. Médecine humaine et pathologie. Université de Grenoble, 2013. Français. NNT : 2013GRENS011 . tel-00949090

HAL Id: tel-00949090

<https://theses.hal.science/tel-00949090>

Submitted on 19 Feb 2014

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

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

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Sciences Cognitives, Psychologie Cognitive & Neurocognition**

Arrêté ministériel : 7 août 2006

Présentée par

CHLOE GONSETH

Thèse dirigée par **Anne VILAIN & Jean-Luc SCHWARTZ**
et encadrée par **Coriandre VILAIN**

Préparée au sein du **Laboratoire Grenoble, Images, Parole, Signal & Automatique (GIPSA-Lab, UMR 5216)**
et de l'**Ecole Doctorale Ingénierie pour la Santé, la Cognition et l'Environnement**

Multimodalité de la communication langagière humaine :

Interaction geste/parole et encodage de la distance dans le pointage

Thèse soutenue publiquement le **12 Mars 2013**,
devant le jury composé de :

Noël NGUYEN

Professeur, Université Aix-Marseille, Rapporteur

Jacques VAUCLAIR

Professeur, Université Aix-Marseille, Rapporteur

Jean-Marc COLLETTA

Professeur, Université de Grenoble, Examineur

Aslı ÖZYÜREK

Professeur, Radboud University, Examinatrice

Coriandre VILAIN

Ingénieur de Recherche, GIPSA-Lab, Invité

Jean-Luc SCHWARTZ

Directeur de Recherche, GIPSA-Lab, Directeur de thèse

Anne VILAIN

Maître de Conférence, Université de Grenoble, Co-Directeur de thèse



A Mamie,

• *Personne ne peut faire pour les enfants ce que font les grands-parents.*
Ceux-ci répandent une espèce de poudre d'étoiles sur leur vie. •

Alex Haley.

Remerciements

Ce travail de thèse, financé par une allocation de recherche du Ministère de l'Enseignement supérieur et de la Recherche, a été réalisé au sein du Gipsa-Lab (CNRS UMR 5216) et de l'Ecole Doctorale Ingénierie pour la Santé, la Cognition et l'Environnement. Il est le fruit de rencontres professionnelles et personnelles particulières, et je souhaite saluer toutes les personnes qui, de près ou de plus loin, volontairement ou à l'insu de leur plein gré, ont contribué à l'écriture de ce manuscrit (tout oubli, bien sûr fortuit, sera, je l'espère, pardonné!).

C'est tout d'abord avec beaucoup de fierté que je remercie les membres de mon jury : **Jacques Vauclair** et **Noël Nguyen**, dans le rôle des rapporteurs, **Aslı Özyürek** et **Jean-Marc Colletta** dans le rôle des examinateurs, **Anne Vilain**, **Coriandre Vilain** et **Jean-Luc Schwartz** dans celui des directeurs de thèse.

Je remercie donc chaleureusement **Jacques Vauclair** et **Noël Nguyen** pour leur investissement dans ce travail et leurs remarques avisées. Un immense merci à **Aslı Özyürek**, pour avoir évalué ce travail et fait un tel déplacement, *thank you so much*! Un grand merci également à **Jean-Marc Colletta**, notamment pour avoir présidé cette soutenance, et ce de façon rassurante et bienveillante!

Je remercie chaudement **Jean-Luc Schwartz**, pour sa présence discrète mais réconfortante, ses conseils éclairés et ses nombreux encouragements au cours de ces dernières années.

Enfin, j'aimerais remercier dignement le duo légendaire formé par **Anne Vilain** et **Coriandre Vilain**! Un simple « Merci pour tout, je m'ai bien amusée » ne saurait exprimer pleinement ma reconnaissance! **Anne** et **Coriandre**, mille fois merci pour votre accueil, votre fantaisie et votre enthousiasme; mille fois merci pour vos encouragements et votre soutien, votre confiance et votre aide précieuse (avant, pendant et après la thèse)! Je suis fière d'avoir (tant) appris à vos côtés, et du chemin, mouvementé et passionnant, parcouru depuis le M2, je garde de formidables souvenirs (comme, parmi tant d'autres, celui du 12 Janvier 2013). Quand je serai grande, je veux être enseignant-chercheur... Merci à vous de m'avoir *montré la voie*!

Je remercie également mes collègues du Gipsa-Lab, tous départements confondus, pour m'avoir offert un cadre de travail riche et enrichissant. Je ne peux malheureusement tous les citer mais le cœur y est! Merci à l'ensemble du personnel, et plus particulièrement à **Akila**, **Houria**, **Jessica**, **Nadine** et **Lucia**, non seulement pour leurs compétences mais surtout pour leur indulgence et leur patience!

Un immense merci, bien qu'un peu nostalgique, aux membres du GipsADoc-ex-GAG, pour les moments inoubliables passés ensemble! Pour cette belle aventure associative, merci donc à **Antoine**, **Olha**, **Vincent**, **Yo** (du courage!), **Jérémie**, et tous les autres! Une pensée toute particulière pour **Gabriel**, dit Docteur GRdC, ou Crêpe, qui, je l'espère, n'oublie pas son français si... singulier! Une pensée toute aussi particulière à **Emilie**, dite Miss E, mon Mimi Doc Emilie, parce que 36 kg de foie gras, c'est tout de même un tantinet excessif (comment ça, on a prévu trop de foie gras?)!

Mes remerciements vont également aux membres du Département Parole & Cognition en général, et aux membres de l'équipe Parole, Cerveau, Multimodalité & Développement, en particulier. Mes sincères remerciements à **Benjamin**, notamment pour m'avoir accompagnée au bout du monde, mais aussi pour son extrême gentillesse et son assistance à la fois technique et psychologique !

Je remercie du fond du cœur les Elus du Bâtiment E ! Des personnes extraordinaires, qui m'ont assistée, épaulée, encouragée... Merci à chacune d'entre elle. Une pensée émue pour les adhérents de l'ALCCCJN (l'Association de Lutte Contre la Consommation de Cassoulet le Jour de Noël), ayant courageusement lutté, mais point vaincu, à mes côtés. Un grand merci à **Christophe** (gardien de la BAF, tragiquement dévalisée courant décembre), **Nathalie** (vive la Savoie et l'tarte al chuc), **Giovanni**, **Carole**, **Silvia**, **Didier** et **Marc**. Un grand merci aux ex-Stendhaliens, mon **Jeanne** (TMTC, je t'm trop bb), **Tom**, **Lionel**, **Krystyna** et **Hien**, ma très chère (ex) cobureau. Un grand merci également aux stagiaires et doctorants, auxquels je souhaite beaucoup de courage pour la rédaction de leur chef-d'œuvre et de passion dans leurs recherches ! Je pense notamment à **Diane**, **Ibrahima-Doki** et son coq aux yeux verts, **Léa**, **Mailyn**, **Manon**, **Rita**, **Yan** et **Xuan** !

Spéciale cace-dédi et mille mercis à **Avril** et à mon **Lu** (dont la particularité est de ne faire qu'un avec Ernest, son vélo), sans oublier **Miette**. Une belle amitié, née, entre autres, de notre passion commune pour l'absurde, les pingouins et les lutins, les Klix et les théories fumeuses (petit clin d'œil, discret, à **Laura**)... De quoi alimenter de longues discussions, si ce n'est intellectuelles, à coup sûr complètement ou-ouf !

J'en profite pour saluer chocolatement le JECL, dont les membres, sus-cités, se reconnaîtront, et dont le symbole pourrait bien être un macaron au bord d'une piscine (ou une chatte).

Enfin, et parce qu'il est unique en son genre, qu'il roule en trottinette et ride la peuf (prononcer [pyf]), je remercie **Rosario**, mon co-bureau-préférée-adoré ! Ce fut un honneur de partager l'E011 avec toi, *carissimo amico mio*. *Grazie mille* ! Que de bons moments passés ensemble... et à venir, je l'espère ! *Coraggio* (et non *corragio*), Rosamio ! *L'avvenire è pieno di sorprese* !

Je remercie de tout cœur chacun de mes participants, petits et grands, et notamment mon neveu Gaspard et ma nièce Iana ! Famille, amis ou collègues, ils ont bravement revêtu, par amour de la Science (si, si), leur costume de "cobaye" ! De beaux souvenirs, entre Chupa Chups et paillettes... Une mention spéciale aux plus jeunes pour leur joie de vivre et leur curiosité, mais également pour leur courage... L'infirmière Chloe a (temporairement) terminé son travail !

Je remercie chaleureusement mes collègues de l'UPMF, et plus particulièrement ceux du Laboratoire de Psychologie et NeuroCognition, ainsi que, j'en profite, mes étudiants de Psychologie et de Sciences du Langage !

Un grand merci, entre autres, à ma tutrice de monitorat, **Françoise Bonthoux** ainsi qu'à **Elsa Spinelli**. Mes pensées vont également à **Muriel**, **Solène**, **Lucie**, **Benoît**, **Nicolas**, **Anne**, **Jennifer** et **Mathilde**.

Un merci plus personnel à **Benoit**, DCE 2009, ou devrais-je dire, Docteur Musel, B. ! Jouons donc de la flûte (sur un air de *Felicità*), paraît que ça revitalise ! Merci pour tout ce que l'on a partagé, des cafés aux doutes, en passant par de nombreux coups de gueule et éclats de rire !

Je remercie solennellement mon Bestah, le Docteur **Stéphane**, dit Stéphan. Un grand merci également aux Docteurs du 52, **Stéphane**, dit Zouave, **Aurélien**, dit Poual, **Sylvain**, dit Bou, ainsi qu'à mes co-rédacteurs, **Pierre**, dit Grand et **Rémi**, dit Boulou. Je n'oublie pas le futur Docteur **Christophe**, dit le Prez ! La rédaction d'une thèse, c'est un peu comme une partie de Catane... Un peu long, certes, mais au final ce sera toi le grand vainqueur ! Merci à tous, scientifiques durs que vous êtes... Je ne suis pas peu fière de représenter les sciences semi-molles au sein de l'ultime crew ! Une pensée également pour **Alex**, **Camille** et **Céline** (♡ LMF ♡).

Mille milliards de mille mercis à mon coupain **Guillaume-GL**, dit le Perceptron, ainsi qu'à mon **Guillaume-GG** et mon **Anne-Lise-AL** ! Merci pour votre présence-à-distance et vos délicates attentions ! Je vous chéritise fortement mes Sciences-Co.

Un immense merci à mon Numero Uno, **Eliane**, pour être une amie d'exception, une confidente, et bien plus encore. Moi, Childe de Croÿ Chanel, te remercie sincèrement d'être à mes côtés, et de mon côté, depuis toutes ces années. Mes pensées vont également à **Anaïs**, loin des yeux mais dans mon cœur, ainsi qu'à **Céline**, mon petit chou.

Je tiens aussi à remercier toutes les personnes présentes, ne serait-ce que par la pensée, le 12 Mars 2013 ! Qu'il est doux d'être (aussi bien) entourée dans un moment comme celui-ci ! Un grand merci à **Boub'**, pour sa venue toute en surprise, ainsi qu'à **Philippe**, **Marc** et **Natacha**.

Je remercie de tout mon cœur ma fratrie, dont je suis si fière. Un grand merci à mon frère aîné **Guillaume**, mon dareuf, à ma grande sœur **Mathilde**, ma bonne fée, ainsi qu'à mon grand frère **Gauthier**, qui a eu la congruence de lire intégralement cette thèse [tousotement]. Vous êtes mes plus beaux exemples... Merci de veiller sur moi, chacun à votre façon.

Un grand merci à ma sœur jumelle, ma dizygote, ma **Clémentine**. « *Si tu sautes moi je saute, pas vrai ?* », me chuchota-t-elle du haut du col de l'utérus. Et elle sauta. Depuis ce mercredi d'Octobre 85, j'ai la chance de t'avoir à mes côtés, quoi qu'il advienne... Comme le chante Bruno Mars (oui, Bruno Mars ; quoi ?), « *you can count on me 'cause I can count on you* ».

Enfin, un grand merci à mon petit frère **Jean**, parce que *Baby, you are my firework* ! N'oublie pas que *c'est en faisant n'importe quoi que l'on devient n'importe qui*, Bro' ! Non, je déconne... Va bosser tes maths !

Mes pensées vont également à mes belles-sœurs et beaux-frères, **Cécile**, **Nicolas**, **Géraldine** et **Christophe**, ainsi qu'à mes adorables neveux et nièces : **Iana** et **Bjanca**, **Alric**, **Aély**s et **Elouan**, **Gaspard**, **Gustave** et **Gédéon**, ainsi que la petite dernière, **Rose** (un jour, mon Rosinou, je te raconterai les points communs entre cette thèse et... toi ! Mais pas demain, demain, t'as Yoga).

Je remercie **Monique**, ma maman, dite **Moute**, sans laquelle, il est grand temps de l'écrire noir sur blanc, je ne serais rien. Je remercie également **Bernard**, mon papa, dit **Papou**, pour lequel je suis devenue chercheur, et de renommée internationale s'il vous plaît, le jour de mon inscription en première année de thèse. Merci donc, mes petits parents, de croire en moi et de m'aimer plus que je n'oserais le faire... Merci pour votre soutien et votre fierté inconditionnels !

Si l'on en croit Céline Dion (oui, bon), « *la famille c'est une richesse incroyable, ça donne des outils pour pouvoir affronter les moments extraordinaires, les moments plus difficiles, les hauts, les bas* ». Je mesure la chance que j'ai d'être ainsi entourée, choyée, soutenue, par cette joyeuse tribu que j'aime plus que tout au monde. Merci à tous !

J'adresse également mes remerciements à la famille Brunet-Law Bo Kang-Leclère pour ses pensées ensoleillées et fruitées de l'Ile de la Réunion ! **Renée, Jacques, Julie, Olivier** et **Maya**, 974 lé la ! Mèrsi zot tout !

J'adresse, amoureusement, tous mes remerciements à mon coach mental, **Laurent**, dit Docteur Lolo (ou Brugnion). Je te remercie de tout mon cœur pour ta patience, ton humour décalé (lire, "à deux balles") et tout l'amour que tu m'apportes. Parce que tu me donnes la force de surmonter l'insurmontable, à tes côtés tout devient possible (et simple comme un tacos). *Té, mi aime a ou mon ti sinois péi (té) (té) !*

Enfin, merci à toi, cher lecteur, de t'aventurer au-delà de ces remerciements, pour quelques pages ou quelques chapitres !

J'espère sincèrement que cette thèse pourra, d'une façon ou d'une autre, venir en aide à quelque(s) futur(s) docteur(s) en péril, comme l'ont fait pour moi, entre autres, celles de **Benjamin**, de **Mathilde** et de **Krystyna**...

Bonne lecture !

Docteur Chlox (et fière de l'être).

Table des matières

Introduction générale	3
1 Le geste manuel dans la phylogenèse du langage	3
1.1 Origine vocale et/ou manuelle du langage humain	3
1.2 Gestes et vocalisations dans l'émergence de la communication référentielle	5
1.3 Une même latéralisation des contrôles vocaux et manuels	6
1.4 Vocalisations, gestes et communication réunis par les neurones miroirs . .	9
2 Problématique et objectifs de ce travail de thèse	11
Bibliographie	13
 1 Langage, geste manuel et parole	 19
1 Le geste manuel comme objet d'étude	19
1.1 La classification des gestes communicatifs de Kendon	19
1.2 La classification des gestes coverbaux de McNeill	21
1.3 Structure interne du geste	23
2 Liens geste/parole dans le langage moderne	24
2.1 Connexion corticale	25
2.2 Connexion temporelle	27
2.3 Connexion structurelle	41
3 Modèles de production gestes/parole	45
3.1 Génération pré-linguistique du geste manuel	46
3.2 Génération post-linguistique du geste manuel	50
3.3 Génération linguistique du geste manuel	52
3.4 Synthèse des modèles de co-production geste/parole	57
4 Conclusion	58
Annexe : Le modèle de production de parole de Levelt (1989)	60
Bibliographie	64
 2 La deixis spatiale, un rendez-vous entre geste, parole et langage	 73
1 Le geste déictique : Nature et développement	74
2 L'expression déictique : On montre de la voix !	77
2.1 Catégories déictiques	78
3 Langage et représentations spatiales	78
3.1 Représentations spatiales perceptives	78
3.2 Représentations spatiales linguistiques	79
3.3 Lien entre représentations spatiales perceptives et linguistiques	81
3.4 Représentations spatiales linguistiques et symbolisme phonétique	87
4 Conclusion	90
Bibliographie	92

3	Interaction geste, parole, langage et encodage de la distance chez l'adulte	99
1	Expérience 1	100
1.1	Méthodologie	100
1.2	Acquisition des données	103
1.3	Mesures	104
1.4	Analyses statistiques & prédictions	108
1.5	Résultats de l'Expérience 1	112
1.6	Résumé des principaux résultats de l'Expérience 1	117
2	Expérience 2	120
2.1	Méthodologie	121
2.2	Acquisition des données	123
2.3	Mesures	123
2.4	Analyses statistiques & prédictions	123
2.5	Résultats de l'Expérience 2	124
2.6	Résumé des principaux résultats de l'Expérience 2	129
3	Implications théoriques	130
3.1	Encodage moteur de la distance dans le pointage vocal et manuel	130
3.2	Interaction voix/geste, une coopération inter-modalités	132
4	Conclusion	135
	Bibliographie	137
4	Interaction geste, parole, langage et encodage de la distance chez l'enfant	141
1	Evolution du couplage voix/geste au cours du développement	142
1.1	Un couplage oro-manuel moteur présent dès la naissance	142
1.2	L'émergence d'un couplage parole/geste fonctionnel	148
1.3	L'évolution des conduites motrices et verbales à l'âge scolaire	151
2	Expérience 3	155
2.1	Méthodologie	155
2.2	Acquisition des données	158
2.3	Mesures	159
2.4	Analyses statistiques et prédictions	160
2.5	Résultats de l'Expérience 3	161
2.6	Implications théoriques	175
3	Conclusion	180
	Annexe : Les tâches de conservation piagésiennes	181
	Bibliographie	182
5	Couplage voix/geste dans la perception du langage	189
1	Mécanismes de perception du langage	190
1.1	Théorie acoustico-auditive <i>versus</i> théorie motrice	190
1.2	Interactions sensorimotrices en jeu dans la perception de la parole	191
1.3	Vers une approche sensorimotrice des unités sonores de parole	193
1.4	Interactions sensorimotrices dans le traitement des unités sémantiques	194
2	Expérience 4	197
2.1	Méthodologie	197

Table des matières

2.2	Mesures	203
2.3	Analyses statistiques & prédictions	204
2.4	Résultats de l'Expérience 4	206
2.5	Implications théoriques	210
3	Conclusion	213
	Bibliographie	215
Conclusion générale		225
4	Synthèse des principaux résultats	225
4.1	Axe de recherche 1 : Production du langage	225
4.2	Axe de recherche 2 : Développement tardif du langage	227
4.3	Axe de recherche 3 : Perception du langage	228
5	Perspective : Phylogénèse de la multimodalité	229
	Bibliographie	232
Résumé/Abstract		235

Table des figures

1	Arbre phylogénétique des primates	4
2	Pointage paume ouverte chez le chimpanzé	5
3	Lobes et principales structures cérébrales impliqués dans le traitement du langage	7
4	Résultats de Meunier et collab., chez le babouin, 2012	8
1.1	Continuum de Kendon (1988)	20
1.2	Structure hiérarchique du geste (Kendon, 1980)	23
1.3	Résultats de Holender, 1980	29
1.4	Protocole expérimental de Levelt, 1985	33
1.5	Hypothèse de travail de Rochet-Capellan, 2007	39
1.6	Clés du code LPC	41
1.7	Modèle interactif de Krauss et collab., 2000	47
1.8	Modèle Sketch de de Ruiter, 1998	49
1.9	Modèle d'interface de Kita & Özyürek, 2003	53
1.10	Modèle GSA d'Hostetter & Alibali, 2008	56
1.11	Synthèse des modèles de production de gestes et de parole	57
1.12	Étapes de traitement de la production de parole	60
1.13	Modèle de production de parole de Levelt, 1989	60
2.1	Résultats de Matasaka, 2003	75
2.2	Résultats de Meunier et collab. chez l'enfant, 2012	75
2.3	La représentation spatiale	79
2.4	Protocole proposé par Coventry et collab., 2008	81
2.5	Résultats de Coventry et collab., 2008	83
2.6	Mode d'articulation de différents types de voyelles	88
2.7	Expressions relatives à la notion <i>petit</i>	88
2.8	Expressions relatives à la notion <i>grand</i>	89
2.9	Résultats de Kapatsinski, 2010	89
3.1	Dispositif expérimental (Exp.1)	101
3.2	Position des diodes (Exp.1)	104
3.3	Valeurs formantiques des voyelles du français	105
3.4	Paramètres cinématiques (Exp.1)	107
3.5	Ouverture des lèvres selon la Condition et la Distance (Exp.1)	112
3.6	F1 selon la Condition et la Distance (Exp.1)	113
3.7	F0 selon la Condition et la Distance (Exp.1)	115
3.8	Intensité selon la Condition et la Distance (Exp.1)	116
3.9	Paramètres cinématiques selon la Distance (Exp.1)	117
3.10	Dispositif expérimental (Exp.2)	121
3.11	Ouverture des lèvres selon la Condition et la Distance (Exp.2)	125
3.12	F1 selon la Condition et la Distance (Exp.2)	126
3.13	Intensité selon la Condition et la Distance (Exp.2)	127

3.14	Durée totale et durée du plateau selon Condition et la Distance (Exp.2)	128
3.15	Pic de vitesse et amplitude selon la Condition et la Distance (Exp.2)	128
3.16	Modèle d'interface (Kita & Ozyürek, 2003) revisité	135
4.1	L'émergence dynamique (Iverson & Thelen, 1999) : Seuil et activation	143
4.2	Développement ontogénétique du langage (Volterra et collab., 2005)	147
4.3	Production conjointe de geste et de parole chez l'enfant d'âge préscolaire	148
4.4	Réponse discordante lors d'une tâche de conservation piagétienne	150
4.5	Dispositif expérimental (Exp.3)	156
4.6	Position des diodes (Exp.3)	159
4.7	Ouverture des lèvres selon la Condition et la Distance (Exp.3)	162
4.8	Ouverture des lèvres selon la Condition et le Groupe d'âge (Exp.3)	163
4.9	F2 selon la Condition et la Distance (Exp.3)	165
4.10	Intensité selon la Condition et la Distance (Exp.3)	166
4.11	Amplitude selon la Condition et la Distance (Exp.3)	167
4.12	Amplitude selon la Distance et le Groupe d'âge (Exp.3)	169
4.13	Durée plateau selon la Condition et la Distance (Exp.3)	169
4.14	Durée du plateau selon la Condition et le Groupe d'âge (Exp.3)	171
4.15	Durée du plateau selon la Distance et le Groupe d'âge (Exp.3)	171
4.16	Pic de vitesse selon la Condition et la Distance (Exp.3)	172
4.17	Synthèse : Effet de la Distance (Exp.3)	173
4.18	Synthèse : Effet de la Condition (Exp.3)	174
4.19	Schémas observés chez l'enfant, relativement au modèle théorique adulte (Exp.3)	175
4.20	Exemple de tâche de conservation spatiale piagétienne	181
5.1	Déroulement temporel d'un essai type selon la condition (Exp.4)	203
5.2	Temps de réponse selon la Condition (Exp. 4)	207
5.3	Temps de réponse selon l'Item (Exp. 4)	207
5.4	Temps de réponse selon la Condition et l'Item (Exp. 4)	208
5.5	Temps de réponse selon la Distance (Exp. 4)	209
5.6	Temps de réponse selon l'Item I' (Exp. 4)	209
5.7	Temps de réponse selon la Distance et l'Item I' (Exp. 4)	210

Liste des tableaux

3.1	Liste des items (Exp.1)	102
3.2	Variables indépendantes (Exp.1)	103
3.3	Variables dépendantes (Exp.1)	108
3.4	Principaux résultats : Données articulatoires (Exp.1)	113
3.5	Principaux résultats : F1 (Exp.1)	114
3.6	Principaux résultats : F0 (Exp.1)	115
3.7	Principaux résultats : Intensité (Exp.1)	116
3.8	Principaux résultats : Données cinématiques (Exp.1)	117
3.9	Variables indépendantes (Exp.2)	123
3.10	Variables dépendantes (Exp.2)	123
3.11	Principaux résultats : Données articulatoires (Exp.2)	125
3.12	Principaux résultats : F1 (Exp.2)	126
3.13	Principaux résultats : Intensité (Exp.2)	127
3.14	Principaux résultats : Données cinématiques (Exp.2)	129
4.1	Calendrier développemental (Capone & McGregor, 2004)	151
4.2	Age des enfants (Exp.3)	156
4.3	Variables indépendantes (Exp.3)	158
4.4	Variables dépendantes (Exp.3)	159
4.5	Variables inter-sujets (Exp.3)	161
4.6	Principaux résultats : données articulatoires (Exp.3)	163
4.7	Principaux résultats : F1 (Exp.3)	164
4.8	Principaux résultats : F2 (Exp.3)	165
4.9	Principaux résultats : Intensité (Exp.3)	167
4.10	Principaux résultats : Amplitude (Exp.3)	168
4.11	Principaux résultats : Durée du plateau (Exp.3)	170
4.12	Principaux résultats : Pic de vitesse (Exp.3)	172
5.1	Critères de sélection des items (Exp.4)	199
5.2	Variables indépendantes (Exp. 4)	203

Introduction générale

« *Etre là, c'est déjà un geste* » Alessandro Baricco (Châteaux de la colère)

La communication langagière humaine est clairement multimodale : parler, oui, mais pas seulement ! Nous disposons d'un ensemble diversifié d'outils communicatifs, nous permettant d'enrichir le contenu et l'impact de nos interactions. Bateson et collab. [5]¹ définissent ainsi la communication comme un « *processus social permanent intégrant de **multiples modes de comportement** : la parole, le geste, le regard, la mimique, l'espace interindividuel, etc.* » ([5], p.24). L'étude de la communication langagière rassemble de fait de nombreuses disciplines, de la psychologie à l'anthropologie, en passant par la neurophysiologie, les sciences cognitives, etc., et constitue en cela un domaine de recherche aussi vaste que passionnant. Naturellement, ce travail de thèse n'a pas la prétention de pouvoir s'y atteler de manière exhaustive mais sera consacré à ce que **nous** considérons comme deux modalités langagières à part entière : la parole et le geste manuel. Celui-ci est en effet omniprésent dans la communication, et de fait, la nature du lien qui l'unit à la parole concerne la nature même du langage et de sa structure.

La question de l'interaction entre geste manuel et parole est au centre de ce travail de thèse et sera explorée à travers la production, la perception et le développement ontogénétique du langage. Mais pour commencer, et donner une idée de la portée de cette thématique, prenons-la à son origine : l'interaction geste manuel/parole est au cœur d'un débat crucial autour de l'émergence et de l'évolution du langage. Nous proposons donc, en guise d'introduction, une revue de la littérature consacrée à la place du geste manuel dans l'émergence de la communication langagière humaine ; parce qu'elles mettent en évidence des connexions particulières entre les modalités vocale et manuelle, ces données constituent un point de départ pertinent pour l'étude de l'interaction entre geste manuel, parole et langage moderne.

1 Le geste manuel dans la phylogénèse du langage

1.1 Origine vocale et/ou manuelle du langage humain

La question de l'origine du langage fait débat depuis des siècles et reste à l'heure actuelle très controversée. En l'absence de preuve physique et matérielle, concrète, concernant l'évolution de la communication langagière humaine, elle laisse place à de nombreuses hypothèses² ; une possible intervention du geste manuel dans la phylogénèse du langage est notamment au cœur des débats.

Certains auteurs (e.g. Lieberman [37], MacNeilage [38]) proposent une théorie purement vocale de l'origine du langage. Différentes variantes de cette approche existent : certaines supposent que

1. Les normes bibliographiques utilisées dans ce manuscrit sont conformes aux règles de présentation énoncées dans l'ouvrage de Malo [39].

2. Une analyse poussée (historique, épistémologique et critique) est disponible dans les travaux de Boë et collab. [6] ainsi qu'une revue détaillée des théories de l'émergence du langage dans le manuscrit de Moulin-Frier [49].

des modifications de la bouche et du pharynx, couplées à une augmentation du volume cérébral, auraient conduit, il y a environ 100 000 ans, au contrôle volontaire des productions vocales, qui étaient jusqu'alors plutôt des cris instinctifs, des réflexes. D'autres postulent que le langage humain émergerait du développement des onomatopées, soit de l'imitation des bruits de l'environnement. MacNeilage [38] a quant à lui développé la théorie *Frame/Content* (Cadre/Contenu), qui propose une évolution à partir des cyclicités ingestives (mâcher, lécher, téter), présentes chez tous les mammifères, en passant par les cyclicités communicatives visuo-faciales chez les grands singes (claquements de lèvres, de langue et de dents), pour aboutir à la parole articulée moderne.

De nombreuses données issues de l'étude comportementale, neuroanatomique et neurophysiologique des primates non humains (e.g. Arbib [3, 4], Corballis [11, 12], Hewes [20], Masataka [40], Vauclair [65]) remettent toutefois en cause l'hypothèse d'une origine purement vocale de la communication langagière humaine. Certains auteurs favorisent alors l'hypothèse d'une origine gestuelle (e.g. Corballis [11, 12], Vauclair [65]), voire mixte (e.g. Masataka [40]) du langage. Celles-ci postulent notamment que le passage à la bipédie aurait eu pour première conséquence de rendre les membres antérieurs disponibles et fonctionnels pour la communication gestuelle, tandis que la transition vers le langage vocal se serait faite progressivement jusqu'il y a 50 000 ans, libérant définitivement les mains pour d'autres usages, comme en témoigne l'explosion technologique et artistique de cette époque.

Meguerditchian et collab. [43], qui proposent une revue passionnante sur le sujet, soulignent que, pour espérer déterminer les caractéristiques ancestrales de la communication langagière, « *a prime question [...] is to investigate whether evolutionary precursors of language may be found in the communicative behaviours of nonhuman primates*³ » ([43], p.91). La proximité phylogénétique unique entre primates humains et non humains, illustrée Figure 1, suggère en effet que l'étude comparative de la communication et de la cognition puisse offrir de précieux éléments quant aux précurseurs potentiels du langage moderne.

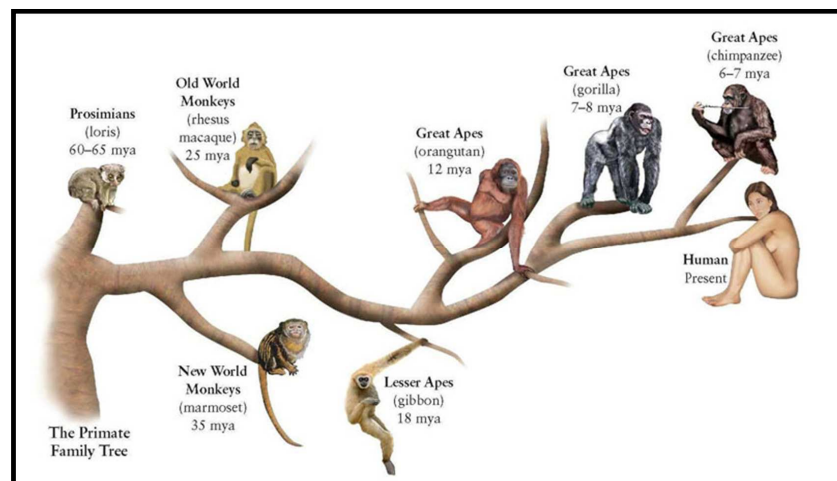


FIGURE 1 – Arbre phylogénétique des primates (source : <http://books.wwnorton.com>)

3. Traduction proposée : Un point central [en primatologie] est de déterminer dans quelle mesure les précurseurs du langage peuvent être observés dans les comportements communicatifs des primates non humains.

1. Le geste manuel dans la phylogénèse du langage

1.2 Gestes et vocalisations dans l'émergence de la communication référentielle

L'observation des primates non humains nous apprend par exemple que ceux-ci possèdent un répertoire de comportements communicatifs étendu et notamment de gestes manuels variés et spécifiques, leur permettant d'interagir avec leurs congénères, et avec l'homme, dans des situations diverses : menace, requête, grooming (i.e. activité de toilettage social ou épouillage mutuel), etc. (e.g. Tomasello et collab. [63]). Dans ce cadre là, le geste de pointage apparaît comme l'un des précurseurs de la communication référentielle : la théorie *Vocalize-to-Localize* (Articuler-pour-Localiser), proposée par Abry et collab. [1], suggère par exemple que le pointage constituerait l'élément pré-langagier clé à l'origine de l'émergence du langage. Plus précisément, l'auteur propose que la synchronisation entre la main et la voix aurait fourni les bases d'association entre une référence (par le geste de pointage) d'une part, et une vocalisation d'autre part. Ainsi, le langage aurait émergé à partir de cette capacité à montrer de la voix, à *articuler pour localiser*.

Néanmoins, l'existence d'un pointage comparable à celui de l'homme, en termes de forme et de fonction, chez le primate non-humain, pose question (Povinelli et Davis [54], Veà et Sabater-Pi [68] vs Krause et Fouts [29], Leavens [31]) : si les grands singes en captivité pointent spontanément et sans entraînement, généralement en configuration paume ouverte (phénomène illustré Figure 2), le pointage de l'index (1) serait quant à lui utilisé spécifiquement par les animaux entraînés au langage (e.g. tâches langagières, apprentissage de signes, etc.), (2) dépendrait du degré d'*enculturation* de l'animal (i.e. intimité avec l'homme) et (3) serait rarement destiné à des pairs (Povinelli et Davis [54], Veà et Sabater-Pi [68]). Mais la **fonction** du pointage (i.e. montrer) ne serait-elle pas plus pertinente que la **forme** du pointage (i.e. index vs paume), comme le suggèrent Krause et Fouts [29] ? A ce propos, Leavens et Hopkins [33] rappellent que le pointage humain est également très variable en forme et peut être réalisé en configuration paume ouverte.



FIGURE 2 – Pointage paume ouverte chez le chimpanzé

Leavens [31] montre également que le pointage chez le grand singe présente tous les critères objectifs de la communication intentionnelle humaine⁴, à savoir : (1) la présence d'un public, (2) l'alternance du regard, (3) la recherche d'attention, et (4) la persistance/élaboration en cas d'échec (Leavens et Bard [32]). Et pourtant, certains auteurs ne leur concèdent, à peine, que le pointage impératif⁵ : « *Although some apes, especially those with extensive human contact, sometimes point imperatively for humans [...], no apes point declaratively ever*⁶ » (Tomasello [62], p.510).

De manière générale, on observe des continuités intéressantes, en termes d'intentionnalité mais aussi de flexibilité d'apprentissage et d'usage (voir notamment les travaux de Meguerditchian et Vauclair [45] ou de Pika [53] pour une revue de la littérature), entre la communication référentielle des primates non humains et celles des humains. Alors dans quelle mesure le système gestuel peut-il être considéré comme précurseur du langage moderne ?

Les travaux sur la langue des signes ont apporté des premiers éléments en faveur de l'implication du geste manuel dans l'émergence du langage : la langue des signes, aussi sophistiquée du point de vue grammatical que les langues orales, solliciterait les mêmes aires du langage que celles-ci (e.g. Hickok et collab. [21]), très proches des aires motrices des bras et de la main. Par ailleurs, l'étude comparative des gestes manuels et oro-faciaux chez le primate (d'un point de vue comportemental) et plus précisément de la latéralisation des contrôles, ainsi que la découverte d'un système de neurones miroirs (d'un point de vue neuroanatomique et neurophysiologique) sont autant d'éléments cruciaux, décrits ci-après, permettant d'enrichir le débat.

1.3 Une même latéralisation des contrôles vocaux et manuels

« *Studying functional laterality in nonhuman primates remains an important way to investigate potential precursors of human hemispheric specialization*⁷ » Vauclair & Meguerditchian ([66], p.183)

Il est connu depuis Paul Broca [7] que le langage humain est un processus hautement latéralisé, c'est-à-dire contrôlé préférentiellement par l'un des deux hémisphères. Ainsi, les fonctions langagières, situées majoritairement au niveau du lobe frontal (e.g. aire de Broca, impliquée notamment dans la production de parole) et temporo-pariétal⁸ (e.g. aire de Wernicke, spécialisée dans la compréhension de parole) - voir Figure 3 - sont latéralisées à gauche pour près de 90% des individus, soit dans l'hémisphère contralatéral à la main préférentielle (près de 90% des individus étant droitiers, e.g. Annett [2]).

4. « *Pointing by apes meets all the objective criteria for intentional communication originally defined with reference to the preverbal communication of human infants* » (Leavens [31], p.391).

5. Le pointage impératif, destiné aux requêtes, implique la manipulation de l'autre mais non la compréhension de son état d'esprit (à l'inverse du pointage déclaratif).

6. Traduction proposée : Bien que certains grands singes, *enculturés*, pointent parfois de manière impérative, pour l'homme, aucun n'a jamais pointé de manière déclarative.

7. Traduction proposée : Etudier la latéralité fonctionnelle chez le primate non humain pourrait nous aider à déterminer les précurseurs potentiels de la spécialisation hémisphérique humaine.

8. Le lobe pariétal intervient dans les processus de perception visuelle, ainsi que dans la coordination oeil/main.

1. Le geste manuel dans la phylogénèse du langage

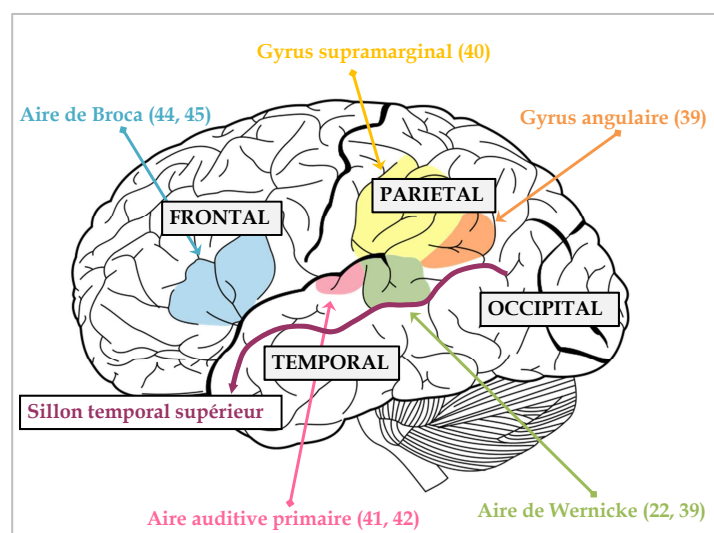


FIGURE 3 – Lobes et principales structures cérébrales impliqués dans le traitement du langage (les numéros réfèrent aux aires de Brodmann [8])

Les données comportementales et neurophysiologiques citées ci-après suggèrent que cette spécialisation hémisphérique du langage humain dérive d'un système communicatif gestuel (au sens de manuo-oro-facial) présent chez l'ancêtre commun du babouin, du chimpanzé et de l'homme (Meunier et collab. [46]).

En effet, à l'instar des humains pour la communication langagière, la majorité des grands singes présentent une préférence manuelle droite : de nombreux gestes communicatifs (dirigés vers l'humain, tels que la demande de nourriture, e.g. Hopkins et Wesley [25] ; mais également vers les pairs, tels que les gestes de menace, salut, quémende, e.g. Meguerditchian et Vauclair [44]) sont produits avec la main droite. Ainsi, selon Hopkins [22], 65% des chimpanzés seraient droitiers. Vauclair et collab. [67], remontant plus loin dans l'échelle phylogénétique, observent une tendance similaire chez les babouins, droitiers à 61%.

Fait intéressant, cette préférence manuelle ne semble valoir que pour les gestes communicatifs, ne concernant pas, par exemple, les gestes d'atteinte et de saisie (Meguerditchian et collab. [42], Vauclair et collab. [67]). Meunier et collab. [46] ont ainsi montré que le babouin pointe préférentiellement de la main droite (et ce, que le référent soit placé à droite ou à gauche de l'axe central de l'individu) tandis que la main utilisée pour saisir un objet dépend clairement de son emplacement (le babouin saisit l'objet de la main droite si ce dernier est placé à droite, mais de la main gauche s'il est placé à gauche). Ce phénomène est illustré Figure 4 ci-après et s'observe également chez l'enfant âgé de 14, 17 et 20 mois (voir Chapitre 2, Section 1).

Hopkins et Cantero [23] montrent également que la proportion de gestes produits avec la main droite augmente de manière significative lors de la production simultanée de vocalisations.

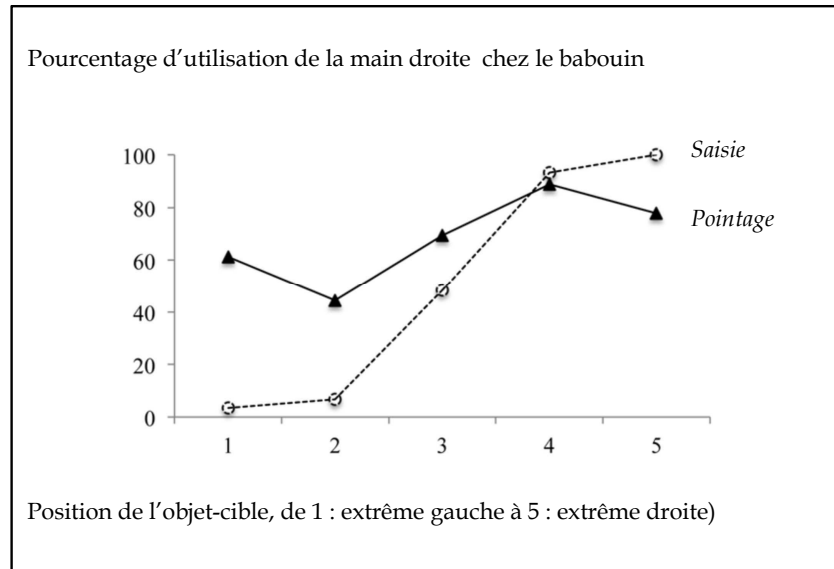


FIGURE 4 – Pourcentage d'utilisation de la main droite chez le babouin, en fonction du type de geste (*pointage* ▲ ; *saisie* ○) et de la position de l'objet par rapport à l'individu (de *gauche* à *droite*) (Meunier et collab. [46], p.2)

Ce lien geste manuel/vocalisation est également étudié d'un point de vue neurophysiologique : Tagliatela et collab. [60] observent, chez le chimpanzé, une implication du gyrus frontal inférieur gauche dans la production de signaux communicatifs multimodaux, gyrus qui chez l'homme joue un rôle essentiel dans la production et la compréhension du langage. Les auteurs mettent ainsi en évidence l'existence d'un système intentionnel (i.e. contrôlé) bimodal chez le chimpanzé : les signaux communicatifs, manuels et/ou oro-faciaux, destinés à l'homme, notamment pour quémander de la nourriture, activent une région homologue de l'aire de Broca, et ce davantage dans l'hémisphère gauche (Tagliatela et collab. [61]).

De fait, si le traitement des gestes communicatifs manuels, chez le primate non humain, est latéralisé de façon similaire aux fonctions langagières humaines, il semble également lié corticalement aux représentations communicatives oro-faciales. C'est donc naturellement que les questions du contrôle et de la latéralisation des **vocalisations** sont posées dans la communauté scientifique. Si l'on en croit les travaux de Hopkins et collab. [24], qui ne font certes pas l'unanimité⁹, la communication vocale des primates non humains serait non seulement contrôlée mais également latéralisée dans l'hémisphère gauche, à l'instar des gestes manuels. Gannon et collab. [18] montrent par exemple chez les gorilles, chimpanzés et orang-outans, que le planum temporal, comprenant l'aire de Wernicke, est plus développé dans l'hémisphère gauche chez 17/18 individus. Dans le même registre, Cantalupo et Hopkins [9] observent une asymétrie gauche de l'aire 44 de Brodman chez 20/26 grands singes. Plus surprenant encore, l'étude de Petersen

9. Voir par exemple les travaux de Lieberman [36] ou Myers [50], qui considèrent les vocalisations comme des signaux involontaires, soit des réponses émotionnelles induites par des stimuli spécifiques, évoquant notamment la peur, ou autres réactions dites primitives.

1. Le geste manuel dans la phylogénèse du langage

et collab. [52] révèle que le traitement perceptif des vocalisations est effectué dans l'hémisphère gauche, uniquement pour les vocalisations de congénères ! Ainsi les macaques japonais réagissent plus rapidement lorsque les stimuli auditifs sont présentés dans l'oreille droite (traités par l'hémisphère gauche), mais seulement s'il s'agit de vocalisations appartenant à leur propre espèce.

L'ensemble de ces données suggère l'existence, chez le primate non humain, d'un système communicatif spécifique, intentionnel et bimodal, latéralisé dans l'hémisphère gauche, et potentiellement à l'origine de la spécialisation hémisphérique des fonctions langagières humaines.

1.4 Vocalisations, gestes et communication réunis par les neurones miroirs

« *Such an observation/execution matching system provides a necessary bridge from **doing** to **communicating***¹⁰ » Rizzolatti & Arbib ([55], p.188).

Découverts initialement chez le macaque rhésus par l'équipe de Rizzolatti [57], les neurones miroirs ont la particularité de s'activer lors de la production d'actions manuelles ou oro-faciales, mais également lors de la perception de ces actions, et ce au niveau de la zone F5 du cortex prémoteur (une zone homologue, en termes cytoarchitectoniques, de l'aire de Broca) et du lobe pariétal inférieur (Fogassi et collab. [16], Gallese et collab. [17], di Pellegrino et collab. [51]).

Plus particulièrement, ces neurones miroirs semblent impliqués dans la production et la perception d'actions manuelles dirigées vers un objet (Rizzolatti et Craighero [56]) mais également de gestes communicatifs oro-faciaux (Ferrari et collab. [15], Rizzolatti et Craighero [56]) : le terme *mouth mirror neurons* désigne ainsi les neurones miroirs qui déchargent lors d'actions relatives aux fonctions ingestives (e.g. attraper de la nourriture avec la bouche) ou communicatives (e.g. claquement des lèvres). Plus récemment, Kohler et collab. [28] ont mis en évidence un phénomène de résonance **audio-visuo-motrice**, illustré par la décharge de neurones miroirs lorsque l'individu observe, effectue ou **entend** le son produit par, une action spécifique. Ces neurones miroirs multimodaux sont également adaptables : Umiltà et collab. [64] et Rochat et collab. [58] observent des activations indépendantes de l'outil utilisé pour réaliser l'action (e.g. attraper de la nourriture avec la main, une pince ou la piquer à l'aide d'un bâton). Ces données suggèrent que « *F5 motor neurons code the goals of motor acts rather than the movements forming them*¹¹ » (Rochat et collab. [58], p.613). De fait, un tel système miroir favoriserait la compréhension et l'imitation de gestes manuels et articulatoires mais également le contrôle et la reconnaissance d'actions audio-visuo-motrices; en cela, il pourrait être considéré comme un médiateur entre gestes manuels et parole, définie comme une séquence complexe de gestes oro-faciaux associés à du son. En revanche, à notre connaissance, l'absence de donnée quant à l'implication du système miroir dans le traitement des gestes communicatifs manuels reste à déplorer : « *There have been no reports of communicative hand mirror neurons in the monkey that respond to the observation and execution of hand movements that do not involve an object*¹² » ([48], p.114).

10. Traduction proposée : Un tel système couplant observation/action pourrait expliquer la transition entre *faire* et *communiquer*.

11. Traduction proposée : Les neurones moteurs situés dans la zone F5 codent le but de l'action plutôt que les mouvements *per se* composant l'action.

12. Traduction proposée : Aucune donnée n'a été rapportée quant à l'existence de neurones miroirs chez le singe,

A l'heure actuelle, les données suggèrent l'existence d'un système miroir équivalent chez l'homme : certains neurones déchargent au niveau du lobe frontal et du lobe pariétal inférieur lors de l'exécution et de l'observation d'actions manuelles dirigées vers un objet (Decety et collab. [13], Iacoboni et collab. [26], Leslie et collab. [34]). On observe également des décharges relatives aux expressions faciales (Montgomery et collab. [47]) mais également aux gestes communicatifs manuels : une étude en Imagerie par Résonance Magnétique fonctionnelle (IRMf), proposée par Montgomery et collab. [48], a en effet montré que les neurones déchargent au niveau des lobes frontal et pariétal inférieur, ainsi qu'au niveau du sillon temporal supérieur (impliqué dans la perception de mouvements biologiques), lors de l'observation, de l'exécution et de l'imitation de gestes manuels dirigés vers un objet, mais également, et de façon similaire, lors de l'observation, de l'exécution et de l'imitation de gestes manuels communicatifs (tels que les gestes déictiques, iconiques, et emblématiques).

L'existence d'un tel système miroir chez l'homme, impliqué dans certaines capacités cognitives complexes, telles que la compréhension des intentions de l'autre, l'empathie, et l'imitation (Arbib [3], Fadiga et Craighero [14], Rizzolatti et Arbib [55]) pourrait jouer un rôle crucial dans le traitement de la parole, en offrant un cadre de références commun aux interlocuteurs. Corballis [12] émet notamment l'hypothèse que l'association entre un geste communicatif perçu et les représentations motrices propres à chacun aurait permis à l'individu de s'approprier le sens du geste. De là auraient émergé les capacités de compréhension de l'intentionnalité et le langage, gestuel dans un premier temps, puis vocal.

Dans ce cadre là, Rizzolatti et Arbib formulent l'hypothèse *MSH* (*Mirror System Hypothesis*¹³ ; [3, 55]) : ils considèrent la reconnaissance d'actions manuelles transitives, vers des objets, comme un tremplin vers le langage moderne. Arbib [4] propose alors sept étapes principales vers l'émergence du langage. Globalement, la capacité d'imitation simple (i.e. d'actions élémentaires) aurait permis à l'ancêtre commun des grands singes et des hommes d'acquérir un répertoire basique de gestes communicatifs *braccio-manuels*. C'est ensuite la capacité d'imitation complexe (i.e. de séquences d'actions élémentaires), qui aurait rendu possible la mise en place d'un répertoire ouvert de signes conventionnels, les *proto-signes*, ou *pantomimes*, nécessaire à l'émergence des propriétés récursives du langage humain. L'acquisition du contrôle du conduit vocal serait apparu progressivement, conjointement aux *proto-signes*, afin de permettre une évolution vers le langage moderne. Arbib [4] propose ainsi une évolution conjointe de deux modalités communicatives et langagières, basée sur notre capacité d'imitation complexe.

relatifs aux gestes communicatifs de la main, et qui répondraient à l'observation et l'exécution de mouvements manuels n'impliquant pas d'objet.

13. Traduction proposée : Hypothèse du Système Miroir.

2 Problématique et objectifs de ce travail de thèse

Si la question de l'émergence du langage, et notamment celle de l'intervention du geste manuel dans ce processus, anime la communauté scientifique, la question du lien entre geste manuel et parole dans le langage moderne est tout aussi épineuse. Deux visions sont classiquement opposées : la première stipule que geste manuel et parole sont deux systèmes communicatifs distincts (Hadar et collab. [19], Krauss et Hadar [30], Levelt et collab. [35]) ; le geste manuel est vu comme un support auxiliaire à la parole et interviendrait préférentiellement lorsque le discours est interrompu ou difficile (e.g. environnement bruyant, distance importante entre interlocuteurs). La deuxième suppose à l'inverse que geste manuel et parole forment un système de communication unifié (Kendon [27], McNeill [41]) ; deux modalités distinctes certes, mais toutes deux liées aux mêmes processus de pensées : « *Gestures and speech [...] share a common fate*¹⁴ » (McNeill [41], p.218). L'objectif de ce travail de thèse est de comprendre et de caractériser au mieux l'interaction voix, geste manuel et langage, par le biais d'une approche contrôlée. Pour des raisons que nous détaillerons par la suite, notre hypothèse générale est que geste manuel et parole sont intégrés au niveau du langage, pour ne former qu'un seul et même système communicatif, par nature multimodal ; en d'autres termes le langage est considéré ici comme un système vocal et gestuel, reposant donc sur l'utilisation conjointe de deux modalités.

Cette hypothèse sur base sur une revue, non exhaustive, de la littérature, consacrée au lien particulier qu'entretiennent le geste manuel et la parole dans le langage moderne, et ce d'un point de vue cortical, temporel et structurel. Ces éléments sont présentés au Chapitre 1, avant la description des principaux modèles actuels de production conjointe de gestes manuels et de parole qui en découlent.

Notre objectif est de comprendre les interactions geste/parole, notamment dans le processus déictique d'encodage de la distance des objets. Le processus déictique, décrit au Chapitre 2, par nature multimodal - on montre avec le doigt certes, mais également avec la voix - nous permettra d'étudier l'encodage de la distance via les propriétés lexicales, acoustiques, articulatoires et cinématiques du pointage. Plus précisément, nous souhaitons montrer (1) que cet encodage est issu d'un phénomène moteur, de bas niveau, dans les deux modalités : désigner un objet éloigné impliquerait un geste plus ample, que celui-ci soit un geste vocal ou manuel ; (2) que cet encodage moteur peut être relié à l'encodage phonologique de la distance dans les langues ; (3) que les deux systèmes, vocal et manuel, interagissent de façon bidirectionnelle.

Dans cette optique, nous avons mis au point un protocole expérimental, présenté au Chapitre 3, où les participants ont pour tâche de désigner une cible plus moins éloignée, nous permettant de comparer les productions unimodales (geste ou parole) aux productions bimodales (geste+parole). Par ailleurs, l'étude de ces conduites chez les enfants âgés de 6 à 12 ans (présentée au Chapitre 4) vise à déterminer si les performances observées chez l'adulte sont liées à des compétences précoces, soit des caractéristiques fondamentales du processus déictique, ou à des compétences plus tardives, soit des fonctions cognitives élaborées, effectives chez le locuteur expert.

14. Gestes et parole partagent un même destin.

Enfin, nous aborderons la question de l'étude de l'intégration geste/parole dans les mécanismes de perception du langage. Cette question est au cœur de débats actuels autour des différentes théories sur l'implication du système moteur dans le traitement sémantique du langage (voir Coello et Bartolo [10] pour une revue), un débat parallèle à celui sur les interactions sensorimotrices dans la perception des unités sonores de la parole (e.g. Schwartz et collab. [59]). Nous chercherons à mettre en évidence ces interactions sensorimotrices dans le processus de la deixis spatiale, en montrant que produire un geste déictique faciliterait la perception d'un terme déictique.

Notre projet global est donc de comprendre l'interaction entre parole et geste manuel, lors de la production et de la perception du langage ainsi qu'au cours de son développement, ce qui nous permettra éventuellement de revenir questionner la problématique de l'émergence du langage humain.

Bibliographie

- [1] Abry, C., A. Vilain et J. Schwarz. 2004, «Vocalize to localize? A call for better crosstalk between auditory and visual communication systems researchers», *Interaction Studies : Social Behaviour and Communication in Biological and Artificial systems*, vol. 5(3), p. 313–325.
- [2] Annett, M. 1985, *Left, right, hand and brain : The right shift theory*, Hillsdale, NJ : Erlbaum.
- [3] Arbib, M. A. 2005, «From monkey-like action recognition to human language : An evolutionary framework for neurolinguistics», *Behavioral and Brain Sciences*, vol. 28, p. 105–167.
- [4] Arbib, M. A. 2005, «Interweaving protosign and protospeech : Further developments beyond the mirror», *Interaction Studies*, vol. 6, n° 2, p. 145–171.
- [5] Bateson, G., Y. Winkin et D. Bansard. 1981, *La nouvelle communication*, Le Temps qui viendra, Seuil.
- [6] Boë, L., J. Schwartz, J. Granat, J. Heim, A. Serrurier, P. Badin, G. Captier et P. Bessière. 2011, «L'émergence de la parole : Aspects historiques et épistémologiques d'une nouvelle réarticulation», *Faits de Langues*, vol. 35, p. 15–67.
- [7] Broca, P. 1865, «Sur le siège de la faculté du langage articulé», *Bulletins de la Société d'Anthropologie de Paris*, vol. 6, p. 377–93.
- [8] Brodmann, K. 1909, *Vergleichende Lokalisationslehre der Grosshirnrinde*, Leipzig : Johann Ambrosius Bart.
- [9] Cantalupo, C. et W. Hopkins. 2001, «Asymmetrical Broca's area in great apes», *Nature*, vol. 414, p. 505.
- [10] Coello, Y. et A. Bartolo. 2012, «Contribution of the action system to language perception and comprehension : Evidence and controversies», dans *Language and Action in Cognitive Neuroscience*, édité par Y. Coello et A. Bartolo, London : Psychology Press, p. 321–342.
- [11] Corballis, M. 2009, «Language as gesture», *Human Movement Science*, vol. 28, p. 556–565.
- [12] Corballis, M. 2010, «Mirror neurons and the evolution of language», *Brain and Language*, vol. 112(1), p. 25–35.
- [13] Decety, J., T. Chaminade, J. Grezes et A. Meltzoff. 2002, «A PET exploration of the neural mechanisms involved in reciprocal imitation», *NeuroImage*, vol. 15, p. 265–272.
- [14] Fadiga, L. et L. Craighero. 2007, «On being moved : From mirror neurons to empathy», dans *Cues on the origin of language : From electrophysiological data on mirror neurons and motor representations*, édité par S. Bråten, Amsterdam : John Benjamins, p. 101–110.
- [15] Ferrari, P., V. Gallese, G. Rizzolatti et L. Fogassi. 2003, «Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex», *European Journal of Neuroscience*, vol. 17, p. 1703–1714.

- [16] Fogassi, L., P. Ferrari, B. Gesierich, S. Rozzi, F. Chersi et G. Rizzolatti. 2005, «Parietal lobe : From action organization to intention understanding», *Science*, vol. 303, p. 662–667.
- [17] Gallese, V., L. Fadiga, L. Fogassi et G. Rizzolatti. 1996, «Action recognition in the premotor cortex», *Brain*, vol. 119, p. 593–609.
- [18] Gannon, P., R. Holloway, D. C. Broadfield et A. Braun. 1998, «Asymmetry of chimpanzee planum temporale : Human-like pattern of Wernicke’s language area homolog», *Science*, vol. 279, p. 220–222.
- [19] Hadar, U., D. Wenkert-Olenik, R. Krauss et N. Soroker. 1998, «Gesture and the processing of speech : Neuropsychological evidence», *Brain and Language*, vol. 62(1), p. 107–126.
- [20] Hewes, G. W. 1973, «Primate communication and the gestural origin of language», *Current Anthropology*, vol. 14, p. 5–24.
- [21] Hickok, G., U. Bellugi et E. Klima. 1998, «The neural organization of language : Evidence from sign language aphasia.», *Trends in Cognitive Sciences*, vol. 2, p. 129–136.
- [22] Hopkins, W. 1995, «Hand preferences for a coordinated bimanual task in 110 chimpanzees : Cross-sectional analysis», *Journal of Comparative Psychology*, vol. 109, p. 291–297.
- [23] Hopkins, W. et M. Cantero. 2003, «From hand to mouth in the evolution of language : The influence of vocal behavior on lateralized hand use in manual gestures by chimpanzees (*Pan troglodytes*)», *Developmental Science*, vol. 6, p. 55–61.
- [24] Hopkins, W., J. Taglialatela et D. Leavens. 2007, «Chimpanzees differentially produce novel vocalizations to capture the attention of a human», *Animal Behaviour*, vol. 73, p. 281–286.
- [25] Hopkins, W. et M. Wesley. 2002, «Gestural communication in chimpanzees (*Pan troglodytes*) : The effect of situational factors on gesture type and hand use», *Laterality*, vol. 7, p. 19–30.
- [26] Iacoboni, M., R. Woods, M. Brass, H. Bekkering, J. Mazziotta et G. Rizzolatti. 1999, «Cortical mechanisms of human imitation», *Science*, vol. 286, p. 2526–8.
- [27] Kendon, A. 2004, *Gesture : Visible Action as Utterance*, Cambridge University Press.
- [28] Kohler, E., C. Keysers, A. Umiltà, L. Fogassi, V. Gallese et G. Rizzolatti. 2002, «Hearing sounds, understanding actions : Action representation in mirror neurons», *Science*, vol. 297, p. 846–848.
- [29] Krause, M. et R. Fouts. 1997, «Chimpanzee (*Pan Troglodytes*) pointing : Hand shapes, accuracy, and the role of eye gaze», *Journal of Comparative Psychology*, vol. 111, n° 4, p. 330–336.
- [30] Krauss, R. et U. Hadar. 1999, «The role of speech-related arm/hand gestures in word retrieval», dans *Gesture, speech, and sign*, édité par R. Campbell et L. Messing, Oxford : Oxford University Press, p. 93–116.

Bibliographie

- [31] Leavens, D. 2004, «Manual deixis in apes and humans», *Interaction studies*, vol. 5, n° 3, p. 387–408.
- [32] Leavens, D. et K. Bard. 2011, «Environmental influences on joint attention in great apes : Implications for human cognition», *Journal of Cognitive Education and Psychology*, vol. 10, p. 9–31.
- [33] Leavens, D. et W. Hopkins. 1999, «The whole-hand point : The structure and function of pointing from a comparative perspective», *Journal of Comparative Psychology*, vol. 113(4), p. 417–425.
- [34] Leslie, K., S. Johnson-Frey et S. Grafton. 2004, «Functional imaging of face and hand imitation : Towards a motor theory of empathy», *NeuroImage*, vol. 21, p. 601–607.
- [35] Levelt, W., G. Richardson et W. L. Heij. 1985, «Pointing and voicing in deictic expressions», *Journal of Memory and Language*, vol. 24, p. 133–164.
- [36] Lieberman, P. 1998, «Speech evolution : Let barking dogs sleep», *Behavioral and Brain Sciences*, vol. 21, p. 520–521.
- [37] Lieberman, P. 2000, *Human language and our reptilian brain : The subcortical bases of speech, syntax, and thought*, Cambridge, MA : Harvard University Press.
- [38] MacNeilage, P. 1998, «The frame/content theory of evolution of speech production», *Behavioral and Brain Sciences*, vol. 21, p. 499–511.
- [39] Malo, M. 1996, *Guide de la communication écrite au cégep, à l'université et en entreprise*, Quebec Amerique.
- [40] Masataka, N. 2008, «The gestural theory of and the vocal theory of language origins are not incompatible with one another», dans *The origins of language : Unraveling evolutionary forces*, édité par N. Masataka, Heidelberg : Springer, p. 1–10.
- [41] McNeill, D. 1992, *Hand and Mind : What Gestures Reveal about Thought*, University of Chicago Press.
- [42] Meguerditchian, A., S. E. Calcutt, E. V. Lonsdorf, S. R. Ross et W. D. Hopkins. 2010, «Captive gorillas are right-handed for bimanual feeding», *American Journal of Physical Anthropology*, vol. 141, p. 638–645.
- [43] Meguerditchian, A., H. Cochet et J. Vauclair. 2011, «From gesture to language : Ontogenetic and phylogenetic perspectives on gestural communication and its cerebral lateralization», dans *Primate Communication and Human Language : Vocalisation, gestures, imitation and deixis in humans and non-humans*, édité par A. Vilain, J. Schwartz, C. Abry et J. Vauclair, Amsterdam : John Benjamins, p. 91–119.
- [44] Meguerditchian, A. et J. Vauclair. 2006, «Baboons communicate with their right hand», *Behavioural Brain Research*, vol. 171, p. 170–174.

- [45] Meguerditchian, A. et J. Vauclair. 2008, «Vocal and gestural communication in nonhuman primates and the question of the origin of language», dans *Learning from animals ? Examining the nature of human uniqueness*, édité par L. Roska-Hardy et E. Neumann-Held, London : Psychology Press., p. 61–85.
- [46] Meunier, H., J. Vauclair et J. Fagard. 2012, «Human infants and baboons show the same pattern of handedness for a communicative gesture», *PLoS ONE*, vol. 7(3), p. 1–3.
- [47] Montgomery, K., M. Gobbini et J. Haxby. 2003, «Imitation, production and viewing of social communication : An fMRI study», *Society for Neuroscience Abstracts*, vol. 128, p. 10.
- [48] Montgomery, M., N. Isenberg et J. Haxby. 2007, «Communicative hand gestures and object-directed hand movements activated the mirror neuron system», *Social, Cognitive, & Affective Neuroscience*, vol. 2, p. 114–122.
- [49] Moulin-Frier, C. 2011, *Rôle des relations perception-action dans la communication parlée et l'émergence des systèmes phonologiques : Etude, modélisation computationnelle et simulations*, thèse de doctorat, Université de Grenoble.
- [50] Myers, R. 1976, «Comparative neurology of vocalization and speech : Proof of a dichotomy», dans *Origins and evolution of language and speech*, édité par S. Harnad, H. Steklis et J. Lancaster, New York : New York Academy of Sciences, p. 745–757.
- [51] di Pellegrino, G., L. Fadiga, L. Fogassi, V. Gallese et G. Rizzolatti. 1992, «Understanding motor events : A neurophysiological study», *Experimental Brain Research*, vol. 91, p. 176–180.
- [52] Petersen, M. R., M. D. Beecher, S. R. Zoloth, D. B. Moody et W. C. Stebbins. 1978, «Neural lateralization of species-specific vocalizations by Japanese macaques (*Macaca fuscata*)», *Science*, vol. 202, p. 324–327.
- [53] Pika. 2008, «Gestures of apes and pre-linguistic human children : Similar or different ?», *First Language*, vol. 28, p. 116–140.
- [54] Povinelli, D. et D. Davis. 1994, «Differences between chimpanzees (*Pan troglodytes*) and humans (*Homo sapiens*) in the resting state of the index finger : Implications for pointing», *Journal of Comparative Psychology*, vol. 108, p. 134–139.
- [55] Rizzolatti, G. et M. A. Arbib. 1998, «Language within our grasp», *Trends in Neurosciences*, vol. 21, p. 188–194.
- [56] Rizzolatti, G. et L. Craighero. 2004, «The mirror-neuron system», *Annual Review of Neuroscience*, vol. 27, p. 169–192.
- [57] Rizzolatti, G., L. Fadiga, V. Gallese et L. Fogassi. 1996, «Premotor cortex and the recognition of motor actions», *Cognitive Brain Research*, vol. 3, p. 131–141.
- [58] Rochat, M., F. Caruana, A. Jezzini, L. Escola, I. Intskirveli, F. Grammont, V. Gallese, G. Rizzolatti et M. Umiltà. 2010, «Responses of mirror neurons in area F5 to hand and tool grasping observation», *Experimental Brain Research*, vol. 204, p. 605–616.

Bibliographie

- [59] Schwartz, J., A. Basirat, L. Ménard et M. Sato. 2012, «The Perception-for-Action-Control Theory (PACT) : A perceptuo-motor theory of speech perception», *Journal of Neurolinguistics*, vol. 25(5), p. 336–354.
- [60] Tagliatela, J., C. Cantalupo et W. Hopkins. 2006, «Gesture handedness predicts asymmetry in the chimpanzee inferior frontal gyrus», *NeuroReport*, vol. 17, p. 923–927.
- [61] Tagliatela, J., J. Russell, J. Schaeffer et W. Hopkins. 2008, «Communicative signaling activates Broca’s homologue in chimpanzees», *Current Biology*, vol. 18, p. 343–348.
- [62] Tomasello, M. 2006, «Why don’t apes point ?», dans *Roots of Human Sociality : Culture, Cognition and Interaction*, édité par N. Enfield et S. Levinson, Oxford & New York : Berg, p. 506–524.
- [63] Tomasello, M., M. Carpenter, J. Call, T. Behne et H. Moll. 2005, «Understanding and sharing intentions : The origins of cultural cognition», *Behavioral and Brain Sciences*, vol. 28, n° 5, p. 675–691.
- [64] Umiltà, M., L. Escola, I. Intskirveli, F. Grammont, M. Rochat, F. Caruana, A. Jezzini, V. Gallese et G. Rizzolatti. 2008, «When pliers become fingers in the monkey motor system», *Proceedings of the National Academy of Sciences*, vol. 105(6), p. 2209–2213.
- [65] Vauclair, J. 2004, «Lateralization of communicative signals in nonhuman primates and the hypothesis of the gestural origin of language», *Interaction Studies : Social Behaviour and Communication in Biological and Artificial Systems*, vol. 5, p. 363–384.
- [66] Vauclair, J. et A. Meguerditchian. 2007, «Perceptual and motor lateralization in two species of baboons», dans *Evolution of hemispheric specialization in primates, special topics in primatology*, édité par W. Hopkins, New York : Academic Press, p. 177–198.
- [67] Vauclair, J., A. Mergueditchian et W. Hopkins. 2005, «Hand preferences for unimanual and coordinated bimanual tasks in baboons (*Papio anubis*)», *Cognitive Brain Research*, vol. 25, p. 210–216.
- [68] Veà, J. et J. Sabater-Pi. 1998, «Spontaneous pointing behaviour in the wild pygmy chimpanzee (*Pan paniscus*)», *Folia Primatologica*, vol. 69(5), p. 289–290.

Langage, geste manuel et parole

L'objectif de ce chapitre est de présenter une revue, non exhaustive, de la littérature, consacrée au lien particulier qu'entretiennent le geste manuel et la parole dans le langage moderne. Nous commencerons par une définition détaillée de notre objet d'étude, le geste manuel à valeur communicative, avant d'aborder les liens corticaux, temporels et structurels qui l'unissent à la parole. Ces liens seront également illustrés via les principaux modèles actuels de génération et de production conjointe de gestes manuels et de parole.

1 Le geste manuel comme objet d'étude

L'étude comportementale et qualitative des gestes manuels communicatifs a permis d'établir différentes typologies de tels gestes. Nous présenterons dans cette section deux classifications majeures et particulièrement influentes dans la littérature actuelle, celles de Kendon [55] et de McNeill [78]. Cela permettra une meilleure lecture des études présentées par la suite dans ce manuscrit.

1.1 La classification des gestes communicatifs de Kendon

A l'exception des gestes ludiques ou auto-centrés, la plupart des gestes manuels présents dans la communication langagière humaine ont une valeur communicative. Kendon [55] en propose une classification, basée sur un continuum comprenant quatre axes (*Kendon's continuum*, pour reprendre les termes de McNeill [78]), présenté Figure 1.1 : les gestes¹ sont ainsi classés selon leur degré de dépendance à la parole, leurs propriétés linguistiques, leur degré de codification, ainsi que leur caractère sémiotique. Kendon [55] distingue de cette façon les gesticulations, des pantomimes, des *language-like*², des emblèmes et des signes.

- **Les gesticulations**

Les gesticulations sont des gestes produits de manière spontanée, dont la particularité est d'accompagner, d'illustrer, voire même de remplacer le discours. Nous y reviendrons en détails par la suite, ces gestes faisant l'objet de la classification proposée par McNeill [78].

1. Le terme "geste" renverra par la suite, sauf exception, aux gestes **manuels**. Il s'agit de simplifier la lecture du manuscrit mais en aucun cas de nier le fait que la parole est composée de gestes communicatifs oraux, ou gestes **vocaux**.

2. Défini par Kendon [55], ce type de gestes ne figure pas dans la description du continuum de Kendon proposée par McNeill [78] ; ils pourraient y être intégrés entre les gesticulations et les emblèmes, au même titre que les pantomimes.

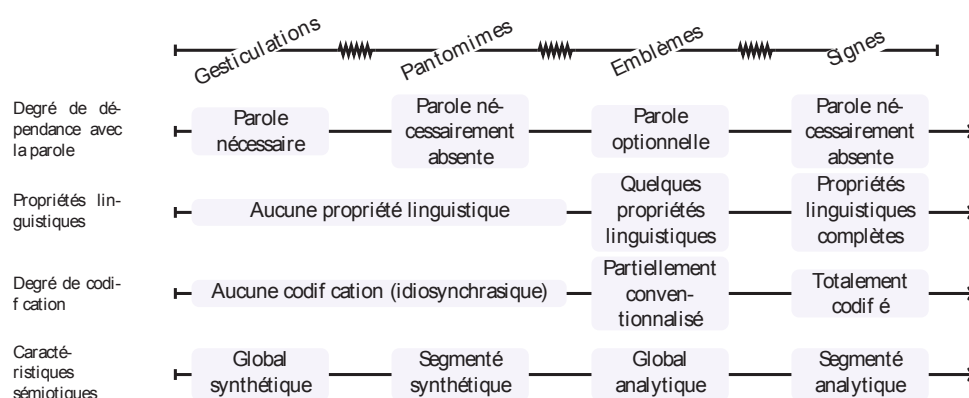


FIGURE 1.1 – Continuum de Kendon (repris par Roustan [88], p.6)

• Les pantomimes

Les pantomimes³ nous permettent de créer une image gestuelle de l'objet, de l'évènement décrit, en mettant sa forme en évidence. S'ils surviennent en l'absence de parole, ils peuvent néanmoins être accompagnés d'onomatopées, soit des interjections émises pour simuler un bruit particulier (McNeill [78]). Il s'agit en fait d'une imitation, d'une représentation analogique du centre d'intérêt : évoquer un nageur par exemple, avec de grands cercles des bras devant soi, paumes ouvertes et orientées vers le bas, ou encore Boucle d'or, en effectuant de rapides rotations descendantes de l'index autour du visage.

• Les language-like

Les *language-like* peuvent être qualifiés de gestes grammaticaux, puisqu'ils remplacent un mot de l'énoncé (voir l'exemple ci-dessous). En cela, ils peuvent être rapprochés des gestes *Butterworths* (McNeill [78]) ou *speech failure* (Freedman et Hoffman [30]), qui apparaissent en cas de difficulté sur le canal verbal, notamment lorsque le locuteur cherche ses mots (ces gestes apparaissent alors en l'absence de parole simultanée).

Exemple : « Je me suis trop penchée pour admirer les... »

[Rotations de l'index et du poignet]

→ Représentation du mot « hélices »

• Les emblèmes

Les emblèmes⁴ sont des gestes conventionnels, propres aux différentes cultures, autonomes, pleinement porteurs de sens et de fait interprétables hors contexte. De Ruiter [89] les évoque en termes de gestes « lexicalisés ». Ils peuvent apparaître conjointement avec la parole, où ils expriment des concepts également exprimés verbalement (redondants donc), mais apparaissent souvent seuls, en remplacement du contenu verbal, notamment lorsque la communication est difficile (distance importante entre les interlocuteurs, environnement bruyant, etc.).

3. Autres appellations : *enactments* au sens d'interprétation (de Ruiter [89]), gestes kinémimiques (Cosnier et Vaysse [20]).

4. Autres appellations : gestes symboliques (Krauss et collab. [66]), quasi-linguistiques (Cosnier et Vaysse [20]), autonomes (Kendon [53]), sémiotiques (Barakat [4]).

1. Le geste manuel comme objet d'étude

Il s'agit par exemple du pouce levé en signe de réussite, du V de la victoire ou du célèbre Shaka Hawaïen, qui consiste à ne tendre que le pouce et l'auriculaire et agiter la main en signe d'amitié et de bienvenue. Il faut noter que les emblèmes, bien qu'indépendants de la parole, ne constituent pas un système linguistique à part entière, dans la mesure où ils n'obéissent pas à des règles grammaticales et combinatoires.

- Les **signes**

Les signes en revanche sont des gestes et postures codifiés, qui composent les langues signées, et peuvent en ce sens s'apparenter aux mots et expressions des langues orales.

1.2 La classification des gestes coverbaux de McNeill

McNeill [78] propose quant à lui une classification plus spécifique, consacrée aux gesticulations de Kendon [55], qu'il désigne sous le terme de gestes *coverbaux*. Ceux-ci dépendent d'une production verbale simultanée et ne peuvent la plupart du temps s'y substituer. Ils sont par nature liés au contexte et ne peuvent être interprétés qu'à partir de celui-ci. Produits de manière spontanée, ils accompagnent la parole à divers niveaux : phonologique, sémantique, syntaxique et pragmatique (voir également McNeill et collab. [81] pour plus de détails). Ces gestes sont divisés en plusieurs catégories ou dimensions, relatives aux entités, actions, espaces et concepts auxquels ils renvoient : les gestes iconiques, métaphoriques, déictiques et de battement.

- Les gestes **iconiques**

Les gestes iconiques⁵ sont liés au contenu sémantique du discours. Ils véhiculent une information complémentaire ou supplémentaire au contenu verbal, en référence à des entités concrètes dont ils dépeignent certains aspects également véhiculés par la parole. Par exemple, l'énoncé « on s'appelle ? » peut être accompagné d'un mouvement de la main près de l'oreille. Finalement, comme le dit si joliment de Ruiter [89], un geste iconique est une « *imaginary sculpture shaped by the speaker's hands*⁶ » ([89], p.2). Ces gestes entretiennent donc une relation visuelle et spatiale avec les informations qu'ils représentent ; Cosnier et Vaysse [20] distinguent ainsi les gestes iconiques *spatiographiques* (liés à la disposition spatiale de l'objet décrit) des gestes iconiques *pictographiques* (liés à leur forme). La forme de ces gestes dépend donc clairement du contenu sémantique qu'ils véhiculent et diffère, naturellement, selon l'entité décrite. De plus, elle varie avec le point de vue du locuteur, qui peut tout aussi bien se placer en acteur, ou en observateur, de la scène.

- Les gestes **métaphoriques**

McNeill [77] donne la définition suivante des gestes métaphoriques⁷ : « *Metaphoric gestures exhibit images of abstract concepts. In form and manner of execution, metaphoric gestures depict the vehicles of metaphors...*⁸ » ([77], p.356).

5. Autres appellations : gestes illustratifs (Cosnier et Vaysse [20]), kinétographes ou pictographes (Ekman et Friesen [25]), kinétographiques ou physiographiques (Efron [24]).

6. Traduction proposée : Le geste iconique est une sculpture imaginaire formée par les mains du locuteur.

7. Autres appellations : idéographes (Cosnier et Vaysse [20], Efron [24], Ekman et Friesen [25]), « underliners » ou gestes spatiaux (Ekman et Friesen [25]).

8. Traduction proposée : Les gestes métaphoriques représentent des concepts abstraits. Ils dépeignent les métaphores, par la forme et la façon dont ils sont exécutés.

En d'autres termes, les gestes métaphoriques diffèrent des gestes iconiques en cela qu'ils décrivent une entité abstraite et non concrète. L'expression « joindre les deux bouts » (i.e. boucler son budget) par exemple peut ainsi être associée à un geste des deux mains, index pointés l'un vers l'autre, se rejoignant.

- **Les gestes déictiques**

Le geste déictique, en tant qu'acte référentiel par excellence, est au centre de cette thèse. Nous en donnons ici une définition succincte, avant de l'aborder de façon détaillée au Chapitre 2, intitulé "*La deixis spatiale, un rendez-vous entre geste, parole et langage*".

Les gestes déictiques, ou de pointage, ont pour fonction d'indiquer, de montrer un objet de référence, sa location et/ou sa direction. Il s'agit en fait d'un fléchage visuel, nous permettant de diriger l'attention de notre interlocuteur et de coordonner ainsi l'attention visuelle de différents intervenants. Les pointages concrets, vers des objets, lieux, événements, personnes présentes dans la situation d'énonciation, sont classiquement distingués des pointages abstraits, vers des entités physiquement absentes ou imaginaires (de Ruiter [89]).

Les gestes déictiques accompagnent généralement la parole mais peuvent également la remplacer. Ces pointages, dits autonomes, surviennent généralement en réponse à une question (Krauss et collab. [66]) : lorsqu'un collègue nous demande où se situe la salle expérimentale, nous pouvons pointer la porte appropriée plutôt que de la décrire oralement, puisque le geste seul répond alors parfaitement à la question ; néanmoins une réponse uniquement gestuelle peut sembler impolie ou brutale, et nous avons, bien sûr, la possibilité de l'accompagner verbalement : « juste ici », ou « celle-ci ».

- **Les gestes de battement**

Les gestes de battement⁹ sont de rapides mouvements biphasiques, généralement d'avant en arrière ou de haut en bas, qui marquent le rythme de la parole. Leur fonction discursive est principalement pragmatique¹⁰, puisqu'ils permettent de faire le lien entre différentes parties du discours, d'introduire un nouvel élément ou d'en mettre certains en évidence. Les gestes de battement semblent particulièrement liés à la prosodie (soit le rythme, l'intonation, la mélodie de la parole ; e.g. Roustan [88]), synchronisés à la production de syllabes accentuées (e.g. Bull et Connelly [12], Krahmer et Swerts [63]). Ces gestes de battement ne sont pas signifiants, au sens où ils seraient liés au contenu sémantique de la parole, et leur forme est de fait indépendante du contenu verbal associé. Une distinction est régulièrement faite entre ces gestes de battement d'une part et les gestes dits représentationnels¹¹ d'autre part (e.g. Alibali et collab. [1]) qui eux véhiculent d'une certaine façon divers aspects du contenu verbal (notamment les gestes iconiques, métaphoriques, et déictiques décrits ci-dessus).

9. Autres appellations : gestes moteurs (Krauss et collab. [66]), paraverbaux (Cosnier et Vaysse [20]), rythmiques ou prosodiques (Ekman et Friesen [25]), bâtons (Efron [24], Ekman et Friesen [25]).

10. Un connecteur pragmatique est une marque linguistique, qui permet notamment d'articuler les éléments du discours afin de rendre leur connexion non ambiguë et univoque (Reboul et Moeschler [86]).

11. Autres appellations : gestes lexicaux (Krauss et collab. [66]), gestes conceptuels (i.e. *ideational*, Hadar et collab. [40], Hadar et Butterworth [41]), illustratifs ou référentiels (Cosnier et Vaysse [20]), gesticulations (Kendon [52]).

1. Le geste manuel comme objet d'étude

Cette classification, inspirée notamment de celles d'Efron [24] et d'Ekman et Friesen [25], est devenue une référence incontournable de la littérature actuelle. Néanmoins, certains auteurs en proposent une version revisitée ; c'est le cas par exemple de de Ruiter [89], qui regroupe les gestes iconiques et métaphoriques en une seule catégorie, de Krauss et collab. [66], qui désignent sous le terme de *gestes lexicaux* non seulement les gestes iconiques et métaphoriques, mais également tous les gestes qui ne sont ni déictiques, ni symboliques (i.e. emblèmes), ni moteurs (i.e. gestes de battement), et de Cosnier et Vaysse [20] qui distinguent les gestes référentiels (i.e. iconiques, métaphoriques et déictiques) des gestes paraverbaux (i.e. de battement).

1.3 Structure interne du geste

Afin d'en décrire le déroulement temporel, Kendon [52] décompose le geste, ou unité gestuelle (*gesture unit*), illustré Figure 1.2, qui débute lorsque le bras s'éloigne du corps et prend fin lorsque celui-ci retrouve une position immobile, en différentes phrases (i.e. *gesture phrase*) : la préparation, le stroke (phrase dite obligatoire), la tenue (*hold*) et le retour (*retraction*). Cette description est citée dans de nombreux travaux actuels.

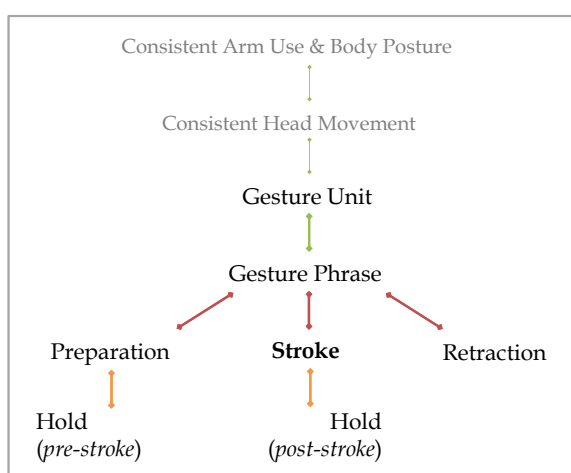


FIGURE 1.2 – Structure hiérarchique du geste proposée par Kendon (vue par McNeill [78], p.82)

Plus précisément, la préparation correspond à l'étape pendant laquelle le bras ou la main se déplace pour atteindre la position initiale du stroke. Le stroke, qui représente l'unité porteuse de sens, est ensuite exécuté ; il peut être défini comme le « *peak of effort in the gesture*¹² » (McNeill [78], p.83), le laps de temps saillant du geste¹³. Deux phases, dites de tenue¹⁴, optionnelles, peuvent être réalisées avant et/ou après l'exécution du stroke, qui précède le retour, phrase finale, en position immobile.

12. Traduction proposée : Le stroke peut être défini comme le pic d'effort du geste.

13. Toutefois, dans le cas particulier du pointage, c'est l'apex (soit l'instant où le geste atteint son but, lorsque les différents effecteurs sont au maximum de leur extension respective), plus que le stroke, qui véhicule le sens du geste.

14. De manière générale, la tenue est une cessation du mouvement au cours du geste, à la différence du repos (McNeill [78], p.83).

De manière générale, les différentes phrases du geste sont déterminées par un changement de direction abrupte et une discontinuité dans le profil de vitesse. Par exemple, l'identification des frontières temporelles du stroke, pendant lequel la force exercée est plus importante, se base sur le profil d'accélération (Kita et collab. [62]).

2 Liens geste/parole dans le langage moderne

Si la présence du geste au cours des interactions communicatives ne peut être contestée, son rôle est en revanche largement controversé. Purement communicatif (avec pour objectif de délivrer un message, de transmettre une information, d'établir une relation avec autrui), soit au service de celui qui le perçoit, le geste rendrait visible en temps réel la structuration du discours (Kendon [53]) ; il faciliterait l'interaction grâce à l'amélioration de la compréhension de l'énoncé. Ce phénomène s'illustre notamment lorsque deux individus tentent de communiquer dans deux langues différentes ! Dans ce cas, le peu de maîtrise de la langue étrangère peut être compensé par une gesticulation exagérée. Toutefois, bon nombre de gestes sont produits alors même qu'ils sont inexploitablement par notre interlocuteur : au téléphone par exemple, ou lorsque des individus aveugles communiquent entre eux (Iverson et Goldin-Meadow [49]). Cela suggère que le rôle du geste pourrait bien être cognitif, soit au service de celui qui le produit : il permettrait la structuration du discours (Alibali et collab. [2], Hostetter et Alibali [47], Kita [58]) et/ou faciliterait l'accès au lexique mental¹⁵ (Krauss [64]). Ces deux rôles n'étant pas mutuellement exclusifs, certains auteurs (e.g. Bavelas [7], Driskell et Radtke [23], Ozyürek [83], de Ruiter [90]) considèrent le geste comme un outil à la fois communicatif et cognitif, qui améliorerait la compréhension de l'énoncé par une amélioration de la production de parole.

Cette relation, qu'entretiennent le geste et la parole, suscite de nombreuses questions : pourquoi, quand et comment produit-on un geste particulier pendant l'acte communicatif ? Le geste et la parole sont-ils deux systèmes communicatifs distincts (Hadar et collab. [42], Krauss et Hadar [67], Levelt et collab. [73]) ou les deux modalités d'un seul et même système, accomplissant un but commun (Kendon [56], McNeill [78]) ?

Cette problématique peut être abordée par le biais d'approches expérimentales diverses, du très écologique au très contrôlé. Le moyen le plus naturel d'étudier la production conjointe de gestes et de parole consiste en l'enregistrement et l'analyse de conversations *in the wild*, à la manière de Kendon [51] : aucune consigne particulière, quant à la façon de bouger ou de parler, n'est donnée aux participants, qui doivent simplement interagir, de manière naturelle et spontanée, sur un thème donné. Dans le même registre, le célèbre paradigme du cartoon (McNeill [78]) consiste à présenter un court dessin animé aux participants, qui doivent ensuite le raconter en détails à leur interlocuteur naïf. Kita et collab. [58, 59] utilisent notamment cette méthodologie dans le but d'étudier les variations cross-linguistiques : observer des participants de langue maternelle différente pour comprendre de quelle façon une structure linguistique particulière peut influencer la production des gestes qui y sont associés.

15. Le lexique mental (Treisman [97]) est un système constitué de représentations lexicales locales et dont les éléments représentent la somme des connaissances orthographiques, phonologiques, morphologiques, sémantiques et syntaxiques qu'une personne a pu mémoriser à propos des mots de sa langue.

2. Liens geste/parole dans le langage moderne

Néanmoins, ces approches dites écologiques présentent un désavantage majeur : aucune consigne n'étant donnée aux participants, elles nécessitent pour être valides l'obtention d'un très grand nombre de données. En effet, le moment où l'individu produit tel ou tel geste, ou ce qu'il exprime par ce geste, sont des éléments particulièrement « instables » : « *gesture is a very noisy phenomenon*¹⁶ » (de Ruiter [89], p.3) ! Cette variabilité, considérable pour un individu donné, l'est d'autant plus entre individus. Le regroupement de gestes en catégories (suivant les classifications de Kendon [55] ou de McNeill [78] par exemple), l'application de statistiques inférentielles, sur la base de critères d'intérêt choisis *a priori*, sur un type de geste particulier (e.g. Ozyürek et collab. [84]) peuvent toutefois réduire le *bruit* qui résulte de cette variabilité intra- et inter-individuelle. Il est donc tout-à-fait pertinent de valider les hypothèses formulées à partir de données écologiques par le biais d'approches plus contrôlées. L'utilisation de paradigmes contrôlés permet en effet de récolter des mesures objectives, nombreuses et précises : les gestes sont produits sur demande et/ou sous différentes conditions (e.g. Rauscher et collab. [85]). Bien que « *the experimental control of spontaneous behavior is a contradiction in terms*¹⁷ » (de Ruiter [89], p.4.), ces méthodes, utilisées pour différents types de gestes, sont plus particulièrement adaptées aux gestes conventionnels (emblèmes et déictiques), qui constituent par définition un système sémiotique partagé par une communauté linguistique : leur variabilité intra- et inter-individuelle est donc nettement plus faible que celle des gestes iconiques et de battement, par exemple.

Nous avons retenu, parmi ce vaste ensemble d'études, certains travaux nous permettant d'appréhender le lien geste/parole en termes de connexion corticale, temporelle et structurelle ; elles soutiennent en cela l'hypothèse d'un **seul** système de communication **multimodal**.

2.1 Connexion corticale

« *Studies in cognitive neuroscience provide evidence for links between language and action in the brain.*¹⁸ » Willems & Hagoort ([99], p.286)

Les données issues d'études en neurologie et neuropsychologie¹⁹ notamment, mettent en évidence des connexions corticales étroites entre langage et action au sens large (voir la revue proposée par Willems et Hagoort [99]), mais également entre parole et gestes communicatifs.

Ainsi les aires du langage sont activées par l'observation (e.g. Buccino et collab. [11], Decety et collab. [21], Gallagher et Frith [33], Grèzes et collab. [39], Skipper et collab. [94]) et la production de gestes manuels (e.g. Gerardin et collab. [34], Kuhtz-Buschbeck et collab. [69]).

16. Traduction proposée : Le geste est un phénomène particulièrement "bruité".

17. Traduction proposée : Quel oxymore que le contrôle **expérimental** de comportements **spontanés**.

18. Traduction proposée : Les études en neurosciences montrent des liens forts entre langage et action dans le cerveau.

19. Les principales technologies utilisées pour étudier l'aspect cérébral du langage sont l'Imagerie par Résonance Magnétique fonctionnelle (IRMf), l'ElectroEncéphaloGraphie (EEG), la Stimulation Magnétique Transcrânienne (TMS), la Tomographie par Emission de Positrons (PET), ainsi que la MagnétoEncéphaloGraphie (MEG) (pour une description détaillée de ces différentes technologies, se référer aux travaux de Dillisenger et Moershel [22] ; pour une description plus générale des structures cérébrales et de leur fonctionnement, consulter l'ouvrage de Carter et collab. [16]).

Decety et collab. [21], par exemple, observent par PET que les actions signifiantes (ici des pantomimes, telles que l'ouverture d'une bouteille) induisent une activation au niveau frontal et temporal de l'hémisphère gauche, tandis que les actions non signifiantes (ici des enchaînements aléatoires de signes de la Langue des Signes Américaine) induisent principalement des activations dans les zones occipito-pariétales de l'hémisphère droit.

Dans le même esprit, Tanaka et Inui [96] montrent en IRMf que l'aire de Broca est davantage impliquée dans l'imitation de la configuration des doigts plutôt que dans l'imitation de postures brachio-manuelles. Les auteurs observent en effet une activation différenciée dans le lobe frontal inférieur gauche en fonction du type d'action considéré. Par ailleurs, les deux types d'action (condition *doigt* vs condition *main*) induisent une activation dans le lobe pariétal (mais de façon bilatérale pour la condition *doigt* et latéralisée à gauche pour la condition *main*).

Löwenbrück et collab. [74, 75, 76] ont quant à eux étudié ces connexions corticales entre gestes communicatifs et parole en situation de production de pointage multimodal : les participants étaient observés en IRMf pendant l'exécution de gestes de pointage digital et oculaire et lors de la production de pointage vocal, prosodique et syntaxique. Les résultats montrent que le lobe pariétal supérieur est activé lors des pointages digitaux, oculaires et prosodiques (un pointage oculaire active davantage les régions antérieures de la zone pariétale qu'un pointage prosodique et qu'un pointage digital). Selon les auteurs, l'absence d'activation en pariétal lors de la production de pointages syntaxiques serait liée à l'intervention d'un processus grammatical. Ces résultats montrent que le lobe pariétal gauche est impliqué dans les différentes modalités du pointage, avec une possible somatotopie, c'est-à-dire une organisation spatiale cérébrale dans laquelle chaque zone est associée à un effecteur de pointage particulier.

Dans un autre registre, l'étude de certaines lésions et pathologies associées, notamment celle de l'aphasie et de l'apraxie²⁰, apporte de nouvelles preuves de connexions particulières entre les systèmes moteurs du geste et de la parole. Heath et collab. [44] montrent ainsi qu'apraxie (idéomotrice²¹ en particulier) et aphasie sont souvent associées chez les patients cérébrolésés : une atteinte au niveau des capacités gestuelles peut donc avoir des retentissements directs sur la parole (avec une fluence et une pertinence verbales défectueuses, comme le montrent Barrett et collab. [6]), et réciproquement, une atteinte au niveau des capacités verbales peut engendrer des difficultés au niveau de la production de gestes manuels coverbaux (Cocks et collab. [19], Hadar et collab. [42]).

L'ensemble de ces données révèle clairement l'existence de liens étroits entre gestes manuels et parole au niveau cortical. Hickok et collab. [45] suggèrent d'ailleurs que l'organisation cérébrale du langage soit indépendante de la modalité langagière recrutée : « *Research to date*

20. L'aphasie est un trouble du langage qui touche notamment l'expression ou la compréhension du langage parlé, en l'absence de déficit sensoriel ou d'un dysfonctionnement de l'appareil phonatoire, et dont les manifestations cliniques dépendent largement de la structure lésée. L'apraxie peut être considérée comme l'équivalent gestuel de l'aphasie, puisque c'est l'incapacité d'effectuer un mouvement, ou une série de mouvements, alors même que les fonctions motrices et sensitives de base sont intactes (les fonctions de coordination et d'adaptation des mouvements volontaires de base, dans le but d'accomplir une tâche donnée, sont, elles, altérées) ; les manifestations cliniques sont ici encore dépendantes de la structure lésée.

21. Il existe une dizaine d'apraxies, dont l'apraxie idéomotrice, causée par une lésion au niveau du gyrus supramarginal dans le cortex pariétal. Cette apraxie se caractérise par l'incapacité d'exécuter un geste sur demande (geste symbolique ou action mimée, comme se peigner) tandis que l'exécution spontanée est possible (dissociation automatico-volontaire).

2. Liens geste/parole dans le langage moderne

*has shown a significant degree of similarity in the neurobiology of signed and spoken languages, suggesting that the neural organization of language is largely modality-independent*²² » ([45], p.129).

2.2 Connexion temporelle

Le lien geste/parole, reflété au niveau de l'organisation cérébrale du langage, est également manifeste dans l'organisation temporelle des productions bimodales : « *Gestures and speech are closely linked in [...] time.* »²³ » (McNeill [78], p.218).

Les différentes phases du geste (décrites Section 1.1) semblent en effet se coordonner temporellement à la parole, ainsi qu'à ses différents niveaux prosodiques²⁴. Leur production obéirait à différentes règles, dites de *synchronie* (McNeill [78]).

La notion de synchronie, un cas particulier de coordination, peut être définie de la façon suivante : sachant que « *deux évènements sont coordonnés lorsque leur réalisation se produit à des instants séparés par un laps de temps constant au fil des productions* », « *deux évènements sont synchrones lorsque leur réalisation se produit au même moment* » (Roustan [88], p.xii).

Ces règles de synchronie s'appliquent à divers aspects de la relation geste/parole (i.e. phonologique, sémantique et pragmatique) :

La règle de synchronie phonologique : Cette règle, énoncée en 1980 par Kendon [52] et reprise par McNeill [78], stipule que le stroke du geste peut précéder ou se terminer au pic phonologique de la syllabe accentuée mais dans tous les cas ne se produit pas après ce pic. Lorsque le stroke est en avance, la main reste en position jusqu'au pic phonologique. McNeill propose donc une relation temporelle stable entre le geste spontané et la parole, le geste pouvant à la fois anticiper ou être synchrone avec la parole, selon la phase du geste considérée (préparation ou stroke).

La règle de synchronie sémantique : Lorsque geste et parole sont produits simultanément, ils doivent être porteurs de la même signification ou tout du moins d'une signification complémentaire. Lorsque des pauses sont insérées entre les mots d'une même expression, formant ainsi une même unité sémantique (hésitation, recherche d'un mot), le stroke est maintenu afin de conserver une cohérence sémantique.

La règle de synchronie pragmatique : Lorsque geste et parole sont produits simultanément, ils doivent avoir une même fonction pragmatique.

22. Traduction proposée : A ce jour, les recherches montrent des similarités neurobiologiques fortes entre langues orales et signées, qui suggèrent que l'organisation cérébrale du langage soit largement indépendante de la modalité utilisée.

23. Traduction proposée : Les gestes et la parole sont étroitement liés dans le temps.

24. Roustan [88] par exemple met en évidence une coordination temporelle entre les deux systèmes de production ainsi que des alignements spécifiques au type de geste produit (déictique, de battement ou non communicatif) ; Chui [18] montre également un décours temporel particulier de la parole spontanée lorsqu'elle est accompagnée de gestes iconiques.

Ces règles s'illustrent notamment lors de la production des phases de tenue, avant et/ou après l'exécution du stroke :

- Pendant la phase de *pre-stroke hold*, la main attendrait que la parole soit produite. Plus précisément cette phase serait liée à la production des connecteurs du discours. Elle permettrait ainsi d'établir une cohésion entre les deux modalités, vocale et gestuelle. Le stroke serait ensuite exécuté en même temps que la syllabe ou mot accentué.
- La phase de *post-stroke hold* permettrait ensuite au geste d'attendre la fin de la production de la parole.

2.2.1 Interférences geste/parole

Un certain nombre d'études s'intéressent à l'organisation temporelle des signaux gestuels et vocaux, et notamment aux interférences geste/parole, définies comme une différence en termes de performances lors de productions unimodales (i.e. simples, n'impliquant qu'une modalité) ou bimodales (i.e. doubles, impliquant geste **et** parole), afin de déterminer de quelle façon les deux systèmes de production interagissent. Ces interférences sont par exemple exploitées pour élaborer des modèles de production conjointe de gestes et de parole, modèles que nous détaillerons dans la suite de ce chapitre.

L'étude d'Holender (1980) À notre connaissance, la première étude consacrée aux interférences geste/parole est celle de Holender [46], qui observe l'interférence entre une réponse vocale et une réponse manuelle associées à un même stimulus, par le biais d'une tâche de nomination et/ou de pression de touche. Les participants sont face à quatre lettres (présentées via écran d'ordinateur), dont le nom commence par le même phonème /ε/ (L, N, R, S), associées à quatre boutons réponse, placés les uns à côté des autres. Les temps de réaction pour nommer la lettre affichée et/ou appuyer sur le bouton correspondant (avec le majeur et l'index, selon l'item, de chacune des deux mains), soit les instants d'initiation (i.e. *onset* vocal et gestuel), sont mesurés. Les performances obtenues lors des deux tâches simples (nomination vs pression) sont comparées à celles obtenues lors de la tâche double (nomination + pression).

Les résultats montrent que lors de tâches simples, la réponse vocale est toujours plus rapide que la réponse manuelle, qui survient environ 475 ms après la présentation du stimulus, soit 75 ms plus tard que la parole. En revanche, lors de tâches doubles, la réponse vocale est retardée d'environ 150 ms tandis que la réponse manuelle n'est pas modifiée par l'ajout d'une seconde modalité (voir Figure 1.3). Ainsi, la parole semble attendre le geste, puisqu'elle survient en moyenne 83 ms après celui-ci.

2. Liens geste/parole dans le langage moderne

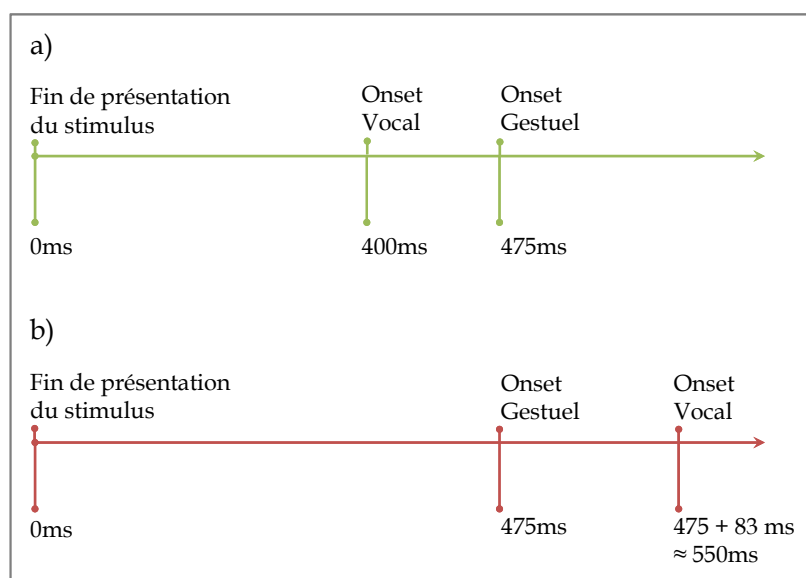


FIGURE 1.3 – Onset vocal et manuel, en tâches simple (a) et double (b)

L'auteur en déduit que la réponse verbale ne peut être initiée qu'après un délai minimum suivant l'initiation de la réponse manuelle et suggère que les participants retardent leur réponse verbale afin de la synchroniser au mieux avec leur réponse manuelle.

Les participants ont ensuite pour consigne, en tâche double, de donner leur réponse verbale avant leur réponse manuelle ou de les synchroniser. S'ils sont capables de donner leurs réponses verbales avant leurs réponses manuelles, celles-ci sont alors considérablement retardées. En revanche, ils ne parviennent pas à synchroniser les deux modalités et donnent spontanément la réponse manuelle en moyenne 70 ms avant la réponse verbale.

Les résultats montrent clairement que la parole s'aligne sur le geste (celle-ci est fortement retardée en tâche double). Selon l'auteur, cette perturbation au niveau du temps de réponse serait la conséquence d'une charge de traitement plus importante en tâche double. Il propose que les réponses de nomination et de pression de touche sont réalisées par des traitements indépendants : lorsque ces deux modalités sont recrutées simultanément, leurs traitements respectifs entrent en compétition²⁵. Ainsi, l'effet de "détérioration" observé sur les temps de réponse verbale témoigne d'une interférence au niveau du traitement des informations nécessaires à la production conjointe de gestes manuels et de parole. Les données d'Holender [46] sont également en faveur d'un lien unidirectionnel entre les deux modalités, le décours temporel du geste manuel n'étant pour sa part pas modifié par l'ajout de parole.

25. « *When used together, these processors compete for a common processing capacity pool* » (Holender [46], p.431).

Il faut toutefois noter une certaine asymétrie, en termes de complexité, entre la tâche verbale et la tâche manuelle : il est possible que la tâche de pression de touche nécessite une charge cognitive plus importante que la tâche de nomination, la lecture étant un processus rapide et (plus ou moins) automatique ; ceci expliquerait l'effet de détérioration unidirectionnel observé par Holender [46] : la charge cognitive élevée induite par la tâche de pression de touche ralentirait le processus de lecture (ce qui entraîne un retard au niveau des performances verbales en tâche double relativement à la tâche simple), tandis que le processus automatique de lecture n'altérerait pas le processus de pression de touche (les performances manuelles sont donc identiques en tâches double et simple). Tester une tâche manuelle moins complexe (en réduisant par exemple le nombre de boutons-réponse utilisés) permettrait de déterminer dans quelle mesure les résultats d'Holender [46] sont liés à une organisation temporelle propre aux productions bimodales ou liés à un biais expérimental (les deux tâches n'étant pas strictement comparables en termes de difficulté). Par ailleurs, comme le fait justement remarquer Roustan [88], il s'agit là d'une coordination entre « *organes moteurs* », i.e. entre deux réponses motrices, dans le sens où geste et parole n'entretiennent pas de lien communicatif. Se pose alors la question de la validité de ces résultats dans un cadre langagier communicatif, où geste et parole auraient pour but commun de transmettre une information, pour quelqu'un.

Caractérisation des interférences parole/geste Plusieurs hypothèses ont été formulées au sujet de ces interférences geste/parole, à l'origine du lien temporel particulier unissant ces deux systèmes de production : comment et à quel niveau de traitement ce lien s'établit-il ? Une première hypothèse (théorie modulaire²⁶, e.g. de Ruiter [89]) postule que le lien geste/parole est établi dès la phase de planification, sans aucune interaction entre les deux systèmes : la coordination serait définie pendant la phase de planification, avant que l'exécution motrice ne soit initiée. Une seconde hypothèse (théorie interactive, e.g. Furuyama et collab. [32]) propose en revanche que le lien geste/parole s'effectue de manière interactive, avec une adaptation mutuelle des deux systèmes, et ce y compris durant la phase d'exécution motrice.

De Ruiter et Wilkins [92] suggèrent que cette synchronisation geste/parole soit caractéristique des productions bimodales et devrait s'observer de façon universelle, indépendamment des différences linguistiques et gestuelles inter-culturelles²⁷ (voir également McNeill [78], Morrel-Samuels et Krauss [82] et de Ruiter [89]). Les auteurs observent des locuteurs néerlandais et arrernte (langue australienne aborigène), filmés en situation d'interview libre, et se focalisent sur les gestes coverbaux de nature spatiale (orientation, localisation, etc.). Ils mesurent l'onset du mouvement, ainsi que l'onset et l'offset du stroke (sans tenir compte de la phase de rétraction, ou retour). Les frontières temporelles de début et de fin de l'affilié lexical (i.e. « *The word or*

26. La théorie modulaire s'inspire de la théorie de l'architecture modulaire énoncée par Fodor [29], selon laquelle l'esprit humain comprendrait un certain nombre de modules spécialisés dans l'exécution de certaines fonctions cognitives. Ces modules fonctionneraient de manière innée, automatiquement, inconsciemment, rapidement, parallèlement et indépendamment les uns des autres, s'opposant en cela au système central conscient, contrôlé mais aussi lent et séquentiel. Un exemple de module proposé par Fodor concerne le traitement du langage, pour lequel il existerait des sous-modules (sémantique, syntaxique, morphologique) indépendants les uns des autres.

27. « *Potential cross cultural differences are presumed not to play a role in the issue of gesture/speech synchronization* » (de Ruiter et Wilkins [92], p.603).

2. Liens geste/parole dans le langage moderne

*phrase with which the gesture is semantically and pragmatically linked*²⁸ », [92], p.604) sont également considérées. Leurs résultats montrent que, bien que les locuteurs arrernte aient une phase de préparation du geste manuel (i.e. intervalle entre le début du mouvement et le début du stroke) plus longue que les locuteurs néerlandais (due à une utilisation plus large de l'espace et notamment des gestes effectués à bout de bras), ceux-ci ne retardent pas pour autant le début de la parole : le timing de la parole ne s'adapte donc pas au timing du geste durant l'exécution. Ces résultats suggèrent une indépendance des systèmes durant l'exécution des programmes moteurs. Par ailleurs, les auteurs observent des patrons temporels similaires à ceux rapportés par Holender [46] : le geste précède la parole, et ce dans 94% des énoncés néerlandais et 88% des énoncés arrernte. Globalement, ces données supportent plutôt une théorie modulaire, selon laquelle chaque système génère ses propres sorties indépendamment de l'autre : la relation geste/parole serait alors établie pendant la phase de planification, au cours de laquelle les deux systèmes auraient accès à une source centrale d'information (i.e. la structure conceptuelle, le message à transmettre). Puis, dès le lancement des programmes moteurs, les opérations d'un module seraient totalement hermétiques aux réactions de l'autre module. La coordination motrice geste/parole serait donc le fruit de décisions centrales prémotrices, impliquant deux systèmes complètement autonomes.

A l'inverse, Furuyama et collab. [32] sont partisans de l'hypothèse interactive, qu'ils tentent de valider par le biais d'un paradigme plus contrôlé, celui du cartoon : la tâche du participant est de raconter, de façon détaillée, un épisode de *Titi & Grosminet* à son interlocuteur, naïf. Mais afin d'élucider le maximum de gestes de pointage, et surtout d'obtenir des gestes variables en termes de durée de phase de préparation (et de fait comparables à ceux étudiés par de Ruiter et Wilkins [92]), les auteurs modifient quelque peu la procédure habituelle : le participant doit pointer du doigt les images appropriées (tirées du dessin-animé), affichées de façon aléatoire sur différents murs de la pièce. Les auteurs mesurent la durée des phases de préparation des gestes de pointage ainsi que la différence temporelle entre le début du stroke et le début de la parole. Ils démontrent le maintien d'une certaine synchronie entre geste et parole, indépendamment de la durée de la phase de préparation. Leurs résultats sont en faveur d'une interaction continue entre les deux systèmes, pour lesquels les relations temporelles sont finement contrôlées. Geste et parole ne sont plus envisagés comme des modules encapsulés mais peuvent atteindre un certain degré d'ajustement mutuel, grâce à une interaction pendant les phases de planification **et** d'exécution des programmes moteurs. Un feedback continu du système gestuel permettrait l'exécution du message verbal associé à un moment bien particulier de l'exécution motrice. Réciproquement, un feedback du système vocal permettrait à l'apex du geste d'être réalisé plus tôt ou plus tard en fonction du moment précis où une expression donnée est formulée. Ces données sont donc contradictoires à celles obtenues par de Ruiter et Wilkins [92] ; selon Furuyama et collab. [32], « *the findings of De Ruiter and Wilkins's experiment [...] resulted from factors other than the modularity of speech-gesture production*²⁹ ».

28. Traduction proposée : Le mot, ou expression, auquel le geste est lié, d'un point de vue sémantique et pragmatique.

29. Traduction proposée : Les résultats obtenus par de Ruiter et Wilkins s'expliquent par d'autres facteurs qu'une modularité des productions parole-geste.

Entre ces deux positions « extrêmes » (modularité totale vs interaction totale), il existe des possibilités intermédiaires. L'objectif des études suivantes est de tester différentes hypothèses (voir notamment l'hypothèse balistique proposée par Levelt et collab. [73], présentée ci-dessous) afin de déterminer, dans un cadre expérimental contrôlé permettant de caractériser le déroulement temporel d'énoncés unimodaux et bimodaux, la nature et le lieu des interférences geste/parole observées par Holender [46].

L'étude de Levelt et collab. (1985) L'étude de Levelt et collab. [73], réalisée en 1985, est une étude expérimentale majeure, ayant suscité, et suscitant toujours, beaucoup d'intérêt. Les auteurs s'intéressent au processus déictique afin de déterminer les mécanismes sous-jacents à la sélection de réponses multimodales, et ce dans un cadre communicatif.

Le but des auteurs est de confronter l'hypothèse interactive, énoncée précédemment et soutenue notamment par les travaux de Furuyama et collab. [32], à une nouvelle hypothèse, dite balistique, afin de mieux comprendre comment s'opère l'interdépendance geste/parole, au cours de la planification et de l'exécution motrice :

- Pour rappel, selon l'hypothèse alternative interactive (Furuyama et collab. [32], McNeill [78]), les canaux vocal et gestuel sont en interaction permanente depuis la planification des mouvements jusqu'à leur réalisation complète : toute perturbation sur l'un des canaux affecte directement l'autre. Les deux systèmes moteurs interagissent constamment pendant les phases de planification et d'exécution motrice de façon à accomplir un but commun.
- Selon l'hypothèse balistique, les deux systèmes interagissent durant la phase de planification mais sont indépendants lors de l'exécution motrice, et fonctionnent alors comme deux modules autonomes.

Quatre expériences successives sont réalisées sur le principe suivant (voir Figure 1.4 ci-après) : les participants sont face à différentes diodes lumineuses (deux ou quatre, selon l'expérience) disposées de gauche à droite le long d'une règle. A chaque enregistrement, l'une de ces diodes s'allume et le participant doit le plus rapidement possible la désigner du doigt et/ou la nommer (avec l'équivalent néerlandais de *cette lampe-ci* et/ou - selon l'expérience - *cette lampe-là*). Notons que la cible peut être située dans le champ contralatéral ou ipsilatéral à la main utilisée pour pointer.

Les auteurs mesurent les trois variables suivantes, nommées à l'origine T_i , T_a , T_v :

- **T[initiation]** correspond à l'intervalle temporel entre l'allumage de la diode et l'initiation du geste manuel ;
- **T[apex]** correspond à l'intervalle temporel entre l'allumage de la diode et l'apex du geste manuel (i.e. son extension maximale) ;
- **T[voix]** correspond à l'intervalle temporel entre l'allumage de la diode et le début de la réponse verbale (i.e. onset du voisement).

De plus, $T[\text{apex}] - T[\text{initiation}]$ leur permet de calculer $T[\text{exécution}]$, soit la durée du stroke. Comme le fait justement remarquer de Ruiter [91], le stroke, tel que défini par Kendon [52] (voir Section 1.3) correspond au pic d'effort du geste, soit la phase porteuse de sens. Mais dans

2. Liens geste/parole dans le langage moderne

le cas du pointage, l'unité porteuse de sens correspondrait plutôt à l'apex, soit le point temporel précis où le bras et l'index sont au maximum de leur extension respective. Il est donc pertinent, dans le cas d'un geste déictique, de considérer le stroke comme l'intervalle entre le début du mouvement et l'apex.

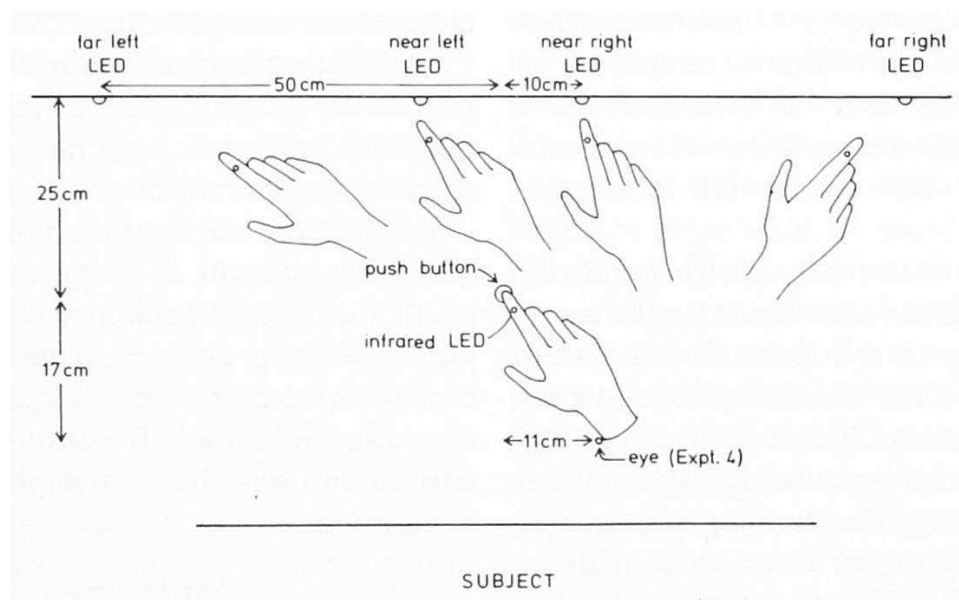


FIGURE 1.4 – Disposition spatiale des diodes lumineuses et positions de la main attendues (Levelt [71], p.136). Le système Selspot (*Selective-Spot-Recognition* : système opto-électronique de capture du mouvement) permet l'enregistrement des propriétés cinématiques du geste manuel ; le bouton poussoir (*push button*), qui indique la position de repos, permet également l'enclenchement des essais ; enfin, le chas (*eye*) est utilisé dans l'Expérience 4 décrite ci-après.

- Expérience 1 :

La première expérience vise à déterminer dans quelle mesure voix et geste interagissent, et ce en observant l'effet de la distance sur l'organisation temporelle de chaque modalité. Les participants ont pour tâche de pointer et simultanément nommer (en utilisant l'un **ou** l'autre déictique : *cette lampe-ci ou cette lampe-là*) l'une des quatre diodes.

Les résultats montrent que $T[\text{apex}]$ augmente avec la distance, puisque le temps mis pour atteindre l'apex est plus grand pour les cibles éloignées (en moyenne 742 ms) que pour les cibles proches (en moyenne 652 ms). Ce résultat est stable puisqu'observé dans les trois autres expériences. Plus précisément, les auteurs observent un effet d'interaction entre la distance d'une part et le champ dans lequel est située la diode d'autre part : ainsi $T[\text{apex}]$ augmente lorsque les cibles sont situées dans le champ contralatéral (le geste est alors "croisé" et l'apex atteint plus tardivement). Par ailleurs, $T[\text{voix}]$ covarie avec $T[\text{apex}]$ puisqu'il est également retardé lorsque la cible est éloignée (611 ms vs 676 ms). Enfin, les résultats concernant l'instant d'initiation du geste confirment les résultats obtenus par Holender [46] : $T[\text{initiation}]$ précède $T[\text{voix}]$, de 341 ms lorsque la cible est proche et de 373 ms lorsqu'elle est éloignée.

- Expérience 2 :

La seconde expérience a pour but de caractériser l'adaptation inter-systèmes : est-elle bien unidirectionnelle comme le proposait Holender [46] ? Levelt et collab. [73] vont pour cela comparer les performances en tâche double (parole+geste, tout comme dans l'expérience 1) aux performances obtenues en tâches simples (parole ou geste). Les auteurs soulignent néanmoins qu'à la différence de la tâche de geste seul, la tâche de parole "seule" sera toujours accompagnée d'une certaine forme de pointage, de la tête ou du regard par exemple. Toutefois, ils estiment que ce pointage gestuel apparaît également en situation de parole+geste et ne devrait donc pas biaiser les résultats.

Les résultats montrent qu'en situation de geste seul, relativement à la tâche double, T[apex] est inchangé. L'exécution du geste n'est donc pas influencée par l'ajout de parole. En revanche T[initiation] est significativement retardé par la présence de parole. T[voix] est quant à lui largement retardé (de 99ms) en situation double et est donc clairement dépendant de la présence du geste³⁰.

Il semble donc que la planification de la parole retarde l'initiation du geste mais n'affecte pas son exécution, tandis que la planification du geste affecte clairement la production de parole. Ainsi, la covariation geste/parole pourrait être expliquée en termes d'adaptation de la parole au geste, et non l'inverse : le système de production de parole s'adapte au système de production du geste. En d'autres termes, la parole attend le geste pour être coordonnée avec celui-ci. L'effet de la main sur la parole serait donc bien unidirectionnel, comme le suggérait Holender [46] ; de fait, les résultats obtenus par Holender [46] semblent liés à la spécificité temporelle des productions geste/parole, plutôt qu'à la différence de charge cognitive induite par les tâches de geste et de parole (respectivement pression de touche et nomination). Levelt et collab. [73] expliquent ces résultats en termes de compétition entre deux systèmes moteurs, sur des ressources communes de traitement (« *Competition for common resources is (...) at stake in our pointing task* », Levelt et collab. [73], p.147). Par ailleurs, leurs données supportent l'hypothèse balistique : la planification des productions dans les deux modalités se ferait de manière interactive, sur un pool de ressources communes, puis leur exécution serait modulaire (c'est pourquoi T[initiation] est significativement retardé par l'ajout de parole, alors que T[apex] reste inchangé). L'espace-temps temporel (en termes d'onset) entre les deux types de productions permettrait de réduire la charge cognitive induite par une telle compétition.

- Expérience 3 :

La troisième expérience consiste à faire varier le nombre d'alternatives gestuelles et verbales, dans le but d'évaluer plus précisément l'éventuelle compétition inter-modalités. Les participants disposent ici d'une ou de deux alternatives verbales et de deux ou de quatre alternatives gestuelles. Selon les auteurs, une compétition pour des ressources communes de traitement impliquerait un effet d'interaction, sur les temps mesurés, entre le nombre d'alternatives gestuelles d'une part et le nombre d'alternatives verbales d'autre part.

30. De plus, comme dans l'expérience 1, T[apex] et T[voix] sont retardés lorsque la cible est éloignée (respectivement de 83 et 47 ms).

2. Liens geste/parole dans le langage moderne

Les résultats pertinents pour cette troisième expérience concernent donc l'effet du nombre d'alternatives vocales et gestuelles, et leur interaction, sur les temps de latence mesurés. Les auteurs montrent que T[initiation] est affecté par le nombre d'alternatives : il est retardé de 25 ms lorsqu'il y a deux déictiques et de 19 ms lorsqu'il y a quatre diodes. De plus, Levelt et collab. observent un effet d'interaction, qualifié de *suradditif* : le retard engendré par deux alternatives verbales est d'autant plus important lorsqu'il y a quatre alternatives gestuelles. T[apex] et T[voix] sont également affectés par le nombre d'alternatives gestuelles (avec un retard respectif de 21 et 23 ms) mais pas par le nombre d'alternatives verbales. Enfin, résultat moins surprenant, T[exécution] ne dépend pas du nombre d'alternatives proposées³¹.

Les auteurs concluent à une planification en parallèle des deux types de production. La compétition pour l'accès à des ressources communes augmente avec le nombre d'alternatives proposées, jusqu'à l'initiation des mouvements (d'où un effet d'interaction uniquement sur T[initiation]). Cette compétition engendrerait par la suite une exécution plutôt sérielle des productions. Ces résultats sont en faveur d'une exécution balistique des productions : les systèmes interagissent jusqu'à l'initiation du mouvement pour devenir autonomes durant la phase d'exécution.

- Expérience 4 :

Une quatrième et dernière expérience est mise en place afin de tester plus précisément l'effet du geste sur la parole, et son décours temporel. Levelt et collab. [73] ajoutent, à l'aide d'un système de ficelle, insérée dans un chas, et de masse (voir Figure 1.4 ci-dessus), une charge sur le poignet du participant, à un moment donné de sa course, provoquant ainsi une perturbation du mouvement.

Les auteurs comparent quatre conditions, impliquant différentes charges (ou lieu de perturbation), par rapport au décours temporel du geste (une charge peut par exemple être appliquée en début ou milieu de course). Les résultats montrent que T[apex] covarie avec le lieu de la perturbation : l'apex est atteint d'autant plus tard que la perturbation survient tard. T[voix] en revanche est retardée uniquement lorsque la perturbation est appliquée en début de course. Sans surprise, T[initiation] est quant à lui indépendant de toute perturbation³².

En observant ainsi l'effet des perturbations obtenues sur les réponses des participants, Levelt et son équipe concluent que geste et parole interagissent jusqu'à un certain point de l'exécution motrice (entre 300 et 370 ms avant le début du voisement), mais qu'ensuite une perturbation du geste n'a plus d'influence sur la parole. Ces résultats sont clairement en faveur d'une théorie balistique : la parole s'adapte au geste jusqu'à l'initiation du mouvement puis les deux systèmes moteurs deviennent totalement indépendants pendant la phase d'exécution motrice.

31. De plus, l'effet de la distance est de nouveau observé, avec un retard au niveau de T[apex], T[voix] et T[initiation] pour la diode la plus éloignée (respectivement de 96ms, 62ms et 39ms).

32. De plus, les résultats obtenus en condition sans charge permettent de confirmer les résultats précédents, liés à la distance de la cible : T[initiation], T[apex] et T[voix] sont tous trois retardés pour la diode éloignée.

L'étude de Feyereisen (1997) Feyereisen [28] reprend dans une certaine mesure les travaux de l'équipe de Levelt et collab. [73] afin d'en confirmer les résultats relatifs aux gestes déictiques et de les généraliser aux gestes iconiques. Le but de ce travail, réalisé en quatre étapes, est d'apporter des précisions quant à la nature exacte (1) des connexions geste/parole (facilitatrices vs inhibitrices) et (2) du lieu d'interférence entre les deux systèmes de production (en phase conceptuelle vs de planification vs d'exécution).

- Expérience 1 :

Une première expérience de pointage a pour but de reproduire les résultats obtenus par Levelt [73] sur les gestes déictiques, en utilisant une procédure simplifiée. La tâche des participants, assis devant un écran d'ordinateur, est de nommer (*ti* pour un stimulus apparaissant à gauche ; *ta* pour un stimulus apparaissant à droite) et/ou de pointer une croix apparaissant à l'écran. L'auteur mesure les temps de réponse, soit l'onset vocal et gestuel. Notons qu'à la différence de Levelt et collab. [73], l'auteur utilise des pseudo-mots en tâche de parole (soit des séquences de phonèmes pour lesquelles aucune représentation lexicale n'existe en mémoire, mais qui respectent les contraintes phonotactiques de la langue, i.e. les séquences de phonèmes admissibles, et sont donc prononçables, à la différence des non-mots). Les pseudo-mots *ti* et *ta* ont été sélectionnés du fait de leur ressemblance d'avec les déictiques utilisés par Levelt et collab. [73] (i.e. *dit* et *dat*) ; de plus ils commencent tous deux par une consonne occlusive, ce qui faciliterait le traitement des temps de réaction.

Cette première expérience permet de retrouver en partie les résultats de Levelt et collab. [73] : en tâche double, le geste de pointage précède la parole et celle-ci est retardée de 91 ms. Toutefois, Feyereisen montre également un retard du geste, de 42 ms, suite à l'ajout de parole. Ces résultats confirment une interdépendance geste/parole, avec une interférence non plus strictement unidirectionnelle mais toujours plus importante du geste sur la parole.

- Expérience 2 :

Leur seconde expérience vise à généraliser ces résultats aux gestes iconiques. Les sujets doivent alors associer quatre symboles, qui apparaissent à l'écran, à quatre configurations de la main (réponse gestuelle) et/ou quatre mots (réponse verbale) :

- \ : extension oblique de la main + « barre »
- ○ : opposition pouce-index + « boule »
- # : geste de quelqu'un prenant un cube d'environ 5 cm + « bloc »
- ◇ : présentation du dos de la main + « blanc »

On peut toutefois s'interroger sur la nature iconique de certains items : si les trois premiers, relatifs à des formes, sont clairement iconiques, le dernier, relatif à une couleur, l'est beaucoup moins. Cette asymétrie pourrait affecter les temps de réaction, potentiellement plus longs pour le dernier symbole.

Les résultats de cette seconde expérience confirment une anticipation du geste sur la parole mais non l'influence de la parole sur le geste : le geste iconique, tout comme le geste déictique dans l'expérience de Levelt et collab. [73], n'est pas modifié par l'ajout de parole. L'interdépendance serait donc unidirectionnelle (comme démontré initialement par Holender [46] et confirmé

2. Liens geste/parole dans le langage moderne

par Levelt et collab. [73]); l'influence de la parole sur le geste en revanche n'est pas robuste et semble liée à l'utilisation d'un paradigme spécifique (et potentiellement à l'utilisation de pseudo-mots).

- Expérience 3 :

Une troisième expérience vise à déterminer le lieu des interférences geste/parole. L'auteur utilise le paradigme d'amorçage, défini précisément au Chapitre 5, dont le principe est le suivant : le traitement d'un stimulus *amorce* faciliterait le traitement ultérieur d'un stimulus *cible*. L'effet d'interférence, en termes de différence *onset vocal* – *onset gestuel*, pourra ou non être réduit par un effet d'amorçage : si l'interférence se met en place dès les premières étapes de traitement de l'information (soit au niveau de la sélection de la réponse), alors l'effet de facilitation devrait réduire l'effet d'interférence ; en revanche, si l'interférence se met en place plus tardivement (lors de la planification de la réponse), elle sera indépendante de l'effet de facilitation.

La présentation des stimuli-cible est donc ici précédée d'indices visuels (stimuli-amorce) : dans le premier cas, un seul indice est présenté et permet de faciliter le traitement du stimulus-cible puisqu'il permet aux participants d'anticiper celui-ci ; dans le second cas, deux indices sont présentés et ne permettent donc pas aux participants d'anticiper leur réponse. En d'autres termes, l'incertitude augmente avec le nombre d'indices proposés. L'auteur fait également varier le délai entre la présentation de l'indice et celle du stimulus-cible (i.e. *Stimulus Onset Asynchrony*, SOA), qui peut être de 0, 300, 600 ou 900 ms. Ainsi, le traitement du stimulus-cible est facilité lorsque le nombre d'indice est minimum mais également lorsque le SOA augmente (un délai plus long permettant aux participants d'intégrer et de traiter l'indice).

Les résultats concernant la production de gestes montrent un effet principal du SOA et du nombre d'indices : les temps de réaction sont plus courts lorsque le SOA est long et lorsqu'il n'y a qu'un seul indice. Par ailleurs, l'auteur observe un effet d'interaction entre ces deux facteurs : les temps de réponse diminuent lorsque le SOA augmente, et ce d'autant plus lorsqu'il n'y a qu'un seul indice. En revanche, il n'observe pas d'effet de la tâche, le geste n'est donc pas modifié par l'ajout de parole. Les résultats diffèrent pour la production de parole : si l'effet principal du SOA ainsi que l'effet d'interaction sont similaires, le nombre d'indice en revanche n'influence pas les temps de réaction. Par ailleurs, l'auteur observe également un effet significatif de la tâche, la parole étant retardée par l'ajout du geste : les participants, en synchronisant le début de la parole avec l'apex du geste, ne donnent leur réponse verbale qu'une fois le geste lancé. Enfin, aucun effet d'interaction tâche/nombre d'indices et tâche/SOA n'est observé sur les productions gestuelles et verbales.

Les résultats principaux sont donc les suivants :

- Un effet d'interaction entre le SOA et le nombre d'indices sur les temps de réponse du geste et de la parole : il y a bien une facilitation du traitement des stimuli-cible, avec pour conséquence des réponses plus rapides.
- Aucune interaction tâche/indices et tâche/SOA sur les temps de réponse du geste et de la parole : la facilitation ne permet pas de réduire l'effet d'interférence geste/parole.

L'auteur conclut que l'interférence a lieu lors de la phase de planification des mouvements : elle serait le résultat d'une compétition entre les deux systèmes, compétition qui augmente

avec le nombre de modalités impliquées. Selon l’auteur, ces résultats sont en accord avec les conclusions de Levelt et collab. [73], à savoir que les deux systèmes interagissent jusqu’au point où les mouvements sont initiés.

- Expérience 4 :

Enfin, une dernière expérience essaie de déterminer si le traitement des productions manuelles et vocales se fait par deux systèmes distincts qui interagissent entre eux ou si un seul système gère les deux modalités. Il s’agit là de comprendre pourquoi l’interférence est unidirectionnelle. L’hypothèse de l’auteur est que la production d’un geste iconique implique nécessairement le traitement du matériel verbal associé (de fait, le geste iconique n’est pas modifié par l’ajout de parole reliée sémantiquement). En revanche, la réciproque n’est pas vraie : la production de parole n’impliquerait pas nécessairement le traitement du matériel visuo-spatial associé. Afin de tester cette hypothèse, l’auteur propose que le traitement de matériel verbal incongruent (i.e. non relié sémantiquement) devrait augmenter l’interférence parole-geste alors que le traitement de matériel visuo-spatial incongruent ne devrait pas modifier l’interférence geste-parole.

Deux tâches sont donc proposées aux participants, avant le traitement du stimulus (les stimuli sont identiques à ceux utilisés dans les expériences 2 et 3) : (1) charge cognitive verbale : mémorisation de paires de mots non reliés (de type *cheval - camion*) ; (2) charge cognitive gestuelle : mémorisation de matrices de croix 3*3. Les participants donnent ensuite leur réponse verbale ou gestuelle correspondant au stimulus-cible. De cette façon, l’auteur impose une charge cognitive forte sur la sélection à effectuer en parole ou en geste.

Les résultats montrent un effet principal de la condition identique pour le geste et la parole (absence d’interaction) : les deux types de production sont plus rapides en condition contrôle qu’en condition de charge cognitive, verbale ou visuo-spatiale. L’auteur conclut donc que l’expérience ne parvient pas à démontrer une dissociation entre les deux systèmes de sélection de réponse et ne permet pas d’expliquer l’unidirectionnalité de l’interférence en termes d’activation automatique de matériel verbal lors de la production de gestes iconiques.

De manière générale, les résultats de ces quatre expériences montrent un comportement balistique des productions : les deux systèmes, vocal et manuel, interagissent pendant la planification et jusqu’à l’initiation du mouvement. Feyereisen a donc pu généraliser les conclusions de Levelt et collab. [73] aux gestes iconiques et déterminer de manière précise le lieu (stade de planification) et la nature (compétitive) de l’interférence geste/parole.

2.2.2 Coordination geste manuel/geste articulatoire : l’étude de Rochet-Capellan (2007)

Rochet-Capellan [87] propose une approche intéressante afin de caractériser le lien temporel entre geste et parole et s’intéresse à la coordination temporelle du système main/doigt et de la mâchoire, soit un lien entre gestes manuels et gestes articulatoires plutôt qu’entre gestes manuels et sons. La relation geste/parole n’est donc plus évaluée en termes d’interférences, contrairement aux études citées précédemment (Feyereisen [28], Holender [46], Levelt et collab. [73]), mais en termes de coordination entre modalités.

2. Liens geste/parole dans le langage moderne

La tâche proposée aux participants est une tâche de production de pointage : pointer un personnage à l'écran tout en prononçant son nom ou réaliser chacune de ces tâches indépendamment. Les noms associés aux personnages sont des séquences bi-syllabiques CVCV, telles que *papa* ou *tapa*, qui requièrent un cycle de mâchoire par syllabe. D'autre part, chacune des syllabes peut être ou non la cible d'accentuation ('CVCV : pa.pa vs CV'CV : ta.pa).

Rochet-Capellan utilise un système opto-électronique précis de capture de mouvements afin d'observer l'organisation temporelle des productions de parole et de geste. En observant les paramètres articulatoires (liés aux mouvements de la mâchoire), acoustiques et temporels des productions, l'auteur met en évidence un « rapprochement » entre les deux systèmes.

L'hypothèse de départ, illustrée Figure 1.5, est celle d'une synchronisation entre la position basse de la mâchoire (pour la voyelle accentuée) et l'apex du pointage. Toutefois, les résultats obtenus correspondent davantage à un rapprochement qu'à une synchronisation stricte entre les deux événements : l'apex du geste semble attiré par la syllabe accentuée. Ceci est d'autant plus vrai lorsque l'accent est situé sur la première syllabe. 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 pointage.

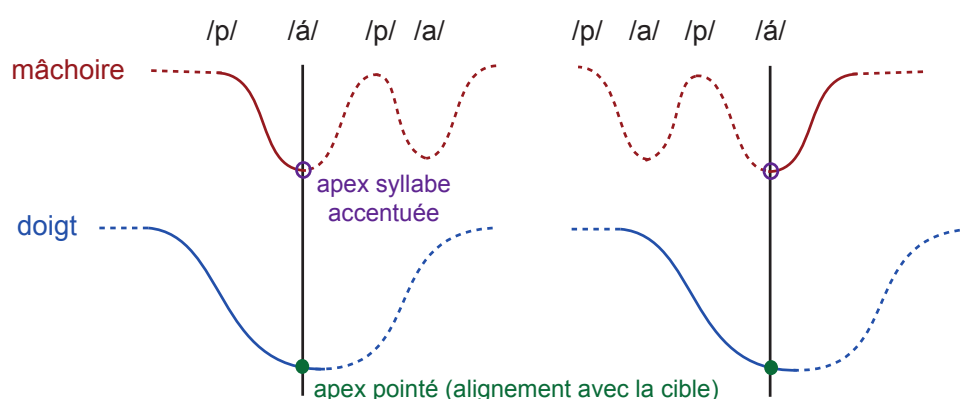


FIGURE 1.5 – Coordination geste manuel/geste articulatoire : en théorie, « l'apex du geste de pointé devrait advenir, à proximité de la position basse du cycle de la mâchoire pour la voyelle accentuée. » [87], p.91

La coordination entre les deux systèmes permet donc de réaliser l'accent de la parole pendant la tenue du geste de pointage : (1) quand l'accent est situé sur la première syllabe, la tenue du pointage est incluse dans le premier cycle de mâchoire (i.e. elle débute autour de la position basse du cycle et se termine après cette position), (2) quand l'accent est situé sur la deuxième syllabe, elle se décale vers le deuxième cycle de la mâchoire (i.e. elle débute autour de la position haute de la mâchoire entre les deux cycles et prend fin autour de la position basse du deuxième cycle). Ces profils de coordination sont relativement stables puisqu'ils s'observent quelles que soient la séquence prononcée et la position de la cible.

Le système de production de parole s'adapte donc au système de production de gestes, avec une réponse de la mâchoire qui débute toujours après le début du pointage manuel et qui s'adapte à la position de l'accent : le cycle accentué tend à se rapprocher de la phase de tenue du pointage. Par ailleurs, le pointage manuel s'adapte au système de parole de manière à se coordonner avec la réalisation de l'accent, avec une durée de geste plus importante (le retour du geste est retardé jusqu'à la réalisation de la partie accentuée) lorsque l'accent est situé sur la deuxième syllabe. Ces résultats, répliqués par Laurent [70] (celui-ci montre que le mouvement de la main commence toujours avant la syllabe accentuée et se termine toujours après la prononciation de la syllabe accentuée) sont à rapprocher des résultats de de Ruiter [89], qui observe une corrélation positive importante ($r : 0.61$, soit 37% de variance commune) entre le début de l'apex et l'onset de la syllabe accentuée.

2.2.3 Le cas particulier du LPC : L'étude d'Attina (2005)

L'étude d'Attina [3] est à notre connaissance la première à s'intéresser à la coordination temporelle entre les différents articulateurs du Langage Parlé Complété (LPC) : les lèvres et la main. Le LPC est un augment de la lecture labiale, composé de clés digitales (voir Figure 1.6 ci-après) : le principe est d'associer chaque voyelle à une position particulière de la main autour du visage et chaque consonne à une configuration particulière des doigts. Ceci permet de désambigüiser des syllabes de type CV et permet l'accès, visuel, à toute l'information de la chaîne parlée, difficilement accessible aux malentendants (en particulier les mots de liaisons ou la fin des mots). C'est donc l'utilisation conjointe des lèvres et de la main qui permet de rendre intelligibles, et d'identifier, les sons de la langue. Notons que le cadre du LPC est particulier dans le sens où geste et parole ne sont pas liés au niveau sémantique (au sens où le geste manuel est arbitraire, en termes de position de la main et de configuration des doigts) mais sont liés au niveau phonologique (puisque c'est l'association geste manuel/geste labial qui représente le phonème), soit un processus de plus bas niveau.

Attina [3] étudie donc l'organisation temporelle des augments manuels en relation avec l'organisation temporelle des lèvres et l'acoustique de la parole. Les codeurs LPC sont enregistrés lors de la production de séquences CV.CV.CV.CV et portent un Cyberglove (gant de données, muni de 18 capteurs angulaires placés au niveau des articulations et entre les doigts).

Le patron temporel des productions multimodales observé dans cette étude peut être résumé ainsi : le mouvement de la main commence avant le début acoustique de la syllabe et atteint sa cible en début de consonne. La main commence donc à se mettre en place pour former la consonne au cours de la réalisation de la voyelle précédente. Les informations vocaliques transmises par la main et par les lèvres ne sont donc pas synchrones : le geste de la main anticipe celui des lèvres. En revanche, l'information consonantique véhiculée par la main est produite en synchronie avec le début acoustique de la consonne.

2. Liens geste/parole dans le langage moderne

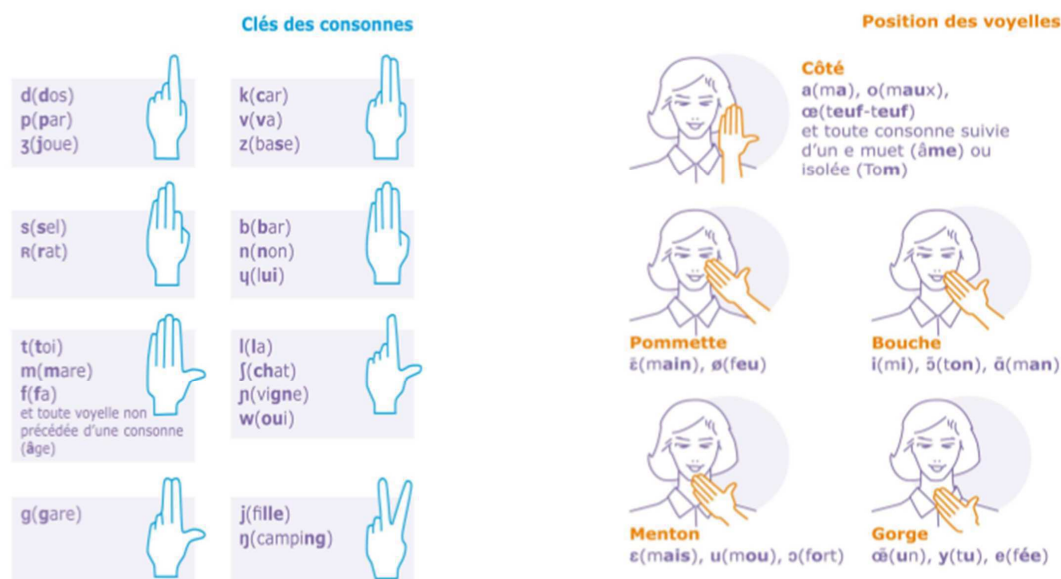


FIGURE 1.6 – Les clés du code LPC

Cette étude montre une interdépendance entre le geste manuel et le geste articulatoire, avec une anticipation du geste manuel, dans le cas particulier du LPC. Le patron temporel observé pour les gestes déictiques et iconiques notamment, i.e. une anticipation du geste sur la parole de moins d'une seconde, apparaît comme une règle générale dans les relations geste/parole et serait le résultat d'un processus de compétition entre ces deux activités motrices. Il y aurait une temporisation entre les deux systèmes (mouvement de la main anticipé ou mouvement des lèvres retardé) afin d'atteindre un même but phonologique. Attina [3] interprète ce phénomène comme un couplage fonctionnel optimal entre la main et les lèvres qui permettrait d'incorporer ces deux articulateurs dans un espace cérébral commun.

2.3 Connexion structurelle

Les études présentées jusqu'alors ont permis de mettre en évidence des connexions corticales et temporelles entre les modalités vocales et gestuelles. Certaines études mettent également en évidence un rôle fonctionnel direct du geste dans la production de parole : produire un geste nous permettrait de structurer notre discours (en termes de formes de surface, voire même de pensées). Ces études portent donc non plus sur le décours temporel des gestes, mais sur leur nature, en lien avec la parole. De plus, il existerait une influence réciproque, de plus bas niveau, entre les paramètres cinématiques du geste manuel d'une part et les paramètres articulatoires de la parole d'autre part.

2.3.1 Rôle du geste dans la production de parole

« *Gestures, together with language, help constitute thoughts*³³ » McNeill ([78], p.245)

Le geste manuel pourrait jouer un rôle fonctionnel dans la production de parole, et plus particulièrement dans la structuration d'énoncés de nature spatio-motrice (Kita [58], Krauss et Hadar [67]). La production de gestes semble en effet liée à l'expression de termes spatiaux et concrets, tels que *dessus* ou *cube*, plutôt que non spatiaux ou abstraits, tels que *chanson* ou *paix* (Krauss [64], Krauss et Morsella [68], Rauscher et collab. [85]). Dans cette optique, le geste est vu comme « *an overt manifestation of spatiomotoric thought*³⁴ », (Hostetter et Alibali [47], p.1/6) et serait clairement lié à la production de parole associée ; mais à quel niveau cette association se produit-elle ?

D'aucuns considèrent que la production d'un geste faciliterait l'accès au lexique et le choix des formes de surface appropriées (*Lexical Access Hypothesis*, i.e. Hypothèse de l'Accès Lexical, soutenue notamment par Krauss et collab. [65]). Comme le soulignent Krauss et Morsella [68], « *evidence from a variety of sources suggests that the movements people make while speaking contribute to speech production by facilitating lexical retrieval*³⁵ » (voir notamment la revue proposée par Krauss et collab. [65], mais aussi les travaux de Frick-Horbury et Guttentag [31] et de Beattie et Coughlan [8] qui invalident cette hypothèse). Par exemple, le nombre de gestes produits semble dépendre de la nature improvisée vs préparée de l'énoncé (Chawla et Krauss [17]) mais également de la nature abstraite vs concrète de l'objet décrit (e.g. Graham et Argye [38]) : ainsi le nombre de gestes augmente lorsque les participants cherchent en direct les termes appropriés ou lorsqu'ils doivent décrire un objet difficilement verbalisable. Par ailleurs, la fluence verbale est altérée lorsque les participants n'ont pas le droit de produire des gestes manuels (e.g. Krauss et Morsella [68]) et ce d'autant plus lorsque le discours est relatif à du contenu spatial (e.g. Rauscher et collab. [85]). Dans le même registre, les patients aphasiques présentant des troubles d'accès au lexique produisent plus de gestes que les participants contrôle du même âge, ce qui n'est pas le cas des autres patients (Hadar et collab. [40]).

D'autres en revanche suggèrent que le geste interviendrait plus tôt dans le traitement de la parole, lors de la conceptualisation même du message (*Information Packaging Hypothesis*, i.e. Hypothèse de Présentation de l'Information, soutenue notamment par Kita [58]). Kita [58] suggère en effet qu'une facilitation du processus d'accès lexical puisse s'expliquer par une facilitation antérieure au niveau du processus de conceptualisation.

Alors, la production du geste facilite-t-elle l'accès au lexique *per se* ou bien la conceptualisation du message, avec pour conséquence directe un accès plus rapide aux formes de surface ?

Hostetter et Alibali [47] tentent de confronter ces deux hypothèses, en manipulant la difficulté conceptuelle de la tâche, tout en gardant la difficulté d'accès au lexique constante. Les auteurs proposent une tâche de description de points, qui peuvent être interprétés comme un ensemble de formes géométriques concrètes, par exemple un carré sur lequel est posé un triangle. Dans

33. Traduction proposée : Les gestes, associés au langage, **nous permettent de structurer nos pensées** (mis en valeur dans le texte original).

34. Traduction proposée : Une manifestation directe et visible de la pensée spatio-motrice.

35. Bon nombre d'études suggèrent que les gestes qui accompagnent le discours contribuent à la production de parole par une facilitation de l'accès lexical.

2. Liens geste/parole dans le langage moderne

la condition de conceptualisation simple, les quatre points formant le carré, ainsi que les trois points formant le triangle, sont reliés entre eux par des lignes (les formes apparaissent alors clairement) ; en revanche en condition de conceptualisation difficile, les points ne sont pas reliés entre eux. Néanmoins, la description verbale de ces patterns devrait être identique dans les deux conditions. Ces auteurs soutiennent l'Hypothèse de Présentation de l'Information (Kita [58]) et supposent donc que le geste joue un rôle primordial dans la structuration de la pensée, avant même le processus de formulation du message. De fait, cette hypothèse prédit que la production de gestes devrait augmenter lorsque la conceptualisation du message est difficile. En revanche, l'Hypothèse d'Accès Lexical prédit que le nombre de gestes devrait être identique dans les deux conditions, la difficulté d'accès au lexique étant maintenue constante.

Tout d'abord, les résultats ne montrent aucune différence en termes de contenu verbal entre les deux conditions, ce qui permet d'affirmer que la difficulté de l'étape de formulation est bien indépendante de la condition. Par ailleurs, le nombre de gestes représentationnels augmente en condition de conceptualisation difficile : selon les auteurs, « *gestures help speakers organize their knowledge that is spatio motoric in nature and put it into a verbalizable form. Gesture is thus a mode of thinking*³⁶ » (Hostetter et Alibali [47], p.2/6). Ces résultats soutiennent clairement l'hypothèse d'un rôle du geste dans la structuration de la pensée spatio-motrice et sont par ailleurs observés chez des enfants âgés de 4.11 à 6.2 ans³⁷ (Alibali et collab. [2]).

2.3.2 Influence structurelle de bas niveau : l'étude de Bernardis & Gentilucci

L'étude de Bernardis et Gentilucci [9] met en évidence une influence structurelle entre les systèmes de production de parole et de gestes, mais de plus bas niveau. L'hypothèse des auteurs est que parole et gestes (ici, des emblèmes) font partie d'un seul et même système de communication et sont structurés l'un par rapport à l'autre. En d'autres termes, la production d'un geste manuel devrait avoir des répercussions sur la production de la parole, et réciproquement la production de parole devrait avoir des répercussions sur la production de gestes manuels.

Les participants doivent répondre à la présentation des items suivants : les mots *ciao*, *no*, *stop*, le pseudo-mot *lao* et la suite *XXX* en condition contrôle. Dans une première expérience, leurs tâches consistent à :

- a. Prononcer l'item affiché à l'écran ;
- b. Produire le geste correspondant au mot affiché à l'écran (par exemple agiter la main pour *ciao*) ;
- c. Produire le geste + le mot correspondant ;
- d. Produire un geste dénué de sens (i.e. oscillations de haut en bas, impliquant des mouvements du poignet, du coude et des épaules, tout comme les gestes signifiants) + le mot affiché à l'écran. Ce geste doit également être produit lorsque la suite *XXX* apparaît à l'écran.

36. Traduction proposée : Les gestes aident le locuteur à organiser ses connaissances spatio-motrices et à les traduire en forme verbalisable. Le geste est ainsi un mode de pensée.

37. La notation utilisée pour l'âge des individus dans ce manuscrit respecte le format suivant : [nombre d'années.nombre de mois] (e.g. 4.11 = 4 ans, 11 mois).

Leur seconde expérience ne diffère que par la condition d. : les participants doivent cette fois effectuer le geste correspondant au mot présenté à l'écran mais prononcer simultanément le pseudo-mot.

Concernant la modalité vocale, les auteurs mesurent l'instant d'initiation de la parole, ainsi que la moyenne des deux premiers formants (F1 et F2), la fréquence fondamentale (F0), l'intensité³⁸ et la durée de la voyelle [o] ou [ɔ]. Concernant la modalité gestuelle, ils mesurent la durée totale, la hauteur maximale, ainsi que l'instant d'initiation du geste. Pour les gestes associés aux items *ciao*, *no* et *lao*, les auteurs mesurent également l'amplitude, le pic de vitesse moyen et le nombre de pics de vitesse (ainsi que la déviation standard associée) des oscillations manuelles.

Les résultats montrent une interaction spécifique entre les systèmes de production de parole et de gestes. Plus précisément, les valeurs de F0 et de F2 pour les mots augmentent lorsqu'ils sont accompagnés d'un geste signifiant (le geste non signifiant n'a quant à lui aucune influence sur la production de la parole). Les valeurs de F1, pour le mot *stop* uniquement, augmentent lorsqu'un geste en général (i.e. signifiant ou dénué de sens) est produit en parallèle. De plus, les différentes phases du geste sont plus courtes lorsqu'un mot (et non un pseudo-mot) est produit simultanément (ces données sont en partie répliquées dans l'étude de Barbieri et collab. [5]).

Les auteurs se basent sur l'augmentation du second formant lorsque l'énoncé est accompagné d'un geste signifiant pour conclure que le geste renforce la parole : une commande motrice serait envoyée simultanément au bras et à la langue, qui se positionnerait plus en avant dans la bouche. Cette protrusion (traduite par des valeurs élevées de F2) serait liée à l'intention du locuteur d'interagir. En d'autres termes, le geste renforcerait la parole afin d'attirer l'attention d'un éventuel interlocuteur et d'entrer en interaction avec lui. A l'inverse, les auteurs mettent en évidence une inhibition de la parole sur le geste : la production d'un mot entraîne une réduction temporelle des différentes phases du geste. La quantité d'informations véhiculées par le geste serait moindre en cas de production conjointe de parole, cette dernière, d'une durée plus courte, étant disponible pour transmettre en partie l'information. En d'autres termes, la parole inhiberait le geste par un mécanisme de feedback négatif, mis en place une fois le mot prononcé.

Néanmoins, l'interprétation proposée par les auteurs pose question : une augmentation des valeurs de F2 pourrait traduire non pas un renforcement, au sens d'hyperarticulation, mais tout l'inverse, soit un phénomène d'hypoarticulation, au sens de centralisation des productions (e.g. si l'on augmente les valeurs de second formant d'un [o], celui-ci tendra vers un schwa). Dans ce cadre là, l'influence du geste sur la parole peut être interprétée comme une influence inhibitrice, de même nature que l'influence de la parole sur le geste.

Notons que cette étude permet également de confirmer les patterns temporels, observés notamment par Feyereisen [28], Holender [46], Levelt et collab. [73] : l'instant d'initiation de la parole (mots et pseudo-mots) suit l'instant d'initiation des gestes (signifiants et dénués de sens). Cette coordination semble modulée par la nature signifiante vs non signifiante des énoncés : re-

38. Ces quatre notions (F1, F2, fréquence fondamentale et intensité) sont définies au Chapitre 3, Section 1.3.

3. Modèles de production gestes/parole

lativement à la condition mot-geste signifiant, l'intervalle entre les deux types de production diminue lorsque le mot est produit avec un geste non signifiant, mais augmente lorsque le geste est produit avec un pseudo-mot. Toutefois, leurs données ne permettent pas de caractériser les interférences parole/geste, puisque les onsets vocal et gestuel obtenus en tâches simples (a. & b.) ne sont pas comparés à ceux obtenus en tâches doubles (c. & d.).

De manière générale, ces données sont en faveur d'un couplage temporel et structurel spécifique entre un geste et un mot sémantiquement reliés, et d'une interaction bidirectionnelle entre les processus de traitement de l'information et d'exécution des programmes moteurs de la main et de la bouche. Les auteurs suggèrent que geste et parole sont gouvernés par un seul et même système linguistique.

3 Modélisation de la production conjointe de gestes et de parole

C'est naturellement que le geste manuel, étroitement lié, à divers niveaux (i.e. cortical, temporel, structurel), à la parole, a progressivement été intégré aux modèles de production de parole. Le but des modèles de production geste/parole est de pouvoir rendre compte précisément de la génération et de la production conjointe des deux types de signaux.

De manière générale, les modèles de production conjointe de gestes et de parole ont pour objectif de répondre aux questions suivantes :

- A quel niveau s'effectue la génération du geste, relativement à la conceptualisation et à la planification de la parole ?
- Comment le contenu du geste est-il relié (de manière informative et temporelle) au contenu de la parole co-occurrente ?
- Le geste est-il un support auxiliaire à la parole ou une modalité langagière à part entière ?

Selon le modèle considéré, la planification du geste peut être pré-linguistique, post-linguistique, ou linguistique. Cela implique nécessairement des hypothèses spécifiques quant à l'interaction geste/parole, qui peut alors être unidirectionnelle ou bidirectionnelle. Nous distinguerons ici deux types d'interaction : (1) une interaction structurelle, relative à l'influence de la parole sur la nature et la forme du geste, et/ou inversement ; (2) une influence temporelle, relative à l'influence de la parole sur le déroulement temporel du geste, et/ou inversement.

Une grande partie des modèles décrits ci-après s'inspire du modèle de production de parole de Levelt ([71, 72]), présenté en Annexe.

3.1 Génération pré-linguistique du geste manuel

L'hypothèse de génération pré-linguistique, que Kita et Ozyürek [60] désignent sous le nom de *Free Imagery Hypothesis* (Hypothèse d'Imagerie Libre), postule, comme son nom l'indique, que les gestes manuels sont générés de manière pré-linguistique, c'est-à-dire indépendamment du potentiel de représentation du langage. Plus précisément, les gestes seraient générés à partir de l'imagerie en mémoire de travail. Cette hypothèse prédit de fait que l'information encodée par le geste ne peut être influencée par la façon dont elle pourrait être verbalement exprimée.

Cette hypothèse est notamment soutenue par Krauss et collab. [66], dans leur modèle interactif KCG, et par de Ruiter [89], dans son modèle Sketch : tous deux sont en faveur d'une génération pré-linguistique du geste. Toutefois, le premier conçoit le lien geste/parole comme un *lien préconceptualiseur* (pour reprendre les termes de McNeill [79]), soit deux modules distincts pour la génération de la parole d'une part et du geste d'autre part, alors que le second le conçoit comme un *lien conceptualiseur* (dans les termes de McNeill [79]), soit un module commun pour la génération de la parole et du geste.

3.1.1 Le modèle interactif KCG (Krauss et collab., 2000)

Le modèle de production de gestes et de parole de Krauss et collab. [66], basé sur le modèle de production de Levelt [71, 72] détaillé en Annexe, est un modèle dit interactif, où les systèmes de production impliqués, geste et parole, fonctionnent en parallèle et interagissent (à la différence des processus dits autonomes, qui opèrent de manière indépendante dès leur initiation achevée). Les relations temporelles geste/parole (voir Section 2.2), et notamment le fait que l'onset du geste précède généralement celui de son affilié lexical, le phénomène du *tip-of-the-tongue* (voir Annexe), ainsi que l'altération de la fluence verbale (liée au contenu spatial) lorsque la production de gestes est prohibée, sont autant de phénomènes en faveur de cette interactivité.

Ce modèle s'intéresse à la production des gestes lexicaux, ces gestes qui surviennent uniquement en présence de parole, variables tant au niveau de leur durée que de leur forme, non répétitifs et complexes. Nombre d'entre eux sont liés sémantiquement à la parole qu'ils accompagnent. Plus globalement, cette catégorie regroupe les gestes qui ne sont ni symboliques, ni déictiques, ni moteurs (i.e. gestes de battement).

Les auteurs font l'hypothèse que geste et parole impliquent deux systèmes de production distincts, mais opérant de concert (voir Figure 1.7).

Le Conceptualiseur joue un rôle strictement identique à celui proposé par Levelt [71, 72] (voir Annexe) et s'occupe donc exclusivement de la génération verbale du message. Le lien geste/parole est ici qualifié de préconceptualiseur : selon les auteurs, le contenu du geste ne fait pas partie de l'intention communicative du locuteur (voir Exemple ci-dessous) mais dérive de représentations non propositionnelles liées au concept-source.

3. Modèles de production gestes/parole

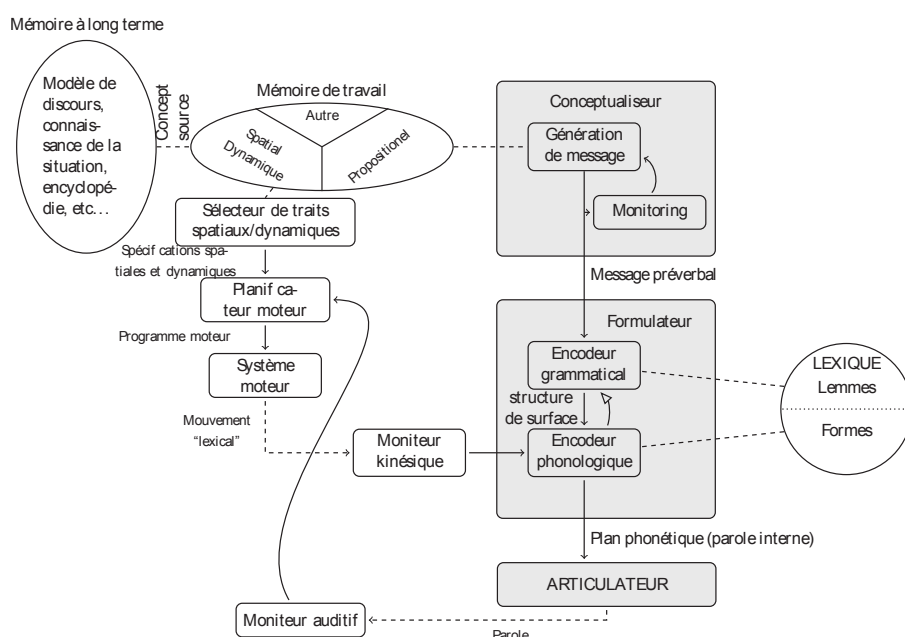


FIGURE 1.7 – Modèle interactif KCG pour la production geste/parole (Krauss et collab. [66], cités par Roustan [88], p.13)

Exemple : « J’ai mangé un gros gâteau »

[Mouvement de l’index décrivant un cercle]

→ Le concept « rond » ne fait pas partie de l’intention communicative.

Les auteurs formulent plusieurs hypothèses quant à l’organisation de la mémoire, véritable base de données de connaissances, et de ses représentations :

- Les connaissances sont stockées en mémoire sous différents formats (non propositionnel et propositionnel) ; la plupart des concepts sont encodés sous plus d’un format de représentation.
- L’activation d’un concept sous un format de représentation donné tend à activer les concepts reliés, encodés sous d’autres formats.
- Les concepts sont plus ou moins adaptés à un format ou un autre ; la représentation mentale complète de certains concepts peut nécessiter la prise en compte de plusieurs formats.
- Certaines représentations dans un format peuvent être traduites dans d’autres formats (par exemple, une description verbale peut donner lieu à une image visuelle, et vice versa).

Selon ce modèle, les gestes sont pris en charge dans un module spécifique, appelé *Sélecteur de Traits Spatio-Dynamiques*. Celui-ci accède aux représentations stockées, sous divers formats, dans la mémoire de travail (on parle de modèle de traits, par opposition au modèle *imagistic*), qui vont par la suite être exprimées par la parole. Son rôle est de transformer l’information stockée en format spatial ou dynamique en un ensemble de *spécifications* spatio-dynamiques, qui représentent les propriétés abstraites du mouvement. Ces spécifications sont ensuite envoyées au Planificateur Moteur, qui génère le programme moteur approprié.

Finalement, le Planificateur Moteur va fournir au Système Moteur les instructions nécessaires à l'exécution du geste. Le geste en sortie est relié à l'Encodeur Phonologique par le Moniteur Kinésique. Ce mécanisme permet aux gestes générés de faciliter l'accès lexical et notamment la récupération des lexèmes. On parle d'amorçage cross-modal, puisque la représentation motrice du geste va faciliter l'accès aux représentations sonores du mot. Par ailleurs, le Planificateur Moteur est relié au Moniteur Auditif. La durée du geste étant intimement liée à l'articulation de son affilié lexical (Morrel-Samuels et Krauss [82]), ce mécanisme, qui nous permet d'entendre l'articulation de l'énoncé, permet au Moniteur Auditif d'envoyer alors un signal au Planificateur Moteur afin que celui-ci mette fin à l'exécution du geste. Le modèle intègre ainsi une coordination temporelle en direct entre les deux modalités.

3.1.2 Le modèle Sketch (de Ruiter, 1998)

De Ruiter [89] se base également sur le modèle de production de parole de Levelt [71, 72], détaillé en Annexe, afin d'expliquer les processus de génération et de production conjointe du geste et de la parole. Il propose lui aussi une architecture modulaire, basée sur une approche de traitement de l'information, à laquelle il préfère le terme de *Representations & Processes Approach*, puisqu'il s'intéresse plus précisément aux représentations qui sous-tendent le traitement du geste. Son modèle permet d'expliquer la production de divers types de gestes, produits durant la parole et reliés sémantiquement à son contenu, mais n'intègre pas les gestes de battement.

La génération du geste se fait de manière pré-linguistique, mais à la différence du modèle KCG (Krauss et collab. [66]), le modèle Sketch propose que le lien geste/parole soit un lien conceptualiseur, où la composante du geste trouve son origine au niveau de la génération du message, dans le Conceptualiseur. De Ruiter conçoit donc le Conceptualiseur comme un module de traitement commun aux gestes et à la parole, qui ne se contente pas de générer un message préverbal mais doit également générer une représentation du geste à effectuer, le Sketch. Par ailleurs, ce modèle postule qu'hors du Conceptualiseur, geste et parole sont traités de manière indépendante et parallèle.

L'intention communicative du locuteur est ici scindée en deux sous-parties (voir Figure 1.8) : une structure propositionnelle, qui sera transformée en message préverbal, ainsi qu'une structure imagée (en d'autres termes, spatio-temporelle), qui sera quant à elle transformée en Sketch.

Le Conceptualiseur doit récupérer les informations pertinentes selon le type de geste à produire et les stocker dans un Sketch. Pour ce faire, il doit accéder à la mémoire de travail, pour en extraire les informations spatio-temporelles, appelées *trajectoires*, et encoder la position relative du locuteur par rapport à ces trajectoires. Le point de vue du locuteur (*perspective*), peut-être égo- ou hétéro-centré, c'est-à-dire centré sur le locuteur lui-même ou sur l'interlocuteur (voir Exemple ci-après).

Exemple : Description d'un itinéraire, où les interlocuteurs sont face à face. Le locuteur peut choisir d'accompagner la phrase « Il te faudra tourner à droite » d'un geste dirigé vers la droite, dans ce cas la perspective est égo-centrée, ou vers la gauche, dans ce cas, la perspective est hétéro-centrée.

3. Modèles de production gestes/parole

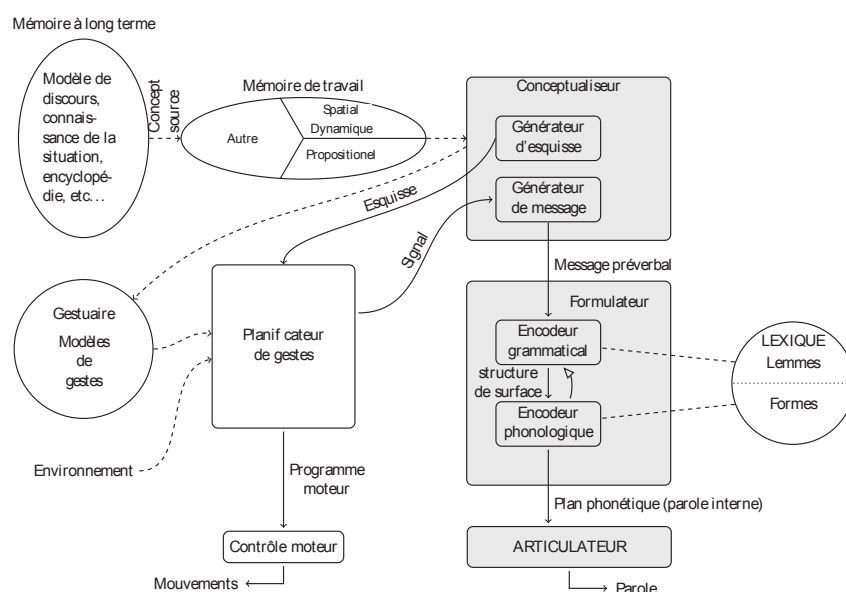


FIGURE 1.8 – Modèle Sketch pour la production geste/parole (de Ruiter [89], cité par Roustan [88], p.14)

Par ailleurs, le Conceptualisateur accède également aux informations relatives aux connaissances motrices (ou schémas moteurs).

Enfin, les gestes conventionnels (gestes déictiques et emblèmes, voir Section 1) nécessitent un traitement supplémentaire, où le Conceptualisateur accède au Gestuaire. Le Gestuaire est une base de données contenant un ensemble de programmes moteurs abstraits, appelés *templates*. Les gestes conventionnels sont composés par définition de paramètres fixés, et ce sont ces paramètres fixés qui sont représentés dans les templates. Prenons l'exemple de l'emblème signifiant « OK » (pouce et index arrondis et joints, autres doigts tendus vers le haut) : la configuration des doigts est entièrement stockée dans le template. Il en va de même pour le geste de pointage, pour lequel la forme de la main est sujette à certaines conventions et pour lequel le Conceptualisateur va également encoder le vecteur associé à la direction et à la location du référent. De manière générale, le Conceptualisateur utilise un système de *pointeurs*, lui permettant de référencer les informations nécessaires (telles qu'un template particulier) dans le Sketch.

De la même façon que le message préverbal est envoyé au Formulateur, le Sketch ainsi formulé au niveau du Conceptualisateur est ensuite envoyé au Planificateur de Gestes, dont le but est de construire le programme moteur approprié sur la base des informations contenues dans le Sketch. Pour cela, il récupère les informations indexées par *pointeur* dans le Sketch. Il doit également tenir compte des parties du corps nécessaires à l'exécution du geste (*body part allocation*). Dans le cas du geste de pointage par exemple, le Sketch contient les informations sur la location du référent mais si les mains ne sont pas disponibles pour réaliser le geste, le Planificateur de Gestes sélectionne un autre effecteur. La génération du Sketch et celle du programme moteur sont bien deux processus distincts. Par ailleurs le Planificateur de Gestes doit également gérer les contraintes imposées par l'environnement. Cela nous permet (en général!) de ne pas

nous heurter les uns aux autres, ou à un quelconque objet. En fonction de ces contraintes, un geste pourra être adapté, voire reporté. Enfin, le Planificateur de Gestes accède au Gestuaire, où comme nous l'avons vu précédemment, sont stockés les templates (soit les paramètres fixés) des gestes conventionnels. Toutefois, si certains paramètres sont fixés, les gestes conventionnels possèdent également un certain nombre de degrés de liberté : la durée du geste et l'orientation de la main notamment sont à définir en fonction de la situation. Le Planificateur de Gestes va donc ajuster ces paramètres libres et les combiner aux informations contenues dans le template, afin d'obtenir un programme moteur entièrement spécifié.

En résumé, le Planificateur de Gestes va prendre en compte les différentes informations contenues dans le Sketch et les combiner, afin de pouvoir envoyer le programme moteur final aux modules de contrôle moteur de bas niveau, qui exécuteront le mouvement. Alors que le plan phonétique est envoyé à l'Articulateur, le programme moteur est envoyé au Contrôle Moteur, qui exécute le mouvement manifeste.

Ce modèle permet d'expliquer plusieurs phénomènes liés à la synchronisation geste/parole, notamment les phases de tenue de geste (*pre-stroke hold* ; *post-stroke hold*, Section 2.2), par un système de signaux de synchronisation, pris en charge au niveau du Conceptualiseur.

De Ruiter émet l'hypothèse que la génération du geste prend moins de temps que celle de la parole, ce qui expliquerait que l'onset du geste précède généralement celui de la parole. Le Sketch est envoyé au Planificateur de Gestes avant que le message préverbal ne soit entièrement spécifié. Le Planificateur de Gestes prépare le programme moteur et envoie le pré-stroke au Contrôle Moteur. Lorsque le message préverbal est finalement envoyé au Formulateur, le Conceptualiseur envoie un signal-résumé au Planificateur de Gestes, qui peut dès lors envoyer la fin du programme moteur au Contrôle Moteur. Le Conceptualiseur envoie un signal-stop au Planificateur de Gestes lorsque la production du message préverbal est terminée, afin que celui-ci envoie le post-stroke au Contrôle Moteur.

Afin d'expliquer une synchronisation plus précise entre mot et geste, notamment dans le cas du pointage, le modèle Sketch prévoit un signal de synchronisation du Contrôle Moteur à l'Encodeur Phonologique : lorsque le Formulateur atteint le stade d'encodage phonologique, il attend un signal-go avant d'envoyer le message à l'Articulateur. Ce signal, en provenance du Contrôle moteur, est envoyé lorsque ce dernier peut approximativement prédire le moment où le geste va atteindre son apex.

Ce modèle permet également d'expliquer que la parole ne s'adapte plus au geste si celui-ci est retardé de plus de 300 ms (voir Section 2.2). Sachant que l'encodage phonologique dure environ 300 ms, une fois que celui-ci est activé, aucune interruption au niveau du geste ne pourra affecter la décours temporel de la parole.

De plus, Kita [57] a montré que lorsque la parole est perturbée, le geste est également interrompu simultanément. Selon de Ruiter, ce phénomène s'explique par le fait que le Conceptualiseur ait la possibilité d'envoyer un signal-stop au Formulateur ainsi qu'au Planificateur de Gestes, qui ralentissent alors leur traitement respectif.

3.2 Génération post-linguistique du geste manuel

L'hypothèse d'une génération post-linguistique du geste, ou *Lexical Semantic Hypothesis* (Hypothèse Lexico-Sémantique, d'après la description de Kita et Ozyürek [60]), suggère que

3. Modèles de production gestes/parole

les gestes soient générés à partir des informations sémantiques contenues dans la parole. Cette hypothèse prédit que les gestes représentationnels n'encodent pas ce qui n'est pas encodé par la parole co-occurrence. Les gestes sont vus comme un système auxiliaire, subordonné au système de parole, tant au niveau informatif que temporel. L'idée que les items lexicaux soient la source des gestes iconiques est proposée pour la toute première fois par Schegloff [93], pour lequel « *various aspects of the talk appear to be sources for gestures affiliated with them*³⁹ » (Schegloff [93], p.273).

Dans cette optique, Butterworth et Hadar [14] considèrent geste et parole comme deux systèmes autonomes. Selon les auteurs, un modèle de production de parole valide impliquerait nécessairement les étapes suivantes (l'ordre est proposé à titre indicatif, et n'est donc pas nécessairement respecté en pratique) :

Etape 1 : Construction pré-linguistique du message ;

Etape 2 : Détermination de la forme grammaticale de la phrase en cours de construction ;

Etape 3 : Sélection des éléments lexicaux sous leur forme abstraite, à partir d'un lexique sémantique ;

Etape 4 : Récupération de la forme phonologique des mots, sur la base de l'étape précédente ;

Etape 5 : Sélection des caractéristiques prosodiques, dont les points d'accentuation de la phrase ;

Etape 6 : Etape phonologique, où les éléments sont ordonnés syntactiquement et les traits prosodiques sont définis ;

Etape 7 : Spécification phonétique complète (phonologie + prosodie), comprenant les paramètres temporels ;

Etape 8 : Instructions envoyées aux articulateurs.

Ces étapes sont globalement similaires à celles proposées par Levelt [71, 72], définies en Annexe : l'étape 1 peut être associée à la Conceptualisation, les étapes 2 à 7 aux différents stades de la formulation, et l'étape 8 à l'articulation. La différence principale ici est l'**étape 3**.

Les auteurs suggèrent que lors de la production conjointe de geste et de parole, les processus de production de parole sont "dominants" : la parole modèle le geste de façon à ce que celui-ci soit conforme aux contraintes qu'elle impose.

Selon les auteurs, chaque étape de ce modèle de parole permet d'expliquer un phénomène gestuel particulier : les gestes iconiques sont générés suite à l'étape 3 (où le sens des mots est disponible), c'est-à-dire à partir de la sélection des items lexicaux sous leur forme abstraite. Autrement dit, la sélection des caractéristiques sémantiques spatiales serait à l'origine des gestes iconiques. Les gestes de battement, liés notamment à la prosodie, sont quant à eux générés suite à l'étape 7 (où l'information précise de la position de l'accent est disponible). Enfin, les changements de postures et les étapes de préparation, coordonnés à l'évolution des tours de parole et des changements de sujet de conversation (Butterworth et Beattie [13], Kendon [54]) relèvent de la première étape.

39. Traduction proposée : Divers aspects de la parole sont la *source* des gestes produits simultanément.

3.3 Génération linguistique du geste manuel

L'hypothèse de génération linguistique du geste, l'*Interface Hypothesis* (Hypothèse d'Interface) dans les termes de Kita et Ozyürek [60], stipule que les gestes sont originaires d'une *représentation d'interface* entre la parole et la *pensée spatiale*. Cette représentation d'interface peut être vue comme une représentation spatio-motrice, structurée en fonction des contraintes linguistiques. Selon cette hypothèse, les gestes encodent les propriétés spatio-motrices du référent (i.e. non linguistiques) et structurent l'information de manière à ce qu'elle soit compatible avec les possibilités d'encodage linguistique. En d'autres termes, cette hypothèse prédit que la forme du geste soit déterminée par les possibilités de formulation du langage (à la différence de l'hypothèse de génération pré-linguistique) et que le geste peut également encoder l'information spatio-motrice qui n'est pas exprimée par la parole (à la différence de l'hypothèse de génération post-linguistique). Cette hypothèse est basée sur la théorie du Growth-Point (GP) de McNeill et Duncan [80], elle-même inspirée de la théorie de Slobin [95], *Thinking-for-Speaking* : notre pensée est structurée de façon à pouvoir être formulée verbalement, conformément aux exigences et contraintes de notre langue : *penser pour s'exprimer*.

La théorie du Growth-Point (McNeill et Duncan [80]) s'applique aux gestes co-expressifs, qui sont synchronisés à la parole : geste et parole participeraient ensemble au dynamisme communicatif du discours, pour faire émerger un sens unique de la production ; ils expriment la même idée sous-jacente mais pas nécessairement les mêmes aspects. Ce modèle s'adresse également aux gestes idiosyncratiques : par opposition aux emblèmes, les gestes idiosyncratiques ne sont pas soumis à des normes mais sont créés sur l'instant par le locuteur. Ils reflètent en quelque sorte ses contenus mentaux et dépendent du sens particulier que le locuteur leur attribue dans une situation particulière. Ces gestes, contrairement à la parole, fonctionnent sur un mode global et analytique. En effet, le sens du geste ne correspond pas à la somme du sens de chacun de ses composants (forme de la main + trajectoire + position) et sa forme doit être interprétée comme un tout. Par ailleurs, une forme particulière de la main ne véhicule pas un sens "unique", au sens d'identique pour chaque geste dans lequel elle se trouve. En revanche, le sens d'une phrase est bien déterminé par la somme du sens de chacun des mots qui la composent et un mot particulier possède une signification "stable" au cours des productions. C'est de la combinaison de ces deux modes sémiotiques qu'un sens peut émerger ; ces combinaisons geste/parole seraient en effet nécessaires à l'obtention de toutes les caractéristiques du concept-source et représentent ainsi, selon les auteurs, « *an enhanced window onto thinking*⁴⁰ » (McNeill et Duncan [80], p.2). Selon McNeill et Duncan [80], les gestes manuels et la parole émergeraient d'un *Growth-Point*, une unité psychologique minimale⁴¹ combinant contenu imagé et contenu linguistique. Un Growth-Point émerge dès qu'une nouvelle idée se forme chez un locuteur. Gestes manuels et parole seraient systématiquement organisés l'un par rapport à l'autre et issus d'un seul et unique système. C'est ce Growth-Point qui permet à deux modes sémiotiques pourtant différents d'être produits en synchronisation et d'être co-expressifs. Selon McNeill et Duncan [80], « *the GP is a mediating link between individual cognition and the language system*⁴² » ([80], p.7).

40. Traduction proposée : Une fenêtre ouverte sur notre pensée.

41. Selon Vygotsky [98], l'unité minimale est l'unité de base conservant les propriétés essentielles d'un objet.

42. Traduction proposée : Le Growth-Point est le médiateur entre cognition individuelle et langage.

3. Modèles de production gestes/parole

3.3.1 Le modèle d'interface (Kita & Özyürek, 2003)

Le modèle d'interface (Kita et Özyürek [60]) se base sur le modèle de Levelt [71, 72] présenté en Annexe, et est illustré Figure 1.9.

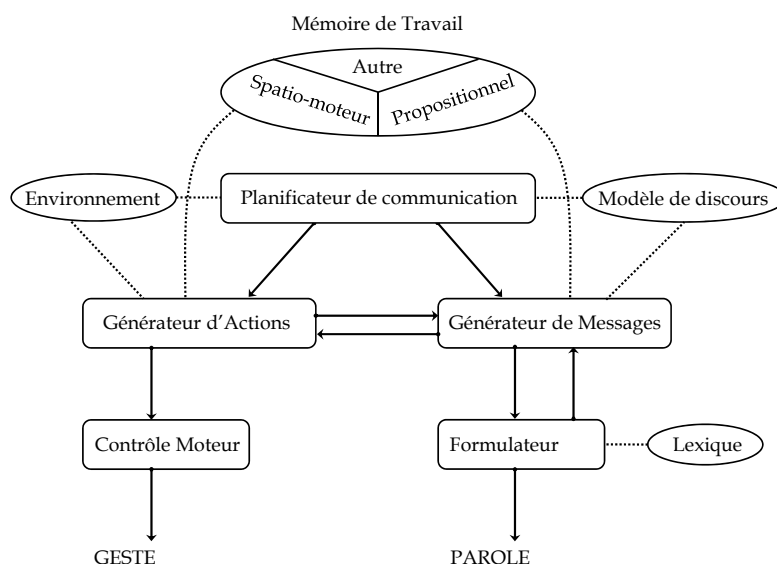


FIGURE 1.9 – Modèle d'interface pour la production geste/parole (Kita et Özyürek [60], cités par Roustan [88], p.16)

Le but principal de ce modèle étant de comprendre comment le contenu du geste est déterminé, il ne permet pas d'expliquer les comportements temporels entre les productions de gestes et parole.

Le Conceptualiseur est ici scindé en deux parties : le Planificateur de Communication et le Générateur de Messages. Le rôle du Planificateur de Communication est de générer l'intention communicative. Il s'occupe de la phase de *macro-planification*, c'est-à-dire qu'il décide de l'information à exprimer, ordonne les différentes parties de l'information et sélectionne les actes de parole appropriés. Par ailleurs, il doit également sélectionner la ou les modalités qui seront utilisées (sans pour autant déterminer quelle information va être exprimée par chacune d'entre elles) et peut éventuellement diviser la charge de travail en conséquence.

D'autre part, le Planificateur de Communication a accès au *Modèle de discours* : il est en permanence informé du déroulement du message (ce qui a été transmis, ce qui doit être transmis) et peut notamment mettre en évidence certaines informations, en fonction du but communicatif. Une fois l'intention communicative générée plutôt grossièrement (le contenu à exprimer n'est pas entièrement spécifié), celle-ci est envoyée en parallèle au Générateur de Messages et au Générateur d'Actions.

Le Générateur de Messages (sous-partie du Conceptualiseur) gère la phase de *micro-planification* : son but est de générer une proposition qui pourra être formulée verbalement, en tenant compte du but communicatif et du contexte discursif de l'énoncé. Cette proposition sera ensuite envoyée au Formulateur (les étapes suivantes sont similaires à celles proposées par Levelt, soit la Formulation puis l'Articulation du message). Le Générateur d'Actions va quant à lui générer

un plan d'action, basé sur les représentations spatio-motrices issues de l'imagerie spatiale et disponibles en mémoire de travail (notamment la courbure de la trajectoire et la direction du mouvement). Comme suggéré par de Ruiter [89] (voir Section 3.1.2), le Générateur d'Actions accède également aux informations relatives à l'environnement afin d'ajuster la forme du geste en fonction des caractéristiques et contraintes physiques de l'environnement (dont la position de l'interlocuteur, de l'objet à saisir ou à pointer, etc.). Le Générateur d'Actions est un mécanisme général pour l'action (et pas seulement à visée communicative), qui est donc relativement autonome lorsqu'il doit sélectionner une information particulière en mémoire de travail ou dans l'environnement (sélection basée sur les affordances, i.e. les possibilités d'action sur un objet donné). Les deux générateurs peuvent donc sélectionner et organiser l'information de façon relativement indépendante, ce qui implique que l'ordre des processus (au niveau des Générateurs) n'est pas fixe ou prédéterminé.

Les deux Générateurs échangent constamment de l'information. Cet échange en ligne et bidirectionnel implique des transformations entre les deux types d'information (propositionnel et spatio-moteur) : le Générateur d'Actions produit une représentation spatio-motrice, qui sera transformée en format propositionnel et transmise au Générateur de Messages ; réciproquement, le Générateur de Messages génère une proposition à formuler, transformée en format spatio-moteur et transmise au Générateur d'Actions. Lorsque la même intention communicative est transmise aux deux Générateurs, les contenus générés par les deux processus tendent à converger, à travers cet échange d'informations.

Par ailleurs, le Générateur de Messages interagit également avec le Formulateur (échange en ligne bidirectionnel) : le message généré par le Générateur de Messages est envoyé au Formulateur. Si la proposition ne peut pas être verbalisée en l'état, le Générateur de Messages reçoit un feedback du Formulateur. Ce dialogue entre le Générateur de Messages et le Formulateur est constamment transmis au Générateur d'Actions, qui pourra adapter le contenu du geste, en fonction des nouvelles propositions. Réciproquement, le Générateur d'Actions indique l'évolution (en termes de structure de l'information) au Générateur de Messages. Il y a donc un échange constant d'informations entre le Générateur d'Actions, le Générateur de Messages, et le Formulateur, jusqu'à ce qu'un point d'équilibre entre les différentes unités soit atteint. L'étape de Formulation peut alors se mettre en place, alors que la représentation spatio-motrice est envoyée au Contrôle Moteur.

C'est précisément cette représentation spatio-motrice, générée au niveau du Générateur d'Actions et déterminée par les possibilités d'encodage linguistique, qui constitue la représentation d'interface.

En résumé, le contenu du geste est déterminé par :

- L'intention communicative générée dans le Planificateur de Communication ;
- Les schéma moteurs (connaissances motrices, sélectionnées sur la base de caractéristiques spatiales imaginées ou réelles) ;
- Un feedback en direct du Formulateur au Générateur de Message.

Ce modèle permet d'expliquer certaines différences au niveau du geste, observées chez des participants de langues maternelles différentes (Kita et Ozyürek [60], Kita et collab. [61]). Par exemple, la description d'un mouvement peut nécessiter de prendre en compte deux notions, la

3. Modèles de production gestes/parole

manière et la trajectoire (*manner & path*). Certaines langues utilisent un seul terme encodant ces deux notions alors que d'autres langues ont besoin de deux termes différents. En français et anglais par exemple, l'action de descendre (*path*) tout en roulant (*manner*) est représentée par une seule expression (en français : *débarouler* ; en anglais : *roll down*). Mais cette même action nécessite deux propositions distinctes, du type « Il descendait en même temps qu'il roulait », en japonais ([*korogat-te*] [*saka-o kudaru*]) et en turc ([*yuvarlan-arak*] [*cadde-den iniyor*]). Cet encodage au niveau linguistique se reflète au niveau du geste : les participants n'utilisant qu'un seul terme utilisent un seul geste alors que les participants utilisant une combinaison au niveau de la parole utilisent également une combinaison au niveau du geste. Ainsi, les gestes produits exhibent une structuration de l'information similaire à celle observée en parole. Le modèle propose ainsi, par le biais de la représentation d'interface, l'existence d'une ressemblance entre la façon dont sont structurés la parole et le geste qui lui est associé.

3.3.2 *Gestures as Simulated Action* (Hostetter & Alibali, 2008)

Hostetter et Alibali [48] proposent une approche originale et novatrice permettant d'expliquer les mécanismes sous-jacents à la production de gestes coverbaux. A la différence des modèles présentés précédemment, le *Gestures as Simulated Action* (GSA) repose sur la théorie de l'*embodiment*, ou cognition incorporée (Wilson [100]).

L'idée sous-jacente est que les mécanismes de perception et d'action sont fortement couplés : « *we must perceive in order to move, but we must also move in order to perceive*⁴³ » (Gibson [36], p.223). Notre perception du monde qui nous entoure nous permettrait de mieux l'appréhender et d'interagir avec ; réciproquement, notre action sur le monde extérieur en faciliterait la perception : « *it is not only that perception determines possible action ; actions also determine what we know and perceive*⁴⁴ » (Hostetter et Alibali [48], p.497). Cette influence circulaire entre perception et action est illustrée Figure 1.10 ci-après.

Ce couplage perception/action joue un rôle essentiel dans nos interactions physiques avec le monde extérieur mais serait également à la base de nos représentations sensori-motrices stockées dans notre imagerie mentale⁴⁵ (e.g. Campos et collab. [15], Ellis et Tucker [26], Gerlach et Paulson [35], Jeannerod [50]). Ce mécanisme perceptivo-moteur sous-tendrait la production de parole, considérée ici comme la capacité d'indexer les objets par rapport au monde réel et d'en simuler les propriétés (e.g. Glenberg et Kaschak [37], Hauk et collab. [43]). Dans ce cadre là, les gestes coverbaux, qui par définition accompagnent la parole, seraient le fruit de simulations perceptivo-motrices, liées à l'activation des représentations de l'imagerie mentale, mises en action lors de la production de parole.

43. Traduction proposée : Nous devons percevoir pour nous déplacer, mais nous devons également nous déplacer pour percevoir.

44. Traduction proposée : Ce n'est pas seulement la perception qui détermine les possibilités d'action ; les actions déterminent également ce que nous savons et percevons.

45. L'image mentale est la représentation analogique d'un objet, ou événement, perceptif ou moteur. On distingue l'imagerie visuelle, qui correspond au fait de visualiser un objet en l'absence d'input sensoriel (représentations analogiques de nos perceptions), de l'imagerie motrice, qui désigne les représentations de notre corps en mouvements (représentations analogiques de l'action).

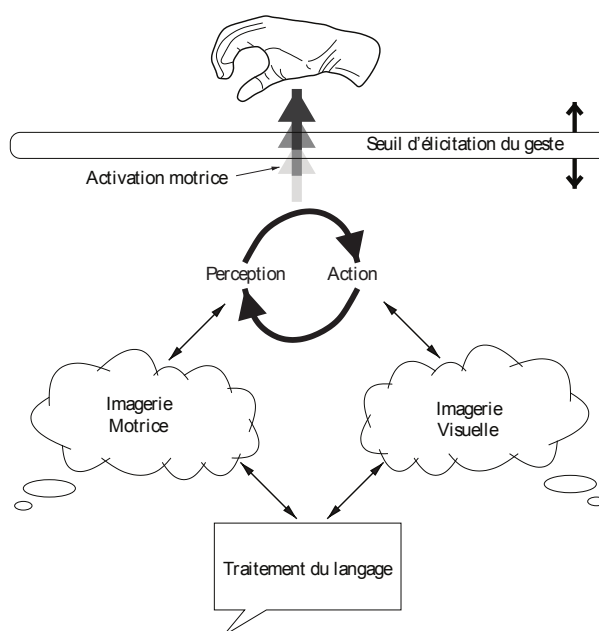


FIGURE 1.10 – Modèle GSA pour la production geste/parole (Hostetter et Alibali [48], cités par Roustan [88], p.18)

Plus précisément, la réalisation effective d'un geste manuel dépendrait de trois facteurs : la force d'activation des simulations, le seuil de geste, ainsi que la production conjointe de parole (i.e. l'engagement simultané du système moteur de la parole).

La force de l'activation dépend de la force de la simulation utilisée lors du traitement linguistique, mais également de l'âge du locuteur, de la clarté du percept simulé, de la nature abstraite vs concrète des pensées au service du langage, ou encore de la stratégie utilisée (individus plus ou moins "visuels"). Plus concrètement, les simulations perceptivo-motrices vont induire l'activation des aires cérébrales liées à la perception et la planification de l'action. Cette activation peut se propager aux aires motrices et aboutir à un mouvement manifeste : « *when this spreading activation occurs, a gesture is born*⁴⁶ » (Hostetter et Alibali [48], p.503).

Le seuil de geste correspond au seuil au delà duquel la production effective d'un geste est réalisée. Il peut varier en fonction des expériences passées, des croyances, mais aussi de facteurs neurologiques (connectivité entre les zones prémotrices et motrices par exemple), de facteurs cognitifs (désir de restreindre la production de gestes, sachant que selon les auteurs, inhiber la propagation de l'activation demanderait plus de ressources cognitives que de produire effectivement le geste), ou de facteurs sociaux.

Ces deux premiers facteurs sont nécessaires mais insuffisants : ils n'expliquent pas pourquoi le geste apparaît le plus souvent **avec** la parole ; comme le soulignent les auteurs « *it is rare*

46. Traduction proposée : Lorsque la propagation de l'activation se produit effectivement, un geste peut apparaître.

3. Modèles de production gestes/parole

[...] *for one to gesture while thinking in the absence of speech*⁴⁷ » ([48], p. 505). Afin d'illustrer ce phénomène, les auteurs proposent que la production de parole joue un rôle essentiel et direct dans les mécanismes de production de geste : **l'engagement simultané du système moteur de la parole** faciliterait l'engagement potentiel du système moteur manuel. Parler est une activité complexe et rapide, de fait l'activation liée aux articulateurs doit nécessairement se propager rapidement du cortex prémoteur au cortex moteur. Une fois le mécanisme de propagation mis en place pour le système de parole, la transmission de messages liés à d'autres commandes motrices, et en particulier aux gestes manuels, est facilitée. En d'autres termes, la production de parole favorise la production de gestes manuels.

Hostetter et Alibali [48] soulignent également que la majorité des gestes accompagnant la parole sont des gestes manuels ; cela suggère un couplage dynamique fort entre les actions orales d'une part, et les actions manuelles d'autre part. Cette coordination particulière entre main et bouche est un argument supplémentaire en faveur de l'existence d'un seul et même système cognitif et communicatif, composé du geste manuel et de la parole.

3.4 Synthèse des modèles de co-production geste/parole

Une synthèse, relative aux principales implications formulées par les différents modèles de production conjointe de gestes et de parole, est proposée Figure 1.11 ci-après. Elle détaille également les types d'interaction, et leur direction, proposés entre le geste et la parole.

	KCG	SKETCH	Butterworth & Hadar	INTERFACE	GSA
Gestes basés sur l'imagerie visuo-spatiale ?	Non	Oui	Non	Oui	Oui
Gestes influencés par des facteurs linguistiques ?	Non	Non	Oui	Oui	Oui
Geste et parole : un seul ou deux systèmes ?	2	2	2	2	1
Type d'interaction					
•Temporelle	Oui ($P \rightarrow G$)	Oui ($G \rightarrow P$)	Oui ($P \rightarrow G$)	Non traité	Oui ($P \rightarrow G$)
•Structurelle	Oui ($G \rightarrow P$)	Non	Oui ($P \rightarrow G$)	Oui ($P \leftrightarrow G$)	Oui ($P \rightarrow G$)

FIGURE 1.11 – Synthèse des modèles de co-production geste (G)/parole (P), à partir des travaux de Hostetter et Alibali [48], auxquels nous avons ajouté les informations écrites en gras

Selon le modèle KCG proposé par Krauss et collab. [66], le geste ne fait pas partie de l'intention communicative du locuteur mais est entièrement déterminé par les traits spatio-dynamiques élémentaires en provenance de l'environnement, le geste n'est donc influencé ni par les représentations visuo-spatiales, ni par des facteurs linguistiques. Geste et parole sont vus comme deux systèmes distincts, qui interagissent pendant la formulation et l'articulation du message verbal. Ce modèle est en faveur d'une interaction structurelle, du geste sur la parole (où le geste permet une facilitation de l'accès lexical), et d'une interaction temporelle, de la parole sur le geste (avec un signal envoyé par le Moniteur Auditif au Planificateur Moteur). Ainsi, ce modèle propose une influence structurelle sémantique à sens unique (geste $\rightarrow$ parole) et une coordination temporelle à sens unique également (parole $\rightarrow$ geste).

47. Traduction proposée : Il est rare de produire des gestes lorsque l'on pense sans parler.

Selon le modèle Sketch proposé par de Ruiter [89], le geste est basé sur les représentations visuo et spatio-temporelles, sans tenir compte de l'information verbale. Geste et parole sont deux systèmes distincts, dont les processus respectifs se chevauchent pendant la conceptualisation. L'interaction temporelle est expliquée par un signal à sens unique du geste sur la parole (envoyé par le Planificateur de Gestes au Formulateur). Ce modèle est en faveur d'une absence d'interaction structurelle et d'une interaction temporelle à sens unique (geste $\rightarrow$ parole).

Pour Butterworth et Hadar [14], le geste est déterminé par des facteurs linguistiques uniquement : geste et parole sont deux systèmes distincts hiérarchiquement organisés, le geste étant subordonné à la parole (système "dominant"). Il y a donc une interaction structurelle et temporelle à sens unique (parole $\rightarrow$ geste).

Pour Kita et Ozyürek [60], le geste et la parole sont issus d'une même représentation d'interface. Le geste est déterminé par les propriétés spatio-motrices du référent et les possibilités de formulation linguistique. Selon les auteurs, geste et parole sont deux systèmes distincts, qui interagissent de façon bidirectionnelle au niveau de la conceptualisation (à haut niveau). Leur modèle propose une interaction de nature structurelle, sémantique et bidirectionnelle (parole $\leftrightarrow$ geste) à haut niveau (entre les deux Générateurs). En revanche ce modèle ne permet pas de faire de prédictions au sujet de l'interaction temporelle, puisque cette thématique n'est pas abordée par les auteurs.

Enfin le modèle GSA proposé par Hostetter et Alibali [48] considère le geste et la parole comme un seul système de production. Les gestes sont déterminés sur la base de l'imagerie visuo-spatiale et à partir de facteurs linguistiques. Les deux systèmes interagissent au niveau structurel et temporel (parole $\rightarrow$ geste).

4 Conclusion

Les données présentées dans ce chapitre mettent en évidence une relation étroite entre les systèmes de production de parole d'une part, et de gestes manuels d'autre part, dans la communication langagière humaine. Ces deux modalités entretiennent des connexions corticales (Section 2.1), temporelles (Section 2.2) et structurelles (Section 2.3), et ce couplage émergerait potentiellement d'une interaction phylogénétique particulière entre gestes oro-faciaux et manuels (voir l'Introduction, Section 1). A quel niveau du processus langagier le geste est-il généré, et dans quelle mesure la génération d'un geste interagit avec et structure-t-elle celle de la parole, sont autant de questions, relatives à la nature exacte des interactions parole/geste (temporelle et/ou structurelle ; uni ou bidirectionnelle) auxquelles tentent de répondre les différents modèles de production conjointe de gestes et de parole (cf. Section 3).

Notre hypothèse, pour ce travail de thèse, est celle d'une intégration geste/parole forte au service du langage, considéré ici comme un système communicatif par nature multimodal et reposant donc sur l'utilisation conjointe de ces deux modalités. La génération du geste et celle de la parole devraient alors s'effectuer au même niveau de traitement, et nous soutenons en cela l'hypothèse linguistique de génération du geste (proposée notamment par Kita et Ozyürek [60])

4. Conclusion

et Hostetter et Alibali [48], inspirés des travaux de McNeill et Duncan [80]).

Plus concrètement, notre objectif est de mettre en évidence une interaction geste $\Leftrightarrow$ parole bidirectionnelle, soit une influence de la parole sur le geste (telle que celle suggérée par Kita et Ozyürek [60]) mais réciproquement, une influence similaire du geste sur la parole : la quantité d'information véhiculée par le geste dépendrait de la quantité d'information véhiculée par la parole, et vice versa. Nous souhaitons également montrer que ces deux modalités entretiennent une relation complémentaire (soit une "inhibition" des productions en situation bimodale) plutôt que redondante (soit un renforcement des productions en situation bimodale), chaque modalité étant ainsi structurée en fonction de la présence/absence de l'autre modalité, et non indépendamment l'une de l'autre.

Dans ce cadre là, notre projet s'intéresse à l'interaction geste, parole et langage dans le cadre de la deixis spatiale, que nous présenterons au chapitre suivant. Il est en effet crucial d'étudier chaque type de geste en particulier, car, comme le soulignent Krahmer et Swerts [63], « *it is conceivable that different kinds of gestures should be integrated in different ways in speech models, although this aspect of speech/gesture interaction is still largely unexplored*⁴⁸ » ([63], p.397).

Le processus déictique nous permettra donc d'étudier non seulement l'interaction entre le geste, la parole et le langage mais également l'encodage de la distance, une fonction déictique particulière, via les propriétés lexicales, phonétiques (i.e. acoustiques et articulatoires) et cinématiques du pointage.

Notre objectif est d'évaluer l'interaction entre les gestes vocaux, les gestes manuels et les structures linguistiques qui y sont associées. Le chapitre suivant nous permettra dans un premier temps de présenter le processus déictique spatial et tout l'intérêt qu'il représente pour l'étude du couplage geste/parole au sein des mécanismes langagiers.

48. Traduction proposée : Les différents types de gestes devraient être intégrés de façon spécifique aux modèles de production de parole, cet aspect de l'interaction parole/geste étant encore largement inexploré.

Annexe : Le modèle de production de parole de Levelt (1989)

Discuter, débattre, raconter... Parler est l'une de nos activités quotidiennes favorites ! Est-elle banale pour autant ? Au cours d'une conversation normale, nous sommes capables de produire deux à trois mots par seconde, sélectionnés parmi les 60 000 mots que nous connaissons. Parler relève donc d'un processus complexe et malgré notre extrême rapidité, nous commettons moins d'une erreur sur 1 000 mots produits. Levelt [71, 72], entre autres, a proposé un modèle capable de rendre compte de ce processus de production de mots : comment, en quelques secondes, un locuteur forme et articule le message souhaité. La production de la parole impliquerait trois étapes principales de traitement, illustrées Figure 1.12 : (1) la préparation conceptuelle du message, (2) sa formulation, elle-même composée de deux sous-étapes, l'encodage morphophonologique et l'encodage phonétique, et finalement (3) son articulation.

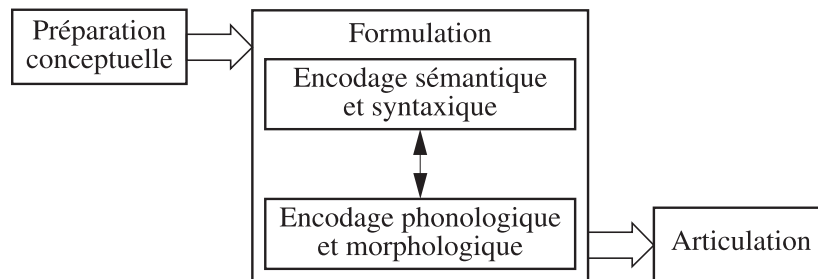


FIGURE 1.12 – Etapes de traitement impliquées dans la production de parole (Ferrand [27], p.28)

Ces différentes étapes sont prises en charge par différents modules encapsulés, nommés respectivement le Conceptualiseur, le Formulateur et l'Articulateur (voir Figure 1.13).

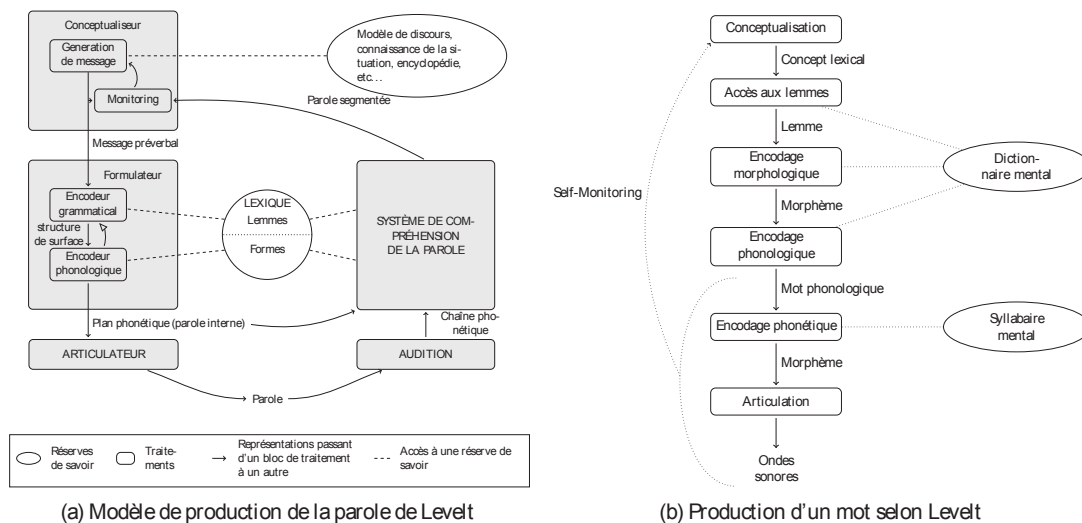


FIGURE 1.13 – Modèle de production de parole et de mot selon Levelt (Levelt [71], cité par Roustan [88], p.12)

-
- Préparation conceptuelle du message préverbal

La première étape de génération de mots consiste en la préparation conceptuelle du message via le Conceptualiseur, c'est-à-dire la préparation des intentions, idées, que le locuteur souhaite exprimer. Cette intention communicative peut être divisée en buts et sous-buts, auxquels vont correspondre différents concepts. La phase de macro-planification consiste à sélectionner le concept approprié pour chaque but et sous-but. Ces concepts devront par la suite être ordonnés (selon le principe de linéarisation) pour constituer une structure propositionnelle (phase de micro-planification) : le concept lexical. Bien que la production intentionnelle d'un mot, opposée à la récitation ou la lecture notamment, requiert toujours l'activation de son concept lexical, il n'existe pas de connexion fixe, unique, entre un référent et son concept. C'est ce que Levelt désigne sous le terme de *perspective taking* : l'activation d'un concept lexical, et la façon dont ses différents composants s'ordonnent dans un format propositionnel, sont influencées par le contexte pragmatique et sémantique de l'énoncé. Il résulte de cette étape un message non linguistique, ou préverbal, constitué d'un ou de plusieurs concepts pour lesquels nous avons des mots correspondants, définis dans notre lexique mental.

- Formulation du message

Le rôle du Formulateur est de transformer le message non linguistique, issu du Conceptualiseur, en structure linguistique, au sens de syntaxique, et contenant cette fois les mots, traduisant les concepts préalablement choisis. Cette étape de formulation fait intervenir différents processus.

- Sélection lexicale (*grammatical encoding*)

La sélection lexicale permet en premier lieu de récupérer les lemmas dans le lexique mental, étant donné un concept à exprimer. Les lemmas correspondent aux propriétés sémantiques et syntaxiques des mots. La récupération d'un lemme est rapide (environ 125 ms) et se fait par un processus d'activation-sélection : chaque concept lexical envoie de l'information au lemme qui lui est attribué, ainsi qu'aux lemmes correspondant aux mots sémantiquement proches. Le lemme approprié sera sélectionné en fonction de son degré d'activation, par un mécanisme de convergence.

- Encodage morphophonologique

Chacun d'entre nous a probablement déjà vécu l'étrange sensation d'avoir un mot particulier sur le bout de la langue (*tip-of-the-tongue*)... Nous disposons de toutes les informations concernant ce mot (genre, définition, caractéristiques, autrement dit nous en avons récupéré le lemme), mais nous restons pourtant incapables de le prononcer ! Et pour cause, il nous manque une information essentielle : une fois le lemme activé, le locuteur doit en effet récupérer les informations phonologiques qui y sont associées. Il s'agit là du lexème, ou forme sonore du mot. Trois types d'information sont nécessaires pour récupérer la forme du mot : les lexèmes sont porteurs d'une information (1) morphémique, (2) segmentale et (3) métrique.

La première correspond à la présence d’affixes et racines des mots dits ouverts⁴⁹ ; l’information segmentale renvoie quant à elle à la structure phonémique du mot (i.e. sa décomposition en phonèmes consonantiques et vocaliques)⁵⁰ ; enfin l’information métrique est relative au nombre de syllabes et au patron accentuel du mot⁵¹. La trame métrique d’un mot se combine à celle des mots adjacents dans le flux de parole par un processus incrémental de syllabation, formant ainsi les formes phonologiques des mots, ne tenant plus compte de leur frontière lexicale⁵². Les séquences segmentales sont ensuite rattachées aux séquences phonologiques correspondantes pour former une forme syllabique phonologique de l’énoncé⁵³.

– Encodage phonétique

Le locuteur peut désormais récupérer la forme syllabique phonétique (ou articulatoire) correspondant à la forme syllabique phonologique précédemment construite. Chaque syllabe phonologique est envoyée au Syllabaire, où aura lieu la sélection de la syllabe phonétique appropriée. Il s’agit d’un répertoire mental, dans lequel sont stockés les programmes moteurs les plus fréquents, associant une forme phonologique particulière à sa forme phonétique. Globalement ces programmes moteurs (ou *gestural score*) renvoient aux tâches effectuées par les différents articulateurs mis en jeu dans la production de parole (lèvres, langue, mâchoire) et sont calculés à partir de différents *tiers* (i.e. niveaux) articulatoires (oral, glottal, nasal). Toutefois, bien que le *gestural score* indique par exemple que les lèvres doivent se fermer pour produire le phonème [p], il ne spécifie pas comment les articulateurs doivent se coordonner pour exécuter cette tâche. Chaque *gestural score* possède un seuil d’activation, fonction de sa fréquence : une syllabe phonétique sera donc choisie en fonction de son degré d’activation, par un mécanisme de convergence sur un geste unique. L’ensemble des syllabes phonétiques ainsi activées forme alors une séquence articulatoire syllabique (ou plan phonétique), la parole interne.

• Articulation

L’Articulateur, en contrôlant la coordination d’un grand nombre de muscles, exécute finalement la séquence articulatoire syllabique. Autrement dit, parole interne devient parole ouverte, ou manifeste. La séquence articulatoire n’est à ce stade pas entièrement spécifiée puisqu’elle correspond à un ensemble de *gestural scores*. Elle n’indique donc pas avec précision comment articuler les différentes syllabes (ce modèle n’a pas pour but d’expliquer le processus d’articulation mais le lecteur pourra se référer notamment aux travaux de Bonnot [10] pour une revue des modèles articulatoires capables d’expliquer comment, à partir d’une représentation minimale du geste articulatoire, l’articulation exacte va émerger).

49. Mots ouverts : Classe de mots pour lesquels l’ajout de morphèmes est possible, notamment par combinaison, dérivation, néologisme, emprunt lexical ou déclinaison.

50. Le mot *ami* est composé de trois segments phonologiques, VCV : [a] [m] [i].

51. Sachant que σ représente une syllabe, l’information métrique du mot *ami* est $\sigma.\sigma$.

52. *mon + ami* devient *mon.na.mi*, où . ne représente plus la frontière syllabique lexicale mais phonologique.

53. *mon ami + $\sigma.\sigma.\sigma$* = [mõ.na.mi].

-
- Self-monitoring et Traitement sériel

« *The person to whom we listen most is ourself*⁵⁴ » Levelt ([72], p.6)

Selon ce modèle, nous sommes capables de contrôler, et de corriger si nécessaire, nos output (i.e. parole interne et parole manifeste).

Exemple : « J’aime beaucoup cette ju... robe. »

→ Erreur contrôlée sur la parole interne, probablement détectée avant l’onset de l’articulation.

Exemple : « J’aime beaucoup cette jupe... robe. »

→ Erreur contrôlée sur la parole manifeste.

Selon Ferrand [27], le modèle proposé par Levelt [71, 72] est strictement sériel : la production des mots implique une série d’étapes de traitement, correspondant à un niveau de représentation spécifique (unités conceptuelles, morphophonologiques, phonétiques) et le traitement d’un type particulier d’information doit être effectué avant que ne commence le traitement de l’information associée au niveau suivant.

Il semble effectivement que le traitement soit automatique et que chaque module travaille indépendamment des autres (comme l’illustre l’effet *tip-of-the-tongue*). Toutefois, notre capacité de self-monitoring suggère qu’un traitement parallèle est envisageable : l’encodage grammatical peut débuter dès lors qu’un concept est disponible, suivi par l’activation du mot approprié et le début d’une construction de phrase, envoyé à l’encodage phonologique et phonétique. Le locuteur peut donc articuler les syllabes initiales, alors que la suite de l’énoncé est toujours en cours de traitement.

54. Traduction proposée : La personne que nous écoutons le plus n’est autre que nous-même.

Bibliographie

- [1] Alibali, M., D. Heath et H. Myers. 2001, «Effects of visibility between speaker and listener on gesture production : Some gestures are meant to be seen», *Journal of Memory and Language*, vol. 44, p. 169–188.
- [2] Alibali, M., S. Kita et A. Young. 2000, «Gesture and the process of speech production : We think, therefore we gesture», *Language and Cognitive Processes*, vol. 15, p. 593–613.
- [3] Attina, V. 2005, *La Langue française Parlée Complétée : Production et perception*, thèse de doctorat, Institut National Polytechnique de Grenoble.
- [4] Barakat, R. 1973, «Arabic gestures», *Journal of Popular Culture*, vol. 6, p. 749–792.
- [5] Barbieri, F., A. Buonocore, R. Dalla Volta et M. Gentilucci. 2009, «How symbolic gestures and words interact with each other», *Brain and Language*, vol. 110, p. 1–11.
- [6] Barrett, A., L. Dore, K. Hansell et K. Heilman. 2002, «Speaking while gesturing : The relationship between speech and limb praxis», *Neurology*, vol. 58(3), p. 499–500.
- [7] Bavelas, J. 1994, «Gestures as part of speech : Methodological implications», *Research on Language & Social Interaction*, vol. 27(3), p. 201–221.
- [8] Beattie, G. et J. Coughlan. 1998, «An experimental investigation of the role of iconic gestures in lexical access using the tip-of-the-tongue phenomenon», *British Journal of Psychology*, vol. 90, p. 35–56.
- [9] Bernardis, P. et M. Gentilucci. 2006, «Speech and gesture share the same communication system», *Neuropsychologia*, vol. 44, n° 2, p. 178–190.
- [10] Bonnot, J. 1990, *Production de la parole et coarticulation : Une analyse critique des principaux modèles*, vol. 20, Strasbourg : Travaux de l'Institut de Phonétique de Strasbourg.
- [11] Buccino, G., F. Binkofski, G. Fink, L. Fadiga, L. Fogassi, V. Gallese, R. Seitz, K. Zilles, G. Rizzolatti et H. Freund. 2001, «Action observation activates premotor and parietal areas in a somatotopic manner : An fMRI study», *European Journal of Neuroscience*, vol. 13, p. 400–404.
- [12] Bull, P. et G. Connelly. 1985, «Body movement and emphasis in speech», *Journal of Nonverbal Behavior*, vol. 9, p. 169–187.
- [13] Butterworth, B. et G. Beattie. 1978, «Gesture and silence as indicators of planning in speech», dans *Recent advances in the psychology of language : Formal and experimental approaches*, édité par R. Campbell et E. Smith, 4, London : Plenum, p. 347–360.
- [14] Butterworth, G. et U. Hadar. 1989, «Gesture, speech, and computational stages : A reply to McNeill», *Psychological Review*, vol. 96, p. 168–174.
- [15] Campos, J. J., D. Anderson, M. Barbu-Roth, E. Hubbard, M. Hertenstein et D. Witherington. 2000, «Travel broadens the mind», *Infancy*, vol. 1, p. 149–219.

Bibliographie

- [16] Carter, R., S. Aldrige, M. Page et S. Parker. 2010, *Le grand Larousse du cerveau*, Paris : Larousse.
- [17] Chawla, P. et R. Krauss. 1994, «Gesture and speech in spontaneous and rehearsed narratives», *Journal of Experimental Social Psychology*, vol. 30, p. 580–601.
- [18] Chui, K. 2005, «Temporal patterning of speech and iconic gestures in conversational discourse», *Journal of Pragmatics*, vol. 37, p. 871–887.
- [19] Cocks, N., L. Dipper, R. Middleton et G. Morgan. 2011, «What can iconic gestures tell us about the language system? A case of conduction aphasia», *International Journal of Language & Communication Disorders*, vol. 46(4), p. 423–436.
- [20] Cosnier, J. et J. Vaysse. 1997, «Sémiotique des gestes communicatifs», *Nouveaux Actes Sémiotiques*, vol. 52, p. 7–28.
- [21] Decety, J., J. Grezes, N. Costes, D. Perani, M. Jeannerod, E. Procyk, F. Grassi et F. Fazio. 1997, «Brain activity during observation of actions influence of action content and subject's strategy», *Brain*, vol. 120(10), p. 1763–1777.
- [22] Dillisenger, J. et E. Moershel. 2009, *Guide des technologies de l'imagerie médicale et de la radiothérapie : Quand la théorie éclaire la pratique*, Masson.
- [23] Driskell, J. et P. Radtke. 2003, «The effect of gesture on speech production and comprehension», *Human Factors : The Journal of the Human Factors and Ergonomics Society*, vol. 45(3), p. 445–454.
- [24] Efron, D. 1941, *Gesture and environment*, Morningside Heights, NY : King's Crown Press.
- [25] Ekman, P. et W. Friesen. 1969, «The repertoire of nonverbal behavioral categories», *Sémiotica*, vol. 1, p. 49–98.
- [26] Ellis, R. et M. Tucker. 2000, «Micro-affordance : The potentiation of components of action by seen objects», *British Journal of Psychology*, vol. 91, p. 451–471.
- [27] Ferrand, L. 2002, «Les modèles de la production de la parole», dans *Production du langage. Traité des Sciences Cognitives*, édité par M. Fayol, Paris : Hermès, p. 27–44.
- [28] Feyereisen, P. 1997, «The competition between gesture and speech production in dual-task paradigms», *Journal of Memory and Language*, vol. 36, p. 13–33.
- [29] Fodor, J. 1983, *The modularity of mind : An essay on faculty psychology*, Cambridge MA : MIT Press.
- [30] Freedman, N. et S. Hoffman. 1967, «Kinetic behaviour in altered clinical states», *Perceptual and Motor Skills*, vol. 24, p. 527–539.
- [31] Frick-Horbury, D. et R. Guttentag. 1998, «The effects of restricting hand gesture production on lexical retrieval and free recall», *American Journal of Psychology*, vol. 111, p. 43–63.

- [32] Furuyama, N., D. McNeill et M. Park-Doob. 2002, «Is speech-gesture production ballistic or interactive?», dans *The congress of the International Society for Gesture Studies*.
- [33] Gallagher, H. et C. Frith. 2004, «Dissociable neural pathways for the perception and recognition of expressive and instrumental gestures», *Neuropsychologia*, vol. 42, p. 1725–1736.
- [34] Gerardin, E., A. Sirigu, S. Lehericy, J. Poline, B. Gaymard, C. Marsault, Y. Agid et D. Le Bihan. 2000, «Partially overlapping neural networks for real and imagined hand movements», *Cerebral Cortex*, vol. 10, p. 1093–1104.
- [35] Gerlach, I., C. Law et O. Paulson. 2002, «When action turns into words : Activation of motor-based knowledge during categorization of manipulable objects», *Journal of Cognitive Neuroscience*, vol. 14, p. 1230–1239.
- [36] Gibson, J. 1979, *The ecological approach to visual perception*, Hillsdale, NJ : Erlbaum.
- [37] Glenberg, A. et M. Kaschak. 2002, «Grounding language in action», *Psychonomic Bulletin & Review*, vol. 9, p. 558–565.
- [38] Graham, J. et M. Argyle. 1975, «A cross-cultural study of the communication of extra-verbal meaning by gestures», *International Journal of Psychology*, vol. 10, p. 57–67.
- [39] Grèzes, J., N. Costes et J. Decety. 1998, «Top-down effect of strategy on the perception of human biological motion : A PET investigation», *Cognitive Neuropsychology*, vol. 15, p. 553–582.
- [40] Hadar, U., A. Burstein, R. Krauss et N. Soroker. 1998, «Ideational gestures and speech : A neurolinguistic investigation», *Language and Cognitive Processes*, vol. 13, p. 56–76.
- [41] Hadar, U. et G. Butterworth. 1997, «Iconic gestures, imagery and word retrieval in speech», *Semiotica*, vol. 115, p. 147–172.
- [42] Hadar, U., D. Wenkert-Olenik, R. Krauss et N. Soroker. 1998, «Gesture and the processing of speech : Neuropsychological evidence», *Brain and Language*, vol. 62(1), p. 107–126.
- [43] Hauk, O., I. Johnsrude et F. Pulvermüller. 2004, «Somatotopic representation of action words in human motor and premotor cortex», *Neuron*, vol. 41, p. 301–307.
- [44] Heath, M., E. Roy, S. Black et D. Westwood. 2001, «Intransitive limb gestures and apraxia following unilateral stroke», *Journal of Clinical and Experimental Neuropsychology*, vol. 23(5), p. 628–642.
- [45] Hickok, G., U. Bellugi et E. Klima. 1998, «The neural organization of language : Evidence from sign language aphasia.», *Trends in Cognitive Sciences*, vol. 2, p. 129–136.
- [46] Holender, D. 1980, «Interference between a vocal and a manual response to the same stimulus», dans *Tutorials in motor behavior*, édité par G. Stelmach et J. Requin, Amsterdam : North-Holland, p. 421–431.

- [47] Hostetter, A. et M. Alibali. 2004, «On the tip of the mind : Gesture as a key to conceptualization», dans *The 26th Annual Conference of the Cognitive Science Society*, édité par K. Forbus, D. Gentner et T. Regier, Chicago, IL : Cognitive Science Society, p. 589–594.
- [48] Hostetter, A. et M. Alibali. 2008, «Visible embodiment : Gestures as Simulated Action», *Psychonomic Bulletin & Review*, vol. 15, n° 3, p. 495–514.
- [49] Iverson, J. et S. Goldin-Meadow. 1998, «Why people gesture when they speak», *Nature*, vol. 396, p. 228.
- [50] Jeannerod, M. 2001, «Neural simulation of action : A unifying mechanism for motor cognition», *NeuroImage*, vol. 14, p. S103–S109.
- [51] Kendon, A. 1972, «Some relationships between body motion and speech», dans *Studies in Dyadic Communication*, édité par A. Seigman et B. Pope, Pergamon Press, Elmsford, NY, p. 177–216.
- [52] Kendon, A. 1980, «Gesticulation and speech : Two aspects of the process of utterance», dans *The relationship of verbal and nonverbal communication*, édité par M. Key, Walter de Gruyter, p. 207–227.
- [53] Kendon, A. 1983, «Gesture and speech : How they interact», *Nonverbal Interaction*, p. 13–45.
- [54] Kendon, A. 1985, «Some uses of gesture», dans *Perspectives on Silence*, édité par D. Tannen et U. Saville-Troike, Norwood, N J : Ablex Publishing, p. 215–234.
- [55] Kendon, A. 1988, «How gestures can become like words», dans *Crosscultural Perspectives in Nonverbal Communication*, édité par F. Poyatos, Toronto : C. J. Hogrefe, p. 131–141.
- [56] Kendon, A. 2004, *Gesture : Visible Action as Utterance*, Cambridge University Press, Cambridge.
- [57] Kita, S. 1993, *Language and thought interface : A study of spontaneous gestures and Japanese mimetics*, thèse de doctorat, University of Chicago.
- [58] Kita, S. 2000, «How representational gestures help speaking», dans *Language and Gesture : Window into Thought and Action*, édité par D. McNeill, Cambridge Univ. Press, p. 162–185.
- [59] Kita, S. 2003, *Pointing : Where Language, Culture, and Cognition Meet*, Mahwah, NJ, US : Lawrence Erlbaum Associates Publishers.
- [60] Kita, S. et A. Ozyürek. 2003, «What does cross-linguistic variation in semantic coordination of speech and gesture reveal? : Evidence for an interface representation of spatial thinking and speaking», *Journal of Memory and Language*, vol. 48, n° 1, p. 16–32.
- [61] Kita, S., A. Özyürek, S. Allen, A. Brown, R. Furman et T. Ishizuka. 2007, «Relations between syntactic encoding and co-speech gestures : Implications for a model of speech and gesture production», *Language and Cognitive Processes*, vol. 22, n° 8, p. 1212–1236.

- [62] Kita, S., I. Van Gijn et H. Van der Hulst. 1998, «Movement phases in signs and co-speech gestures, and their transcription by human coders», dans *Gesture and Sign Language in Human-Computer Interaction : Lecture Notes in Artificial Intelligence*, édité par I. Waschmuth et M. Fröhlich, Berlin : Springer-Verlag, p. 23–35.
- [63] Krahmer, E. et M. Swerts. 2007, «The effects of visual beats on prosodic prominence : Acoustic analyses, auditory perception and visual perception», *Journal of Memory and Language*, vol. 57, n° 3, p. 396–414.
- [64] Krauss, R. 1998, «Why do we gesture when we speak?», *Current Directions in Psychological Science*, vol. 7(2), p. 54–60.
- [65] Krauss, R., Y. Chen et P. Chawla. 1996, «Nonverbal behavior and nonverbal communication : What do conversational hand gestures tell us?», *Advances in Experimental Social Psychology*, vol. 28(3), p. 89–450.
- [66] Krauss, R., Y. Chen et R. Gottesman. 2000, «Lexical gestures and lexical access : A process model», dans *Language and Gesture*, Cambridge University Press, New York, p. 261–283.
- [67] Krauss, R. et U. Hadar. 1999, «The role of speech-related arm/hand gestures in word retrieval», dans *Gesture, speech, and sign*, édité par R. Campbell et L. Messing, Oxford : Oxford University Press, p. 93–116.
- [68] Krauss, R. et E. Morsella. 2002, «Movement facilitates speech production : A gestural feedback model», Columbia University.
- [69] Kuhtz-Buschbeck, J., C. Mahnkopf, C. Holzknicht, H. Siebner, S. Ulmer et O. Jansen. 2003, «Effector-independent representations of simple and complex imagined finger movements : A combined fMRI and TMS study», *European Journal of Neuroscience*, vol. 18, p. 3375–3387.
- [70] Laurent, D. 2007, *Pointage et démonstratifs : Exploration de la grammaire par la coordination geste-parole*, mémoire de maîtrise, Université Stendhal.
- [71] Levelt, W. 1989, *Speaking : From Intention to Articulation*, MIT Press, Cambridge, MA.
- [72] Levelt, W. 1999, «Models of word production», *Trends in Cognitive Science*, vol. 3, n° 6, p. 223–232.
- [73] Levelt, W., G. Richardson et W. L. Heij. 1985, «Pointing and voicing in deictic expressions», *Journal of Memory and Language*, vol. 24, p. 133–164.
- [74] Løevenbruck, H., M. Baciú, C. Segebarth et C. Abry. 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*, vol. 18, n° 3, p. 237–258.
- [75] Løevenbruck, H., C. Vilain, F. Carota, M. Baciú, C. Abry, L. Lamalle, C. Pichat et C. Segebarth. 2007, «Cerebral correlates of multimodal pointing : An fMRI study of prosodic focus, syntactic extraction, digital- and ocular-pointing», dans *The 16th International Congress of Phonetic Sciences*, p. 1861–1864.

Bibliographie

- [76] Løevenbruck, H., C. Vilain, F. Carota, M. Baciù, C. Abry, L. Lamalle, C. Pichat et C. Segebarth. 2008, «Left parietal activation during the production of pointing in several modalities : Prosodic focus, syntactic extraction, digital- and ocular-pointing», dans *155th Meeting of The Acoustical Society of America (Acoustics 08)*, vol. 123(5), p. 3320.
- [77] McNeill, D. 1985, «So you think gestures are nonverbal», *Psychological Review*, vol. 92(3), p. 350–371.
- [78] McNeill, D. 1992, *Hand and Mind : What Gestures Reveal about Thought*, University of Chicago Press.
- [79] McNeill, D. 2000, «Catchments and contexts : Non-modular factors in speech and gesture production», dans *Language and Gesture*, édité par D. McNeill, Cambridge University Press, p. 313–328.
- [80] McNeill, D. et S. Duncan. 2000, «Growth points in thinking-for-speaking», dans *Language and Gesture*, Cambridge University Press, p. 141–161.
- [81] McNeill, D., E. Levy et L. Pedelty. 1990, «Speech and gesture», dans *Advances in psychology : Cerebral control of speech and limb movements*, édité par G. Hammond, Amsterdam : Elsevier/North Holland Publishers, p. 203–256.
- [82] Morrel-Samuels, P. et R. Krauss. 1992, «Word familiarity predicts temporal asynchrony of hand gestures and speech», *Journal of Experimental Psychology : Learning, Memory, and Cognition*, vol. 18 (3), p. 615–622.
- [83] Ozyürek, A. 2002, «Do speakers design their co-speech gestures for their addressees ? The effects of addressee location on representational gestures», *Journal of Memory and Language*, vol. 46(4), p. 688–704.
- [84] Ozyürek, A., S. Kita, S. Allen, R. Furman et A. Brown. 2005, «How does linguistic framing of events influence co-speech gestures ? Insights from crosslinguistic variations and similarities», *Gesture*, vol. 5 (1/2), p. 219–240.
- [85] Rauscher, F., R. Krauss et Y. Chen. 1996, «Gesture, speech and lexical access : The role of lexical movements in speech production», *Psychological Science*, vol. 7 (4), p. 226–231.
- [86] Reboul, A. et J. Moeschler. 1998, *Pragmatique du discours. De l'interprétation de l'énoncé à l'interprétation du discours*, Paris : Armand Colin.
- [87] Rochet-Capellan, A. 2007, *De la substance à la forme : Rôle des contraintes motrices orofaciales et brachiomanuelles de la parole dans l'émergence du langage*, thèse de doctorat, Institut National Polytechnique de Grenoble.
- [88] Roustan, B. 2012, *Étude de la coordination gestes manuels/parole dans le cadre de la désignation*, thèse de doctorat, Université de Grenoble.
- [89] de Ruiter, J. 1998, *Gesture and speech production*, thèse de doctorat, Universiteit Nijmegen.

- [90] de Ruiter, J. 2006, «Can gesticulation help aphasic people speak, or rather, communicate?», *Interactional Journal of Speech-Language Pathology*, vol. 8(2), p. 124–127.
- [91] de Ruiter, J. 2007, «Some multimodal signals in humans», dans *Proceedings of the Workshop on Multimodal Output Generation*, édité par I. In Van der Sluis, M. Theune et E. E. Reiter, E. & Krahmer, p. 141–148.
- [92] de Ruiter, J. et D. Wilkins. 1998, «The synchronization of gesture and speech in Dutch and Arrernte (an australian aboriginal language) : A cross-cultural comparison», dans *Oralité et Gestualité*, édité par S. Santi, I. Guaitella, C. Cavé et G. Konopczynski, Paris : L'Harmattan, p. 603–607.
- [93] Schegloff, E. 1984, «On some gestures' relation to talk», dans *Structures of social action*, édité par J. Atkinson et J. Heritage, Cambridge : Cambridge University Press, p. 266–298.
- [94] Skipper, J., S. Goldin-Meadow, H. Nusbaum et S. Small. 2007, «Speech-associated gestures, Broca's area, and the human mirror system», *Brain and Language*, vol. 101(3), p. 260–277.
- [95] Slobin, D. 1996, «From "thought and language" to "thinking for speaking"», dans *Rethinking linguistic relativity*, vol. 17, édité par J. Gumperz et S. Levinson, Cambridge : Cambridge University Press, p. 70–96.
- [96] Tanaka, S. et T. Inui. 2002, «Cortical involvement for action imitation of hand/arm postures versus finger configurations : An fMRI study», *Neuroreport*, vol. 13(13), p. 1599–1602.
- [97] Treisman, A. 1960, «Contextual cues in selective listening», *Quarterly Journal of Experimental Psychology*, vol. 12, p. 242–248.
- [98] Vygotsky, L. 1998, *The collected works of L.S. Vygotsky*, New York, Plenum.
- [99] Willems, R. et P. Hagoort. 2007, «Neural evidence for the interplay between language, gesture, and action : A review», *Brain and Language*, vol. 101, p. 278–289.
- [100] Wilson, M. 2002, «Six views of embodied cognition», *Psychonomic Bulletin & Review*, vol. 9, p. 625–636.

La deixis spatiale, un rendez-vous entre geste, parole et langage

« *Deixis is the domain par excellence where language and reality meet.*¹ »

Weissenborn & Klein ([64], p.3)

Notre étude porte sur la deixis spatiale. Nous avons choisi de travailler spécifiquement sur ce processus communicatif, qui, comme nous le verrons, tient une place toute particulière dans la communication langagière humaine, étant universel et omniprésent dans les interactions quotidiennes. Il apparaît alors comme le cadre idéal pour l'étude du couplage geste/parole au sein des mécanismes langagiers.

La deixis, du grec ancien *deiknunai* - *montrer*, est un processus communicatif de désignation, monstration et pointage. Selon la définition du Larousse, le terme « déictique » renvoie à *tout élément linguistique dont la fonction consiste à articuler l'énoncé sur la situation particulière dans laquelle il est produit ou à l'inscrire dans un discours* ; en d'autres termes, l'utilisation de déictiques nous permet d'indiquer ce dont on parle, de partager de l'information avec autrui en attirant l'attention de notre interlocuteur sur un objet d'intérêt. Cet objet d'intérêt, ou référent, est l'élément de la réalité auquel renvoie le déictique et ne peut être défini qu'en relation avec les interlocuteurs impliqués dans la situation d'énonciation ; il peut s'agir aussi bien d'une personne, d'un objet ou d'un événement, et parfois même d'une entité abstraite, voire invisible.

La particularité de ce processus réside en sa **multimodalité** : on peut montrer du doigt mais également de la voix, et nous combinons d'ailleurs bien souvent ces deux modalités (Kendon [41]), comme l'illustre l'exemple ci-dessous.

Exemple : « *La **voici** !* », *s'écria Jeoffrey, le **doigt** tendu.*

Nous commencerons donc par définir précisément ce que l'on entend par geste et mot déictiques, avant d'aborder la notion d'encodage spatial, une fonction déictique particulière.

1. Traduction proposée : La deixis est le domaine par excellence où langage et réalité s'entremêlent.

1 Le geste déictique : Nature et développement

« *Comment nommer les êtres et les choses, sinon par les gestes que nous faisons vers eux ou par les mouvements qui les conduisent à nous ?* » *Pages d'écriture*, Jean Tardieu

Défini par Kita [42], le pointage canonique (au sens de naturel, instinctif) est un mouvement **communicatif** effectuant une courbe partant du corps vers l'extérieur et indiquant à notre interlocuteur l'emplacement souhaité. Le pointage brachio-manuel par exemple consiste à tendre l'index et le bras en direction du référent, tout en gardant les autres doigts repliés sous la main, le pouce étant orienté vers le bas et le côté. Le pointage ne se restreint toutefois pas au système main-bras : du pointage oculaire à celui du menton ou des sourcils, en passant par le pointage des lèvres (Wilkins [65]) et même celui du nez (Cooperrider et Nunez [13]) dans certaines cultures, nous disposons d'un grand nombre d'effecteurs, et ce dès le plus jeune âge. Le pointage est ainsi l'un des premiers outils utilisés pour désigner le monde qui nous entoure (Kita [42]) : les premiers pointages oculaires apparaissent dès l'âge de 8 mois, suivis de près par les pointages brachio-manuels aux alentours de 9-11 mois ; à l'âge de 12 mois, 60% des gestes produits par les enfants sont des gestes de pointage.

Mais le pointage manuel est-il un outil communicatif par nature ou rien de plus qu'un mouvement de saisie abrégé, comme le suggérait Wundt [67] il y a un siècle déjà ? Cette hypothèse, qui suppose que le pointage de l'index dériverait du mouvement d'atteinte et de saisie et en serait une réalisation économique et référentielle, est aujourd'hui encore répandue. Vygotsky [63] supposait ainsi que le pointage se développerait en dehors de l'interaction, lorsque l'enfant tente vainement de saisir un objet ; c'est la succession d'échecs de saisie qui conduirait l'enfant à pointer l'objet désiré, et les succès occasionnels, c'est-à-dire lorsque l'objet est apporté à l'enfant, l'amènerait à prendre conscience de l'utilité communicative de ce type de geste.

Masataka [47] propose en revanche une toute autre théorie : le pointage de l'index émergerait non pas de tentatives d'atteinte mais d'extensions de l'index. Une étude longitudinale, réalisée auprès de huit enfants japonais, âgés de 3 à 16 mois, montre que (1) les productions de vocalisations et d'extensions de l'index sont fortement corrélées, et ce dès l'âge de 3 mois ; (2) le nombre d'extensions de l'index croît au fur et à mesure du développement pour atteindre un pic à l'âge de 11-12 mois, période des premiers pointages ; (3) lorsque les premiers pointages sont en place, la fréquence d'extensions de l'index chute brusquement, alors que celle des atteintes évolue à peine au cours de la période considérée (voir Figure 2.1).

De fait, l'extension de l'index serait liée non pas au désir de saisir un objet mais plutôt à la nécessité d'explorer et de partager son environnement. Il existerait donc une continuité développementale entre l'extension de l'index d'une part, et le pointage de l'index d'autre part : lorsqu'un élément capture l'attention de l'enfant et que celui-ci souhaite le partager avec autrui, l'extension de l'index s'accompagne d'une extension intentionnelle du bras pour devenir pointage.

1. Le geste déictique : Nature et développement

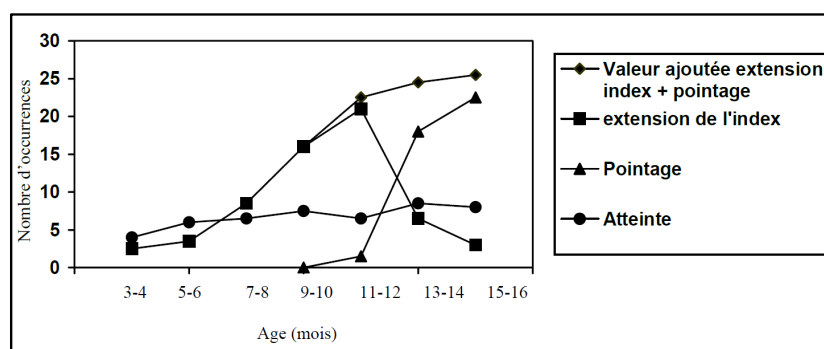


FIGURE 2.1 – Extensions de l'index (■), pointages (▲) et mouvements d'atteinte (●). Les occurrences additionnées d'extensions et de pointages de l'index (◆) ont été calculées par Ducey-Kaufmann afin de « *montrer que les gestes du pointage prennent le relais en continuité des gestes d'extension de l'index, lorsque ceux-ci sont peu à peu remplacés par le pointer* » ([23], p.153)

Certaines différences notables entre pointage et mouvement d'atteinte et de saisie vont également à l'encontre de la théorie vygotskienne. Le geste d'atteinte est par exemple destiné quasiment exclusivement aux objets proches, ce qui n'est pas le cas du geste de pointage² (Franco et Butterworth [30], cités par Butterworth et Grover [9]). Simon et collab. [55] observent également des activations cérébrales spécifiques au type de geste produit (atteinte/saisie vs pointage), principalement au niveau du lobe pariétal. Dans le même registre, Meunier et collab. [48] montrent que l'utilisation de la main droite est fonction du type de geste produit (voir Figure 2.2) : l'enfant âgé de 14 à 20 mois utilisera préférentiellement sa main droite pour pointer un objet, que celui-ci soit placé sur sa droite ou sur sa gauche, mais utilisera ses deux mains de manière relativement indifférenciée pour saisir un objet (la main droite pour saisir un objet placé sur sa droite, la gauche pour saisir un objet placé sur sa gauche).

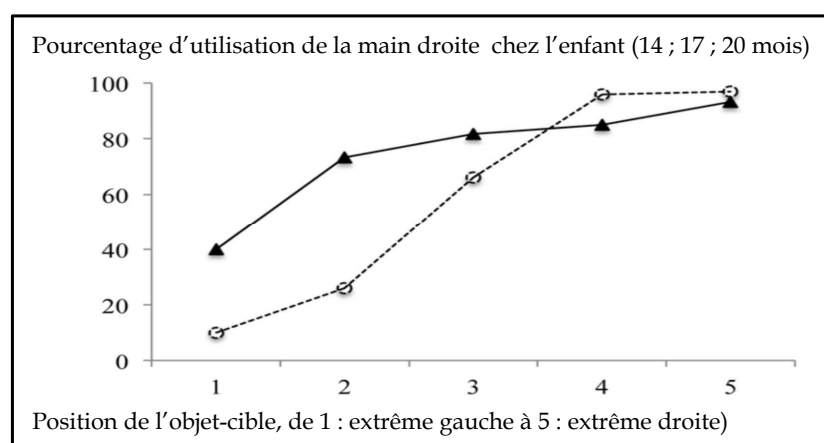


FIGURE 2.2 – Pourcentage d'utilisation de la main droite chez l'enfant âgé de 14, 17 et 20 mois, en fonction du type de geste (*pointage* ▲ ; *saisie* ○) et de la position de l'objet par rapport à l'individu (de *gauche* à *droite*) (Meunier et collab. [48], p.2)

2. Le pointage vers les objets proches apparaît environ deux mois avant le pointage vers les objets plus éloignés (Carpenter et collab. [11]).

Le geste déictique n'apparaît plus comme un acte manqué de saisie mais bien comme un acte référentiel et communicatif (Butterworth et Grover [9], Fogel et Hannan [28], Fogel et Thelen [29]); Franco et Butterworth [31] suggèrent d'ailleurs une opposition entre l'atteinte, utilisée principalement dans un contexte impératif/instrumental, et le pointage, utilisé principalement dans un contexte déclaratif/référentiel. Le geste de pointage apparaît ainsi comme un geste complexe à part entière, qui implique non seulement une coordination précise entre l'extension bras-index et l'objet d'intérêt, mais également la prise en compte d'un interlocuteur, supposé localiser et diriger son attention vers celui-ci.

Par ailleurs, le rôle du pointage dans l'émergence de la communication intentionnelle et l'établissement d'un cadre d'attention conjointe (Butterworth [7], Diessel [18]) lui confère un statut particulier. La communication intentionnelle est le fruit de trois étapes successives (Tomasello et collab. [56]) : **l'engagement dyadique** tout d'abord, aux alentours de 3 mois, correspond à la capacité de partager les comportements et émotions de l'autre. Puis aux alentours de 9-12 mois, le cadre d'attention conjointe à proprement parler peut se mettre en place : l'interaction est alors conçue comme une triade, constituée d'un locuteur, d'un interlocuteur et d'un objet de référence ; les individus partagent alors un même centre d'intérêt, concentrent leur attention et coordonnent leur regard sur un même objet. C'est **l'engagement conjoint passif**, avec partage des buts et des perceptions, qui aboutit à l'âge de 12-15 mois à **l'engagement conjoint collaboratif** ou coordonné (avec partage des intentions et attentions, via des activités collaboratives). Tomasello et collab. [56] estiment que cette faculté d'entrer dans une relation triadique fait partie du jeu des compétences sociales fondamentales, nécessaires au développement du langage. Cette compétence émerge alors même qu'apparaissent les premiers pointages déclaratifs, suivis de près par les premiers mots.

Alors, le pointage peut-il être vu comme un tremplin vers le langage, comme la *voie royale* menant de la communication pré-verbale au langage parlé (Butterworth [8]) ? C'est en tout cas ce que suggèrent nombre d'études (e.g. Bates et Dick [1], Capirci et collab. [10], Goldin-Meadow [32], Volterra et Caselli [61]), qui mettent en évidence des parallèles étonnants entre l'apparition et la production du pointage et certains éléments clés dans le processus d'acquisition du langage, tels que l'émergence de la morphosyntaxe (Capirci et collab. [10], Goldin-Meadow et Butcher [33], Morford et Goldin-Meadow [49], Volterra et collab. [62]), la construction du lexique (Bruner [6]) ou encore l'apprentissage de vocabulaire (Kalagher et Yu [38])³. Il semble que la façon dont les enfants utilisent le pointage puisse nous renseigner sur leur développement langagier futur.

3. Le Chapitre 4, consacré à l'étude des productions déictiques chez l'enfant, nous permettra de voir certains de ces éléments plus en détails.

2 L'expression déictique : On montre de la voix !

« *Le premier métier [des mots], c'est de désigner les choses.* »

La grammaire est une chanson douce, Erik Orsenna

Le processus déictique passe donc par l'utilisation de la modalité gestuelle mais également par celle de la modalité vocale. Nous n'aborderons pas l'utilisation de variations prosodiques (domaine de l'intonation, du rythme et du phrasé) dans la désignation, et notamment celle du focus, qui consiste à accentuer un ou plusieurs éléments du discours afin de le(s) mettre en évidence⁴, mais nous nous intéresserons ici exclusivement à l'emploi de termes déictiques, qui par bien des aspects peuvent être considérés comme l'équivalent verbal direct du pointage manuel.

De toute première importance dans la communication langagière humaine, les termes déictiques font partie des premiers mots de l'enfant mais également des plus fréquents (Diessel [19]). Ils sont par ailleurs universels (Diessel [15, 16, 18], Dixon [21], Himmelmann [37]), à la différence d'autres classes grammaticales, telles que les articles définis ou encore les auxiliaires.

D'autre part, les termes déictiques sont bien souvent accompagnés d'un geste de pointage (Diessel [18], Enfield [24]). Cette combinaison parole+geste est typique de la deixis dite **exophorique**. On parle de deixis exophorique lorsque la fonction du déictique est d'attirer l'attention des interlocuteurs sur les entités concrètes qui les entourent (Levinson [45]) ; les termes déictiques ont alors exactement la même fonction que le pointage manuel : tous deux « montrent » l'emplacement de l'objet d'intérêt, relativement au centre déictique⁵ et tous deux permettent en cela l'établissement d'un cadre d'attention conjointe. Pointage vocal et pointage manuel sont donc ici intimement liés, à tel point que selon certains auteurs, l'utilisation canonique d'un démonstratif implique nécessairement la production conjointe d'un pointage manuel⁶ (Bühler [3], Levinson [45], Lyons [46]).

On distingue classiquement cette fonction exophorique, principale, de la fonction **endophorique** ; celle-ci apparaît plus tard dans l'acquisition du langage et est divisée en deux sous-classes : la deixis anaphorique et la deixis discursive. La première renvoie aux référents déjà cités au moment de l'énonciation et permet ainsi de garder une trace des éléments précédemment évoqués ; la seconde renvoie aux propositions du discours, aux éléments linguistiques qui le constituent, et permet d'établir des liens entre les différents composants (ou *chunks*) de l'énonciation en cours (Fillmore [27], Himmelmann [37]). Mais indépendamment de leur fonction (1) exophorique ou (2) endophorique, les déictiques permettent bien d'établir un cadre d'attention conjointe, puisqu'ils vont diriger l'attention de l'interlocuteur sur (1) les entités concrètes du monde physique, ou (2) les éléments linguistiques de l'énonciation.

4. Le lecteur intéressé pourra se référer aux travaux de Dohen et Roustau, tels que Dohen [22], Roustau [51].

5. De manière générale, l'utilisation des déictiques est relative au centre déictique (*deictic center* ou *origo*), qui fixe un point de repère, généralement défini par rapport au locuteur, permettant ainsi d'interpréter la référence de manière adéquate (Bühler [3], Diessel [19]).

6. Dans certaines langues, telles que le Kilivila, une langue de la pointe papoue, ou le Goemai, une langue du Nigeria, l'utilisation d'un démonstratif vers une entité concrète du monde physique est *obligatoirement* accompagnée d'un pointage manuel (Hellwig [36], Senft [54]).

2.1 Catégories déictiques

Les principaux types déictiques grammaticalisés, qui existent dans tous les systèmes linguistiques (e.g. Fillmore [27], Levinson [44]), concernent la personne, le temps et l'espace.

Deixis de personne : Les déictiques de personne permettent de désigner les individus directement impliqués (locuteur, destinataire), ou simplement mentionnés, dans la situation d'énonciation ; il s'agit de pronoms personnels (*je, tu, il*, etc.), de pronoms et adjectifs possessifs (*son, le sien, mon*, etc.), de prénoms et noms propres, ainsi que d'adjectifs démonstratifs (*ce, cette*, etc.), qui véhiculent notamment des informations de genre, de nombre et d'inclusivité.

Deixis de temps : Les déictiques de temps renvoient au moment de l'énonciation ; les adverbes et locutions adverbiales de temps (*aujourd'hui, demain, hier, maintenant, plus tard*, etc.) ainsi que l'utilisation du passé, présent, futur, permettent ainsi de renseigner sur les coordonnées temporelles de la situation d'énonciation. Fillmore [27] distingue l'*encoding time*, qui correspond au moment où l'énonciation est produite, du *decoding time*, celui où elle est perçue. Bien que souvent futile (la production et la réception d'un message oral étant la plupart du temps quasi simultanées), cette distinction devient pertinente lorsqu'il s'agit de correspondance ou de rediffusion.

Deixis spatiale : Les déictiques spatiaux font référence à l'emplacement des intervenants et du référent ; les adverbes et locutions adverbiales de lieu (*ici, là-bas, devant, dessus, voici, loin*, etc.) ainsi que les pronoms démonstratifs (*ceci, cela*, etc.) renseignent ici sur les coordonnées spatiales de la situation d'énonciation.

3 Langage et représentations spatiales

3.1 Représentations spatiales perceptives

Les compétences spatiales, telles que la reconnaissance et la recherche d'objet, la navigation et l'orientation dans l'espace, sont existentielles dans le règne animal en général et chez l'homme en particulier. Notre capacité de percevoir les objets, leurs emplacements et déplacements, suscite donc de nombreux travaux dans la communauté scientifique. Selon Landau et Jackendoff [43], la reconnaissance d'objets impliquerait deux systèmes de traitement spécifiques : le système du *What/quoi* et celui du *Where/où*. Cette hypothèse s'appuie notamment sur l'existence chez le singe d'aires cérébrales spécialisées dans l'identification des objets ou dans leur **localisation**, et situées respectivement au niveau du cortex temporal inférieur et du lobe pariétal (Ungerleider et Mishkin [60]). On observe un parallèle similaire chez l'homme au niveau du système visuel, composé du système parvocellulaire, de la voie ventrale (du lobe occipital au lobe temporal inférieur), impliqué notamment dans la reconnaissance de la couleur et de la forme des objets, et du système magnocellulaire, de la voie dorsale (du lobe occipital au lobe pariétal), impliqué dans la reconnaissance du mouvement, de la profondeur et de la localisation (Ungerleider et Haxby [59], cités par Kemmerer [40]). Farah et collab. [25] mettent également en évidence une dissociation *quoi/où* chez des patients, lésés au niveau de l'aire temporelle inférieure : ceux-ci montrent des déficits dans la reconnaissance des formes mais non dans leur localisation.

3. Langage et représentations spatiales

L'ensemble de ces données suggère donc l'existence d'un système particulier, spécifique au traitement des informations spatiales.

Concernant plus précisément le traitement de ces informations spatiales, il semble également qu'il existe au niveau cérébral des mécanismes distincts pour la représentation de l'espace péri-personnel, i.e. à portée de main, qui serait traité par la voie dorsale, et de l'espace extrapersonnel, i.e. hors de portée, qui serait traité par la voie ventrale (voir les travaux de Bjoertomt et collab. [4] ainsi que la revue de littérature proposée par Kemmerer [40]). Ces mécanismes distincts donnent nécessairement lieu à des représentations perceptives spécifiques à l'espace proche et lointain. La négligence spatiale unilatérale, ou hémignégligence, qui est l'incapacité de détecter, de s'orienter vers, ou de répondre à des stimuli présentés dans l'hémiespace contralésionnel (e.g. Heilman et Valenstein [35]), illustre parfaitement ce phénomène : certains patients présentent une hémignégligence relative à l'espace péripersonnel, sans trouble particulier pour les objets situés dans l'espace extrapersonnel, ou inversement.

Par ailleurs, l'homme a la capacité d'exploiter ces diverses représentations spatiales, formées sur la base d'informations visuelles, mais également auditives et haptiques (voir Figure 2.3), afin d'exprimer son vécu spatial ; nous disposons pour cela d'un « langage spatial » (Landau et Jackendoff [43]), constitué d'un ensemble de déictiques spatiaux.

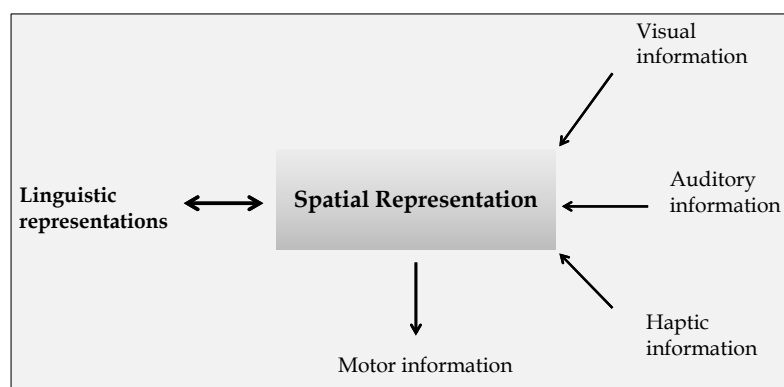


FIGURE 2.3 – La représentation spatiale (Landau et Jackendoff [43], p.218)

3.2 Représentations spatiales linguistiques

L'utilisation de termes spatiaux nous permet, comme nous l'avons vu précédemment, de renseigner les coordonnées spatiales de la situation d'énonciation, mais également de scinder l'espace qui nous entoure en différents secteurs, afin de localiser au mieux, en termes de précision, les éléments dans l'espace. Diessel [17] s'est intéressé à cet encodage lexical de la distance, via l'utilisation de démonstratifs, afin d'établir une classification des structures linguistiques, en fonction du contraste de distance qu'elles véhiculent. Cette classification concerne 234 langues, réparties finalement en cinq catégories⁷, listées ci-après.

7. Cette classification permet de fait de réfuter une théorie longtemps répandue, selon laquelle il ne pourrait exister plus de trois catégories de démonstratifs au sein d'un même système linguistique (e.g. Fillmore [26]).

1. Aucun contraste de distance ;
2. Contraste binaire ;
3. Contraste ternaire ;
4. Quatre termes ;
5. Cinq termes et plus.

Diessel [17] montre ainsi qu’une majorité des systèmes linguistiques, 127 langues, soit 54% de l’échantillon, présente un contraste de distance binaire. En d’autres termes, les démonstratifs sont classés en démonstratifs proximaux, utilisés préférentiellement pour désigner les objets proches du locuteur, et démonstratifs distaux, pour désigner les objets plus éloignés. C’est le cas notamment de l’anglais, qui utilise le terme proximal *this* (ceci) et le terme distal *that* (cela). Par ailleurs, 88 langues, soit près de 38% de l’échantillon, montre un contraste ternaire. Cette catégorie est divisible en deux sous-catégories : le contraste de distance peut être de type *distance-oriented* (i.e. orienté-distance) ou *person-oriented* (i.e. orienté-personne). Dans le premier cas, on ajoute à la distinction classique proximal/distal une dimension médiale ; c’est le cas notamment de l’espagnol où les termes *este*, *aquel* et *ese* renvoient respectivement aux objets situés près, loin, ou à mi-chemin du locuteur. Dans le second cas, l’un des trois démonstratifs au moins renvoie à la distance du référent par rapport à l’interlocuteur ; c’est le cas notamment du japonais, où l’on distingue les référents (1) proches du locuteur, (2) proches de l’interlocuteur, (3) loin des deux intervenants⁸. Les catégories restantes sont nettement moins bien représentées ; huit langues présentent un système à quatre termes et quatre langues un système à cinq termes ou plus (parmi ces distinctions plus rares, certaines prennent en compte le fait que l’objet soit visible ou qu’il appartienne au locuteur). A notre connaissance, le nivkh (langue paléo-sibérienne, parlée en Russie) bat tous les records avec non moins de 15 catégories relatives à la distance du référent (Gruzdeva [34]). Enfin, sept langues sont classées en « aucun contraste », car leur système de démonstratifs est neutre (*deictically non-contrastive* ou *distance-neutral*). C’est le cas par exemple de l’allemand ou du français. La stratégie alors utilisée pour véhiculer un contraste de distance est d’ajouter un suffixe adverbial au démonstratif (voir les Exemples ci-dessous).

Exemple 1 : *ce* ou *celui* [Démonstratif] + *ci* [Adverbe locatif proximal] = *ceci* ou *celui-ci* ;

Exemple 2 : *ce* ou *celui* [Démonstratif] + *là* [Adverbe locatif distal] = *cela* ou *celui-là*.

Mais qu’il soit binaire ou plus complexe, un tel encodage lexical de la distance pose question : dans quelle mesure une structure linguistique particulière peut-elle être liée aux mêmes représentations mentales que les mécanismes de perception/action qui sous-tendent notre vécu spatial ?

8. Il semble que le système orienté-distance soit mieux représenté que le système orienté-personne ; le pourcentage exact n’est pas précisé dans l’article de Diessel [17], mais l’auteur avait déjà établi le même type de classification sur un échantillon restreint (80 langues) et montré que les systèmes ternaires orientés-distance représentaient 30% de l’échantillon contre 12% pour les systèmes ternaires orientés-personne (Diessel [15]).

3. Langage et représentations spatiales

3.3 Etudes expérimentales du lien entre représentations spatiales perceptives et linguistiques

Bien qu'un certain nombre de structures linguistiques encodent la notion d'espace de manière plus complexe, Coventry et collab. [14] suggèrent que les représentations linguistiques de l'espace basées sur une distinction proximale - distale, soient intimement liées aux représentations perceptives et visuelles de l'espace, soit une distinction espace proche (au sens de péripersonnel, à portée de main) - loin (au sens d'extrapersonnel, hors de portée). De fait, si les représentations linguistiques sont déterminées par nos représentations perceptives, l'utilisation de termes proximaux et distaux devrait être étroitement corrélée à l'emplacement de l'objet dans l'espace.

Afin de tester cette hypothèse, les auteurs proposent une tâche de production de pointage à des participants anglais (dont le système démonstratif est binaire : *this*, proximal ; *that*, distal) et espagnols (dont le système démonstratif est ternaire : *este*, proximal ; *ese*, médial ; *aquel*, distal). Les participants sont assis face à une table longue de 320 cm, illustrée Figure 2.4, divisée en quatre secteurs. A chaque nouvel essai correspond une carte *who-which-where*, indiquant *qui* doit placer *quoi* et *où*. Un objet, parmi neuf de formes différentes, est alors placé, soit par le participant, soit par l'expérimentateur, sur l'un des 12 marqueurs de couleurs. Le participant doit alors le désigner par un geste de pointage, associé à une phrase type : Démonstratif + Couleur + Forme, telle que *This/that red triangle* (i.e. ce triangle rouge).

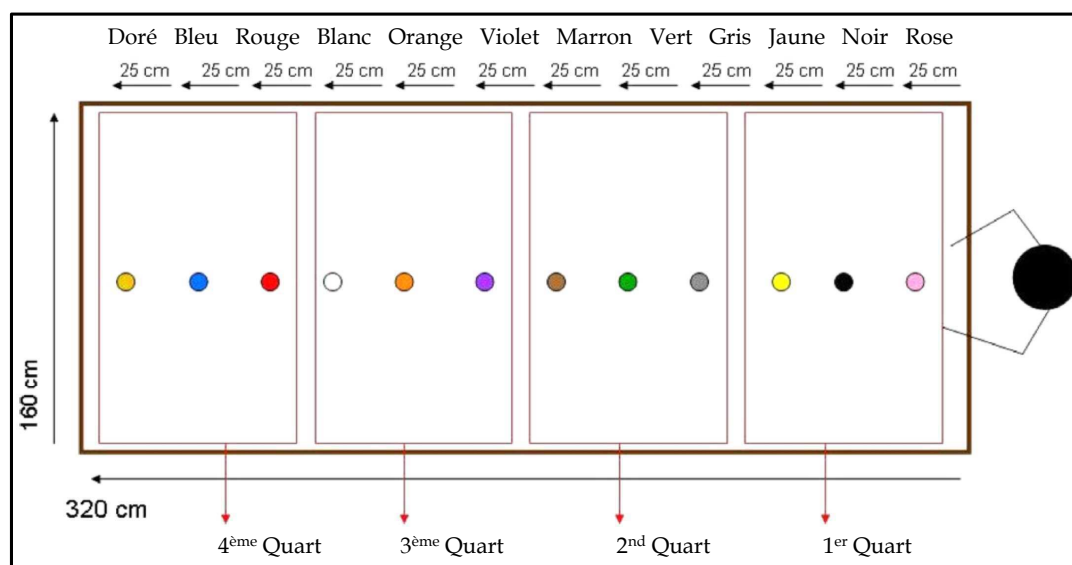


FIGURE 2.4 – Marqueurs de couleur et de position de l'objet relativement aux participants (Coventry et collab. [14], p.891)

Les auteurs manipulent différentes variables, dont l'interaction avec l'objet (qui place l'objet) ainsi que la position de l'expérimentateur (qui peut être face à ou à côté du participant). Ces deux variables sont testées afin de déterminer plus précisément la nature ternaire du système démonstratif espagnol (orienté-distance vs orienté-personne) et ne seront pas donc pas détaillées ici. Nous nous intéressons en revanche aux deux autres variables, soit la distance participant-objet (qui peut varier de 25 à 300 cm, seuls les objets situés dans le premier quart sont alors à portée de main) ainsi que le type de pointage : les participants pointent manuellement ou à l'aide d'un bâton long de 70 cm ; dans ce cas là, les objets situés dans les deux premiers quarts sont atteignables. Les auteurs mesurent notamment la proportion de déictiques proximaux (i.e. *this* et *este*), avec l'hypothèse que celle-ci devrait être plus importante pour les objets accessibles, soit les objets situés dans le premier quart en situation de pointage sans outil, et dans les deux premiers quarts en situation de pointage avec outil. Selon les auteurs, l'utilisation du bâton devrait en effet induire un *remapping* de l'espace initialement extrapersonnel en espace péripersonnel, comme démontré précédemment par Berti et Frassinetti [2].

Les résultats, présentés Figure 2.5 ci-après, montrent une correspondance assez nette entre le démonstratif utilisé et l'emplacement de l'objet : la proportion de déictiques proximaux chute brusquement pour les objets situés hors de portée. De plus l'effet d'interaction entre la distance et l'outil est significatif : le pourcentage de déictiques proximaux pour les objets situés dans le second quart (à priori hors de portée) est plus élevé lorsque les participants utilisent le bâton que lorsqu'il pointent manuellement.

Les auteurs concluent que les représentations linguistiques spatiales sont largement influencées par nos représentations perceptives ; le choix d'un déictique proximal dépendrait principalement des caractéristiques physiques de l'environnement, et plus précisément du caractère accessible de l'objet. En d'autres termes, c'est une structure cognitive de base qui nous permettrait de représenter nos interactions avec les objets qui nous entourent. L'utilisation d'un terme médial nous permettrait d'interpréter le monde selon le point de vue de notre interlocuteur (avec une distinction de type mon espace / ton espace).

3. Langage et représentations spatiales

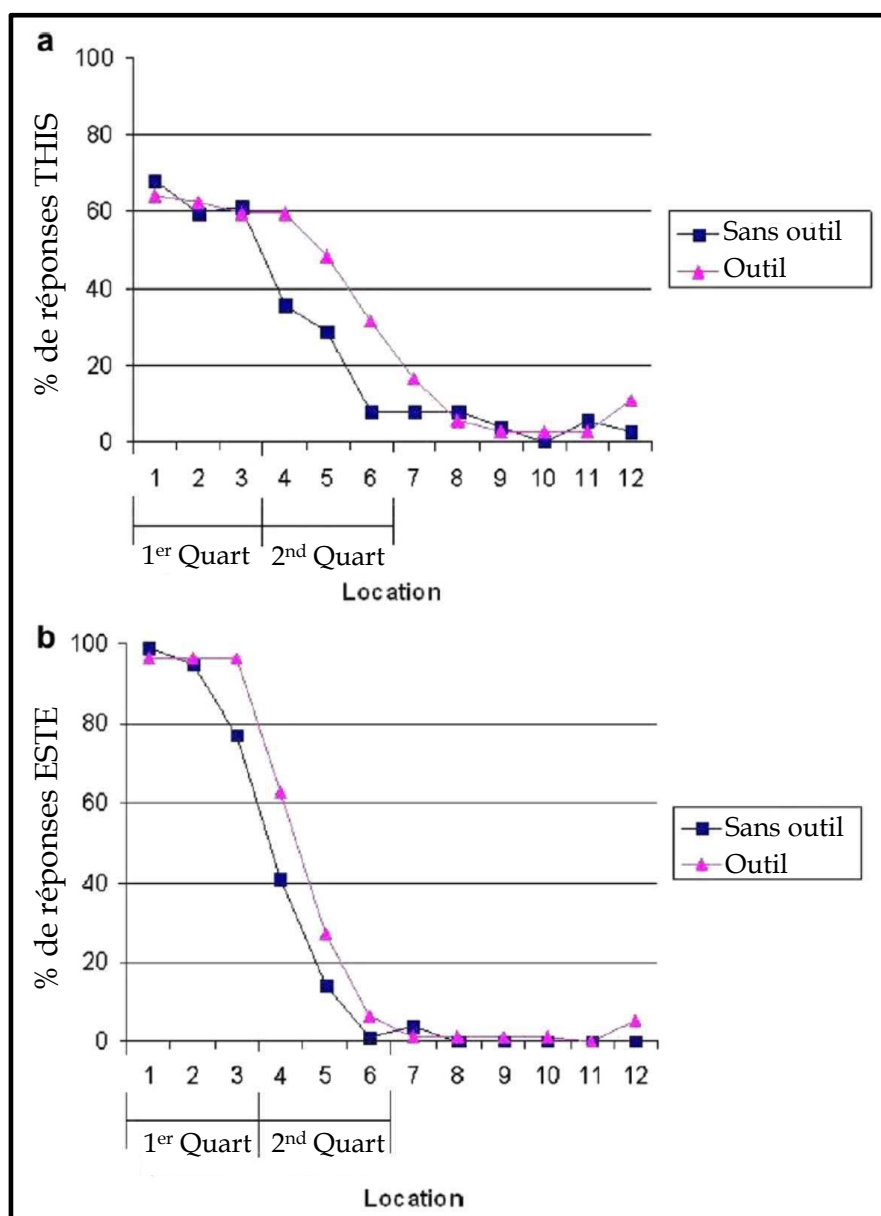


FIGURE 2.5 – (a) Pourcentage d'utilisation de THIS; (b) Pourcentage d'utilisation de ESTE en fonction de la distance de l'objet et de la présence (▲) vs absence (■) d'un outil (Coventry et collab. [14], p.893)

Kemmerer [40] propose en revanche une toute autre théorie : selon lui, bien que des ressemblances certaines existent entre les représentations perceptives et linguistiques de l'espace, il n'y aurait pas de correspondance stricte entre l'utilisation de termes proximaux/distaux et l'espace proche/loin. L'utilisation de termes spécifiques serait plutôt liée à leurs propriétés sémantiques et pragmatiques ; de fait, elle ne dépendrait pas de facteurs concrets (i.e. spatiaux au sens strict du terme, c'est-à-dire objectivement mesurables), tels que des frontières physiques ou des possibilités d'actions sur l'objet, mais serait déterminée par les propriétés abstraites et subjectives du contexte discursif et situationnel. De cette façon, le contraste de distance véhiculé par les démonstratifs ne serait pas absolu mais relatif. Le système déictique constituerait en cela un *language-internal system*, au sens saussurien du terme, soit un système langagier spécifique, pour lequel chaque notion est définie non pas en rapport à l'espace physique réel mais en opposition à une autre notion (Saussure [53]). Par exemple, *celui-ci* signifie simplement « plus proche du centre déictique que celui-là », sans pour autant être nécessairement à portée de main, et inversement, *celui-là* signifie « plus loin du centre déictique que celui-ci ». La flexibilité pragmatique de ces termes nous permettrait de véhiculer un contraste proximal/distal pour toute distinction spatiale que l'on souhaiterait exprimer.

Dans un registre similaire, certains travaux se sont penchés sur la problématique geste / parole / distance, à savoir le lien entre un encodage de la distance au niveau linguistique d'une part, et au niveau gestuel d'autre part. Nous allons présenter deux études, réalisées par Bonfiglioli et collab. [5] et Chieffi et collab. [12], dont le but est de caractériser les interférences entre une représentation linguistique particulière (véhiculant une information relative à la distance d'une cible) et la production des gestes manuels qui y sont associés.

Bonfiglioli et collab. [5] s'intéressent à l'influence des démonstratifs italiens *questo/a*_[masc-fem] et *quello/a*_[masc-fem], équivalents respectifs des termes *ceci* et *cela*, sur la production du mouvement d'atteinte et de saisie. Le but initial des auteurs est de mettre en évidence une interaction particulière entre geste et parole, où la production d'un geste serait déterminée par des facteurs situationnels et linguistiques. Pour cela, ils souhaitent montrer que la distinction proximal/distal, véhiculée par les démonstratifs, ne correspond pas à la distinction perceptive entre un espace péripersonnel et un espace extrapersonnel (ce qui était proposé par Coventry et collab. [14]), mais dépendrait plutôt de l'éloignement relatif des objets entre eux et par rapport au centre déictique ; de fait, un démonstratif distal pourrait être associé à un objet situé dans l'espace péripersonnel, si celui-ci est relativement plus éloigné qu'un autre (comme suggéré par Kemmerer [40]).

La tâche des participants est d'écouter une instruction auditive, contenant un démonstratif proximal (*questo/a*), distal (*quello/a*) ou neutre (*lo/a*, que l'on pourrait traduire par *ça*) :

- Proximal : « *Prendi questo/a* » - Attrape ceci ;
- Distal : « *Prendi quello/a* » - Attrape cela ;
- Neutre « *Prendi lo/a* » - Attrape ça.

Une gomme ou un bouchon est ensuite placé devant eux, à 12 ou 30 cm, soit dans l'espace péripersonnel proche ou péripersonnel lointain (dans les deux cas, la cible est à portée de main, mais l'objet placé à 30 cm est relativement plus éloigné du participant que celui placé à 12 cm).

3. Langage et représentations spatiales

Le paradigme utilisé est celui du Go/No-Go, où deux classes d'items sont définies : les items-cible d'une part, pour lesquels une réponse est attendue, et les items *fillers* (ou de remplissage) d'autre part, pour lesquels aucune réponse n'est attendue. Les participants doivent ici répondre, c'est-à-dire effectuer un mouvement d'atteinte et de saisie, uniquement s'il y a congruence entre le genre véhiculé par l'instruction auditive (masculin vs féminin) et l'objet (un bouchon vs une gomme), comme l'illustrent les exemples ci-dessous.

Exemple GO : « Prendi quello » ou « Prendi lo » + [un bouchon]
« Prendi questa » ou « Prendi la » + une gomme ;

Exemple NO-GO : « Prendi questo » ou « Prendi lo » + [une gomme]
« Prendi quella » ou « Prendi la » + [un bouchon]

Cette procédure permet de s'assurer que les participants écoutent la totalité de l'instruction auditive et non la syllabe finale uniquement, qui, en italien, permet à elle seule de déterminer le genre de l'objet ; il leur est également demandé dans ce même but de répéter l'instruction, dans 20% des cas.

Les auteurs cherchent à déterminer dans quelle mesure la congruence/incongruence entre le positionnement de la cible (péripersonnel proche vs péripersonnel lointain) et le démonstratif utilisé (proximal vs distal) va influencer les propriétés du geste manuel. En d'autres termes, ils cherchent à montrer qu'une représentation linguistique particulière (e.g. proximale) va activer préférentiellement un geste plutôt un autre (e.g. un geste destiné à un objet proche). En effet, si le geste et la parole sont en interaction au sein d'un même système communicatif, les productions manuelles devraient être altérées en situation incongruente, et ce bien que les deux objets soient situés dans le même espace perceptif : l'encodage lexical de la distance pourrait dans un premier temps, c'est-à-dire au moment de l'écoute, permettre au participant de programmer un geste en particulier, mais en cas d'incongruence entre le démonstratif perçu et la distance réelle de l'objet, le geste devrait être reprogrammé, et donc retardé. De tels résultats montreraient que le système linguistique permet d'activer un programme moteur spécifique.

Les auteurs utilisent le système de capture de mouvements opto-électronique Qualisys⁹ afin d'enregistrer les propriétés cinématiques du mouvement manuel (les profils d'amplitude, de vitesse et d'accélération) ; ils calculent également le temps de réaction, les temps d'initiation et de fin de mouvement, ainsi que la durée totale du geste.

Leurs résultats montrent, qu'à l'exception du temps de réaction, chaque paramètre est influencé par l'emplacement de la cible, soit plus élevé ou retardé pour les objets les plus éloignés. Par ailleurs, s'ils n'observent pas d'effet d'interaction entre l'emplacement de la cible et le déictique perçu sur les paramètres cinématiques et la durée du geste, ils l'observent sur les temps de réaction : ceux-ci sont significativement retardés en situation incongruente (i.e. démonstratif proximal + objet loin ; démonstratif distal + objet proche). Ces résultats mettent en évidence une interférence entre le système linguistique d'une part, et le système moteur manuel d'autre part. Les participants développent des attentes quant à la position de la cible à partir du démonstratif perçu et planifient leur geste manuel en conséquence. Ce n'est qu'une fois l'emplacement

9. <http://www.qualisys.com>

de la cible révélé qu'ils vont mettre en correspondance la position attendue et la position réelle de l'objet. En situation congruente, le mouvement peut être déclenché immédiatement, mais en situation incongruente, celui-ci doit être reprogrammé, ce qui induit des temps de réaction nécessairement plus longs. Ces résultats, qui sont observés de manière indifférenciée pour les différents type de démonstratifs, suggèrent que les termes proximaux et distaux ne sont pas strictement associés à l'espace péripersonnel et extrapersonnel mais sont liés à la distance relative des objets entre eux et par rapport au centre déictique ; de fait, le contraste proximal/distal véhiculé par les démonstratifs active des représentations motrices distinctes, y compris pour deux objets situés à portée de main, dans le même espace perceptif.

Cette première étude suggère que les gestes manuels sont planifiés non seulement par rapport aux contraintes spatio-motrices (i.e. liées à l'emplacement de l'objet) mais également par rapport à des facteurs linguistiques (i.e. l'information de distance véhiculée par les démonstratifs) ; elle montre ainsi une influence du système linguistique sur la forme du geste. Néanmoins, parce qu'elle est consacrée aux gestes d'atteinte/saisie (un geste produit "pour soi"), cette étude ne permet pas de rendre compte de l'aspect communicatif de l'interaction geste/parole.

Chieffi et collab. [12] ont quant à eux étudié l'interaction entre adverbes de lieu et gestes de pointage. L'objectif des auteurs est de déterminer dans quelle mesure la production d'un mot déictique peut influencer celle d'un geste déictique, et réciproquement. Ils vont pour cela tester l'interférence entre geste et parole lorsque leur contenu respectif sont congruents ou incongruents.

Dans une première expérience, la tâche des participants est de lire à haute voix les adverbes *qua* (*ici*) ou *la* (*là*), imprimés sur un jeton, situé près ou "loin" d'eux (à 8 et 68 cm, soit dans l'espace péripersonnel). Dans le même temps, ils doivent effectuer un geste de pointage vers leur propre corps, lorsque le jeton est proche, ou vers l'horizon, lorsqu'il est éloigné. Deux variables sont ainsi manipulées : le déictique [proximal - distal] et la direction du pointage [proche - loin]. Ainsi, la relation entre le contenu de la parole et celui du geste est congruente lorsque le déictique proximal est imprimé sur le jeton proche (et le déictique distal imprimé sur le jeton éloigné) mais incongruente lorsqu'il est imprimé sur le jeton éloigné (et lorsque le déictique distal est imprimé sur le jeton proche). Les auteurs ajoutent une condition contrôle, dans laquelle les inscriptions *XXX* et *XX* remplacent le déictique ; dans ce cas, les participants se contentent de pointer (vers eux si le jeton est proche, vers l'horizon s'il est éloigné). Dans une seconde expérience, les participants doivent s'abstenir de pointer, le paradigme étant par ailleurs similaire.

Cependant, bien que ces deux expériences permettent en théorie de comparer les productions unimodales (Expérience 1 : geste seul ; Expérience 2 : parole seule) aux productions bimodales (Expérience 1 : geste+parole), cet aspect n'est pas exploité par les auteurs. Ces derniers s'intéressent uniquement aux interférences entre deux messages contradictoires (soit aux interférences entre le geste et la parole quand leur contenu respectif sont incohérents). Les tâches de geste seul et de parole seule ne sont ici que des tâches contrôle : la tâche de geste seul leur permet de s'assurer que les effets obtenus sont liés à l'incongruence geste/parole plutôt qu'à la seule présence du jeton (un potentiel distracteur visuel qui pourrait perturber les productions) ; parallèlement, la tâche de parole seule leur permet de s'assurer que les effets obtenus sont liés à l'incongruence geste/parole plutôt qu'à l'incongruence entre la position du jeton et l'inscription notée sur le jeton).

3. Langage et représentations spatiales

Les auteurs formulent trois hypothèses :

- Un effet de congruence/incongruence sur la production de la parole et celle du geste serait en faveur d’une interaction bidirectionnelle entre les deux systèmes de production ;
- Un effet sur l’un des deux systèmes uniquement serait en faveur d’une interaction à sens unique ;
- L’absence d’effet serait en faveur d’une indépendance entre les deux systèmes.

Les productions manuelles sont enregistrées par le système de capture de mouvements opto-électronique Elite¹⁰ afin de mesurer le déplacement de l’index et les profils de vitesse et d’accélération. Les productions orales sont également enregistrées, afin de mesurer les valeurs moyennes du premier et du second formant de la voyelle [a] (notions définies au Chapitre 3, Section 1).

Les auteurs observent un effet de la congruence/incongruence geste/parole sur les propriétés cinématiques du geste et sur les propriétés acoustiques de la parole. Plus précisément, le pic d’accélération ainsi que le pic de vitesse sont plus bas en situation incongruente (jeton proche + déictique loin ; jeton loin + déictique proche) qu’en situation congruente (jeton proche + déictique proche ; jeton loin + déictique loin) ou contrôle, tout comme les valeurs de second formant (plus basses en situation incongruente, relativement aux situations congruente ou contrôle). De manière générale, les productions semblent inhibées lorsque les deux modalités transmettent des informations contradictoires.

Les résultats de cette étude, compatibles avec la première hypothèse des auteurs, suggèrent une influence bidirectionnelle entre les systèmes de production du geste et de la parole, chaque modalité étant influencée par la congruence/incongruence des productions (i.e. le geste est altéré lorsque la parole transmet simultanément un message incohérent, et vice versa). Toutefois, les cibles du pointage manuel (le participant lui-même ou l’horizon) ne sont pas liées aux cibles du pointage vocal (l’inscription sur le jeton) ; de fait même en situation congruente le geste et la parole ne véhiculent pas de but commun et leur fonction est de fait très différente. Outre la complexité d’une telle tâche, cette étude ne permet pas plus que la précédente d’évaluer l’aspect communicatif de l’interaction geste/parole.

3.4 Représentations spatiales linguistiques et symbolisme phonétique

Quatre cents ans avant Jésus-Christ, Platon dans son *Cratylus* tentait déjà de percer le mystère des mots : sont-ils arbitraires ou bien le reflet naturel de ce qu’ils désignent ?

On sait aujourd’hui que de manière générale, la relation entre le son et le sens d’un mot est arbitraire (Saussure [53]) mais il existe néanmoins des cas pour lesquels le son *correspond* au sens, et dont l’exemple le plus frappant est celui des onomatopées. Ce lien motivé, systématique, entre phonétique et sémantique, est désigné par le terme de symbolisme phonétique.

Ohala [50] montre par exemple un parallèle étonnant entre la taille d’un objet et les propriétés acoustiques des termes utilisés pour le désigner. Ainsi, les voyelles fermées, antérieures¹¹

10. <http://www.eliteopto.com>

11. Les voyelles antérieures, telles que [e], sont produites avec la langue positionnée vers l’avant, tandis que les voyelles postérieures, telles que [u], sont produites avec la langue positionnée vers l’arrière. Les voyelles fermées, [i],

(voir Figure 2.6 ci-après) et dont la fréquence fondamentale¹² est élevée (par exemple [i], [ɪ], [y], [e]) sont associées à la notion de petite taille, tandis que les voyelles ouvertes, postérieures (voir Figure 2.6) et dont la fréquence fondamentale est basse (telles que [ɑ], [ʌ], [ɔ], [o]) sont associées à la notion de grande taille. Les Figures 2.7 et 2.8 illustrent ces phénomènes (mais le lecteur pourra trouver des contre-exemples dans Diffloth [20]). De manière générale, il semble que la fréquence d'un son soit inversement proportionnelle à la taille de l'objet qu'il désigne, phénomène qu'Ohala [50] appelle *The Frequency Code* (code-fréquence).

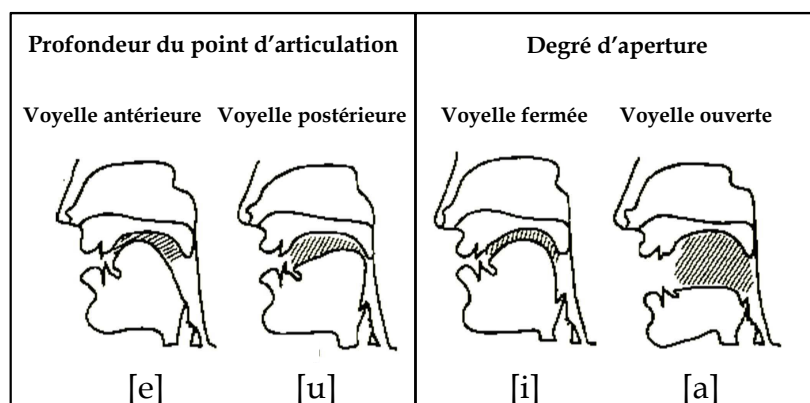


FIGURE 2.6 – Articulation des voyelles antérieures/postérieures et fermées/ouvertes

Language	Word, Morpheme	Translation
Ewe	[kítsíkítsí]	‘small’
Yoruba	[bírí]	‘be small’
Yoruba	[kpénkpé]	‘small’
Spanish	[ˈtʃiko]	‘small’
Greek	[mikros]	‘small’
English	[-i]	diminutive suffix
Spanish	[-it-]	diminutive suffix
Irish	[-in]	diminutive suffix
French	[pətit]	‘small’

FIGURE 2.7 – Expressions relatives à la notion *petit* (Ohala [50], p.2)

sont réalisées avec la langue placée en haut, tandis que les voyelles ouvertes, [a], sont réalisées avec la langue placée en bas.

12. La fréquence fondamentale, que nous définirons plus précisément au Chapitre 3, est la fréquence de base de vibration des cordes vocales, liée à la sensation de hauteur de la voix (aïgue ou grave).

3. Langage et représentations spatiales

Language	Word, Morpheme	Translation
Ewe	[$\widehat{g}\widehat{b}\widehat{a}\widehat{g}\widehat{b}\widehat{a}$]	‘large’
Yoruba	[$\widehat{b}\widehat{i}\widehat{r}\widehat{i}$]	‘be large’
Yoruba	[$\widehat{g}\widehat{b}\widehat{\eta}\widehat{g}\widehat{b}\widehat{\eta}$]	‘large’
Spanish	[$^l\text{gordo}$]	‘fat’
Greek	[makros]	‘large’
French	[$g\text{ʁ}\check{a}$]	‘large’

FIGURE 2.8 – Expressions relatives à la notion *grand* (Ohala [50], p.2)

Kapatsinski et Johnston [39] mettent également en évidence une relation particulière entre la taille d’un objet et le type de voyelles utilisées pour le désigner. Des objets inconnus, de différente taille, sont présentés aux participants ; ceux-ci ont pour tâche de leur choisir un nom parmi 20 mots et pseudo-mots contenant les voyelles [i], [u], [a] ou [o]. Leurs résultats, illustrés Figure 2.9, montrent que les mots ou pseudo-mots contenant des voyelles fermées tendent à être associés aux petites créatures, tandis que les mots ou pseudo-mots contenant des voyelles ouvertes tendent à être associés aux grandes créatures. Sapir [52] en 1929 observait déjà une association semblable, puisque les non-mots contenant la voyelle ouverte [a] étaient associés à de grands objets, tandis que les mêmes non-mots contenant la voyelle fermée [i] étaient associés à de petits objets, et ce dans 80% des cas.

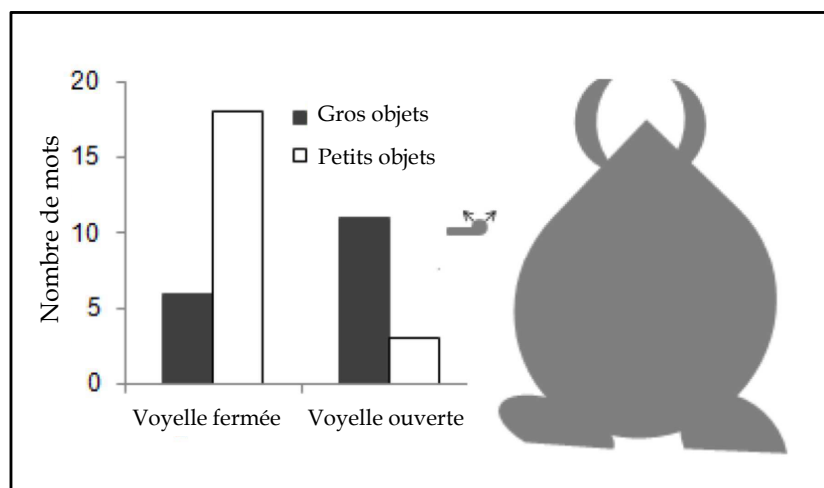


FIGURE 2.9 – Association entre la taille des objets et le type de voyelle utilisé (Kapatsinski et Johnston [39], p.5)

On observe également ce type de relation non arbitraire entre phonétique et sémantique au niveau des termes spatiaux encodant la distance des objets. Traunmüller [57], qui étudie ce lien d'un point de vue acoustique, montre que le F², qui correspond à l'ensemble des formants au-delà du premier formant¹³, tend à être plus élevé dans les formes proximales des déictiques relativement aux formes distales. Cette tendance est observée sur 32 des 37 langues de son échantillon, soit 86.5%. Dans le même registre, Woodworth [66] observe que les formes proximales, relativement aux formes distales, contiennent plutôt des voyelles à second formant élevé, corrélé à une position antérieure de la langue. Néanmoins, cette tendance n'est observée que pour 13 des 26 langues considérées, soit 50% de l'échantillon ; parmi les 13 langues restantes, deux seulement présentent le pattern opposé (soit une association entre second formant élevé et formes distales), les onze autres ne montrant aucun résultat clair.

Ultan [58] étudie quant à lui ce lien phonétique/sémantique d'un point de vue articulatoire ; il montre pour 33.1% de son échantillon, constitué de 136 langues, une tendance à associer voyelles antérieures fermées et non arrondies, telles que [i], et formes proximales, dans les démonstratifs.

Bien que relativement variables, ces données suggèrent l'existence d'une tendance, au niveau des déictiques spatiaux, à associer certaines voyelles à certaines tailles et distances. Plus précisément, on observe un contraste vocalique particulier, qui se manifeste par une tendance à associer des voyelles antérieures fermées, possédant un second formant élevé et un premier formant bas, aux objets petits et/ou proches, et des voyelles postérieures ouvertes, possédant un second formant bas et un premier formant élevé, aux objets ou distances plus larges. Selon Sapir [52], cette tendance pourrait s'expliquer de deux façons : « (1) *acoustically, certain vowels have greater "volume" than others* ; (2) *kinesthetically, the one "spatially extended gesture" (i.e., tongue position and resonance cavity) is symbolic of a larger reference*¹⁴ » ([52], p.225).

4 Conclusion

Le processus déictique se révèle être le candidat idéal pour étudier l'intégration geste/parole au niveau du langage. C'est notamment par sa fonction exophorique spatiale que la nature communicative et multimodale de ce processus prend tout son sens : geste et parole ont alors précisément le même but : structurer, exprimer et partager notre vécu spatial.

Les termes déictiques spatiaux présentent la particularité d'encoder la distance de leur référent : on distingue ainsi généralement les termes proximaux, destinés aux objets proches du centre déictique, des termes distaux, destinés aux objets plus éloignés.

Bon nombre d'études cherchent à comprendre les déterminants d'un tel encodage lexical : est-il lié à nos représentations perceptives de l'espace, c'est-à-dire déterminé par les propriétés concrètes et objectives du monde physique, et notamment à nos possibilités d'action sur les objets, ou dépend-il de propriétés abstraites et subjectives, liées au contexte discursif et situa-

13. La notion de formant sera détaillée au Chapitre 3.

14. Traduction proposée : (1) d'un point de vue acoustique, certaines voyelles présentent un "volume" plus important que d'autres ; (2) d'un point de vue kinesthésique, un geste étendu dans l'espace (i.e. position de la langue et cavité) symbolise un référent plus large.

4. Conclusion

tionnel particulier dans lequel ils sont produits ? Par ailleurs, on observe un parallèle étonnant au niveau de ces termes déictiques spatiaux : leurs propriétés sémantiques (i.e. proximale vs distale) semblent étroitement liées à leurs propriétés phonétiques (i.e. acoustiques et articulatoires) ; ce phénomène se manifeste par une tendance à associer les objets proches aux voyelles fermées et/ou antérieures et les objets éloignés aux voyelles ouvertes et/ou postérieures.

Dans ce cadre là, nous proposons d'étudier l'encodage de la distance via les propriétés lexicales du pointage mais également via ses propriétés acoustiques, articulatoires et cinématiques. Notre objectif général étant de mettre en évidence un couplage fonctionnel et structurel entre le geste et la parole dans le langage, nous souhaitons montrer une interaction forte entre les gestes manuels, les gestes vocaux et les structures linguistiques qui y sont associées. Plus concrètement, notre hypothèse est que l'encodage de la distance observé au niveau lexical soit issu d'un phénomène moteur de bas niveau : désigner un objet éloigné impliquerait un geste plus ample, que celui-ci soit un geste manuel ou vocal. Ces comportements moteurs auraient progressivement été intégrés à la phonologie des langues. De fait, nous devrions pouvoir observer, en fonction de la distance de l'objet pointé, (1) des variations acoustiques et/ou articulatoires intra-catégorielles : une même voyelle, indépendamment de sa nature et de son contexte (lexical ou non), devrait être réalisée en configuration plus ouverte pour un objet plus éloigné ; (2) des variations cinématiques, et notamment un geste plus ample pour un objet plus éloigné.

Le chapitre suivant présentera notre étude expérimentale, dont le but est d'évaluer, dans un cadre contrôlé, la nature de l'interaction geste/parole et de l'encodage de la distance, et ce par la comparaison de productions unimodales (pointage gestuel ou vocal) et bimodales (pointage gestuel et vocal), destinées à des cibles plus ou moins éloignées.

Bibliographie

- [1] Bates, E. et F. Dick. 2002, «Language, gesture, and the developing brain», *Developmental Psychobiology*, vol. 40, p. 293–310.
- [2] Berti, A. et F. Frassinetti. 2000, «When far becomes near : Remapping of space by tool use», *Journal of Cognitive Neuroscience*, vol. 12, p. 415–420.
- [3] Bühler, K. 1934, *Sprachtheorie : Die Darstellungsfunktion der Sprache*, Jena : Fischer.
- [4] Bjoertomt, O., A. Cowey et V. Walsh. 2002, «Spatial neglect in near and far space investigated by repetitive transcranial magnetic stimulation», *Brain*, vol. 125(9), p. 2012–2022.
- [5] Bonfiglioli, C., C. Finocchiaro, B. Gesierich, F. Rositani et V. Massimo. 2009, «A kinematic approach to the conceptual representations of this and that», *Cognition*, vol. 111, p. 270–274.
- [6] Bruner, J. 1983, *Child Talk*, Norton, New York.
- [7] Butterworth, G. 1998, «What is special about pointing in babies?», dans *The Development of Sensory Motor and Cognitive Capacities in Early Infancy : From Perception to Cognition*, édité par F. Simion et G. Butterworth, Hove, England : Psychology Press/Erlbaum, p. 171–190.
- [8] Butterworth, G. 2003, «Pointing is the royal road to language for babies», dans *Pointing : Where Language, Culture, and Cognition Meet*, édité par S. Kita, Mahwah, NJ, US : Lawrence Erlbaum Associates Publishers, p. 9–33.
- [9] Butterworth, G. et L. Grover. 1990, «Joint visual attention, manual pointing, and pre-verbal communication in human infancy», dans *Attention and Performance XIII, Motor Representation and Control*, édité par M. Jeannerod, Lawrence Erlbaum Publishers, p. 605–624.
- [10] Capirci, O., J. Iverson, E. Pizzuto et V. Volterra. 1996, «Gesture and words during the transition to two-word speech», *Journal of Child Language*, vol. 23, p. 645–673.
- [11] Carpenter, M., K. Nagell et M. Tomasello. 1998, «Social cognition, joint attention, and communicative competence from 9 to 15 months of age.», *Monographs of the Society for Research in Child Development*, vol. 63 (4, Serial No. 255).
- [12] Chieffi, S., C. Secchi et M. Gentilucci. 2009, «Deictic word and gesture production : Their interaction», *Behavioural Brain Research*, vol. 203(2), p. 200–206.
- [13] Cooperrider, K. et R. Nunez. 2012, «Nose-pointing : A facial gesture of the Yupno Valley, Papua New Guinea», dans *The 11th Conceptual Structure, Discourse, and Language Conference*, Vancouver, Canada.
- [14] Coventry, K. R., B. Valdés, A. Castillo et P. Guijarro-Fuentes. 2008, «Language within your reach : Near-far perceptual space and spatial demonstratives», *Cognition*, vol. 108, p. 889–895.

Bibliographie

- [15] Diessel, H. 1999, *Demonstratives : Form, Function, and Grammaticalization*, John Benjamins, Amsterdam.
- [16] Diessel, H. 2003, «The relationship between demonstratives and interrogatives», *Studies in Language*, vol. 27, p. 581–602.
- [17] Diessel, H. 2005, «Distance contrasts in demonstratives», dans *World Atlas of Language Structures*, édité par M. Haspelmath, M. Dryer, D. Gil et B. Comrie, Oxford : Oxford University Press, p. 170–173.
- [18] Diessel, H. 2006, «Demonstratives, joint attention, and the emergence of grammar», *Cognitive Linguistics*, vol. 17, p. 463–489.
- [19] Diessel, H. 2012, «Deixis and demonstratives», dans *An International Handbook of Natural Language Meaning*, édité par C. Maienborn, K. von Stechow et P. Portner, Berlin : Mouton de Gruyter, p. 1–25.
- [20] Diffloth, G. 1994, «I : Big, a : Small», dans *Sound Symbolism*, édité par L. Hinton, J. Nichols et J. Ohala, Cambridge University Press : Cambridge, p. 107–114.
- [21] Dixon, R. 2003, «Demonstratives : A cross-linguistic typology», *Studies in Language*, vol. 27, p. 61–122.
- [22] Dohen, M. 2005, *Deixis prosodique multisensorielle : Production et perception audiovisuelle de la focalisation contrastive en français*, thèse de doctorat, Institut National Polytechnique de Grenoble.
- [23] Ducey-Kaufmann, V. 2007, *Le cadre de la parole et le cadre du signe : Un rendez-vous développemental*, thèse de doctorat, Université Stendhal.
- [24] Enfield, N. J. 2003, «Demonstratives in space and interaction : Data from lao speakers and implications for semantic analysis», *Language*, vol. 79, p. 82–117.
- [25] Farah, J., K. Hammond, D. Levine et R. Calvanio. 1988, «Visual and spatial mental imagery : Dissociable systems of representation», *Cognitive Psychology*, vol. 20, p. 439–462.
- [26] Fillmore, C. 1982, «Towards a descriptive framework for spatial deixis», dans *Speech, Place, and Action : Studies in Deixis and Related Topics*, édité par R. Jarvella et W. Klein, Chichester : John Wiley, p. 31–59.
- [27] Fillmore, C. 1997, «Coming and going», dans *Lectures on Deixis*, Stanford : CSLI Publications, p. 77–102.
- [28] Fogel, A. et T. Hannan. 1985, «Manual actions of nine- and fifteen-weekold human infants during face-to-face interaction with their mothers», *Child Development*, vol. 56, p. 1271–1279.
- [29] Fogel, A. et E. Thelen. 1987, «Development of early expressive and communicative action : Reinterpreting the evidence from a dynamic systems perspective», *Developmental Psychology*, vol. 23, p. 747–761.

- [30] Franco, F. et G. Butterworth. 1988, «The social origins of pointing in human infancy», dans *The Annual Conference of the Developmental Psychology Section of the British Psychological Society*. Colleg Harlech, Wales.
- [31] Franco, F. et G. Butterworth. 1996, «Pointing and social awareness : Declaring and requesting in the second year», *Journal of Child Language*, vol. 23(2), p. 307–336.
- [32] Goldin-Meadow, S. 1999, «The development of gesture with and without speech in hearing and deaf children», dans *Gesture, Speech, and Sign*, édité par L. Messing et R. Campbell, Oxford University Press, New-York, p. 117–132.
- [33] Goldin-Meadow, S. et C. Butcher. 2003, «Pointing toward two-word speech in young children», dans *Pointing : Where Language, Culture, and Cognition Meet*, édité par S. Kita, Mahwah, NJ, US : Lawrence Erlbaum Associates Publishers, p. 85–107.
- [34] Gruzdeva, E. 2007, «Challenging theory : Spatial deixis in Nivkh», dans *Proceedings of Conference on Language Documentation and Linguistic Theory*, édité par P. Austin, O. Bond et D. Nathan, London, SOAS, p. 103–113.
- [35] Heilman, K. et E. Valenstein. 1979, «Mechanisms underlying hemispatial neglect», *Ann Neurol*, vol. 5, p. 166–170.
- [36] Hellwig, B. 2003, *The Grammatical Coding of Postural Semantics in Goemai (a West Chadic Language of Nigeria)*, Nijmegen : Max Planck Series in Psycholinguistics.
- [37] Himmelmann, N. 1997, *Deiktikon, Artikel, Nominalphrase. Zur Emergenz syntaktischer Struktur*, Tübingen : Narr.
- [38] Kalagher, H. et C. Yu. 2006, «The effects of deictic pointing in word learning», dans *The 5th International Conference of Development and Learning*. Kyoto, Japan.
- [39] Kapatsinski, V. et L. Johnston. 2010, «Investigating phonotactics using xenolinguistics : A novel word-picture matching paradigm», dans *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*, édité par S. Ohlsson et R. Catrambone, Austin, TX : The Cognitive Science Society, p. 2010–2015.
- [40] Kemmerer, D. 1999, «"Near" and "far" in language and perception», *Cognition*, vol. 73, p. 35–63.
- [41] Kendon, A. 2004, *Gesture : Visible Action as Utterance*, Cambridge University Press, Cambridge.
- [42] Kita, S. 2003, *Pointing : Where Language, Culture, and Cognition Meet*, Mahwah, NJ, US : Lawrence Erlbaum Associates Publishers.
- [43] Landau, B. et R. Jackendoff. 1993, «"What" and "where" in spatial language and spatial cognition», *Behavioral and Brain Sciences*, vol. 16, p. 217–266.
- [44] Levinson, S. 1983, *Pragmatics*, Cambridge : Cambridge University Press.

Bibliographie

- [45] Levinson, S. 2004, «Deixis and pragmatic», dans *The Handbook of Pragmatics*, édité par L. Horn et G. Ward, Oxford : Blackwell Publishing Ltd, p. 97–121.
- [46] Lyons, J. 1977, *Semantics (Vol. 1-2)*, Cambridge : Cambridge University Press.
- [47] Masataka, N. 2003, «From index-finger extension to index-finger pointing : Ontogenesis of pointing in preverbal infants», dans *Pointing : Where Language, Culture, and Cognition Meet*, édité par S. Kita, Mahwah, NJ, US : Lawrence Erlbaum Associates Publishers, p. 69–84.
- [48] Meunier, H., J. Vauclair et J. Fagard. 2012, «Human infants and baboons show the same pattern of handedness for a communicative gesture», *PLoS ONE*, vol. 7(3), p. 1–3.
- [49] Morford, M. et S. Goldin-Meadow. 1992, «Comprehension and production of gesture in combination with speech in one-word speakers», *Journal of Child Language*, vol. 19(3), p. 559–580.
- [50] Ohala, J. 1994, «The frequency code underlies the sound-symbolic use of voice pitch», dans *Sound Symbolism*, édité par L. Hinton, N. J. et J. Ohala, Cambridge University Press : Cambridge, p. 325–347.
- [51] Roustan, B. 2012, *Étude de la coordination gestes manuels/parole dans le cadre de la désignation*, thèse de doctorat, Université de Grenoble.
- [52] Sapir, E. 1929, «A study in phonetic symbolism», *Journal of Experimental Psychology*, vol. 12, p. 225–39.
- [53] Saussure, F. d. 1916/1983, *Course in general linguistics*, Duckworth : London.
- [54] Senft, G. 2004, «Aspects of spatial deixis in Kilivila», dans *Deixis and Demonstratives in Oceanic Languages*, édité par G. Senft, Canberra : Australian National University, p. 59–80.
- [55] Simon, O., J. Mangin, L. Cohen, D. Le Bihan et S. Dehaene. 2002, «Topographical layout of hand, eye, calculation, and language-related areas in the human parietal lobe», *Neurosciences*, vol. 33, p. 475–487.
- [56] Tomasello, M., M. Carpenter, J. Call, T. Behne et H. Moll. 2005, «Understanding and sharing intentions : The origins of cultural cognition», *Behavioral and Brain Sciences*, vol. 28, n° 5, p. 675–691.
- [57] Traunmüller, H. 2000, «Sound symbolism in deictic words», dans *Tongues and Texts Unlimited : Studies in Honour of Tore Janson, Stockholms Universitet*, édité par H. Aili et A. Trampe, p. 213–234.
- [58] Ultan, R. 1978, «Size sound symbolism», dans *Universals of Human Language. Vol. 2, Phonology*, édité par J. Greenberg, C. Ferguson et E. Moravcsik, Stanford : Stanford University Press, p. 525–568.
- [59] Ungerleider, L. et J. Haxby. 1994, «"What" and "where" in the human brain», *Current Opinion in Neurobiology*, vol. 4, p. 157–165.

- [60] Ungerleider, L. et M. Mishkin. 1982, «Two cortical visual systems», dans *Analysis of Visual Behavior*, édité par D. Ingle, M. Goodale et R. Mansfield, Cambridge MA : MIT Press, p. 549–586.
- [61] Volterra, V. et M. Caselli. 1986, «First stages of language acquisition through two modalities in deaf and hearing children», *Journal of Neurological Sciences*, vol. 5, p. 109–115.
- [62] Volterra, V., M. Caselli, O. Capirci et E. Pizzuto. 2005, «Gesture and the emergence and development of language», dans *Elizabeth Bates : A Festschrift*, édité par M. Tomasello et D. Slobin, Mahwah, N.J. : Lawrence Erlbaum Associates, p. 3–40.
- [63] Vygotsky, L. 1998, *The collected works of L.S. Vygotsky*, New York, Plenum.
- [64] Weissenborn, J. et W. Klein. 1982, *Here and There : Cross-Linguistic Studies on Deixis and Demonstration*, John Benjamins Publishing, Pragmatics and Beyond, III : 2-3.
- [65] Wilkins, D. 2003, «Why pointing with the index finger is not a universal (in sociocultural and semiotic terms)», dans *Pointing : Where Language, Culture, and Cognition meet*, édité par S. Kita, Mahwah, NJ, US : Lawrence Erlbaum Associates Publishers, p. 171–215.
- [66] Woodworth, N. 1991, «Sound symbolism in proximal and distal forms», *Linguistics*, vol. 29, p. 273–299.
- [67] Wundt, W. 1912, *Elemente der Völkerpsychologie. Grundlinien einer psychologischen Entwicklungsgeschichte der Menschheit.*, Leipzig : Barth.

Interaction geste, parole, langage et encodage de la distance chez l'adulte

"An experimental study of speech/gesture interactions & distance encoding"

Gonseth et collab. [13] - **Speech Communication** - 2013

L'objectif du présent chapitre est de tester expérimentalement, et dans un cadre contrôlé, l'interaction entre le geste et la parole, que nous considérons comme deux modalités langagières à part entière. Dans cette optique, l'étude expérimentale présentée ci-après vise à déterminer la nature de l'interaction geste/parole et de l'encodage de la distance, par le biais d'un protocole contrôlé, dans lequel les participants auront pour tâche de désigner une cible, située à différentes distances, via la production de pointages manuels et/ou vocaux.

Notre première hypothèse concerne la nature de l'interaction geste/parole : nous souhaitons mettre en évidence une interaction structurelle et bidirectionnelle, au sens où le **contenu**, la **forme** et la **quantité d'informations** véhiculée par le pointage vocal pourraient être influencés par le **contenu**, la **forme** et la **quantité d'informations** véhiculée par le pointage manuel, et **vice versa**. En d'autres termes, les productions de chaque modalité seraient générées relativement aux productions de l'autre modalité et non indépendamment les unes des autres. Cette influence, de nature complémentaire plutôt que redondante, devrait se traduire par une inhibition du pointage, en termes de propriétés physiques (i.e. articulatoires, acoustiques et cinématiques), lorsque deux modalités plutôt qu'une seule sont disponibles pour transmettre l'information (et parallèlement, par un renforcement du pointage lorsqu'une seule modalité est disponible pour transmettre le message).

Notre seconde hypothèse concerne l'encodage déictique de la distance via les propriétés lexicales, acoustiques, articulatoires et cinématiques du pointage : nous souhaitons montrer un lien fort entre les gestes manuels, les gestes vocaux et les structures linguistiques qui y sont associées ; autrement dit, l'encodage lexical de la distance pourrait être associé à une tendance à produire des gestes plus larges pour désigner des objets éloignés. Ces comportements moteurs de bas niveau auraient progressivement été intégrés à la phonologie des langues, sous la forme d'une distinction entre termes proximaux/distaux.

1 Expérience 1

L'objectif de notre expérience est de caractériser au mieux la relation geste/parole, en déterminant de quelle manière leur contenu informatif, lié à la distance de la cible, est distribué sur chaque modalité. De fait, il nous paraît essentiel d'étudier cette interaction lorsque les deux modalités sont porteuses d'une même fonction communicative. Les participants seront donc testés en situation de production de pointage manuel et/ou vocal, vers des cibles situées à différentes distances. En lien avec notre hypothèse générale, i.e. un couplage geste/parole dans le langage, nous souhaitons montrer que la production des signaux vocaux dépend fortement de celle des signaux manuels qui y sont associés, et inversement, c'est-à-dire une influence bidirectionnelle entre ces deux systèmes. Nous suggérons que l'interaction geste/parole soit de nature structurale et complémentaire, au sens d'une "coopération" inter-modalités : des productions inhibées en situation bimodale, l'information étant partagée sur l'ensemble des modalités disponibles. Par ailleurs, nous souhaitons mettre en évidence un encodage moteur de la distance, soit des variations phonétiques et cinématiques en fonction de la distance de l'objet pointé, indépendamment de la lexicalité du terme utilisé ou du type de voyelle qu'il contient : des gestes vocaux et manuels plus larges pour désigner des objets plus éloignés, soit une hyperarticulation vocale, potentiellement associée à une augmentation du premier formant, ainsi qu'une amplitude plus importante du geste manuel.

Ce travail de thèse s'inscrit dans la continuité d'une étude préliminaire réalisée par Etty Reboud au sein du Gipsa-Lab, dans le cadre de son Master Recherche. L'élaboration du protocole, l'enregistrement des données, ainsi qu'une première série d'analyses ont été pris en charge par E. Reboud, A. Vilain et C. Vilain [26].

1.1 Méthodologie

1.1.1 Participants

Vingt-quatre participants (dont neuf hommes et 15 femmes), âgés de 18 à 57 ans (Moyenne (M) = 31.08, Ecart-type (SD) = 12.05) ont participé, bénévolement, à cette étude. Tous étaient de langue maternelle française, sans déficit visuel non corrigé ou trouble moteur particulier, et droitiers. Recrutés majoritairement par diffusion d'e-mails, les participants étaient pour la plupart étudiants à l'Université de Grenoble.

1.1.2 Matériel

Les participants sont enregistrés individuellement en chambre sourde et sont assis à une table, face à trois diodes lumineuses (LED), placées à trois distances différentes d'eux (voir Figure 3.1) :

- D1 : située à 55 cm, dans l'espace **péripersonnel**, à portée de main ;
- D2 : située à 140 cm, dans l'espace **extrapersonnel proche**, hors de portée ;
- D3 : située à 425 cm (soit la distance maximale autorisée par les dimensions de la chambre sourde), dans l'espace **extrapersonnel lointain**, toujours hors de portée.

1. Expérience 1

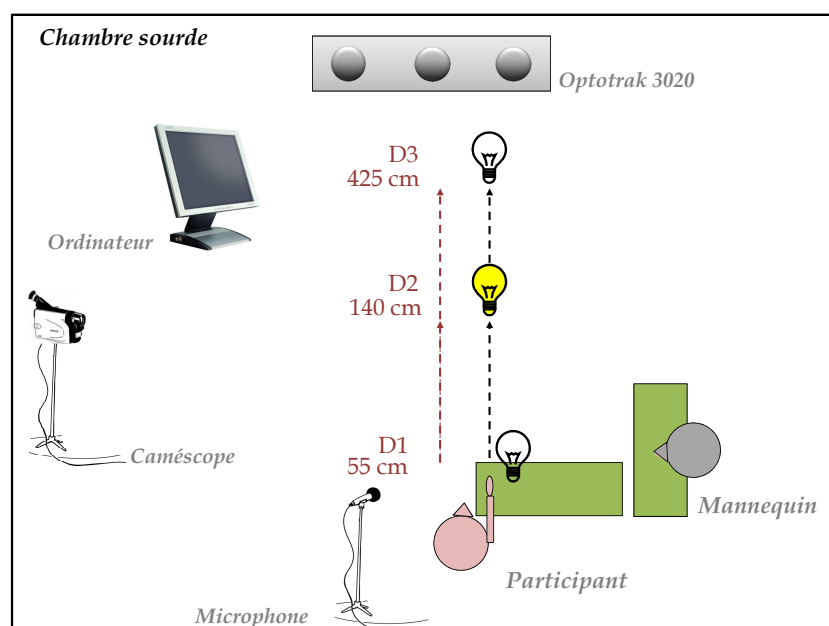


FIGURE 3.1 – Dispositif expérimental

Ces trois distances vont permettre d'étudier une distinction entre deux objets placés dans l'espace péripersonnel et extrapersonnel, mais également entre deux objets situés dans le même espace extrapersonnel. Les deux LED situées hors de portée sont alignées l'une par rapport à l'autre mais également par rapport à l'axe œil-index du participant, qui les ajuste lui-même sur une même ligne ; de cette façon, la direction du pointage manuel ne peut pas les distinguer, et un encodage particulier de la distance, notamment via les propriétés cinématiques du geste, devrait s'avérer nécessaire.

Les instructions sont présentées via un écran d'ordinateur, situé sur la gauche des participants et dont l'interface est contrôlée par le logiciel Presentation¹. Ce logiciel permet notamment l'enregistrement d'un fichier .log, soit un historique séquentiel des événements présentés. Par ailleurs, un microphone AKG C1000S ainsi que le système opto-électronique de capture de mouvements Optotrak 3020 sont utilisés pour l'enregistrement des productions orales et gestuelles (ces deux outils sont définis plus précisément en Section 1.2). Un caméscope numérique permet d'enregistrer l'ensemble de la scène afin de contrôler ultérieurement la validité des tâches effectuées.

1.1.3 Procédure

La tâche des participants, dont la main droite est initialement placée en position de repos sur la table, est de désigner la LED allumée au mannequin situé sur leur droite (à 100 cm et 90° par rapport à la direction de leur regard). L'utilisation d'un tel mannequin est justifiée par le

1. <http://www.neurobs.com>

besoin de créer une situation d'interaction homogène pour chacun des participants, c'est-à-dire non soumise à la variabilité induite par un interlocuteur humain. A la différence d'un interlocuteur réel, qui pourrait avoir des réactions inappropriées, répercutées sur le comportement des participants (e.g. fixer la mauvaise cible), le mannequin aura toujours une même "réaction", adéquate, celle que lui attribuera le participant.

Deux conditions expérimentales sont testées : une situation unimodale (**Parole Seule**) et une situation bimodale (**Parole+Geste**).

- Parole Seule : En situation de parole seule, les participants doivent nommer la LED allumée à l'aide de l'item présenté à l'écran. Il leur est ici formellement interdit d'utiliser leurs mains, qu'ils peuvent, au choix, laisser en position de repos, sur la table ou sur leurs genoux ;
- Parole+Geste : En situation de parole+geste, les participants doivent simultanément nommer et pointer, par un geste de l'index de la main droite, la LED allumée.

Les items à utiliser pour désigner oralement la cible, présentés au Tableau 3.1, sont cinq mono-syllabes, de structure CV, insérés dans la phrase porteuse « c'est [+ item] ».

	Déictiques	Pseudo-Mots
<i>I1</i>	ça	
<i>I2</i>	là	
<i>I3</i>		[pi]
<i>I4</i>		[pa]
<i>I5</i>		[pe]

Tableau 3.1 – Liste des items

Nous avons choisi d'utiliser à la fois des mots déictiques et des pseudo-mots, et de tester, dans ces derniers, plusieurs voyelles différentes. Il est en effet possible que seuls les termes déictiques puissent endosser une fonction déictique, ou qu'à l'inverse, seuls les pseudo-mots puissent varier, les termes déictiques étant potentiellement "figés" lexicalement ; de la même façon, on pourrait s'attendre à des variations plus importantes pour les voyelles [a] et [i], qui sont bien représentées dans les déictiques du français, ou inversement, des variations plus importantes pour la voyelle [e], qui à notre connaissance n'est pas utilisée dans les déictiques du français ; enfin on pourrait également envisager des variations plus fortes pour la voyelle [a] qui, de façon générale, est plus variable que les autres en termes de réalisation phonétique.

Les termes déictiques sélectionnés, *ça* et *là*, sont relativement neutres, dans le sens où ils ne véhiculent pas systématiquement une information proximale/distale, mais peuvent être utilisés de manière relativement indépendante de la distance du référent. Les participants ont pour consigne d'utiliser chacun des items, mots et pseudo-mots, pour désigner la cible, c'est-à-dire de leur attribuer une fonction déictique. Si l'encodage lexical de la distance est effectivement associé à un comportement moteur, nous devrions observer des variations articulatoires et/ou acoustiques en fonction de la distance de l'objet pointé pour chacun des items, indépendamment donc de la nature de la voyelle qu'ils contiennent ou du contexte, lexical ou non, dans lequel elle est réalisée.

1. Expérience 1

Le paradigme utilisé est un paradigme en blocs : l'expérience se déroule en deux parties successives (1 : Parole Seule, 2 : Parole+Geste, l'ordre étant contrebalancé en fonction du participant), précédées par une courte session d'entraînement (2×4 essais, avec possibilité de recommencer l'entraînement si nécessaire). Chaque essai débute par l'allumage, aléatoire, de l'une des trois diodes et l'affichage simultané de l'un des items à l'écran (toujours de façon aléatoire). La couleur de l'item change après trois secondes, signal de départ pour le participant, ce qui lui permet simplement de mémoriser l'item et de recentrer son attention, de l'ordinateur vers la cible.

La durée de l'expérience est d'environ 45 minutes (comprenant l'installation, les consignes ainsi qu'un court debriefing), pendant lesquelles les participants sont confrontés à 300 essais (= 3 Distances $\times$ 2 Conditions $\times$ 5 Items [= 2 mots + 3 pseudo-mots] $\times$ 10 Itérations).

Le plan d'expérience est un plan intra-sujets, de manière à ce que chaque participant soit confronté à chaque modalité de chaque variable (voir Tableau 3.2 pour un récapitulatif des variables indépendantes - ou facteurs d'intérêt - utilisées).

Variables Indépendantes	Modalités	
Distance	55 cm - 140 cm - 425 cm	
Condition	Parole Seule - Parole+Geste	
Items	ça - là	[pi] - [pa] - [pe]
↓		
Lexicalité	Mots	Pseudo-mots

Tableau 3.2 – Récapitulatif des variables indépendantes

1.2 Acquisition des données

L'enregistrement du signal de parole est réalisé par un microphone AKG C1000S, connecté à un enregistreur Marantz-PMD 670 (en voie 1). Les niveaux sont réglés pour chaque participant avant le début de la passation et restent invariables au cours d'une même session. De plus, un bip de synchronisation, généré au début de chaque nouvel essai par un boîtier électronique, développé au sein du Gipsa-Lab et piloté par le port parallèle de l'ordinateur présentant les stimuli, est également enregistré (en voie 2).

L'enregistrement des productions gestuelles est réalisé par un système opto-électronique de capture de mouvements, l'Optotrak 3020², constitué de diodes émettrices infra-rouge. Ce dispositif permet de suivre la trajectoire des diodes au cours du temps, soit toutes les 5 ms, avec une période d'échantillonnage de 200 ms.

2. <http://www.ndigital.com/optotrak-techspecs.php>

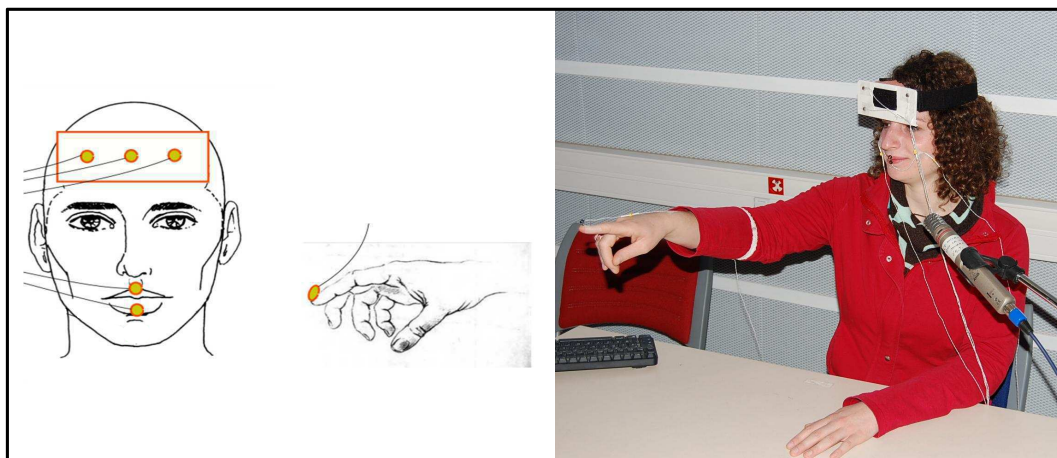


FIGURE 3.2 – Position des diodes émettrices infra-rouge (© C. Gonseth)

Six diodes émettrices, illustrées Figure 3.2, sont mises en place :

- Trois diodes de référence placées dans un plan rigide, au niveau du front du participant³ ;
- Deux diodes placées au centre des contours des lèvres supérieure et inférieure (notées D_{LS} et D_{LI}) ;
- Une diode placée sur l'ongle de l'index droit (notée D_I), collée sur une équerre métallique.

Par ailleurs, une septième diode, fixée sur un pied métallique, est synchronisée au signal sonore, afin de repérer le début de chaque nouvel essai.

1.3 Mesures

Les mesures présentées ci-après concernent 17 des 24 participants, des problèmes techniques, liés notamment à l'utilisation du système opto-électronique et à l'enregistrement des productions, étant survenus pour sept participants.

1.3.1 Données acoustiques

L'ensemble des productions orales est tout d'abord segmenté sous Audacity⁴, afin de récupérer les productions pour chacun des blocs (Parole Seule, Parole+Geste). Ces blocs sont ensuite segmentés de façon automatique par un programme Matlab⁵, qui permet de repérer les signaux sonores annonçant le début de chaque nouvel essai. Une fois segmentés, les fichiers .wav sont créés et nommés d'après la condition, la distance et l'item utilisé. Ces nouveaux fichiers .wav sont ensuite "nettoyés" par un filtre passe-haut (fréquence de coupure fixée à 80 Hz), afin de ne garder que le signal de parole. Finalement, le programme Matlab utilisé ici permet la détection automatique de la seconde voyelle de l'énoncé, soit la voyelle contenue dans l'item ([a], [i] ou [e],

3. Notons que les trois diodes de référence n'ont pas été exploitées, n'étant pas nécessaires pour les mesures effectuées.

4. audacity.sourceforge.net

5. www.mathworks.fr

1. Expérience 1

la première voyelle correspondant au [ɛ] de « c'est », à partir des signaux d'énergie (i.e. repérage du second pic d'énergie), et d'en annoter le début, le milieu et la fin. Ces frontières temporelles peuvent être corrigées manuellement si besoin est, et sont enregistrées dans un fichier .txtGrid, utilisé par le logiciel Praat⁶ afin de mesurer les paramètres acoustiques suivants⁷ :

- *Fréquence Fondamentale :*

La Fréquence fondamentale (F0) est la fréquence de base de vibration des cordes vocales et caractérise donc les parties voisées du signal de parole. Elle est notamment liée à la sensation de hauteur de la voix (aigüe ou grave) et participe ainsi aux variations prosodiques du signal de parole (accentuation, intonation). Afin d'extraire les valeurs de F0 en Hertz (Hz), Praat utilise une méthode temporelle dite d'auto-corrélation.

- *Intensité :*

L'intensité correspond à la variation de l'amplitude de signal de parole. Elle dépend donc de l'amplitude des vibrations (déterminée par l'importance du flux d'air traversant la glotte), mais également du degré de fermeture de la glotte et de la tension des cordes vocales. Elle fournit une mesure de la force sonore de la voix, faible ou forte (i.e. si la fermeture est incomplète et qu'une partie du flux d'air passe sans être transformée en vibration, le son devient plus faible) et joue ainsi un rôle dans les variations prosodiques du signal de parole. L'intensité est ici calculée sous Praat en décibels (dB).

- *Premier & Second Formants :*

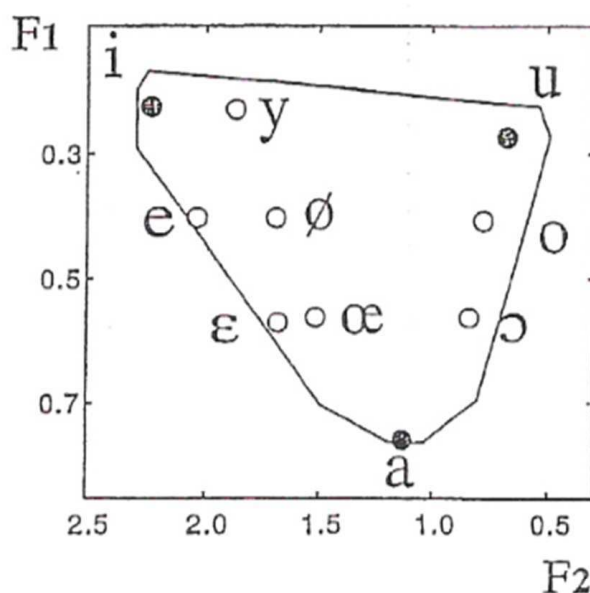


FIGURE 3.3 – Valeurs formantiques (F1, F2 en kHz) des voyelles du français

6. www.fon.hum.uva.nl/praat/

7. Le lecteur trouvera une description détaillée des méthodes de calcul utilisées sous Praat dans l'ouvrage de Boersma [4].

Les voyelles sont caractérisées par des zones d'harmoniques renforcées, appelées formants. Il s'agit de pics d'amplitude dans le spectre d'un son (ou pic formantique). Les valeurs de formants dépendent de la taille et de la configuration des cavités bucco-pharyngales. Quatre dimensions permettent de modifier la forme ou l'accès à ces cavités : le degré d'aperture de la mandibule, la position de la langue, la position des lèvres et la position du velum (autorisant ou non le passage de l'air dans les fosses nasales). De cette façon, chaque voyelle produite occasionne des valeurs de formants spécifiques, illustrées Figure 3.3 ci-dessus.

Le premier formant (F1) est corrélé au degré d'aperture de la mandibule et à la hauteur de la langue (valeur élevée pour une voyelle ouverte, basse). Le deuxième formant (F2) est quant à lui corrélé à la position avant ou arrière de la langue (valeur élevée pour une voyelle antérieure). Ils sont calculés ici à l'aide du logiciel Praat par une analyse LPC (*Linear Prediction Coding* : codage prédictif linéaire), dont l'algorithme est celui de Burg (e.g. Childers [8], Press et collab. [25]).

Les valeurs moyennes de ces différents paramètres ont été calculées entre le début et la fin de la voyelle, mais nous retiendrons ici les valeurs relevées à l'instant t correspondant au milieu de la voyelle, où les formants sont les plus stables⁸ (e.g. Fukushi [9], Lehist et Peterson [18], Vasilescu et collab. [31]).

1.3.2 Données articulatoires

Les trajectoires tridimensionnelles des diodes infra-rouge placées sur les lèvres sont segmentées (par bloc puis par essai, à partir de la diode de synchronisation) puis mises en correspondance avec les données acoustiques (sur la base du fichier .txtGrid). Un programme Matlab permet ensuite le calcul automatique du maximum d'ouverture des lèvres, au niveau de la seconde voyelle, soit la distance euclidienne entre les deux diodes (haute et basse) situées sur les lèvres ($D_{LS} - D_{LI}$). Le programme permet également la visualisation de l'ensemble acoustique/articulatoire, ainsi qu'une correction manuelle en cas de détection erronée.

L'intérêt de l'Optotrak est qu'il fournit des coordonnées millimétrées (mais relativement imprécises, le diamètre de la diode étant de l'ordre de plusieurs millimètres) : l'ouverture maximale est alors une valeur interprétable directement. Par ailleurs, ce système a l'avantage de ne pas être invasif, à la différence notamment de l'Articulographie ElectroMagnétique (EMA), et perturbe donc dans une moindre mesure les productions orales.

1.3.3 Données manuelles

La trajectoire de la diode émettrice infra-rouge positionnée sur l'index est segmentée (par bloc puis par essai, à partir de la diode de synchronisation) puis annotée selon quatre points :

- *Onset* : départ de la position de repos ;
- *Apex* : extension maximale de l'index ;

8. Lehist et Peterson [18] décrivent trois phases à la structure formantique des voyelles : une partie **stable** et deux phases de transition formantique (initiale et finale) ; dans une structure CVC, la transition initiale correspond au relâchement de la consonne au début de la syllabe et la transition finale correspond à la fermeture consonantique à la fin de la syllabe.

1. Expérience 1

- *Retour* : départ de la position d'apex ;
- *Offset* : retour en position de repos.

L'annotation se base sur la détection préalable de deux pics de vitesse, correspondant à l'aller et au retour du geste de pointage (voir Figure 3.4).

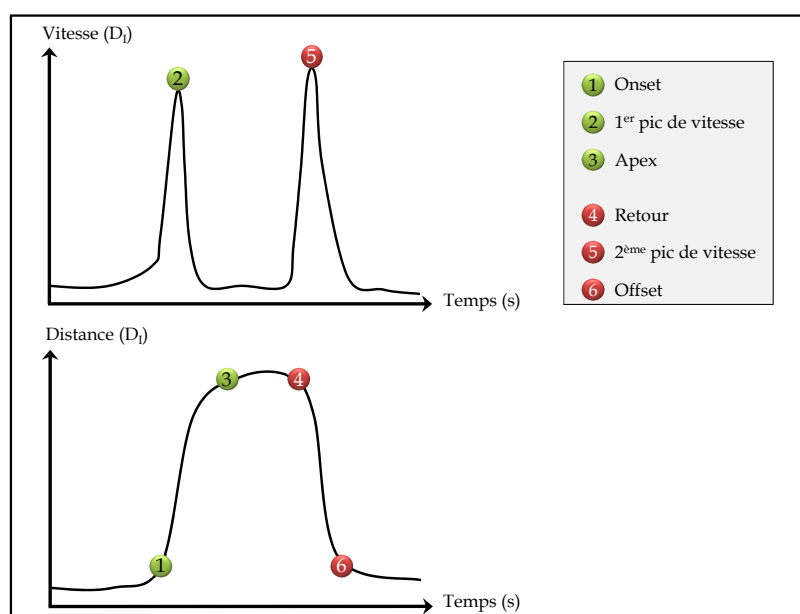


FIGURE 3.4 – Annotation des paramètres cinématiques

L'onset est ainsi défini comme le point où la vitesse atteint 10% de sa valeur maximale avant le premier pic, et l'apex comme le point où la vitesse atteint 10% de sa valeur maximale après le premier pic. De la même façon, le retour est défini comme le point où la vitesse atteint 10% de sa valeur maximale avant le second pic, et l'offset comme le point où la vitesse atteint 10% de sa valeur maximale après le second pic.

A partir de ces différents points d'intérêt, les paramètres cinématiques suivants ont été calculés :

- *Durée totale* (s), définie par l'intervalle [Offset - Onset] ;
- *Pic de vitesse* (mm/s) du premier stroke, soit au niveau de l'intervalle [Apex - Onset] ;
- *Amplitude* (mm) du premier stroke, soit la distance maximale entre la position de l'index à l'onset et la position de l'index à l'apex.

1.3.4 Récapitulatif des variables dépendantes

Le Tableau 3.3, présenté ci-après, propose un récapitulatif des variables dépendantes mesurées.

Variables Dépendantes	Mesures
Données acoustiques	F0 (Hz) - Intensité (dB) - F1 (Hz) - F2 (Hz)
Données articulatoires	Ouverture des lèvres (mm)
Données cinématiques	Durée (s) - Pic de vitesse (mm/s) - Amplitude (mm)

Tableau 3.3 – Récapitulatif des variables dépendantes

1.4 Analyses statistiques & prédictions

Certains essais ont dû être exclus de l'analyse statistique : les erreurs (d'item) ou oublis (d'item et/ou de pointage) mais également les données manquantes liées à des problèmes techniques (masquage des diodes émettrices infra-rouge notamment), soit 0.9% des données acoustiques, 12.25% des données articulatoires et 7.68% des données manuelles. Ce pourcentage d'essais exclus est relativement important pour les données opto-électroniques et correspond aux données articulatoires de deux participants et aux données manuelles d'un participant, exclus de l'analyse statistique finale.

Les analyses statistiques acoustiques portent sur les 17 participants retenus, les analyses articulatoires portent sur 15 participants et les analyses cinématiques sur 16 participants.

Plus précisément, nous avons réalisé des analyses de la variance⁹ (ANOVA, pour *ANalysis Of VAriance*) pour plans à mesures répétées et facteurs intra-sujets :

- Acoustique : $S_{17}[Distance_3 \times Condition_2 \times Item_5] \ \& \ S_{17}[Distance_3 \times Condition_2 \times Lexicalité_2]$
- Articulatoire : $S_{15}[Distance_3 \times Condition_2 \times Item_5] \ \& \ S_{15}[Distance_3 \times Condition_2 \times Lexicalité_2]$
- Cinématique : $S_{16}[Distance_3]$

Les analyses statistiques portent ici sur la médiane, soit la plus petite valeur pour laquelle l'effectif n cumulé dépasse $\frac{n}{2}$ et qui partage ainsi la suite ordonnée des observations en classes de même effectif : la médiane est donc l'observation $x\left(\frac{n+1}{2}\right)$, ou $\frac{x\left(\frac{n}{2}\right) + x\left(\frac{n}{2} + 1\right)}{2}$ dans le cas d'un effectif pair. Elle présente l'avantage d'être plus robuste que la moyenne en présence de valeurs extrêmes et représente donc un meilleur indicateur de position centrale que celle-ci.

L'ANOVA à mesures répétées permet de contrôler les différences individuelles inter-sujets, chaque participant devenant son propre contrôle à travers les différents traitements expérimentaux, sous réserve de sphéricité et de normalité. Dans le cas où ces deux postulats ne sont pas respectés, le risque d'erreur de type I (i.e. l'acceptation erronée de l'hypothèse nulle H_0) est plus fort, il est donc conseillé d'appliquer des transformations et corrections aux données (néanmoins le test est relativement robuste et en général peu affecté par la violation de ces deux pré-requis).

Le test de Mauchly (Mauchly [22]) permet de vérifier la sphéricité des données, c'est-à-dire que les variances et covariances entre les différents niveaux des variables indépendantes sont homogènes (H_0 , rejetée si le test est significatif). Dans le cas contraire, le correctif de Greenhouse-

9. Dont le modèle général s'écrit : $Y_{ijk} = \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}$, où α_i correspond à l'effet du premier facteur, β_j à celui du second facteur, γ_{ij} à leur interaction et ε_{ijk} à l'erreur aléatoire.

1. Expérience 1

Geisser (Greenhouse et Geisser [14]), soit un ajustement des degrés de liberté $(dl)^{10}$, a été appliqué $(dl \times \varepsilon)$. Notons toutefois que si le premier degré de liberté $dl1$ est de 1, aucune correction n'est nécessaire. D'autre part, le test de Shapiro-Wilk (Shapiro [30]) permet de vérifier que l'échantillon considéré est issu d'une population normalement distribuée (H_0 , rejetée si le test est significatif). Dans le cas contraire, une transformation logarithmique (où $Y' = \ln(Y) + \text{une constante si nécessaire}$, i.e. en cas de valeurs négatives) nous a permis de stabiliser les variances et d'obtenir de nouvelles variables transformées normalisées (Y').

Concernant les analyses articulatoires et acoustiques, nous avons testé quatre effets principaux : celui de la Distance, de la Condition, de l'Item et de la Lexicalité (i.e. mots vs pseudo-mots). Nous nous intéresserons exclusivement aux effets de la Distance et de la Condition, au niveau intra-catégoriel. Les effets principaux de l'Item et de la Lexicalité ne sont pas ici pertinents : chaque item induit nécessairement des variations acoustiques et articulatoires qui lui sont propres (liées au type de voyelle et au phénomène de coarticulation, selon lequel la réalisation sonore d'un phonème est influencée par la production des phonèmes précédents et suivants). Concernant les analyses cinématiques, nous avons testé l'effet principal de la Distance.

Nos prédictions sont les suivantes :

- Effet principal de la distance

Un encodage moteur de la distance aurait pour conséquence des valeurs articulatoires, acoustiques et cinématiques plus élevées pour les distances les plus éloignées. Plus précisément, nous attendons une ouverture des lèvres ainsi que des valeurs de F1 et d'intensité plus importantes pour les distances $D2$ et $D3$, situées dans l'espace extrapersonnel. En revanche, aucun effet n'est attendu sur les valeurs de F2, liées à la protrusion de la langue, et aucune hypothèse particulière n'est formulée quant aux valeurs de F0. Au niveau du pointage manuel, l'encodage se traduirait par une amplitude plus importante du premier stroke (i.e. aller du geste) pour les distances $D2$ et $D3$. Par ailleurs, les données de la littérature suggèrent que le pic de vitesse et la durée du geste augmentent lorsque la cible est éloignée (voir notamment les travaux de Gentilucci et collab. [10] et de Bonfiglioli et collab. [5] sur le geste d'atteinte et de saisie et ceux de Rochet-Capellan [27], sur le geste de pointage).

- Effet principal de la condition

Une coopération entre les systèmes de production de la parole et du geste aurait pour conséquence un pointage inhibé en condition bimodale relativement à la condition unimodale (i.e. un pointage renforcé en condition unimodale, relativement à la condition bimodale). En d'autres termes, l'ajout d'une modalité permettrait de répartir l'information sur l'ensemble des modalités disponibles, tandis que l'absence d'une modalité serait compensée par une amplification des divers paramètres acoustiques et articulatoires. Nous attendons une ouverture des lèvres ainsi que des valeurs de F1 plus élevées en condition unimodale par rapport à la condition bimodale. De nouveau, nous n'attendons aucun effet sur les valeurs de F2 et aucune hypothèse particulière n'est formulée quant aux valeurs d'intensité et de F0.

10. Pour une ANOVA à mesures répétées, $dl1$ = Nombre de modalités -1 ; $dl2$ = Nombre de Participants $-1 \times dl1$.

- Effet d'interaction

Concernant les analyses acoustiques et articulatoires, l'ANOVA permet également de tester cinq effets d'interaction (trois interactions dites simples $D * C$, $D * I$, $C * I$, $D * L$, $C * L$ et deux interactions triples $D * C * I$ et $D * C * L$). Les interactions impliquant Item et Lexicalité permettront d'évaluer l'impact du type de voyelle et celui des contraintes lexicales. Un effet de la distance lié à un encodage moteur de bas niveau devrait s'observer indépendamment de la catégorie vocalique considérée et de la lexicalité de l'item utilisé. De fait, les interactions $D * I$ et $D * L$ ne devraient pas être significatives. Par ailleurs, différentes prédictions peuvent être formulées quant à l'interaction Item ou Lexique et Condition : une influence du geste sur la parole liée à l'intention communicative du locuteur, soit une influence de haut niveau, se traduirait par des effets d'interaction $C * I$ et $C * L$ significatifs (i.e. un effet significatif pour les items lexicaux uniquement) ; en revanche une influence geste/parole de plus bas niveau se traduirait par une absence d'interaction $C * I$ et $C * L$ (i.e. un effet de la condition identique pour chacun des items).

Notons que si l'ANOVA permet de mettre en évidence une différence significative entre les modalités d'un facteur, elle ne permet pas en revanche de déterminer précisément quelles modalités diffèrent l'une de l'autre. Les tests de comparaisons *post-hoc* s'avèrent donc nécessaires lorsqu'un effet significatif est observé pour les facteurs à trois modalités ou plus (ici la Distance et l'Item). Nous avons choisi le test *post-hoc* de Bonferroni-Dunn (Abdi [1], Winer et collab. [33]), qui permet de calculer les statistiques t de Student¹¹, en donnant un résultat au seuil α : les tests individuels sont réalisés au seuil *par comparaison* (PC) de la façon suivante : $\alpha_{PC} = \frac{\alpha}{c}$, où c est le nombre de comparaisons à effectuer. De manière générale, $c = \frac{k(k-1)}{2}$, où k est le nombre de modalités du facteur considéré, soit $c = 3$ pour le facteur Distance, $c = 10$ pour le facteur Item. Ce test est dit conservatif, dans la mesure où il présente peu de risques de commettre une erreur de type I (conclure sur des différences non réelles), mais davantage de risques de commettre une erreur de type II (ne pas mettre en évidence une différence qui existe).

11. Pour deux modalités a et b , $t = \frac{\bar{x}_a - \bar{x}_b}{E}$, où $\bar{x}_a$ et $\bar{x}_b$ représentent les moyennes observées dans les deux modalités, et E l'erreur type, calculée à partir du carré moyen (CM) intra-groupe de l'ANOVA ($E^2 = CM_{INTRA} \left(\frac{1}{n_a} + \frac{1}{n_b} \right)$).

1. Expérience 1

Le glossaire statistique présenté ci-dessous liste diverses abréviations utilisées par la suite, relatives aux analyses et résultats statistiques.

GLOSSAIRE STATISTIQUE

- C = Condition
→ $C1$ = Parole Seule - $C2$ = Parole+Geste
- D = Distance
→ $D1$ - $D2$ - $D3$ = 55 - 140 - 425 cm
- I = Item
→ $I1$ - $I2$ - $I3$ - $I4$ - $I5$ = ça - là - [pi] - [pa] - [pe]
- L = Lexicalité
→ $L1$ - $L2$ = mots - pseudo-mots
- ES = Erreur Standard ^a ;
- p = degré de significativité ;
- α = fixé à priori ^b, au minimum $<.05$;
- NS = Non Significatif ;
- F = Fisher ;
- ε = epsilon de Greenhouse-Geisser ;

^a. Estimation de l'écart-type lié à l'erreur de mesure. L'Erreur Standard de la moyenne d'une population X , de n individus, est égale à $\sqrt{\frac{Var(X)}{n}}$.

^b. Seuil de risque de conclusion erronée acceptable.

1.5 Résultats de l'Expérience 1

Cette section présente un descriptif des effets observés pour chaque facteur considéré. Les implications théoriques qui en découlent seront discutées dans la Section 1.6.

Dans la suite de ce manuscrit, les graphiques associés aux analyses à mesures répétées représentent la moyenne des observations.

1.5.1 Propriétés articulatoires du pointage vocal

La figure 3.5 montre l'effet de la Condition et de la Distance sur l'ouverture des lèvres (*OL*). Les voyelles ouvertes [a] et fermées [i, e] sont représentées séparément, dans un souci de lisibilité (les analyses ayant été conduites sur l'ensemble des items).

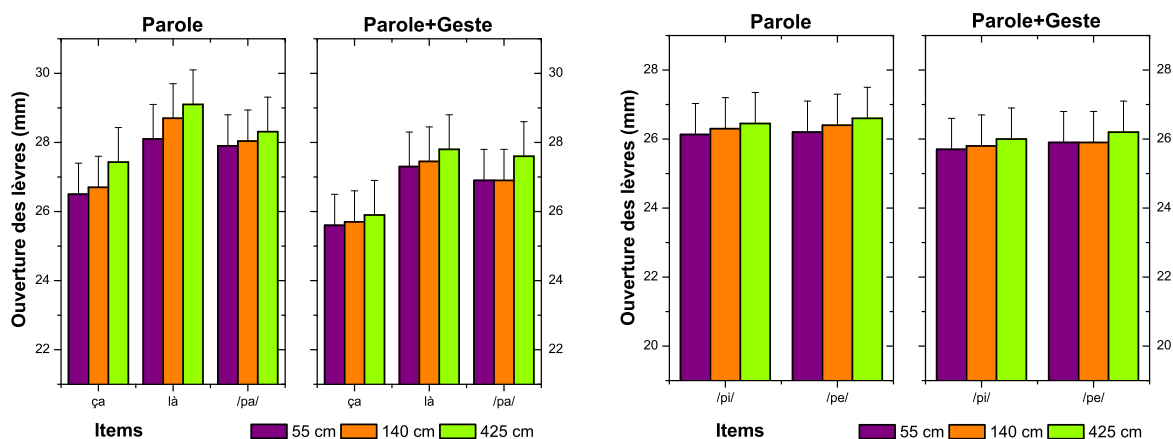


FIGURE 3.5 – *OL* (mm) en fonction de la Condition et de la Distance pour les cinq items

Effets principaux Les résultats montrent un effet significatif de la Condition ($F(1, 14) = 8.25$, $p < .05$) : l'ouverture des lèvres est plus grande en situation unimodale. De plus, on observe un effet significatif de la Distance ($\varepsilon = 0.58$, $p < .05$) : plus la distance entre le locuteur et la cible est grande, plus l'ouverture des lèvres l'est également. Le test *post-hoc* met en évidence des différences significatives ($p < .01$) pour les différentes modalités de la Distance prises deux à deux.

Effets d'interaction Les interactions $D * C$ et $D * I$, ainsi que l'interaction triple $D * C * I$, ne sont pas significatives (respectivement $\varepsilon = 0.70$, $p = .31$, *NS*; $\varepsilon = 0.48$, $p = .35$, *NS*; $\varepsilon = 0.49$, $p = .18$, *NS*). Il en va de même pour les interactions $D * L$ et l'interaction triple $D * C * L$ (respectivement $\varepsilon = 0.76$, $p = .28$, *NS*; $\varepsilon = 0.86$, $p = .21$, *NS*). En revanche l'interaction $I * C$ est significative ($\varepsilon = 0.49$, $p = .01$) : le test *post-hoc* montre que l'effet principal de la Condition dépend de la voyelle produite, il est significatif pour les voyelles ouvertes [a] uniquement ($p < .01$). De la même façon, l'interaction $L * C$ est significative ($F(1, 14) = 7.24$, $p < .05$) : le test *post-hoc* montre que l'effet principal de la Condition n'est significatif que pour les mots ($p < .01$).

1. Expérience 1

A retenir ! Le Tableau 3.4 propose un bref aperçu des principales différences observées sur les données articulatoires.

Ouverture des lèvres	
Distance	$D1 < D2 < D3$
Condition	$C1 > C2$
Interaction	$C * I - L * I$
	→ Effet de la condition significatif pour les voyelles ouvertes uniquement
	→ Effet de la condition significatif pour les mots uniquement

Tableau 3.4 – Principaux résultats - données articulatoires

1.5.2 Propriétés acoustiques du pointage vocal

Premier formant (F1) La figure 3.6 représente l'effet de la Condition et de la Distance sur les valeurs de F1. Les voyelles ouvertes et fermées sont représentées séparément, toujours par souci de clarté.

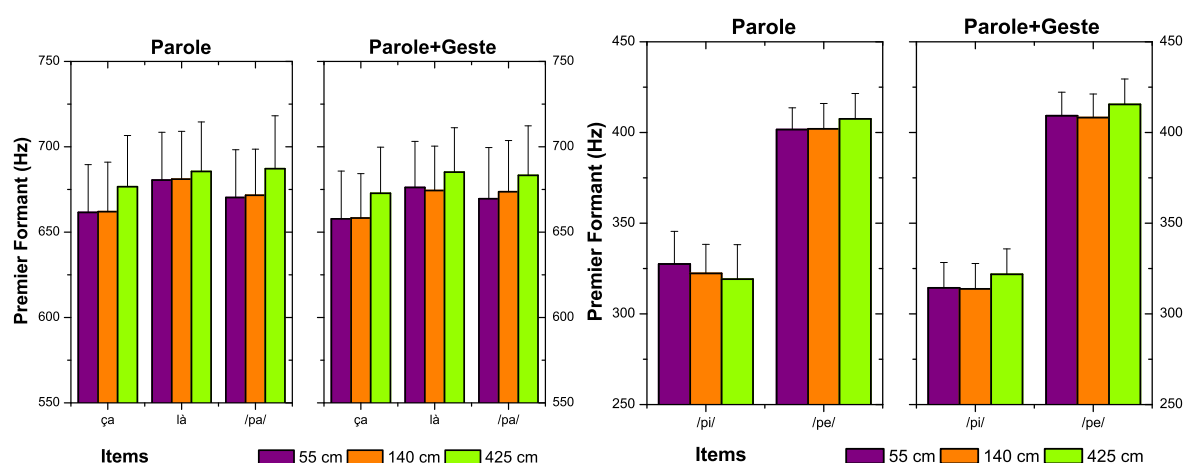


FIGURE 3.6 – $F1(\text{Hz})$ en fonction de la Condition et de la Distance pour les cinq items

Effets principaux Les résultats montrent que la Condition n'influence pas les valeurs de F1 ($F(1, 16) = 0.01, p = .91, NS$). En revanche, l'effet de la Distance est significatif ($\varepsilon = 0.82, p < .01$) : en d'autres termes, une distance éloignée implique des fréquences plus élevées. Le test *post-hoc* montre que les valeurs de F1 diffèrent pour les couples $/D1 D3/$ et $/D2 D3/$ ($p < .01$) mais qu'elles ne diffèrent pas significativement pour le couple $/D1 D2/$ ($p > .05$).

Effets d'interaction Aucun effet d'interaction n'est significatif :

- $C * D$: $\varepsilon = 0.91, p = .05, NS$;
- $C * I$: $\varepsilon = 0.63, p = .33, NS$;
- $D * I$: $\varepsilon = 0.26, p = .57, NS$;

- $C * D * I : \varepsilon = 0.41, p = .15, NS$.
- $C * L : F(1, 16) = 1.25, p = .28, NS$;
- $D * L : \varepsilon = 1.02, p = .61, NS$;
- $C * D * L : \varepsilon = 1.06, p = .48, NS$.

A retenir ! Le Tableau 3.5 propose un bref aperçu des principales différences observées sur les données acoustiques (F1).

F1	
Distance	$D1 = D2 < D3$

Tableau 3.5 – Principaux résultats - données acoustiques : F1

Second formant (F2)

Effets principaux Aucun effet principal n'est observé sur les valeurs de F2 : comme attendu, celles-ci ne sont influencées ni par la Condition ($F(1, 16) = 0.02, p = .89, NS$), ni par la Distance ($\varepsilon = 0.78, p = .37, NS$).

Effets d'interaction De la même façon, aucune interaction n'est significative :

- $C * D : \varepsilon = 0.98, p = .08, NS$;
- $C * I : \varepsilon = 0.75, p = .34, NS$;
- $D * I : \varepsilon = 0.52, p = .73, NS$;
- $C * D * I : \varepsilon = 0.69, p = .85, NS$.
- $C * L : F(1, 16) = 0.15, p = .70, NS$;
- $D * L : \varepsilon = 0.73, p = .33, NS$;
- $C * D * L : \varepsilon = 0.99, p = .85, NS$.

Fréquence fondamentale (F0) La figure 3.7 montre l'effet de la Condition et de la Distance sur les valeurs de F0. Les voyelles ouvertes et fermées sont ici représentées sur un seul et même graphique (l'étendue des valeurs de F0 étant moins importante que celles de l'ouverture des lèvres et des deux premiers formants).

Effets principaux Les résultats ne montrent pas d'effet de la Condition ($F(1, 16) = 3.41, p = .08, NS$), mais un effet significatif de la Distance ($\varepsilon = 0.56, p < .01$). Le test *post-hoc* montre des différences significatives pour les différentes modalités de la Distance prises deux à deux ($p < .01$).

Effets d'interaction Aucun effet d'interaction n'est significatif :

- $C * D : \varepsilon = 0.65, p = .56, NS$;
- $C * I : \varepsilon = 0.37, p = .38, NS$;

1. Expérience 1

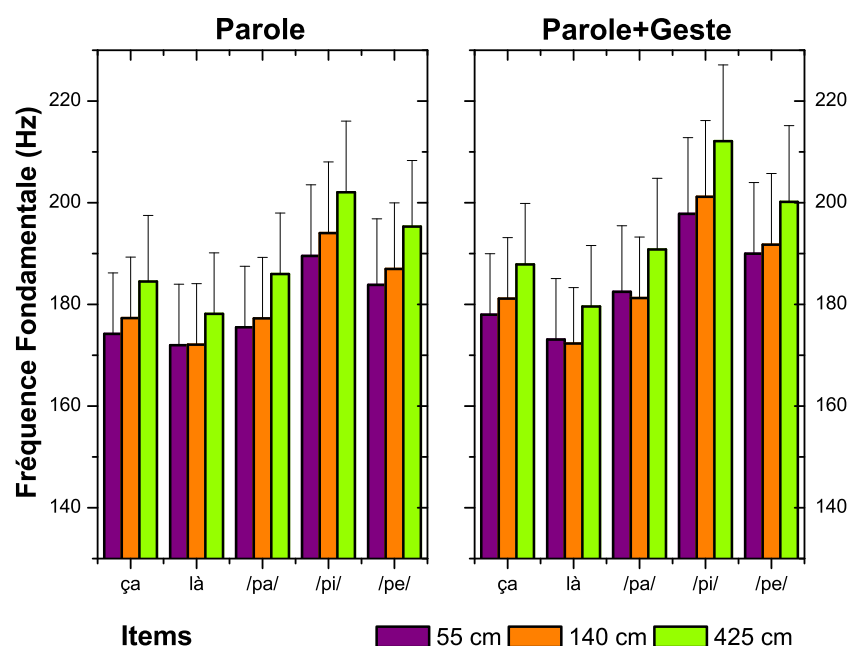


FIGURE 3.7 – F_0 (Hz) en fonction de la Condition et de la Distance pour les cinq items

- $D * I : \varepsilon = 0.36, p = .18, NS$;
- $C * D * I : \varepsilon = 0.48, p = .78, NS$.
- $C * L : F(1, 16) = 0.36, p = .55, NS$;
- $D * L : \varepsilon = 0.67, p = .40, NS$;
- $C * D * L : \varepsilon = 0.97, p = .91, NS$.

A retenir ! Le Tableau 3.6 présente un bref aperçu des principales différences observées sur les données acoustiques (F_0).

F0	
Distance	$D1 < D2 < D3$

Tableau 3.6 – Principaux résultats - données acoustiques : F_0

Intensité La figure 3.8 montre l'effet de la Condition et de la Distance sur l'intensité du signal de parole. Les voyelles ouvertes et fermées sont ici représentées sur un seul et même graphique.

Effets principaux Les résultats mettent en évidence un effet significatif de la Condition ($F(1, 16) = 9.59, p < .01$), l'intensité augmente en situation bimodale. On observe également un effet significatif de la Distance ($\varepsilon = 0.53, p = .01$) : l'intensité augmente avec la distance du référent. Le test *post-hoc* montre des différences significatives pour chaque couple de modalités ($p < .01$).

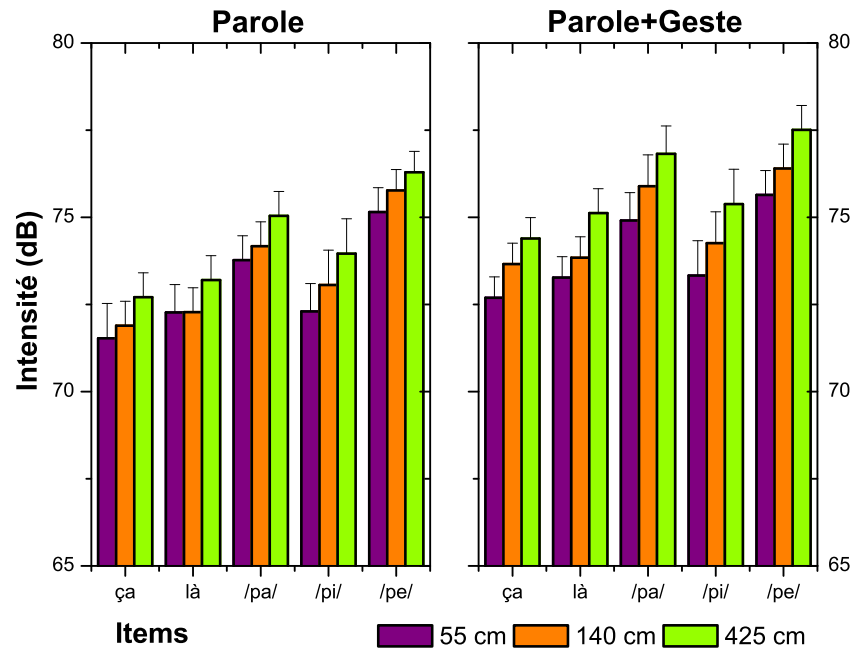


FIGURE 3.8 – *Intensité* (dB) en fonction de la Condition et de la Distance pour les cinq items

Effets d'interaction Aucun effet d'interaction n'est significatif :

- $C * D : \varepsilon = 0.61, p = .16, NS$;
- $C * I : \varepsilon = 0.42, p = .42, NS$;
- $D * I : \varepsilon = 0.58, p = .52, NS$;
- $C * D * I : \varepsilon = 0.46, p = .82, NS$;
- $C * L : F(1, 16) = 0.50, p = .58, NS$;
- $D * L : \varepsilon = 1.08, p = .40, NS$;
- $C * D * L : \varepsilon = 0.89, p = .67, NS$.

A retenir ! Le Tableau 3.7 présente un bref aperçu des principales différences observées sur les données acoustiques (*Intensité*).

	Intensité
Distance	$D1 < D2 < D3$
Condition	$C1 < C2$

Tableau 3.7 – Principaux résultats - données acoustiques : *Intensité*

1.5.3 Propriétés cinématiques du pointage manuel

La figure 3.9 montre l'effet de la Distance sur les paramètres cinématiques du pointage manuel (Pic de Vitesse du 1^{er} Stroke ; Durée totale ; Amplitude du 1^{er} Stroke).

1. Expérience 1

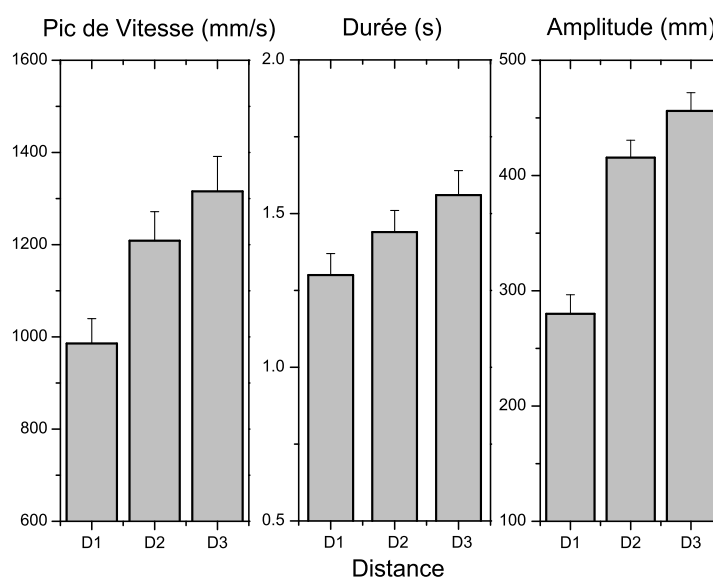


FIGURE 3.9 – *Pic de Vitesse* (mm/s) ; *Amplitude* (mm) ; *Durée* (s) en fonction de la Distance

Les résultats montrent que chacun des paramètres est influencé par la distance de la cible. Le pic de vitesse ($F(2, 30) = 37.62, p < .01$), l'amplitude ($F(2, 30) = 69.64, p < .01$), ainsi que la durée ($F(2, 30) = 34.07, p < .01$) du pointage sont plus élevés pour les distances les plus éloignées. Les tests *post-hoc* montrent des différences significatives pour chaque couple de modalités et ce pour chacun des trois facteurs ($p < .01$).

A retenir ! Le Tableau 3.8 présente un bref aperçu des principaux résultats observés sur les données cinématiques.

	Vitesse	Amplitude	Durée
Distance	$D1 < D2 < D3$	$D1 < D2 < D3$	$D1 < D2 < D3$

Tableau 3.8 – Principaux résultats - données cinématiques

1.6 Résumé des principaux résultats de l'Expérience 1

1.6.1 Encodage moteur de la distance dans le pointage vocal et manuel

Les résultats de cette première expérience suggèrent un encodage particulier de la distance via le pointage vocal et manuel. Nous observons un effet principal de la Distance significatif au niveau des paramètres articulatoires, acoustiques et cinématiques. Plus précisément, l'ouverture des lèvres, les valeurs de F1, de F0 et d'intensité, ainsi que le pic de vitesse, l'amplitude et la durée totale du geste augmentent avec la distance de l'objet pointé. Les valeurs de second formant, associées à une protrusion de la langue, ne sont pas influencées par la distance de la cible.

Ces résultats suggèrent une distinction entre l'espace péripersonnel et l'espace extrapersonnel (seules les valeurs de F1 ne diffèrent pas significativement entre les cibles situées en *D1* et *D2*) mais également entre l'espace extrapersonnel proche et l'espace extrapersonnel lointain (*D2* et *D3*). Le geste et la voix permettent donc de distinguer une cible proche d'une cible lointaine, mais également de désambiguïser deux cibles hors d'atteinte. De manière générale, les participants utilisent donc un pointage plus large et plus soutenu pour désigner une cible plus éloignée : un geste manuel plus ample, mais également plus rapide et maintenu plus longtemps ; de la même façon, un geste vocal plus large mais également plus accentué, puisque des valeurs plus élevées d'intensité, de F0 et de F1 (couplées à un allongement temporel de la syllabe, qu'il serait intéressant de vérifier¹²) marquent en effet une focalisation vocale particulière (Krahmer et Swerts [17]).

Cet encodage moteur de la distance semble être de bas niveau, puisque l'interaction entre la Distance et l'Item d'une part, et la Distance et la Lexicalité d'autre part, ne sont pas significatives. Cela suggère que différents types de voyelle (ouvertes ou fermées ; plus ou moins bien représentées dans les déictiques du français), ainsi que différents items (lexicaux et non lexicaux), peuvent endosser une fonction déictique : ils sont influencés de la même manière par la distance de la cible. On observe donc de manière générale une ouverture des lèvres plus importante ainsi que des valeurs de premiers formants (seules les valeurs de F1 de l'item [pi] montrent une tendance inverse, en situation de parole seule uniquement), d'intensité et de F0 plus élevées, lorsque le pointage vocal est associé à une cible éloignée. Ces résultats sont en faveur d'un encodage moteur de la distance de bas niveau : une tendance à produire des gestes manuels et vocaux plus larges pour désigner des objets éloignés. L'hypothèse suivante peut être formulée : cette tendance pourrait expliquer le patron phonologique observé au niveau des termes spatiaux dans la plupart des langues du monde (voir Chapitre 2, Section 3), à savoir des voyelles antérieures/fermées pour désigner de petits objets ou des objets proches, et des voyelles postérieures/ouvertes pour désigner de gros objets ou des objets éloignés. Cette tendance à associer certaines voyelles à certaines propriétés spatiales des objets pourrait donc s'expliquer, comme le proposait Sapir [29], en termes kinesthésiques (produire un geste plus étendu dans l'espace pour représenter un objet plus éloigné).

1.6.2 Interaction voix/geste

Nos résultats montrent également un effet significatif de la Condition sur l'ouverture des lèvres et l'intensité du signal de parole. Toutefois, si la parole semble être influencée structurellement par la production simultanée d'un geste manuel, les deux effets que l'on observe sont contradictoires. Ils ne permettent donc pas de conclure de façon robuste sur la nature (inhibitrice ou renforçatrice) de l'interaction voix/geste. Les données relatives à l'ouverture des lèvres soutiennent notre hypothèse de coopération : l'ouverture des lèvres est plus importante en situation unimodale, ce qui suggère que les participants utilisent un pointage renforcé pour compenser l'absence de pointage manuel. En revanche, les données relatives à l'intensité montrent tout l'inverse : l'intensité augmente en situation bimodale, lorsque deux modalités sont disponibles pour transmettre le message. Ce résultat est cohérent avec ceux de Bernardis et Gentilucci [3]

12. Des analyses temporelles préliminaires, effectuées dans le cadre d'un stage d'ingénieur 2A (Chadœuf [6]), montrent que la durée de la parole augmente pour les distances les plus éloignées.

1. Expérience 1

et de Krahmer et Swerts [17] qui montrent que le signal de parole est renforcé par la production conjointe d'un geste manuel.

Par ailleurs, l'interaction $C * I$ est significative : l'effet de la condition sur l'ouverture des lèvres (une hyperarticulation en situation unimodale) n'est observé que sur les voyelles ouvertes (pour les items *ça*, *là* et [pa]). De la même façon, l'interaction $C * L$ est significative : l'effet de la condition n'est significatif que pour les mots. Toutefois, cette dernière peut être expliquée par la répartition des voyelles ouvertes et fermées dans notre corpus, puisque ce dernier contient plus de voyelles ouvertes (trois items sur cinq) que de voyelles fermées (deux items sur cinq). De fait, l'interaction entre Condition et Lexicalité est très probablement directement liée à l'interaction entre Condition et Item ; elle montre une influence du geste sur la parole de bas niveau, mais associée aux voyelles peu contraintes phonétiquement.

Par ailleurs, nous attendions un effet significatif de la Condition sur les valeurs de premier formant. L'absence de variation au niveau du F1 pourrait être expliquée par un nombre important d'items testés (soit des stratégies spécifiques aux items, qui ont pu "noyer" l'effet au niveau global) ou par l'utilisation d'un protocole spécifique.

Enfin, le problème majeur de cette première expérience est qu'elle ne permet pas d'étudier la direction de l'influence geste/parole : nous n'avons en effet pas la possibilité de comparer les productions manuelles simples aux productions manuelles accompagnées de parole. L'absence d'une condition Geste Seul est donc clairement à déplorer. Par ailleurs, les analyses statistiques cinématiques ont été restreintes à l'effet principal de la Distance ; pourtant, étudier l'effet de la Lexicalité sur les différents paramètres du geste aurait pu apporter des éléments de réponse supplémentaires (e.g. des différences de durée et de vitesse en fonction de la nature de l'item pour déterminer dans quelle mesure l'influence de la parole sur le geste est de bas niveau ou de plus haut niveau).

1.6.3 Elaboration d'un nouveau protocole

Nous avons choisi d'élaborer un nouveau protocole afin de pallier les déficits évoqués ci-dessus et de déterminer dans quelle mesure ces premiers résultats sont liés à (1) un encodage moteur de la distance et une interaction particulière entre deux systèmes de production ; (2) l'utilisation d'un protocole spécifique. L'élaboration d'un nouveau protocole permettra de confirmer et de préciser nos premiers résultats, par l'amélioration de la méthodologie existante.

Tout d'abord, il est essentiel de simplifier le protocole afin de réduire la durée de l'expérience et de limiter la charge cognitive de la tâche. Comme suggéré par Bernardis et Gentilucci [3], « *We preferred to run two experiments instead of a longer one to insure that the task did not become automatic and participants emitted the responses in a spontaneous way like in everyday life*¹³ » ([3], p.180). En d'autres termes, une passation trop longue et/ou une charge cognitive trop importante sont deux biais dont il faut tenir compte. Dans notre expérience, les productions ont pu être automatisées et ne reflètent donc plus l'intention communicative souhaitée. Dans cette optique, il s'avère également essentiel, pour favoriser l'établissement d'une interaction communicative naturelle, de remplacer le mannequin par un interlocuteur réel, qui s'efforcera d'être le

13. Traduction proposée : Il est préférable de conduire deux expériences plutôt qu'une seule, nécessairement plus longue, afin de s'assurer que la tâche ne devienne pas automatique et que les participants émettent leurs réponses de façon spontanée, comme ils le feraient dans la vie de tous les jours.

plus homogène possible avec les participants.

Nous souhaitons également restreindre notre étude à un seul item. Cela nous permettra de tester plus précisément les variations intra-catégorielles, liées à l'ouverture des lèvres et au premier formant, en minimisant l'impact lié à l'utilisation de stratégies articulatoires spécifiques, pour la production successive de différents items. Du fait des résultats obtenus sur l'effet de la distance (i.e. une absence d'interaction $D * I$ et $D * L$) et sur l'effet de la condition (i.e. lié aux voyelles ouvertes), nous testerons l'adverbe locatif *là*, qui présente moins d'impact coarticulatoire que le démonstratif *ça* ; en effet la fricative [s] contraint fortement la position de la langue et limite donc les variations intra-catégorielles du [a].

Par ailleurs, nous testerons uniquement les deux LED situées dans l'espace extra-personnel, puisqu'elles sont clairement distinguées au niveau articulatoire, acoustique et cinématique. Les LED situées à 55 et 140 cm impliquent quant à elles un geste très différent l'une de l'autre (bras plié vs bras tendu) ; il n'y a donc aucune ambiguïté sur la cible désignée, ce qui pourrait expliquer l'absence de variation au niveau du premier formant entre ces deux distances : les gestes articulatoires et manuels sont potentiellement suffisants pour distinguer les deux cibles. En nous focalisant sur les cibles ambiguës, nous serons en mesure de favoriser l'apparition d'un encodage particulier, tant sur le plan cinématique que phonétique.

Enfin, il s'avère nécessaire d'ajouter une condition Geste Seul, qui permettra de tester (1) l'influence de la parole sur la production des gestes manuels et (2) la direction (au sens de réciprocité) de l'influence geste/parole : la parole influence-t-elle le geste de la même façon que le geste influence la parole ?

2 Expérience 2

Notre seconde expérience a donc pour objectif d'améliorer le protocole initial, en limitant le temps de passation et la complexité de la tâche, et en améliorant l'aspect communicatif de la situation. Notre objectif est de confirmer certains de nos résultats (notamment ceux observés sur l'encodage moteur de la distance) mais également de mieux comprendre certaines incohérences (concernant l'effet de la Condition).

Les principaux changements concernent :

- Les conditions : Parole+Geste - Parole Seule - **Geste Seul**
→ Caractériser au mieux l'interaction voix/geste en étudiant (1) l'influence du geste sur la parole, (2) l'influence de la parole sur le geste.
- Les distances : $D1 = 140$ cm - $D2 = 425$ cm
→ Tester l'encodage de la distance sur deux cibles ambiguës, situées dans un même espace.
- Les items : *Là*
→ Ne tester qu'un seul item afin de limiter l'impact lié à l'utilisation de stratégies articulatoires multiples et d'étudier au mieux les variations intra-catégorielles.
- La situation d'interaction : Un interlocuteur **réel**
→ Améliorer la situation d'interaction pour favoriser l'émergence d'une intention communicative.

2. Expérience 2

2.1 Méthodologie

2.1.1 Participants

Vingt-huit participants (dont 11 hommes et 17 femmes), âgés de 18 à 36 ans ($M = 23.6$, $SD = 3.6$) ont participé, bénévolement, à cette étude. Tous étaient de langue maternelle française, sans déficit visuel ou auditif non corrigé, ni trouble moteur particulier. Recrutés majoritairement par diffusion d'e-mails, les participants étaient pour la plupart étudiants à l'Université de Grenoble.

2.1.2 Matériel

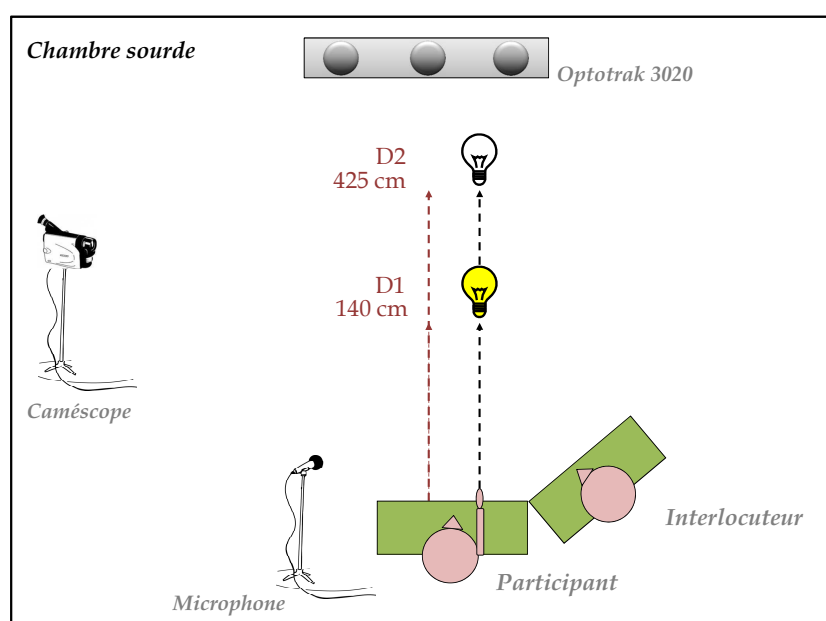


FIGURE 3.10 – Dispositif expérimental

Comme indiqué sur la Figure 3.10, le dispositif utilisé est similaire à celui de la première expérience : les participants sont évalués individuellement en chambre sourde, mais sont à présent assis face à deux LED, placées à deux distances différentes d'eux :

- D1, située à 140 cm, dans l'espace **extra-personnel proche** ;
- D2, située à 425 cm, dans l'espace **extra-personnel lointain**.

Nous avons fait le choix de ne conserver que les deux LED ambiguës, situées hors de portée, qui, au vu de nos précédents résultats, sont clairement distinguées. Ces deux LED, alignées l'une par rapport à l'autre et par rapport à l'axe œil-index du participant, vont permettre de confirmer un encodage particulier de la distance, non seulement via les propriétés cinématiques du geste (l'angle du bras seul ne pouvant pas les dissocier) mais également via les propriétés phonétiques du pointage vocal.

Le microphone AKG C1000S, relié à un enregistreur multipistes Korg D1200, ainsi que le système opto-électronique de capture de mouvements Optotrak 3020 permettent d'enregistrer les productions orales et gestuelles. Un caméscope numérique permet d'enregistrer l'ensemble de la passation, afin de vérifier ultérieurement la validité des tâches effectuées et le déroulement de l'expérience. Néanmoins, ce contrôle vidéo s'avère nettement moins nécessaire que lors de la première expérience, puisque l'expérimentateur effectue un premier contrôle en direct et peut intervenir si la tâche n'est pas réalisée correctement.

2.1.3 Procédure

La tâche des participants, dont la main est initialement placée en position de repos en milieu de table, est de désigner la LED allumée à l'interlocuteur situé sur leur droite, en réponse à la question « C'est où ? », prononcée de manière neutre. La situation d'interaction est ici explicite et plutôt naturelle, puisqu'elle est basée sur un jeu de questions/réponses. L'interlocuteur s'efforce d'être le plus homogène possible : tourné vers le participant lorsqu'il lui pose la question, il attend que celui-ci réponde pour regarder dans la direction du pointage.

Trois conditions expérimentales sont testées : deux situations unimodales (**Parole Seule**, **Geste Seul**) et une situation bimodale (**Parole+Geste**).

- Parole Seule : En situation de parole seule, les participants doivent nommer la LED allumée en utilisant l'adverbe locatif « là », inséré dans la phrase porteuse : « c'est [+ item] ». Il leur est strictement interdit d'utiliser leurs mains, et ils ont donc le choix de les laisser en position de repos, sur la table ou sur leurs genoux ;
- Geste Seul : En situation de geste seul, les participants doivent pointer la cible par un geste de l'index droit ;
- Parole+Geste : En situation de parole+geste, les participants doivent simultanément nommer et pointer la LED allumée.

Le paradigme utilisé est un paradigme en blocs : l'expérience se déroule en six blocs successifs (2×3 Conditions). Le premier bloc de chaque tâche est précédé par une courte session d'entraînement (2×4 essais, avec possibilité de recommencer l'entraînement si nécessaire). La moitié des participants commencent par la condition Geste Seul et l'autre moitié par la condition Parole+Geste (il est en effet probable que la condition Parole Seule soit la moins naturelle, et potentiellement la plus difficile). Chaque essai débute par la question de l'interlocuteur, suivie de l'allumage de l'une des deux diodes, signal de départ pour le participant.

La durée de l'expérience est considérablement réduite par rapport à celle de la première expérience, ce qui limite le développement de comportements automatiques et réduit la charge cognitive : la passation dure à présent 20 minutes (comprenant installation, consigne et debriefing), pendant lesquelles les participants sont confrontés à 60 essais ($= 2 \text{ Distances} \times 3 \text{ Conditions} \times 10 \text{ Itérations}$).

Le plan d'expérience est un plan intra-sujets, de manière à ce que chaque participant soit confronté à chaque modalité de chaque variable (voir Tableau 3.9 pour un récapitulatif des variables indépendantes utilisées).

2. Expérience 2

Variables Indépendantes	Modalités
Distance	140 cm - 425 cm
Condition	Parole Seule - Geste Seul - Parole+Geste

Tableau 3.9 – Récapitulatif des variables indépendantes

2.2 Acquisition des données

L’acquisition des données étant identique à celle de l’Expérience 1, le lecteur pourra trouver les informations utiles en Section 1.2.

2.3 Mesures

Les mesures effectuées sont les mêmes que lors de la première expérience. Nous renvoyons donc le lecteur à la Section 1.3 pour plus de détails. Le Tableau 3.10 propose également un récapitulatif des variables dépendantes.

Toutefois, nous mesurons également la durée de tenue du geste manuel (i.e. durée du plateau), qui permet d’obtenir des informations temporelles plus précises sur la réalisation du pointage. En effet, la durée du plateau est définie par l’intervalle [Retour-Apex] et permet donc d’observer le maintien en position de pointage. En d’autres termes, le plateau correspond au maintien de l’apex, soit au maintien de la phase signifiante du geste, celle qui nous permet de montrer. Mesurer la durée du plateau nous permettra de déterminer si l’effet de la Distance observé sur la durée totale du geste lors de la première expérience (un geste plus long pour les distances les plus éloignées) est lié à une tenue en position de pointage plus longue.

Variables Dépendantes	Mesures
Données acoustiques	F0 (Hz) - Intensité (dB) - F1 (Hz) - F2 (Hz)
Données articulatoires	Ouverture des lèvres (mm)
Données cinématiques	Durée (s) - Durée plateau (s) Pic de vitesse (mm/s) - Amplitude (mm)

Tableau 3.10 – Récapitulatif des variables dépendantes

2.4 Analyses statistiques & prédictions

Certains essais ont dû être exclus de l’analyse statistique : les oublis (d’item et/ou de pointage) mais également les données manquantes liées à des problèmes techniques (masquage des diodes émettrices infra-rouge notamment), soit 3.75% des données acoustiques, 8.21% des données articulatoires et 12.23% des données manuelles. Ce pourcentage d’essais exclus est relativement important et correspond aux données acoustiques d’un participant, aux données articulatoires de deux participants et aux données manuelles de trois participants, exclus de l’analyse statistique finale. Les analyses statistiques acoustiques portent sur 27 participants, les

analyses articulatoires portent quant à elles sur 26 participants, et les analyses cinématiques sur 25 participants. Pour le détail technique et théorique des analyses statistiques en général, le lecteur pourra se référer à la Section 1.4.

Les analyses de la variance pour plans à mesures répétées et deux facteurs intra-sujets,

$$S_{25|26|27}[Distance_2 \times Condition_2],$$

portent sur les médianes. Le test de Shapiro-Wilk (Shapiro [30]) permet de vérifier la normalité des données, une transformation logarithmique est appliquée en cas de besoin. Le test de Mauchly (Mauchly [22]) et le correctif de Greenhouse-Geisser (Greenhouse et Geisser [14]) ne sont pas nécessaires ici puisque $dl1$ est de 1 pour chaque facteur étudié.

Dans le cas présent, L'ANOVA à mesures répétées va nous permettre de tester deux effets principaux, celui de la Distance et celui de la Condition, ainsi qu'un effet d'interaction, $C * D$, pour lequel aucune hypothèse n'est formulée à priori.

Nos prédictions sont les suivantes :

- Effet principal de la distance

Nous souhaitons confirmer l'encodage moteur de la distance au niveau des propriétés acoustiques, articulatoires et cinématiques du pointage. Au vu de nos résultats précédents, nous attendons une ouverture des lèvres plus grande pour la LED la plus éloignée $D2$, ainsi que des valeurs de $F1$, d'intensité et de $F0$ plus élevées (aucun effet n'est attendu concernant les valeurs de $F2$). L'amplitude, le pic de vitesse ainsi que la durée totale et la durée du plateau devraient également augmenter avec la distance de la cible.

- Effet principal de la condition

Nous souhaitons également caractériser de façon précise l'interaction voix/geste, comme une influence inhibitrice et bidirectionnelle entre ces deux systèmes. Ce nouveau protocole va nous permettre d'étudier non seulement l'influence du geste manuel sur la parole mais également l'influence de la parole sur le geste manuel. Une coopération inter-modalités aurait pour conséquence un renforcement des productions en condition unimodale : au niveau du pointage vocal, une ouverture des lèvres plus grande, associée à des valeurs de $F1$ plus élevées ; au niveau du pointage manuel, nous attendons une influence de même nature, soit un pic de vitesse réduit et des durées (totale et de plateau) plus longues. En revanche, aucun effet n'est attendu sur les valeurs de $F2$, de $F0$, et d'amplitude. De la même façon, aucun effet n'est attendu sur les valeurs d'intensité, puisque selon nous, l'effet observé précédemment sur l'intensité du signal de parole pourrait être lié à l'utilisation d'un paradigme particulier, et notamment à l'absence d'un interlocuteur réel ; il devrait donc disparaître avec l'amélioration de la situation d'interaction communicative.

2.5 Résultats de l'Expérience 2

Cette section présente un descriptif des effets observés pour chaque facteur considéré. Les implications théoriques qui en découlent seront discutées dans la Section 2.6.

2. Expérience 2

2.5.1 Propriétés articulatoires du pointage vocal

La figure 3.11 montre l'influence de la Condition et de la Distance sur l'ouverture des lèvres.

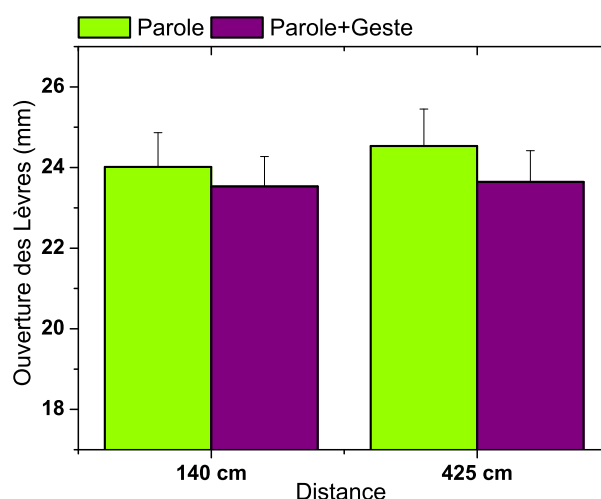


FIGURE 3.11 – OL (mm) en fonction de la Condition et de la Distance

Effets principaux Les résultats montrent un effet significatif de la Condition ($F(1, 25) = 4.96, p < .05$) et de la Distance ($F(1, 25) = 8.21, p < .01$) sur l'ouverture des lèvres, qui augmente en situation unimodale et pour la distance la plus éloignée.

Effet d'interaction L'interaction $C*D$ est significative ($F(1, 25) = 5.72, p < 0.05$) : le test *post-hoc* met en évidence un effet significatif de la distance pour la condition unimodale uniquement ($p < 0.5$).

A retenir ! Le Tableau 3.11 propose un bref aperçu des principales différences observées sur les données articulatoires.

Ouverture des lèvres	
Distance	$D1 < D2$
Condition	$C1 > C2$
Interaction	$C * D$
→ Effet de la distance significatif pour la condition Parole Seule	

Tableau 3.11 – Principaux résultats - données articulatoires

2.5.2 Propriétés acoustiques du pointage vocal

Premier formant F1 La figure 3.12 montre l'influence de la Condition et de la Distance sur les valeurs de F1.

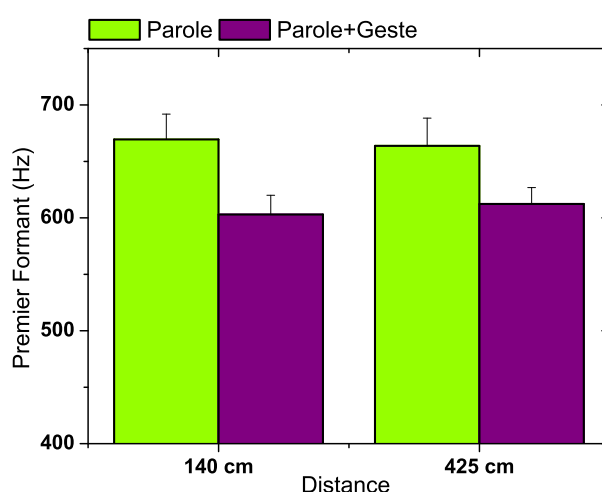


FIGURE 3.12 – $F1(\text{Hz})$ en fonction de la Condition et de la Distance

Effets principaux Les résultats indiquent que la Distance n'influence pas les valeurs de $F1$ ($F(1, 26) = 0.12$, $p = .73$, NS). En revanche, l'effet de la Condition est significatif ($F(1, 26) = 23.41$, $p < .01$) : les valeurs de $F1$ sont plus élevées en situation unimodale.

Effet d'interaction L'interaction $C * D$ n'est pas significative ($F(1, 26) = 1.26$, $p = 0.27$, NS).

A retenir ! Le Tableau 3.12 propose un bref aperçu des principales différences observées sur les données acoustiques ($F1$).

F1	
Condition	$C1 > C2$

Tableau 3.12 – Principaux résultats - données acoustiques : $F1$

Second formant $F2$

Effets principaux Aucun effet principal n'est observé sur les valeurs de $F2$: celles-ci ne sont influencées ni par la Condition ($F(1, 26) = 0.31$, $p = .58$, NS), ni par la Distance ($F(1, 26) = 0.55$, $p = 0.46$, NS).

Effet d'interaction De la même façon, l'interaction $C*D$ n'est pas significative ($F(1, 26) = 2.73$, $p = 0.17$, NS).

2. Expérience 2

Fréquence fondamentale F0

Effets principaux Les résultats ne montrent aucun effet significatif : les valeurs de F0 ne sont influencées ni par la Condition ($F(1, 26) = 1.03$, $p = .32$, *NS*) ni par la Distance ($F(1, 26) = 0.39$, $p = .53$, *NS*).

Effet d'interaction L'effet d'interaction $C * D$ n'est pas significatif ($F(1, 26) = 0.27$, $p = 0.60$, *NS*).

Intensité La figure 3.13 représente l'influence de la Condition et de la Distance sur l'intensité du signal de parole.

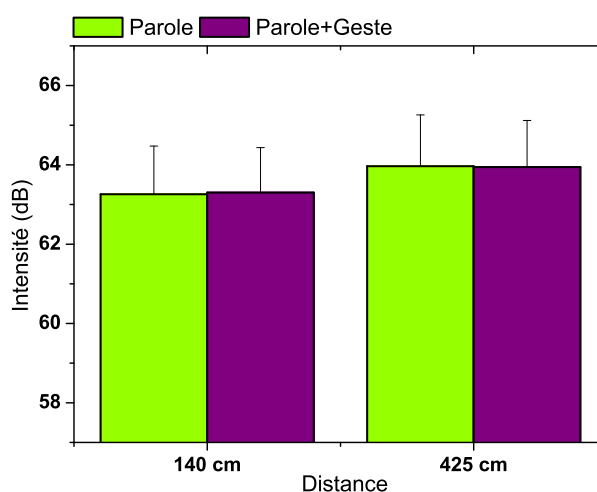


FIGURE 3.13 – *Intensité* (dB) en fonction de la Condition et de la Distance

Effets principaux Les résultats montrent une influence significative de la Distance ($F(1, 26) = 9.07$, $p < .01$) : l'intensité augmente avec la distance du référent. La Condition, en revanche, n'affecte pas l'intensité du signal ($F(1, 26) < 0.01$, $p = .98$, *NS*).

Effet d'interaction L'effet d'interaction $C * D$ n'est pas significatif ($F(1, 26) = 0.04$, $p = 0.85$, *NS*).

A retenir ! Le Tableau 3.13 présente un bref aperçu des principales différences observées sur les données acoustiques (Intensité).

Intensité	
Distance	$D1 < D2$

Tableau 3.13 – Principaux résultats - données acoustiques : Intensité

2.5.3 Propriétés cinématiques du pointage manuel

Les figures 3.14 et 3.15 montrent l'influence de la Distance sur les paramètres cinématiques du pointage manuel (Durée totale ; Durée Plateau ; Pic de Vitesse du 1^{er} Stroke ; Amplitude du 1^{er} Stroke).

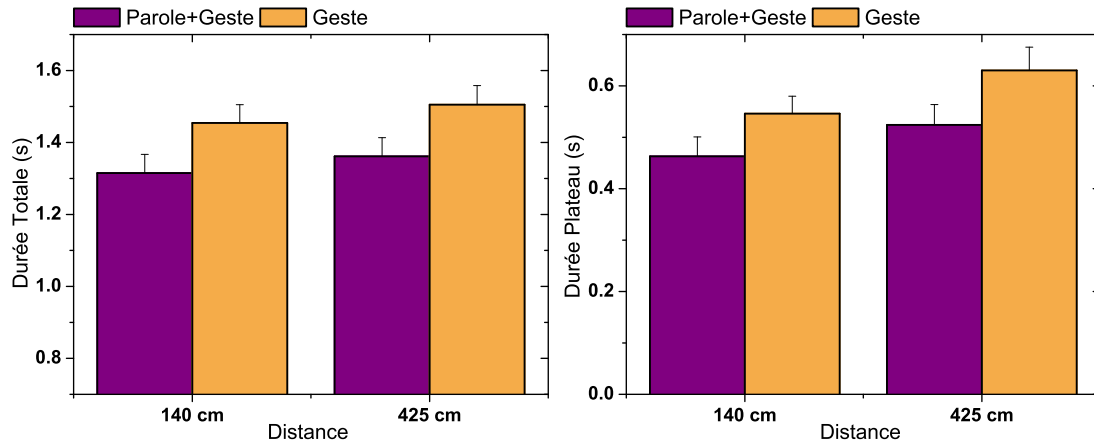


FIGURE 3.14 – *Durée Totale & Durée Plateau (s)* en fonction de la Distance et de la Condition

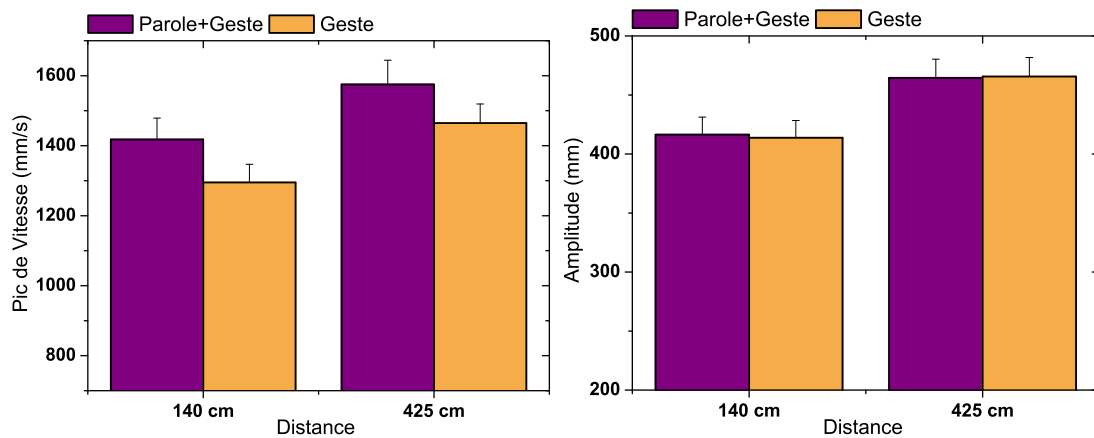


FIGURE 3.15 – *Pic de Vitesse (mm/s) & Amplitude (mm)* en fonction de la Distance et de la Condition

Effets principaux Les résultats indiquent que chacun des paramètres est influencé par la distance de la cible : le pic de vitesse ($F(1, 24) = 16.88, p < .01$), l'amplitude ($F(1, 24) = 13.98, p < .01$), ainsi que la durée totale ($F(1, 24) = 6.74, p = .01$) et celle du plateau ($F(1, 24) = 14.85, p < .01$) sont plus élevées pour les distances les plus éloignées. D'autre part, nous observons une influence significative de la Condition sur le pic de vitesse, la durée totale et celle du plateau (respectivement, $F(1, 24) = 16.49, p < .01$; $F(1, 24) = 47.79, p < .01$; $F(1, 24) = 26.30, p < .01$).

2. Expérience 2

La vitesse est réduite en situation unimodale, tandis que le geste dure et est maintenu plus longtemps.

Effet d'interaction L'interaction $C * D$ n'est significative pour aucun des facteurs :

- Vitesse : $F(1, 24) = 0.34$, $p = .56$, *NS*
- Amplitude : $F(1, 24) = 0.81$, $p = .37$, *NS*
- Durée Totale : $F(1, 24) < 0.01$, $p = .98$, *NS*
- Durée Plateau : $F(1, 24) = 0.04$, $p = .85$, *NS*

A retenir ! Le Tableau 3.14 présente un bref aperçu des principales différences observées sur les données cinématiques.

	Vitesse	Amplitude	Durée	Plateau
Distance	$D1 < D2$	$D1 < D2$	$D1 < D2$	$D1 < D2$
Condition	$C1 < C2$	ϕ	$C1 > C2$	$C1 > C2$

Tableau 3.14 – Principaux résultats - données cinématiques

2.6 Résumé des principaux résultats de l'Expérience 2

2.6.1 Encodage moteur de la distance dans le pointage vocal et manuel

Cette seconde expérience permet tout d'abord de confirmer l'encodage moteur de la distance au niveau du pointage. L'effet de la Distance est significatif sur les différents paramètres cinématiques du pointage manuel : la durée totale, la durée du plateau, la vitesse et l'amplitude du geste sont plus élevées pour la distance la plus éloignée. De plus, l'effet de la Distance est significatif au niveau du pointage vocal : l'ouverture des lèvres et l'intensité du signal de parole augmentent pour la distance la plus éloignée, ce qui n'est pas le cas des valeurs de F0, F1 et de F2. Ces résultats sont globalement cohérents avec ceux de l'Expérience 1, mais ne permettent pas de répliquer l'effet de la distance observé sur les valeurs de F0 et de F1 ; de plus l'interaction $C * D$ est à présent significative au niveau de l'ouverture des lèvres (i.e. l'effet de la distance est significatif en situation unimodale uniquement). Ces résultats seront discutés en Section 3.

2.6.2 Interaction voix/geste

Par ailleurs, nos résultats sont en faveur d'une coopération entre les systèmes de production de parole et de geste. Les productions vocales sont renforcées en situation unimodale, tandis que l'ajout d'un geste manuel inhibe la parole ; réciproquement, l'ajout de la parole inhibe la production d'un geste manuel. Plus précisément, l'ouverture des lèvres est plus large et est associée à des valeurs de F1 plus élevées en situation unimodale ; de la même façon le geste, plus lent, est maintenu plus longtemps en situation unimodale. Les productions manuelles et vocales sont générées l'une par rapport à l'autre, et non indépendamment l'une de l'autre. Les deux modalités coopèrent en situation bimodale, où l'information est distribuée de manière équitable sur le canal gestuel et sur le canal vocal.

Néanmoins, il faut noter certaines incohérences entre nos deux expériences : (1) nous n'observons plus l'effet significatif de la Condition sur l'intensité, qui était contraire à notre hypothèse ; (2) nous observons des variations au niveau du premier formant. Ces résultats seront discutés en Section 3.

3 Implications théoriques

L'objectif général de cette étude était d'évaluer au mieux l'interaction entre le geste et la parole en mettant en évidence (1) un encodage moteur de la distance, de bas niveau, via les propriétés acoustiques, articulatoires et cinématiques du pointage, et ce indépendamment de la voyelle utilisée ou du contexte, lexical ou non, dans lequel elle est réalisée ; (2) une influence geste/parole structurelle et bidirectionnelle, sous la forme d'une coopération entre les caractéristiques phonétiques de la parole d'une part, et les propriétés cinématiques du geste manuel d'autre part.

3.1 Encodage moteur de la distance dans le pointage vocal et manuel

L'ensemble de nos résultats montrent un effet significatif et stable (i.e. observé dans les deux expériences) de la distance sur les propriétés cinématiques du pointage manuel. Le pointage manuel est plus ample, il dure et est maintenu plus longtemps, mais il est également plus rapide, pour les cibles les plus éloignées. Les participants adoptent une stratégie particulière pour désigner de manière explicite une cible éloignée, ce qui leur permet de faire la distinction entre deux objets situés dans l'espace péripersonnel et extrapersonnel, mais également entre deux objets ambigus, situés dans un même espace perceptif. Ces données sont cohérentes avec celles habituellement rapportées dans la littérature (Bonfiglioli et collab. [5], Gentilucci et collab. [10], Rochet-Capellan [27]), à savoir que les paramètres cinématiques (durée, vitesse, amplitude) du geste d'atteinte/saisie et de pointage sont plus élevés pour un objet plus éloigné.

Cet encodage ne se restreint pas au système manuel : les propriétés articulatoires du pointage vocal, ainsi que l'intensité du signal de parole, sont également influencées de façon robuste par la distance de la cible. L'ouverture des lèvres et l'intensité augmentent avec la distance de l'objet pointé.

Toutefois, nous observons certaines incohérences entre nos deux expériences :

- Les variations au niveau de la fréquence fondamentale, qui augmente avec la distance de la cible, ne sont observées que dans l'Expérience 1.
De plus, l'interaction $C * D$ sur l'ouverture des lèvres (i.e. l'effet de la distance est significatif en situation unimodale uniquement) n'est observée que dans l'Expérience 2.

Notre hypothèse pour expliquer ces deux résultats est la suivante : ils seraient liés à l'utilisation d'un protocole spécifique, et plus précisément à la présence du mannequin dans l'Expérience 1, remplacé par un interlocuteur réel dans l'Expérience 2.

3. Implications théoriques

Plus précisément, lors de l'Expérience 1, c'est-à-dire en l'absence d'un interlocuteur réel, les participants étaient dans une situation d'interaction inhabituelle. Cela a probablement altéré leurs productions et induit une accentuation exagérée du signal de parole, soit des variations au niveau de la fréquence fondamentale. En revanche, la présence d'un interlocuteur réel lors de l'Expérience 2 a pu améliorer la situation d'interaction et favoriser la production d'un pointage vocal plus naturel.

De la même façon, l'interaction observée sur l'ouverture des lèvres (i.e. un effet de la distance significatif en condition Parole Seule) lors de l'Expérience 2 peut s'expliquer par la présence d'un véritable interlocuteur : en situation bimodale, l'interlocuteur bénéficie d'une information visuelle non négligeable, le geste de l'index ; en revanche, en situation unimodale, le pointage est clairement plus ambigu et occasionne une distinction $D1 - D2$ plus importante, au niveau articulatoire. Cette variation articulatoire permet aux participants de compenser l'absence d'information visuelle, visiblement jugée cruciale pour être compris de leur interlocuteur. La présence du mannequin lors de l'Expérience 1 ne permettait pas aux participants de faire une telle distinction, d'où la production de gestes articulatoires renforcés dans les deux conditions.

- Les variations au niveau du premier formant, qui augmente avec la distance de la cible (pour les objets situés dans l'espace extra-personnel), ne sont observées que dans l'Expérience 1.

L'abaissement de la mâchoire pour l'encodage de la distance (soit une ouverture plus grande), aurait dû avoir des conséquences acoustiques (notamment au niveau du premier formant) dans les deux expériences. Or, notre hypothèse est qu'il a pu être compensé, lors de l'Expérience 2, par une réorganisation de la position de la langue, avec pour conséquence l'absence de variation au niveau du premier formant. En effet, comme suggéré par Lindblom et collab. [19], la production de la parole est un processus "compensatoire". Maeda [21] en donne la définition suivante : « *compensation means that a deviation in the position of one articulator can be counteracted by a readjustment of other articulator(s) to minimize the deviation in the acoustic pattern*¹⁴ » ([21], p. 328). En d'autres termes, une forte variabilité articulatoire est compensée par une faible variabilité acoustique, et ce dans le but de produire des patterns acoustiques relativement invariants, dépendants uniquement de l'identité de la voyelle.

Maeda [21] illustre ses propos par une étude dans laquelle il examine les patterns temporels, correspondants aux positions de la mâchoire et de la langue, lors de la production de deux voyelles non arrondies [i] et [a], insérées dans différents contextes phonétiques. Ses résultats montrent que les valeurs articulatoires varient significativement plus que les valeurs acoustiques correspondantes (i.e. les valeurs de premier et second formants) : la variabilité articulatoire est deux fois plus importante que la variabilité acoustique ! Notre hypothèse est que lors de l'Expérience 1, les participants n'ont pu développer de telles stratégies de compensation du fait du nombre important de voyelles testées ([a, i, e]). La variabilité acoustique au niveau du premier formant est donc observable. En revanche, cette compensation peut se manifester dans l'Expérience 2, où l'on ne teste que la voyelle [a], et la variabilité acoustique "disparaît".

14. Traduction proposée : La compensation est le fait qu'une déviation au niveau de la position d'un articulateur peut être compensée par un réajustement de la position des autres articulateurs, afin de minimiser la déviation au niveau acoustique.

Nos deux expériences montrent que le second formant n'est pas influencé par la distance de la cible : nos résultats sont donc clairement en faveur d'un lien entre la distance et l'ouverture du conduit vocal, plutôt qu'entre la distance et le déplacement antérieur-postérieur de la langue dans la bouche. Chieffi et collab. [7] observaient quant à eux des variations liées au second formant (voir leur étude présentée en début de chapitre) : ils montraient une diminution des valeurs de F2 lorsqu'un déictique proximal était associé à un geste de pointage vers l'horizon. Néanmoins, ces variations peuvent être liées à la congruence/incongruence de la situation plutôt qu'à un encodage spécifique de la distance de la cible. En effet, les auteurs observent une augmentation du second formant (soit l'effet inverse), pour le déictique distal (soit l'autre version du démonstratif) lorsque celui-ci est inscrit sur un objet proche, et donc toujours en situation incongruente. Néanmoins, comme évoqué au Chapitre 2, Section 3.3, la validité de la situation dite congruente pose question, puisque les cibles du pointage manuel (vers soi/vers l'horizon) ne sont pas liées aux cibles du pointage vocal (inscription sur un jeton proche/loin).

Ces premiers résultats indiquent donc que l'encodage de la distance se manifeste dans le pointage par des variations cinématiques mais également articulatoires : des gestes manuels et vocaux plus larges pour désigner des objets plus éloignés. Ce résultat s'observe indépendamment du type de voyelle (ouverte, fermée, fréquemment utilisée dans les déictiques ou non) et du type d'item (mot, pseudo-mot). L'encodage de la distance apparaît donc comme une fonction motrice de bas niveau, pouvant être à l'origine d'une distinction phonologique entre les termes proximaux et distaux. Le pattern phonétique que l'on observe est en effet cohérent avec le pattern phonologique observé dans un grand nombre de langues, à travers le monde (voir le Chapitre 2, Section 3) : les versions proximales des termes spatiaux tendent à contenir davantage de voyelles fermées et/ou antérieures que les versions distales.

Toutefois, il est possible que cet encodage ne soit pas un phénomène purement moteur mais le reflet du pattern lexical de notre langue maternelle. En effet, les participants de nos deux expériences sont de langue maternelle française, langue qui présente justement un contraste ouvert/fermé au niveau des déictiques (e.g. *là-bas* vs *ici*). De fait, il est envisageable que la distinction articulatoire, observée au niveau intra-catégoriel, soit liée à une distinction linguistique inter-catégorielle : en d'autres termes, les participants pourraient simplement reproduire le pattern phonétique auquel ils sont habitués. Afin de vérifier cette hypothèse, il serait pertinent d'étudier les productions de participants dont la langue maternelle suit un pattern différent, voire opposé. Les démonstratifs grecs, par exemple, présentent un contraste postérieur/antérieur : [eðo, eki] (respectivement *ici* et *là*) ; si l'encodage de la distance est effectivement un phénomène moteur, les mêmes résultats devraient être observés pour les participants grecs que pour les participants français, soit des voyelles plus ouvertes pour désigner un objet plus éloigné.

3.2 Interaction voix/geste, une coopération inter-modalités

Nos résultats sont également en faveur d'une interaction de nature structurelle et bidirectionnelle entre le geste et la parole, qui coopèrent pour véhiculer une seule et même information. Autrement dit, ces deux modalités sont utilisées de manière complémentaire plutôt que redondante, lorsqu'elles véhiculent une fonction linguistique particulière, la deixis.

3. Implications théoriques

Nos résultats sont en faveur d'une influence inhibitrice de la production du geste manuel sur la production de la parole : l'ouverture des lèvres est plus large en situation unimodale et permet ainsi de compenser l'absence de la modalité manuelle ; en revanche, l'ajout du geste permet de répartir le contenu informatif sur les deux modalités disponibles (les productions bimodales, relativement aux productions unimodales, sont donc inhibées). Les variations du premier formant sont moins robustes : les valeurs de F1 n'augmentent en situation unimodale que lors de la seconde expérience. L'amélioration du protocole, et plus particulièrement l'utilisation d'un nombre limité d'items a, comme attendu, favorisé l'émergence d'un effet significatif. De la même façon, les variations d'intensité observées dans la première expérience, qui allaient à l'encontre de notre hypothèse de coopération (avec une intensité plus élevée en situation bimodale), ne sont pas répliquées lors de la seconde expérience et peuvent s'expliquer par l'utilisation d'un protocole spécifique, non adapté à l'étude d'un processus communicatif (situation d'interaction "non naturelle", induite par la présence du mannequin). Ces résultats, en faveur d'une influence inhibitrice du geste sur la parole sont contraires aux observations de Bernardis et Gentilucci [3] (présentées au Chapitre 1, Section 2.3). Les auteurs montrent une augmentation du second formant en situation bimodale et l'interprètent comme un renforcement des productions. Ils concluent que le geste renforce la parole. Cependant, l'hypothèse émise par les auteurs est que l'effet observé sur les valeurs de F2 soit lié à l'intention du participant d'interagir avec l'autre¹⁵. Il est donc surprenant qu'il ne soit observé qu'en situation bimodale. Par ailleurs, une augmentation du second formant ne traduit pas nécessairement un renforcement, au sens où nous l'entendons (i.e. une hyperarticulation), mais plutôt une hypoarticulation, soit une inhibition (au sens de centralisation des productions, comme évoqué au Chapitre 1, Section 2.3) des productions. Dans ce cas, le geste inhibe la parole de la même manière que la parole inhibe le geste. Leur résultat sur le F2 peut également s'expliquer par le type de geste étudié, un geste symbolique, conventionnel, lexicalisé, qui apparaît le plus souvent seul (voir le Chapitre 1), et qui diffère donc clairement du geste déictique. Il semble donc nécessaire d'étudier les différents types de gestes indépendamment les uns des autres. Comme suggéré par Krahmer et Swerts [17] et par Roustan [28], chaque type de geste pourrait entretenir une relation spécifique à la parole qu'il accompagne.

Notre seconde expérience nous a permis de tester l'influence de la parole sur le geste manuel et donc de caractériser la direction de l'interaction geste/parole. Nos résultats sont en faveur d'une influence inhibitrice de la production de parole sur la production du geste manuel, de même nature donc que l'influence du geste sur la parole. Ces résultats sont cohérents avec ceux de Bernardis et Gentilucci [3] (présentées au Chapitre 1, Section 2.3) et de Barbieri et collab. [2], qui montraient une réduction des différentes phases du geste lorsque celui-ci était accompagné de parole. Dans notre expérience, les participants produisent un pointage manuel plus visible, afin de compenser l'absence de la modalité vocale : le geste manuel est renforcé en situation unimodale, où il dure plus longtemps, est maintenu plus longtemps et est plus lent. En revanche, l'ajout de la parole permet de nouveau d'équilibrer le contenu informatif sur les deux modalités disponibles (la production de geste est donc inhibée en situation bimodale).

15. « *This effect expresses the intention to interact closely with other individuals* » ([3], p.188).

Ces résultats mettent donc en évidence une distribution de l'information en fonction des modalités disponibles : le geste vocal varie en fonction de la présence/absence du geste manuel, et **vice-versa**. Nos données sont ainsi en faveur d'une influence structurelle, de nature coopérative, et réciproque, entre les deux systèmes de production. Toutefois, cette inhibition, observée en situation bimodale, pourrait être liée à une charge cognitive plus importante lorsque deux effecteurs sont contrôlés simultanément plutôt qu'un seul. Nous pouvons en effet supposer qu'une charge cognitive élevée occasionnerait une difficulté plus importante et donc une inhibition au niveau des productions. Or, nous soutenons ici Hostetter et Alibali [15], qui suggèrent que « *gesturing is the norm, and inhibiting gesture is an effortful task*¹⁶ » ([15], p.511). Plusieurs études montrent en effet que si l'on interdit aux participants de produire des gestes, lors d'une tâche d'explication (de type problème mathématique), leurs performances lors d'une seconde tâche (de type rappel verbal) sont altérées (e.g. Goldin-Meadow et collab. [12], Wagner et collab. [32]) ; cela suggère que s'empêcher de faire un geste monopolise en partie nos ressources cognitives. C'est donc la situation Parole Seule qui devrait induire la charge cognitive la plus importante et donc résulter en des productions inhibées. Nos résultats, qui montrent au contraire un renforcement des productions en situation unimodale, sont donc clairement en faveur d'une interaction structurelle particulière (sous la forme d'une coopération, où la quantité d'information véhiculée par la parole dépend de la quantité d'information véhiculée par le geste, et vice-versa) plutôt que la conséquence d'une charge cognitive trop importante.

Nos résultats, liés au geste de pointage, peuvent être intégrés aux modèles de production conjointe de gestes et de parole. Pour l'heure, la plupart de ces modèles s'intéressent à la production de gestes représentationnels, et plus particulièrement de gestes iconiques, métaphoriques et de pantomimes. Il nous paraît donc essentiel d'y intégrer progressivement la production du pointage et d'établir ainsi plus clairement son rôle dans la production de la parole et dans la communication en général. Nos résultats semblent compatibles avec le modèle d'interface, proposé par Kita et Ozyürek [16] et présenté au Chapitre 1. Ce modèle est le seul à envisager un échange d'informations **bidirectionnel**, de haut niveau, entre le geste et la parole. Plus précisément, leur modèle est basé sur l'existence d'un Planificateur de Communication, qui (1) génère l'intention communicative, (2) sélectionne la (ou les) modalité(s) qui seront utilisées, et (3) divise éventuellement la charge de travail en conséquence. L'intention communicative est ensuite envoyée aux Générateurs d'Actions et de Messages, qui vont échanger constamment de l'information, jusqu'à qu'une représentation d'interface émerge. De plus, le Générateur de Messages et le Formulateur sont inter-connectés, ce qui permet au geste manuel d'être façonné, en ligne, par les possibilités de formulation, ou d'encodage, linguistique (et par les propriétés spatio-motrices du référent). Ce modèle permet ainsi d'expliquer une ressemblance certaine entre la structure de la parole et la structure des gestes manuels qui l'accompagnent. Nos résultats montrent que le geste est déterminé par la parole, et sont, dans ce sens, compatibles avec le modèle de Kita et Ozyürek [16]. L'intérêt de nos travaux sur le pointage est qu'ils mettent en évidence une influence bidirectionnelle entre le geste et la parole, de plus bas niveau : certes, le geste est façonné en direct par les possibilités de formulation linguistique, mais la parole est également façonnée en direct par les possibilités de "formulation" manuelle.

16. Traduction proposée : Produire un geste est la norme, l'inhiber représente un effort.

4. Conclusion

Afin d'expliquer ce phénomène, nous proposons d'ajouter un système d'échange bidirectionnel direct permettant au Contrôle Moteur et au Formulateur d'échanger des informations (voir Figure 3.16).

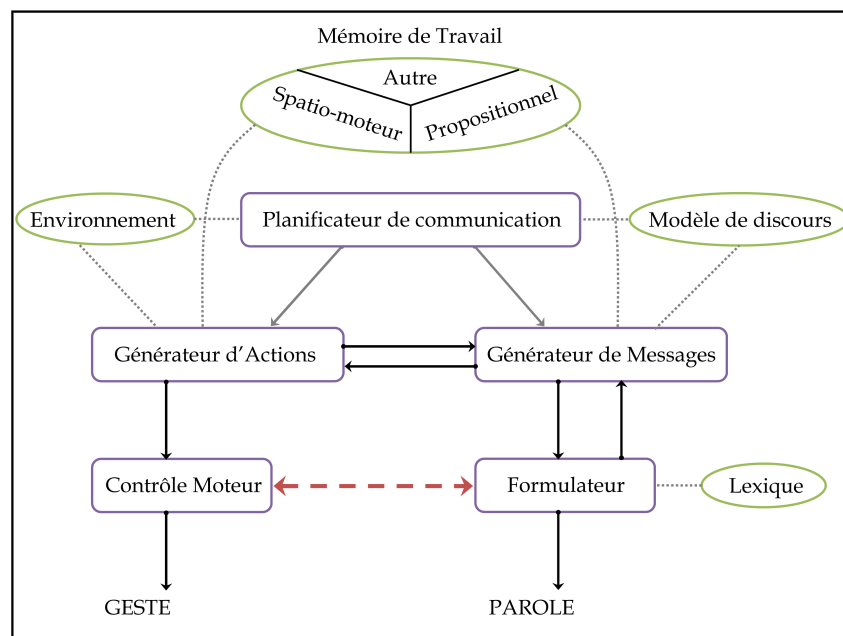


FIGURE 3.16 – Modèle d'interface pour la production geste/parole, adapté de Kita et Ozyürek [16], p.28. La flèche pointillée que nous avons ajoutée représente l'échange d'information bidirectionnel direct entre le Contrôle Moteur et le Formulateur

4 Conclusion

Les résultats de notre étude sont en faveur d'un couplage particulier entre geste/parole, à l'œuvre dans les mécanismes de production du langage, et plus précisément au niveau du processus déictique. Geste et parole apparaissent comme deux modalités langagières à part entière, étroitement liées, et sous l'influence d'une interaction structurelle de bas niveau, bidirectionnelle, et de nature inhibitrice/coopérative. Par ailleurs, le système moteur, vocal et manuel, pourrait bien être impliqué dans nos représentations linguistiques, liées à l'encodage lexical de la distance.

Néanmoins, ces résultats pourraient être améliorés, en augmentant le nombre de participants (afin de limiter la variance liée aux différences inter-sujets), et par le biais de procédures de normalisation sur nos données. En effet, bien que les analyses à mesures répétées permettent de contrôler en théorie les différences inter-individuelles, certaines procédures de normalisation permettraient de réduire considérablement la variance de nos échantillons et de favoriser l'apparition d'effets significatifs plus robustes. Les données articulatoires, par exemple, pourraient être normalisées par rapport à la distance de base entre les diodes situées sur les lèvres inférieure et supérieure, c'est-à-dire en soustrayant, à chacune des valeurs, la distance $D_{LS} - D_{LI}$ au repos. Dans le même registre, certaines procédures de normalisation sur les données acoustiques

permettent de prendre en compte les différences anatomiques entre les locuteurs (et notamment entre hommes et femmes), qui ont des répercussions importantes au niveau de la fréquence fondamentale et des valeurs de formants (i.e. des valeurs moyennes plus élevées lorsque le conduit vocal du locuteur est plus court). Nguyen et Espesser [24] ont évalué la qualité de trois procédures de normalisation acoustique, celles de Gerstman [11], Lobanov [20] et de Nearey [23], en calculant le rapport entre la dispersion intra-catégories et la dispersion inter-catégories. Selon les auteurs, une bonne procédure tendrait à minimiser ce rapport, c'est-à-dire qu'elle permettrait non seulement de réduire les variations observées au sein d'une même voyelle, mais également de préserver les différences entre des voyelles distinctes. Leurs résultats suggèrent que la procédure de Lobanov, qui consiste à rapporter chaque valeur de fréquence à sa moyenne et à son écart-type (afin que ces deux indicateurs soient identiques pour chaque locuteur et chaque formant), soit de manière générale la plus adaptée (toutefois, leurs données suggèrent que la qualité d'une procédure de normalisation dépende du système phonologique étudié).

Plusieurs perspectives, et notamment de nouvelles expériences, peuvent également être envisagées. Comme nous l'avons évoqué précédemment, de nouvelles analyses statistiques pourraient être effectuées. Il serait par exemple intéressant de tester l'effet de la Lexicalité sur les paramètres cinématiques du geste, à partir des données de notre première expérience ; ceci pourrait permettre de déterminer si les mots ont la même influence que les pseudo-mots sur la production du geste de pointage (l'étude de Bernardis et Gentilucci [3], par exemple, montre que seuls les items lexicaux influencent la production des gestes reliés sémantiquement). Il serait également intéressant de conduire des analyses en temporel sur nos deux corpus (e.g. comparer le durée de la voyelle à celle du plateau, comparer les temps d'initiation et de fin de parole à ceux du geste manuel), ce qui permettrait de proposer un modèle de pointage plus complet (seules des analyses préliminaires ont été effectuées à ce jour, dans le cadre d'un stage d'ingénieur 2A, Chadœuf [6]). Une expérience similaire pourrait également être proposée à des participants de diverses langues maternelles (plus particulièrement des langues pour lesquelles l'encodage lexical de la distance est basé sur un contraste autre que contraste voyelle ouverte/fermée), afin de tester plus précisément l'aspect moteur de l'encodage de la distance.

Quoi qu'il en soit, nos données mettent en évidence une interaction particulière entre parole, geste et langage, chez l'adulte. On peut donc se demander comment de tels comportements (i.e. un encodage moteur de la distance via les propriétés cinématiques et phonétiques du pointage ; une coopération inter-modalités) évoluent au cours du développement de l'enfant. En d'autres termes, ces comportements sont-ils des propriétés intrinsèques du langage ou bien des outils communicatifs disponibles chez le locuteur expert ? L'étude de ces comportements chez l'enfant permettrait l'élaboration d'un "agenda développemental" du pointage linguistique, dans les modalités vocale et manuelle et de faire ainsi le lien entre les données observées chez le tout-petit (au cours des deux premières années de vie) et les données observées chez l'adulte. Le chapitre suivant a donc pour objectif de tester notre second protocole expérimental chez l'enfant âgé de 6 à 12 ans, pour déterminer dans quelle mesure les performances observées chez l'adulte sont liées à des compétences précoces, soit des caractéristiques fondamentales du processus déictique, ou à des compétences plus tardives, soit des fonctions cognitives plus élaborées.

Bibliographie

- [1] Abdi, H. 2007, «Bonferroni and Sidàk corrections for multiple comparisons», dans *Encyclopedia of Measurement and Statistics*, édité par N. Salkind, Thousand Oaks, CA : Sage, p. 103–107.
- [2] Barbieri, F., A. Buonocore, R. Dalla Volta et M. Gentilucci. 2009, «How symbolic gestures and words interact with each other», *Brain and Language*, vol. 110, p. 1–11.
- [3] Bernardis, P. et M. Gentilucci. 2006, «Speech and gesture share the same communication system», *Neuropsychologia*, vol. 44, n° 2, p. 178–190.
- [4] Boersma, P. 2001, «Praat, a system for doing phonetics by computer», *Glott International*, vol. 5, n° 9/10, p. 341–345.
- [5] Bonfiglioli, C., C. Finocchiario, B. Gesierich, F. Rositani et V. Massimo. 2009, «A kinematic approach to the conceptual representations of *this* and *that*», *Cognition*, vol. 111, p. 270–274.
- [6] Chadœuf, A. 2010, «Analyse de la production gestuelle d'agents en interaction», Mémoire de stage d'ingénieur 2A, Université de Grenoble.
- [7] Chieffi, S., C. Secchi et M. Gentilucci. 2009, «Deictic word and gesture production : Their interaction.», *Behavioural Brain Research*, vol. 203(2), p. 200–206.
- [8] Childers, D. 1978, *Modern spectrum analysis*, IEEE Press : New-York.
- [9] Fukushi, M. 2008, *Le système vocalique chez les enfants bilingues français/japonais : Comparaison avec les monolingues*, mémoire de maîtrise, Université du Québec à Montréal.
- [10] Gentilucci, M., U. Castiello, M. Corradini, M. Scarpa, C. Umiltà et G. Rizzolatti. 1991, «Influence of different types of grasping on the transport component of prehension movements», *Neuropsychologia*, vol. 5, p. 361–378.
- [11] Gerstman, L. 1968, «Classification of self-normalized vowels», *IEEE Transactions of Audio Electroacoustics*, vol. AU-16, p. 78–80.
- [12] Goldin-Meadow, S., H. Nusbaum, S. Kelly et S. Wagner. 2001, «Explaining math : Gesturing lightens the load», *Psychological Science*, vol. 12, p. 516–522.
- [13] Gonseth, C., A. Vilain et C. Vilain. 2013, «An experimental study of speech/gesture interactions and distance encoding», *Speech Communication*, vol. 55(4), p. 553–571.
- [14] Greenhouse, S. et S. Geisser. 1959, «On methods in the analysis of profile data», *Psychometrika*, vol. 24, p. 95–112.
- [15] Hostetter, A. et M. Alibali. 2008, «Visible embodiment : Gestures as Simulated Action», *Psychonomic Bulletin & Review*, vol. 15, n° 3, p. 495–514.
- [16] Kita, S. et A. Ozyürek. 2003, «What does cross-linguistic variation in semantic coordination of speech and gesture reveal ? : Evidence for an interface representation of spatial thinking and speaking», *Journal of Memory and Language*, vol. 48, n° 1, p. 16–32.

- [17] Krahmer, E. et M. Swerts. 2007, «The effects of visual beats on prosodic prominence : Acoustic analyses, auditory perception and visual perception», *Journal of Memory and Language*, vol. 57, n° 3, p. 396–414.
- [18] Lehiste, I. et G. Peterson. 1961, «Transitions, glides, and diphthongs», *The Journal of the Acoustical Society of America*, vol. 33, p. 268–277.
- [19] Lindblom, B., J. Lubker et T. Gay. 1979, «Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation», *Journal of Phonetics*, vol. 7, p. 147–161.
- [20] Lobanov, B. 1971, «Classification of Russian vowels spoken by different speakers», *Journal of the Acoustical Society of America*, vol. 49, p. 606–608.
- [21] Maeda, S. 1991, «On articulatory and acoustic variabilities», *Journal of Phonetics*, vol. 19, p. 321–331.
- [22] Mauchly, J. 1940, «Significance test for sphericity of a normal n -variate distribution», *The Annals of Mathematical Statistics*, vol. 11, p. 204–209.
- [23] Nearey, T. 1977, *Phonetic Feature Systems for Vowels*, thèse de doctorat, University of Alberta.
- [24] Nguyen, N. et R. Espesser. 2004, «Méthodes et outils pour l’analyse acoustique des systèmes vocaliques», *Bulletin PFC (Phonologie du Français Contemporain)*, vol. 3, n° 3, p. 77–85. 2182 2182 Programme interdisciplinaire CNRS TCAN.
- [25] Press, W., S. Teukolsky, W. Vetterling et B. Flannery. 1992, *Numerical recipes in C : The art of scientific computing*, Cambridge University Press.
- [26] Reboud, E. 2008, «*Etude des relations voix/geste dans le pointage déictique, effet de la distance de l’objet désigné*», mémoire de master, Institut National Polytechnique de Grenoble.
- [27] Rochet-Capellan, A. 2007, *De la substance à la forme : Rôle des contraintes motrices orofaciales et brachiomanuelles de la parole dans l’émergence du langage*, thèse de doctorat, Institut National Polytechnique de Grenoble.
- [28] Roustan, B. 2012, *Étude de la coordination gestes manuels/parole dans le cadre de la désignation*, thèse de doctorat, Université de Grenoble.
- [29] Sapir, E. 1929, «A study in phonetic symbolism», *Journal of Experimental Psychology*, vol. 12, p. 225–39.
- [30] Shapiro, S. 1965, «An analysis of variance test for normality (complete samples)», *Biometrika*, vol. 52, p. 591–611.
- [31] Vasilescu, I., M. Candea et M. Adda-Decker. 2004, «Hésitations autonomes dans 8 langues : Une étude acoustique et perceptive», dans *Identification des Langues et des Variétés Dialectales par les Humains et par les Machines*, p. 25–30.

Bibliographie

- [32] Wagner, S., H. Nusbaum et S. Goldin-Meadow. 2004, «Probing the mental representation of gesture : Is handwaving spatial ?», *Journal of Memory and Language*, vol. 50, p. 395–407.
- [33] Winer, B., R. Brown et K. Michels. 1991, *Statistical Principles in Experimental Design*, New York : McGraw-Hill.

Interaction geste, parole, langage et encodage de la distance chez l'enfant

L'objectif du présent chapitre est d'étudier l'évolution du couplage geste/parole au cours du développement de l'enfant. Si les chapitres précédents nous ont permis de mettre en évidence une interaction particulière entre parole, geste et langage chez l'adulte, nous cherchons désormais à comprendre ce qu'il en est chez le bébé, puis chez l'enfant plus âgé, et quel peut être le rôle d'une telle interaction dans la mise en place et l'acquisition du langage.

La littérature récente sur ce thème distingue, au cours du développement de l'enfant, un couplage voix/geste purement moteur, présent dès la naissance, d'un couplage voix/geste fonctionnel, mis en place progressivement au cours des premières années de vie. Ces éléments sont inclus dans (1) le modèle développemental d'Iverson et Thelen [36], qui décrit l'évolution du couplage voix/geste au cours des 18 premiers mois du bébé, et (2) la littérature consacrée aux enfants plus âgés, et notamment à l'étude de leurs productions bimodales.

En revanche, peu de données concernant l'interaction parole/geste chez l'enfant d'âge scolaire, soit de 6 ans et plus, sont disponibles. Le couplage continue-t-il d'évoluer pendant l'enfance ? Et comment évolue-t-il jusqu'à l'âge adulte ? Nous formulerons certaines hypothèses sur la base de changements observés, aux alentours de 7-8 ans, dans la modalité vocale d'une part, et dans la modalité manuelle d'autre part ; quelles conséquences ces changements au niveau unimodal peuvent-ils avoir sur l'interaction entre le geste et la parole ?

Notre objectif est d'appliquer notre protocole expérimental, testé précédemment auprès d'adultes, à l'enfant, âgé de 6 à 12 ans, afin de combler un "vide" certain entre les études réalisées chez le tout-petit et celles réalisées chez l'adulte. Plus précisément, nous souhaitons déterminer dans quelle mesure les performances observées chez l'adulte dans les Expériences 1 & 2, présentées au Chapitre 3, sont liées à des compétences précoces, soit des caractéristiques fondamentales du processus déictique, ou à des compétences plus tardives, soit des fonctions cognitives plus élaborées. Ceci permettra, à long terme, d'élaborer un "agenda développemental" du pointage linguistique, via les modalités vocale et manuelle.

1 Evolution du couplage voix/geste au cours du développement

1.1 Un couplage oro-manuel moteur présent dès la naissance : Modèle de l'émergence dynamique d'Iverson & Thelen (1999)

Le geste de pointage semble jouer un rôle clé dans l'émergence du langage : lié notamment à l'établissement d'un cadre d'attention conjointe, il sous-tend la construction du lexique (e.g. Bruner [8]) ainsi que l'émergence de la morphosyntaxe (e.g. Volterra et collab. [64]) et constitue en cela une modalité privilégiée d'entrée dans le langage. Mais le lien geste/parole semble s'étendre bien au-delà du processus déictique. Nombre d'études, détaillées dans la revue proposée par Iverson [33], mettent en effet en évidence une relation toute particulière entre le développement moteur en général et le développement langagier : globalement, l'émergence de nouvelles capacités motrices pourrait faciliter le développement de capacités communicatives et langagières (notamment par la pratique de mouvements nécessaires ultérieurement au langage parlé, ou par l'apparition de nouvelles possibilités d'action sur notre entourage).

Iverson et Thelen [36] proposent un modèle développemental permettant d'expliquer la mise en place du lien geste/parole au cours du développement par un couplage précoce (dès la naissance), dynamique et évolutif (jusqu'à 18 mois) entre le système manuel d'une part, et le système oral d'autre part. Leur modèle se base sur la théorie des systèmes dynamiques pour le contrôle moteur, relative à la production d'actions adaptatives (i.e. patterns ou motifs d'action), réalisées grâce à la coordination de différents composants, ou oscillateurs, au sein d'un même système (voir, entre autres, les travaux de Kugler et Turvey [41] et de Kelso [40]). Selon Von Holst [65], une telle coordination, entre les mouvements rythmiques de différents composants d'un même système, serait régie par différents principes¹ :

1. Fréquence préférentielle : Chaque oscillateur présente un rythme de base lorsqu'il entre en action seul ;
2. Effet de superposition (*Superposition effect*) : L'oscillation (ou le rythme) de l'un des oscillateurs peut être décelée dans l'oscillation du second oscillateur ;
3. Effet d'aimantation (*Magnet effect*) : Chaque oscillateur tente d'entraîner l'autre oscillateur, en lui imposant son propre rythme de mouvement. Il en résulte un rythme dit coopératif, correspondant à un équilibre entre les deux rythmes initialement concurrents ;
4. Effet de maintenance (*Maintenance effect*) : Chaque oscillateur tente de maintenir sa fréquence préférentielle lorsqu'il est couplé au second. Cela conduit à des variations autour du rythme coopératif moyen. Cet effet illustre la nature dynamique du couplage inter-oscillateurs : il existe une tension entre, d'une part, le maintien du rythme de base et, d'autre part, la force d'entraînement par l'autre oscillateur.

Adaptée à l'homme, cette théorie a pour but d'expliquer la façon dont les différentes parties de notre corps se coordonnent pour réaliser diverses actions. En d'autres termes, la question est de savoir « *How do people [...] so precisely move their limbs in time and space to walk, run, or manipulate objects ?*² » (Iverson et Thelen [36], p.27).

1. L'auteur s'est basé sur l'observation des oscillations des nageoires chez le poisson.

2. Traduction proposée : Comment les êtres humains [...] arrivent-ils à déplacer leurs membres si précisément dans le temps et l'espace, de façon à pouvoir marcher, courir, ou manipuler des objets ?

1. Evolution du couplage voix/geste au cours du développement

Kugler et Turvey [41] montrent par exemple que les mouvements d'oscillation de chaque bras s'adaptent et s'entraînent les uns les autres pour produire un comportement coordonné, dont le rythme final est un compromis entre les rythmes initiaux de chacun des deux bras (i.e. *Magnet effect*, selon la définition de Von Holst [65]). Iverson et Thelen [36] proposent de la même façon que le système manuel et le système oral soient régis par des principes similaires, soit par un entraînement mutuel, dynamique et flexible ; les auteurs considèrent ainsi la main et la voix comme deux oscillateurs, ou deux composants, au service d'un seul et même système.

Plus précisément, le modèle d'Iverson et Thelen [36] a pour objectif d'expliquer comment les systèmes de productions manuelles et vocales évoluent conjointement pour aboutir au modèle adulte, c'est-à-dire à un couplage entre deux modalités, sous la forme de productions bimodales synchrones (voir à ce sujet les travaux de McNeill [48] et la Section 2.2 du Chapitre 1). Cette progression est illustrée Figure 4.1 ci-dessous.

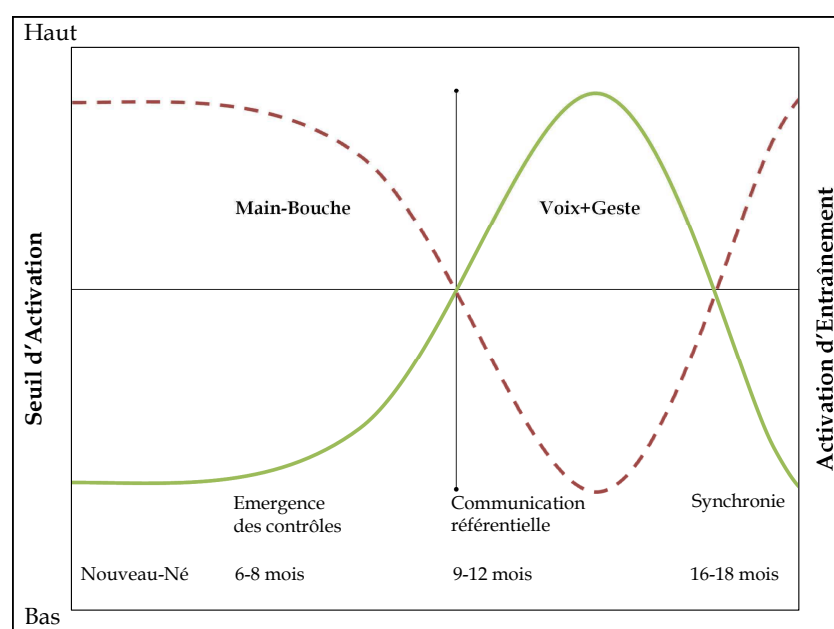


FIGURE 4.1 – Seuil (*trait plein*) et force d'activation d'entraînement (*trait pointillé*) dans le système oro-manuel au cours des premières années (Iverson et Thelen [36], p.31)

Deux concepts sont ici essentiels : le seuil d'activation (représenté en trait plein sur la Figure 4.1) et l'activation d'entraînement (représentée en trait pointillé).

Le seuil d'activation représente la facilité d'exécution d'un comportement donné et est à priori lié à sa fréquence d'occurrence. Ainsi, un comportement fréquent, utilisé dans des contextes variés, aura un seuil bas, tandis qu'un comportement nécessitant des efforts (notamment un comportement rare ou nouveau) aura un seuil haut. L'activation d'entraînement (ou niveau, force d'activation) varie quant à elle de manière inverse : haute pour un comportement stable (au sens de maîtrisé), basse pour un comportement plus complexe. L'entraînement mutuel entre les deux systèmes, c'est-à-dire le couplage dynamique, ne sera possible que pour des niveaux d'activation élevés, soit pour des comportements fréquents et/ou contrôlés.

L'évolution du couplage geste/parole se fait en quatre étapes, que nous allons détailler et illustrer par la suite : (1) les liens initiaux, présents dès la naissance, (2) l'émergence du contrôle, aux alentours de 6-8 mois, (3) l'émergence des premiers gestes et mots (ou couplage flexible), aux alentours de 9-14 mois, puis (4) l'émergence de la synchronisation entre geste et parole (ou couplage synchrone), aux alentours de 16-18 mois. La première étape correspond à un couplage purement moteur entre le système oral et le système manuel, tandis qu'un couplage fonctionnel entre voix et geste apparaît progressivement au cours de la seconde année (mais nous y reviendrons dans la section suivante, consacrée à l'étude des productions bimodales).

1.1.1 Lien initiaux (*Initial linkages*)

Les liens initiaux, présents dès la naissance, correspondent à un couplage moteur entre les systèmes de productions oral et manuel, comme en témoigne notamment le réflexe de Babkins : une pression sur la paume du nourrisson induit une ouverture de sa bouche. Par ailleurs, les productions conjointes et spontanées d'activités orales et manuelles, sous forme de contact main-bouche (*hand-to-mouth*), sont fréquentes (Butterworth [10]) : à l'âge de 2 mois par exemple, le bébé commence à porter les objets à la bouche pour les explorer oralement (Lew et Butterworth [43]), une activité qui occupera d'ailleurs la majorité de son temps, et ce tout au long de la première année !

La fonction de ces co-activations main-bouche va petit à petit évoluer et celles-ci vont apparaître dans des situations d'interaction sociale, aux alentours de 2-3 mois. Fogel et Hannan [20] observent par exemple, chez des bébés âgés de 9 à 15 semaines, une relation systématique entre certains comportements manuels d'une part et certaines activités orales d'autre part : l'apparition conjointe de mouvements manuels et de vocalisations est notable, et l'extension de l'index a par exemple de fortes chances d'apparaître en même temps que les mouvements buccaux ou vocalisations.

Ces co-occurrences de comportements oraux et manuels constituent alors à cet âge la majeure partie du répertoire comportemental du bébé ; elles témoignent ainsi d'un couplage précoce entre les activités manuelles discrètes, d'une part, et les activités orales, d'autre part. Le couplage main/bouche est donc considérable bien avant l'émergence des premiers gestes volontaires, au sens de gestes contrôlés, et des premiers mots³. De fait, au cours des premiers mois de la vie, le seuil de ces activités main-bouche est bas (puisque ce sont des comportements fréquents) et leur activation d'entraînement est en conséquence élevée (cf. Figure 4.1 ci-dessus). Ce pattern évoluera progressivement pour s'inverser aux alentours de 9-12 mois, période de l'émergence de la communication référentielle (ou stade du couplage flexible) ; seule la pratique permettra, à terme, de (ré)inverser cette tendance (stade du couplage synchrone).

3. Iverson et Thelen [36] formulent, non sans précaution, l'hypothèse d'un établissement phylogénétique du lien précoce main-bouche : il résulterait de mécanismes couplant manipulation et alimentation, deux fonctions bien différentes du langage ! C'est également ce que suggèrent Bates et collab. [5], pour qui « *language [...] must be viewed as a new machine built out of old parts* » ([5], p.35) ; i.e. le langage [...] devrait être considéré comme une machinerie moderne, constituée de pièces anciennes.

1. Evolution du couplage voix/geste au cours du développement

1.1.2 Emergence du Contrôle (*Emerging control*)

Vers l'âge de 3-4 mois, on observe une réorganisation du système oro-manuel, avec l'émergence du contrôle volontaire des productions, du système main-bras et du système oro-vocal. L'utilisation des mains et de la bouche devient peu à peu adaptative et l'on voit apparaître des activités rythmiques dans les deux modalités.

A cette période, les gestes manuels deviennent en effet plus contrôlés, comme l'illustrent les mouvements d'atteinte et de saisie, dirigés vers un but. Un changement similaire peut également s'observer au niveau des productions orales, avec l'apparition du *cooing*, ou Stade II, selon la classification de van Beinum et collab. [7]⁴, qui correspond à la production d'« *interrupted phonation within one breath unit without any articulatory movement*⁵ », et la production des premières voyelles, notamment pendant les interactions sociales.

D'autre part, la production de mouvements rythmiques du bras et de la main (tels que agiter, balancer, taper) et de babillage manuel (ou "silencieux"), observé chez l'enfant sourd et entendant (Meier et Willerman [49], Petitto et Marentette [57]), augmente considérablement aux alentours de 26-28 semaines. Le babillage manuel consiste en la production de mouvements manuels répétitifs, effectués dans une zone limitée, située à l'avant du corps (qui correspond d'ailleurs à l'espace phonétique des langues signées) et qui respectent les schémas rythmiques propres aux langues orales (Petitto et collab. [56])⁶. A cette même période, une activité similaire apparaît dans la modalité vocale, le babillage canonique, soit la production rythmique de syllabes redupliquées, telles que [bababababa], qui survient en moyenne à 27 semaines (Oller et Eilers [52]).

L'apparition du babillage canonique semble donc liée dans le temps à l'évolution des schémas rythmiques, observée dans la modalité manuelle : le taux d'oscillations des mains est relativement faible chez l'enfant pré-babilleur, augmente chez le babilleur novice, et finalement décline chez le babilleur expert, tout en demeurant plus élevé que chez le pré-babilleur (Locke et collab. [45]). Selon Iverson et Thelen [36], les activités rythmiques orales et manuelles devraient alors s'entraîner mutuellement, puisque le seuil de leur co-activation est bas et leur niveau d'activation d'entraînement est élevé (Ejiri [18], cité par Iverson et Thelen [36], montre à ce propos que 40% des activités rythmiques manuelles apparaissent en présence de babillage canonique, 75% du babillage canonique apparaît conjointement aux activités rythmiques manuelles) ; en effet la durée des voyelles est plus longue lorsque la main est active (Ejiri et Masataka [19]), preuve d'un compromis entre les deux modalités (au minimum, ces données nous permettent d'affirmer que la parole est entraînée par la main, influencée par la présence du geste).

4. Il existe bon nombre de classifications des stades de développement de la parole au cours de la première année de vie, dont Gillis et collab. [22] proposent une revue. La classification de van Beinum et collab. [7] se base sur une approche sensori-motrice plutôt que phonético-linguistique ; en d'autres termes, elle ne décrit pas les productions enfantines relativement à la phonologie adulte, mais se base sur la description des mouvements phonatoires d'une part, et articulatoires d'autre part.

5. Traduction proposée : Phonation interrompue au cours d'un même cycle respiratoire, sans mouvement articulatoire.

6. Plusieurs hypothèses sont formulées quant au babillage manuel, qui consiste par exemple en un contact rythmique de l'index droit sur la paume gauche : peut-il être considéré comme l'analogue du babillage vocal et ainsi témoigner d'une faculté d'acquisition du langage amodale (Petitto et Marentette [57]) ? Ou est-il un "simple" comportement rythmique transitoire parmi tant d'autres (Meier et Willerman [49]) ?

Afin d'expliquer ce phénomène, Iverson et Thelen [36] proposent que la production d'activités répétitives et rythmiques manuelles favoriserait la production d'activités rythmiques en général, i.e. plurimodales (par exemple celles du tronc, des jambes, etc.), et donc d'activités rythmiques orales. Ainsi la production d'activités rythmiques manuelles faciliterait la production d'oscillations mandibulaires et, de fait, de babillage canonique.

Ces activités rythmiques peuvent également être considérées comme des activités transitoires, permettant l'émergence d'un contrôle plus élaboré des productions manuelles et vocales, contrôle nécessaire ultérieurement pour la production de gestes et de mots signifiants. En d'autres termes, la production de babillage, dans les deux modalités, va permettre à l'enfant d'affiner progressivement le contrôle de ses "articulateurs" oro-manuels. Dans cet esprit, Petitto et Marentette [57] proposent que « *the similarities between manual and vocal babbling suggest that babbling is a product of an amodal, brain-based language capacity under maturational control, in which phonetic and syllabic units are produced by the infant as a first step toward building a mature linguistic system*⁷ » ([57], p.1493).

1.1.3 Couplage flexible (*Flexible couplings*)

A la fin de la première année, les activités rythmiques vocales et manuelles, dont le babillage, laissent peu à peu place à des activités plus différenciées. On observe alors l'apparition de gestes communicatifs (tels que le pointage et la requête) mais également de mots, ou du moins de sons de parole dits *word-like* (c'est-à-dire des sons de parole systématiquement utilisés par un enfant particulier pour désigner un objet particulier, comme par exemple [apo] pour « chapeau »). On passe donc de comportements rythmiques et répétitifs à des comportements plus contrôlés et à visée clairement communicative.

A ce stade, les enfants produisent leurs premiers gestes référentiels plusieurs semaines avant de produire leurs premiers mots ; plus précisément, ils produisent un geste pour un objet particulier environ trois mois avant de produire le mot correspondant à cet objet (Bates et collab. [4], Caselli [12]). Les gestes sont donc plus nombreux que les mots dans le répertoire communicatif de l'enfant, qui semble montrer une nette "préférence" pour la modalité gestuelle lors d'interactions spontanées (Iverson et collab. [34]). Iverson et Goldin-Meadow [35], ainsi que Volterra et collab. [64]⁸, montrent effectivement une prédominance du geste aux alentours de 10-14 mois [35] et de 16 mois [64], suivie d'une prédominance de la parole, aux alentours de 17-24 mois et de 20 mois (phénomène illustré Figure 4.2 ci-après). Le fait qu'un objet soit désigné en premier lieu avec un geste permet de prédire l'apparition d'items lexicaux dans le répertoire vocal de l'enfant ; autrement dit, « *gesture may thus serve as a transitional device in early lexical development*⁹ » (Iverson et Goldin-Meadow [35], p.369).

7. Traduction proposée : L'analogie entre babillage vocal et babillage manuel suggère que le babillage, en général, est le produit d'une capacité langagière amodale, en cours de développement, où la production d'unités phonétiques et syllabiques permettrait à l'enfant de construire progressivement un système linguistique abouti.

8. Iverson et Goldin-Meadow [35] et Volterra et collab. [64] utilisent la méthode longitudinale (notamment décrite par Tourrette [63]), basée sur l'observation répétée d'échantillons appariés, à différents âges, sur les mêmes variables dépendantes (l'âge étant la variable indépendante principale). Elle permet donc de comparer un individu à lui-même, mais nécessite de longues périodes d'observation (dont le problème majeur est la "mort expérimentale", due par exemple à l'abandon de certains participants).

9. Traduction proposée : Le geste pourrait servir d'outil transitionnel, au cours du développement lexical précoce.

1. Evolution du couplage voix/geste au cours du développement

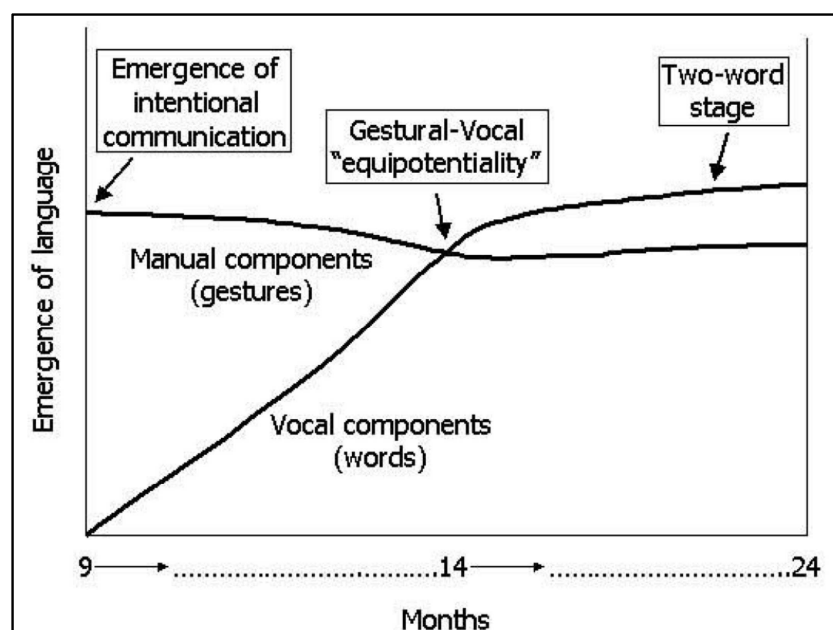


FIGURE 4.2 – Développement ontogénétique du langage (Volterra et collab. [64], p.22) : la stratégie vocale gagne peu à peu du terrain sur la stratégie manuelle, sans pour autant que cette dernière soit supprimée du répertoire de l'enfant.

Cette asymétrie au niveau de l'activation et du contrôle de ces deux effecteurs (i.e. les gestes communicatifs sont fréquents, tandis que la parole se développe plus lentement) peut s'expliquer par un contrôle des "articulateurs" manuels plus avancé que celui des articulateurs vocaux ; en d'autres termes, la parole serait retardée, relativement au geste, car elle nécessite plus d'efforts et, de fait, une charge cognitive plus élevée (Volterra et collab. [64]). Le seuil des gestes est donc à cette période relativement faible, mais, à l'inverse, le seuil de la parole est relativement élevé (cf. Figure 4.1 présentée en début de section).

Néanmoins, malgré cette différence en termes de seuil et de niveau d'activation, les deux systèmes restent intimement liés. On observe ainsi, entre 9 et 13 mois, des gains dans le développement du langage liés à la production de gestes manuels : Bates et collab. [4] montrent par exemple que les enfants qui gesticulent le plus sont également les plus précoces en termes d'acquisition de mots nouveaux. A cette même période, la pratique (i.e. entraîner l'enfant à faire des gestes) favorise d'ailleurs l'amélioration des performances langagières (Goodwyn et Acredolo [25, 26]), et parallèlement, un retard au niveau du langage (en termes de pauvreté de vocabulaire) peut être prédit par un nombre plutôt faible de gestes communicatifs (Thal et Tobias [62]).

1.1.4 Couplage synchrone (*Synchronous coupling*)

L'apprentissage de nouveaux mots, ainsi que la pratique de comportements communicatifs dans les deux modalités, va progressivement faire diminuer le seuil des comportements verbaux et augmenter la force de leur activation d'entraînement. Petit à petit, les productions parole/geste vont donc se synchroniser temporellement, pour devenir proches de celles des adultes, aux alentours de 16-18 mois.

Plus précisément, les productions se synchronisent lorsque l'enfant combine un geste et un mot sémantiquement reliés. Butcher et Goldin-Meadow [9] montrent en effet que les premières "combinaisons" voix/geste (i.e. vocalisations non significatives + geste) sont réalisées plutôt séquentiellement et ne présentent pas de profil temporel clairement établi ; en revanche, dès lors que geste et parole forment un système sémantique unifié (mot signifiant + geste communicatif), les deux types de production vont être synchronisés. Autrement dit, la synchronisation temporelle entre geste et parole pourrait bien être liée à la mise en place du système sémantique.

Pour Iverson et Thelen [36], la période d'apprentissage de mots nouveaux est donc cruciale, puisqu'elle va permettre un entraînement mutuel entre le geste et la parole (i.e. un couplage parole/geste synchrone). En effet, c'est à cette période que la force d'activation d'entraînement de la parole atteint son paroxysme (notamment par la pratique de nouveaux mots et l'effort que celle-ci implique), ce qui lui permet de capturer et d'activer le système gestuel simultanément. Cette co-activation résulte en la production de combinaisons synchrones : deux effecteurs distincts, mais au service d'une seule et même intention communicative.

1.2 L'émergence d'un couplage parole/geste fonctionnel

Le couplage geste/parole, initialement moteur, évolue progressivement vers un couplage plus fonctionnel. Les combinaisons bimodales (dont un exemple est illustré Figure 4.3) semblent en effet jouer un rôle particulier dans certains mécanismes cognitifs, tels que le langage ou les capacités de raisonnement.



FIGURE 4.3 – Production conjointe de geste et de parole chez l'enfant d'âge préscolaire : utilisation d'une combinaison complémentaire « chat » + [Pointage vers le chat] (© C. Gonseth)

Volterra et collab. [64] ont établi une typologie de ces combinaisons geste/parole :

- Combinaisons **équivalentes ou redondantes** : geste et parole véhiculent une seule et même information ; e.g. « Oui » + [Hochement de tête]
- Combinaisons **complémentaires** : le geste permet la désambiguïsation du message (procédé de *nominalisation*) mais n'ajoute pas d'information ; e.g. « Ça » + [Montre un objet]

1. Evolution du couplage voix/geste au cours du développement

- Combinaisons **supplémentaires** : le geste permet l'extension du message (procédé de *prédication*), en ajoutant une information non véhiculée par la parole ; e.g. « Tombé » + [Montre un objet]. Les combinaisons supplémentaires permettent de *dire quelque chose à propos de quelque chose* et constituent en cela un premier pas vers la syntaxe.

De nombreuses études se sont intéressées à ces productions, dans le contexte linguistique particulier de l'énoncé à un mot (aux alentours de 10-14 mois) et à deux mots (aux alentours de 17-23 mois). Ce dernier met en évidence la faculté de combiner deux symboles linguistiques et est ainsi une étape particulièrement importante du processus d'apprentissage du langage. Cette progression, d'un à deux symboles, s'observe indépendamment de la culture et du langage ambiant et serait ainsi un trait universel d'apprentissage du langage (Caselli et collab. [13]). Ces énoncés à deux mots surviennent en moyenne deux-trois mois après les combinaisons non redondantes (complémentaires et supplémentaires) ; alors, si de telles combinaisons parole/geste précèdent l'entrée dans la syntaxe, peut-on en déduire qu'elles l'annoncent ?

Goldin-Meadow et Butcher [23] suggèrent que ces combinaisons non redondantes forment potentiellement un "pont transitionnel" vers la parole à deux mots. Les auteurs observent des enfants longitudinalement, entre 12 et 27.5 mois, et montrent une corrélation positive, forte et fiable ($r = .82$; $r^2 \approx 0.67$, soit $\approx 67\%$ de variance expliquée) entre l'âge de début des combinaisons geste/parole supplémentaires et l'âge de début des énoncés à deux mots ; en d'autres termes, les premiers enfants à produire des combinaisons supplémentaires sont également les premiers à produire des énoncés à deux mots. Une telle corrélation permet de prédire le comportement langagier des enfants, à un âge donné, à partir de leurs productions bimodales. Un lien particulier entre le geste et la parole devient alors annonciateur de changements au niveau du système linguistique.

Iverson et Goldin-Meadow [35] confirment ces résultats. Les auteurs montrent que (1) les combinaisons redondantes surviennent quatre-cinq mois avant les énoncés à deux mots, et (2) les combinaisons supplémentaires surviennent environ deux mois avant. La corrélation entre l'apparition de combinaisons supplémentaires et celle d'énoncés à deux mots est de nouveau positive, forte et fiable ($r = .94$; $r^2 \approx 0.88$, soit $\approx 88\%$ de variance expliquée) ; en revanche, elle ne l'est pas pour les combinaisons redondantes ($r = .24$, *NS*) ! C'est donc bien l'association syntaxique entre un geste et un mot, et non la "simple" combinaison geste/parole, qui permet de prédire l'âge d'apparition des énoncés multi-mots.

Dans le même registre, Özçalışkan et Goldin-Meadow [53] montrent que le geste permet à l'enfant d'élargir son répertoire communicatif et de véhiculer des idées complexes avant de pouvoir le faire en parole seule. Les auteurs montrent en effet que les combinaisons complexes de types argument + prédicat sont présentes dès 18 mois sous forme de combinaisons geste/parole mais n'apparaissent en parole seule qu'aux alentours de 22 mois.

Par ailleurs, le couplage geste/parole semble également jouer un rôle dans les processus d'acquisition de connaissances et de raisonnement.

Church et Goldin-Meadow [14], ainsi qu'Alibali et Goldin-Meadow [1], ont par exemple étudié les productions de geste et de parole d'enfants d'âge scolaire (respectivement 28 enfants âgés de 6 à 8 ans, et 90 enfants âgés d'en moyenne 10 ans), pendant la résolution de tâches de conservation piagétienne (illustrées en Annexe) et d'équivalence mathématique. Outre le nombre

important de gestes iconiques, les auteurs observent, dans leurs études respectives, deux types de combinaisons geste/parole :

- Les combinaisons **concordantes**, où geste et parole véhiculent une seule et même information ;
- Les combinaisons **discordantes**, où le geste véhicule une information différente, voire incohérente/contradictoire, de celle véhiculée simultanément par la parole. Par exemple, l'enfant exprime verbalement une inégalité (« *il y a plus de jetons dans la première rangée* ») mais relie, par un geste, les jetons de chaque rangée deux à deux (cf. Figure 4.4) ; son message verbal est incorrect alors que son message manuel est correct.

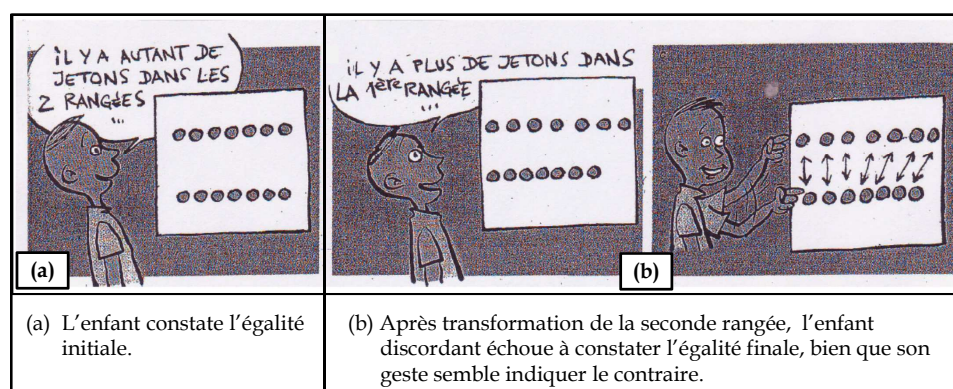


FIGURE 4.4 – Exemple de réponse discordante lors de la résolution d'une tâche de conservation piagétienne (tiré du cours de M.P. Lebaud, Université de Rennes 1)

De manière générale, ces enfants, discordants, semblent bénéficier davantage des explications d'un adulte : ils s'améliorent, ajoutent de nouvelles explications d'équivalence à leur répertoire (Church et Goldin-Meadow [14]) et sont également capables de généraliser leurs connaissances à des problèmes similaires et de les maintenir dans le temps (Alibali et Goldin-Meadow [1]). La discordance entre le geste et la parole apparaît alors comme une étape majeure dans le développement cognitif de l'enfant et serait une caractéristique générale du nouvel apprenant (Alibali et Goldin-Meadow [1]) : elle montre que celui-ci est sur le point de franchir une nouvelle étape, qu'il est prêt à recevoir les explications d'un adulte pour progresser (« *child's readiness to learn*¹⁰ », Kelly et collab. [38], p.8.). Elle serait, selon Alibali et Goldin-Meadow [1], la conséquence de l'activation simultanée de plusieurs stratégies en compétition (un phénomène d'ailleurs caractéristique des phases de transition développementale). Cette activation multiple devrait engendrer un traitement cognitif plus coûteux et, de fait, moins de ressources attentionnelles résiduelles ; c'est le cas puisque les enfants discordants, relativement aux enfants concordants, réalisent de moins bonnes performances lors de tâches concurrentes (e.g. tâche de rappel verbal, dans l'étude de Goldin-Meadow et collab. [24]). Les auteurs font alors l'hypothèse d'un modèle d'apprentissage en trois étapes successives : (1) la formulation d'une hypothèse initiale incorrecte, (2) une transition avec de multiples hypothèses, véhiculées par les modalités vocales et manuelles (plus précisément, l'enfant véhicule par le geste ce qu'il n'est encore pas capable de formuler verbalement), et finalement (3) une hypothèse finale correcte.

10. Traduction proposée : Propension de l'enfant à apprendre.

1. Evolution du couplage voix/geste au cours du développement

Bien que l'ensemble de ces études semblent indiquer un rôle particulier du couplage geste/parole dans certains processus cognitifs, tels que les capacités langagières, de raisonnement et d'apprentissage, la question de la nature d'un tel couplage chez l'enfant d'âge scolaire, et notamment celle de l'interaction inter-modalités dans le processus déictique, reste assez largement inexplorée.

1.3 L'évolution des conduites motrices et verbales à l'âge scolaire

Les données présentées dans les sections précédentes ont mis en évidence l'existence d'un couplage moteur, précoce et dynamique entre le système oral et le système manuel chez le tout-petit. Ce couplage évolue progressivement vers un couplage geste/parole plus fonctionnel, qui semble impliqué dans la mise en place et le développement de certaines capacités cognitives.

Capone et McGregor [11] proposent un calendrier développemental, pour la production de gestes en lien avec la parole, qui résume relativement bien les données citées jusqu'à présent (en couvrant la période 10 mois - âge scolaire) ; nous l'avons toutefois complété, en y ajoutant deux points, concernant (1) les premiers mois de vie et (2) la période scolaire (en caractères gras et en italique dans le Tableau 4.1 ci-dessous).

Période considérée	Comportements main/bouche
<i>0-18 mois</i>	<i>Couplage évolutif et dynamique des systèmes oro-manuel</i>
10-13 mois	Apparition du pointage et prédiction des premiers mots
12-13 mois	Emergence des premiers mots ; fonction complémentaire du geste
15-16 mois	Préférence gestuelle ou vocale
28-20 mois	Préférence pour les mots parlés, utilisation de combinaisons bimodales, augmentation significative des mots et des pointages combinés à la parole
2-5 ans	Intégration geste/parole
Age scolaire	Le geste comme transition vers l'acquisition des concepts <i>Hypothèse envisagée : évolution des productions vers un modèle de production adulte, par acquisition et maîtrise progressive de nouvelles stratégies, i.e. "peaufinage".</i>

Tableau 4.1 – Calendrier développemental pour la production de gestes et de parole (adapté de Capone et McGregor [11], p.183)

Si les données quant à l'interaction parole/geste et son évolution sont nettement plus rares lorsqu'il s'agit d'enfants d'âge scolaire¹¹, on pourrait supposer néanmoins que les conduites multimodales évoluent tout au long de l'enfance et de l'adolescence, en direction des usages attestés chez l'adulte.

11. « Tant que le petit enfant est dans la période préverbale, on s'intéresse à l'ensemble de ses conduites : ses premiers mots, certes, mais aussi ses cris et vocalises, ses regards, ses mimiques, et ses premiers gestes. [...] Mais à peine l'enfant devient-il un homo loquens, à peine est-il doté de la parole qu'on perd de vue les aspects non verbaux et contextuels de ses conduites pour ne plus s'intéresser qu'au développement lexical, morphologique et syntaxique. [...] Comme si la parole n'était que de la langue, dépourvue de substance sonore et visuelle ». Colletta [15] met ici en évidence une « lacune concernant le développement tardif de la parole chez l'enfant âgé de 6 à 11 ans et scolarisé à l'école élémentaire » (p.13-14).

Une telle hypothèse se base notamment sur le fait que divers aspects des performances motrices, d'une part, et linguistiques, d'autre part, semblent évoluer, de façon non uniforme, entre l'âge de 6 et 11 ans. Cette évolution serait possible grâce à l'intégration progressive de nouvelles stratégies, comme l'illustrent les études présentées ci-après (dont la grande majorité est consacrée au pointage). Par ailleurs, l'évolution des capacités cognitives, discursives et communicatives semble liée à celle des gestes produits, en termes de variété, de degré d'abstraction et de contenu communicatif (Batista et Le Normand [6], Colletta [15, 16], McNeill [48]).

Concernant l'évolution des performances motrices, Badan et collab. [3] étudient par exemple le développement du contrôle visuo-moteur, via la production de pointage manuel, chez des enfants âgés de 6 à 10 ans et chez l'adulte. Les auteurs manipulent la difficulté de la tâche de pointage en faisant varier le nombre, la taille et l'espacement des cibles. Leurs résultats montrent que les paramètres temporels des séquences motrices (i.e. les temps de réaction, de mouvement et de pause, soit le temps passé sur chaque cible), ainsi que les paramètres spatiaux (en termes de précision), évoluent chez l'enfant en fonction de son âge, mais sans atteindre, à 10 ans, les valeurs observées chez l'adulte. Plus précisément, les auteurs observent une diminution significative des valeurs temporelles, en fonction du groupe d'âge, ainsi qu'une amélioration de la précision, ou de l'exactitude, du pointage (mesurée par rapport au centre de la cible). Par ailleurs, ils notent une variation particulière des performances chez l'enfant de 7 ans. Les résultats de ces enfants sont ainsi comparables à ceux des enfants de 8 ans lorsque la tâche est simple, mais comparables à ceux des enfants de 6 ans lorsque la tâche est plus complexe ; différentes stratégies seraient donc disponibles chez ces enfants. Selon les auteurs, « *age 7 appears to be a critical transition*¹² » ([3], p.56), caractérisée par la maîtrise de certaines facultés motrices (telles que la vitesse des mouvements, déjà comparable à celle des enfants plus âgés) et, parallèlement, l'émergence de nouvelles stratégies (notamment de planification), plus élaborées, et donc potentiellement plus coûteuses. Le développement des performances motrices serait donc graduel et permettrait à l'enfant de peaufiner les stratégies dont il dispose à un âge donné, en vue des conduites adultes.

Dans le même registre, nous pouvons citer des travaux plus anciens, tels que ceux de Hay (e.g. [28, 30]), qui propose également une revue de littérature consacrée à l'étude de l'ontogenèse des mouvements dirigés vers un objet [29]. L'auteur observe les productions manuelles d'enfants âgés de 4 à 11 ans, et plus précisément celles de pointage en situation de boucle ouverte (i.e. sans contrôle visuel) et montre une régression des performances (notamment en termes de précision), ainsi qu'une instabilité des productions (caractérisée par des mouvements plus hésitants) aux alentours de 7 ans, âge que l'on pourrait de nouveau qualifier d'"âge charnière". Selon l'auteur, cette discontinuité développementale traduirait un conflit entre les différentes composantes du mouvement (la programmation initiale et les processus rétroactifs d'assistance et de guidage). Autrement dit, elle révèle une transition vers un mode de contrôle moteur plus élaboré : les enfants utiliseraient tout d'abord une stratégie pleinement balistique, avant de pouvoir intégrer les informations proprioceptives, nécessaires au contrôle en ligne de leurs productions. Par ailleurs, l'auteur montre que les enfants sont capables de trouver un compromis optimal vitesse/précision de la réponse vers l'âge de 9 ans, ce qui ne semble pas être le cas chez les enfants plus jeunes.

12. Traduction proposée : L'âge de 7 ans apparaît comme une transition cruciale [dans le développement du contrôle visuo-moteur].

1. Evolution du couplage voix/geste au cours du développement

De la même manière que précédemment, ces données attestent une amélioration croissante des productions, les caractéristiques cinématiques du mouvement étant petit à petit optimisées.

Des résultats similaires ont également été observés par Connolly et collab. [17]. Les auteurs montrent, dans une tâche de pointage alternatif de deux cibles, que la vitesse augmente chez l'enfant âgé de 8 à 10 ans, ce qui n'est pas le cas chez l'enfant âgé de 6 ans à 7 ans. Pellizzer et Hauert [55] observent quant à eux (dans une tâche de pointage avec consigne de vitesse et de précision), une augmentation des erreurs jusqu'à l'âge de 8 ans (qui diminuent par la suite), ainsi qu'une diminution des temps de réaction et de mouvement avec l'âge, excepté pour les enfants de 8 ans, pour lesquels ces paramètres (ré)augmentent. Ces données sont donc clairement en faveur de changements quantitatifs et qualitatifs (en termes de régression des performances) aux alentours de 8 ans, phénomène qui, selon les auteurs, serait lié à la maturation corticale et, plus précisément, à une réorganisation du mapping sensori-moteur à cette période. A ce propos, Gesell et Ilg [21] mettaient déjà en évidence, il y a près de 70 ans, une phase transitionnelle aux alentours de 7-9 ans, âge pour lequel ils notaient l'apparition d'habiletés motrices plus fines ainsi qu'une meilleure inhibition motrice générale (voir également les travaux de Keller et collab. [37] et la revue proposée par Ripoll et collab. [59] consacrés aux mouvements de saisie).

Concernant l'évolution des performances linguistiques, Hazen et collab. [31] et Allen et collab. [2] ont étudié les compétences verbales, et plus précisément l'utilisation d'expressions spatiales, telles que les déictiques, chez des enfants âgés respectivement de 3, 4, et 5 ans et de 6, 9 et 11 ans. Les auteurs ont dans ce cadre là étudié les productions verbales associées à la description de labyrinthes, de trajets entrée-sortie, de trajets inverses, etc. De manière générale, ils observent une amélioration progressive des facultés de navigation et d'orientation dans l'espace, et de fait une amélioration des productions verbales, en termes d'exactitude et de précision de la réponse, et ce jusqu'à l'âge de 9 ans.

Plus récemment, Colletta [15] a étudié le développement de la parole entre l'âge de 6 et 11 ans. Son analyse porte sur 334 séquences de conduites langagières, enregistrées chez des enfants de CP, CE1, CE2, CM1 et CM2, correspondant à des explications, des descriptions, des récits et des séquences argumentatives (ces dernières impliquant au minimum deux enfants). L'auteur mesure la durée de chaque séquence, la quantité d'informations verbales (par comptage syllabique et info-syllabique¹³), ainsi que la quantité d'informations discursives (soit le nombre de propositions, de connecteurs et d'opérateurs énonciatifs). Chacune de ces valeurs est également mesurée par seconde, afin d'estimer la densité informationnelle, soit le débit verbal (brut et corrigé¹⁴) et le débit discursif. L'intérêt principal de cette étude est qu'elle tient compte de l'aspect multimodal de la parole ; l'auteur a donc également mesuré la quantité d'informations et le débit non verbaux, en termes de kinésie coverbale (i.e. mimo-gestualité, nombre de gestes paraverbaux, de gestes déictiques et de gestes iconiques) et de regards phatiques (i.e. nombre de contacts oculaires avec l'interlocuteur). Les enfants ont été regroupés en deux classes d'âge, estimées à 6-8 ans (classes préparatoires et élémentaires) et 9-10 ans (classes moyennes), afin d'évaluer la progression dans le temps des comportements dans les deux modalités.

13. Le comptage info-syllabique consiste à ne retenir que les syllabes informatives, i.e. à supprimer les hésitations, faux-départs et répétitions, phénomènes relativement fréquents dans les productions orales spontanées.

14. Débit verbal brut = nombre de syllabe / seconde ; Débit verbal corrigé = nombre d'info-syllabes / seconde.

Les résultats montrent une évolution assez nette de l'ensemble des comportements, puisque chacun des paramètres considérés augmente significativement entre les deux groupes d'âge. Plus précisément, l'auteur observe un allongement des conduites monologuées, qui sont également plus riches et plus denses en informations verbales et non verbales, généralement à partir de 9 ans. Autrement dit, l'évolution des conduites verbales s'accompagne d'une évolution parallèle des conduites non verbales. Selon Colletta [15], « *la simultanéité de ces phénomènes est à remarquer et suggère un développement multimodal des conduites langagières* » ([15], p.227).

La dernière étude que nous avons retenue est consacrée au processus déictique et s'intéresse plus particulièrement aux productions conjointes de geste et de parole. Elle permet en cela d'obtenir des informations relatives au couplage temporel de ces deux modalités, au cours du développement. Pechmann et Deutsch [54] ont ainsi proposé une tâche de "communication référentielle", où les participants, des enfants âgés de 2, 4, 5, 6 et 9 ans, ainsi que des adultes, devaient indiquer un objet à leur interlocuteur, parmi des ensembles de huit objets (Expérience 1, comparaison des enfants de 2, 6, 9 ans et des adultes) ou de quatre objets (Expérience 2, comparaison des enfants de 4-5 ans et des adultes). Dans la première expérience, le pointage manuel est qualifié d'*ineffectif*, dans le sens où il ne peut aider l'interlocuteur à identifier sans erreur le référent ; dans la seconde expérience, la disposition des objets est telle qu'un pointage seul suffit en théorie à les distinguer les uns des autres. Les auteurs observent, dans l'Expérience 1, des changements quantitatifs au cours du développement, et notamment une diminution des pointages *ineffectifs* entre 2 et 9 ans (seule la différence entre 9 ans et l'âge adulte n'est pas significative), associée à un nombre croissant de productions verbales adéquates (la différence est cette fois significative entre chaque groupe d'âge considéré). L'utilisation fréquente du pointage chez les enfants de 2 et 6 ans, alors même que celui-ci n'est pas adapté à la situation, indique que ces enfants compensent l'absence d'autres solutions (verbales notamment), qui ne sont alors pas encore disponibles dans leur répertoire. Dans l'Expérience 2 en revanche, lorsque la situation est non ambiguë, les enfants de 4-5 ans et les adultes produisent un nombre similaire de pointages et de descriptions verbales correctes. Par ailleurs, les auteurs notent des phénomènes temporels réguliers au niveau des productions conjointes de parole et de geste :

- Le recouvrement des productions vocales et manuelles, dans 92% des cas.
- L'avance de l'onset du mouvement sur celui de la parole, dans 95% des cas (rappelons que ce phénomène est également observé par Holender [32] et Levelt et collab. [42] chez l'adulte, voir Chapitre 1, Section 2.2).
- La production du pointage manuel lors de la prononciation des traits distinctifs de l'objet, dans 73% des cas.

Les auteurs mettent ainsi en évidence une relation temporelle stable entre les systèmes de production de gestes et de parole, et ce dès l'âge de 4 ans, puisque ces phénomènes sont observés de manière indifférenciée chez les enfants de 4-5 ans et chez les adultes.

De manière générale, ces études mettent en évidence une évolution progressive, et potentiellement conjointe, des conduites motrices et verbales ; les performances des enfants atteindraient celles des adultes aux alentours de 9 ans, après une période de transition, caractérisée par l'émergence de nouvelles stratégies, aux alentours de 7-8 ans.

2 Expérience 3

L'objectif de notre Expérience 3 est de caractériser au mieux la transition entre les performances du tout-petit et celles de l'adulte, en évaluant certains aspects de l'interaction parole/geste chez l'enfant d'âge scolaire. Pour cela, nous souhaitons appliquer le protocole initialement testé chez l'adulte, présenté au Chapitre 3, à l'enfant âgé de 6 à 12 ans. La question principale est de savoir si les comportements observés chez l'adulte, soit un encodage de la distance, via les propriétés phonétiques et cinématiques du pointage, ainsi qu'une interaction structurelle et bidirectionnelle, de nature coopérative, entre voix et geste, sont des caractéristiques fondamentales du processus déictique ou bien des compétences plus tardives, apparaissant progressivement dans le répertoire communicatif des locuteurs.

Certaines hypothèses peuvent être formulées à partir des études citées précédemment. On pourrait par exemple s'attendre, chez le jeune enfant, à une certaine asymétrie entre l'utilisation du geste (préférentielle) et celle de la parole, suivie d'une transition progressive (aux alentours de 8 ans), qui résulterait en un équilibre entre ces deux modalités. De fait, si l'exploration spatiale est initialement portée par le geste, on pourrait envisager que (1) l'information de distance soit tout d'abord véhiculée par le pointage manuel, avant d'être encodée de la même façon via le pointage vocal, et (2) seule la production du geste manuel puisse influencer celle de la parole, avant qu'une coopération mutuelle entre les deux modalités ne soit mise en place.

2.1 Méthodologie

2.1.1 Participants

Vingt-trois enfants (dont 15 garçons et huit filles¹⁵), âgés de 6 à 12 ans (Moyenne (M) = 8.4, Ecart-type (SD) = 2; voir Tableau 4.2 ci-après) et accompagnés d'un adulte, ont participé à cette étude. Tous étaient de langue maternelle française, sans déficit visuel ou auditif non corrigé, ni trouble moteur particulier. L'échantillon compte par ailleurs 21 droitiers pour deux gauchers. Les enfants ont été recrutés majoritairement par diffusion d'annonces, auprès de la communauté scientifique et du grand public, et ont été remerciés de leur participation par un *Diplôme du Meilleur Participant*, et quelques friandises!

Notre objectif initial était d'obtenir un maximum de données, auprès d'un maximum d'enfants, afin d'étudier leurs performances par tranches d'âge minimale, c'est-à-dire d'une année. Toutefois, la taille de l'échantillon étant petite, nous avons dû répartir les enfants en deux classes d'âge inégales, afin de pouvoir (1) conduire des analyses de groupe sur un nombre suffisant (et équivalent) de données, et (2) observer une potentielle transition aux alentours de 8 ans.

Par ailleurs, afin de pouvoir comparer les productions enfantines aux productions adultes, le protocole général est identique à celui de l'Expérience 2 (présenté au Chapitre 3); le lecteur pourra donc s'y reporter en cas de besoin. Simplement, nous avons ajouté coussins, nappes et couvertures de couleurs, afin de rendre la chambre expérimentale plus conviviale et la passation plus agréable.

Nous présentons ici la description simplifiée du protocole expérimental, afin de faciliter la lecture indépendante de ce chapitre.

15. L'étude d'Allen et collab. [2] montre une absence d'interaction $Sexe * VI$ chez l'enfant âgé de 6 à 11 ans.

Groupe 1 : [6-8 ans[$M = 6.7$; $SD = 0.6$		Groupe 2 : [8-12 ans[$M = 9.8$; $SD = 1.6$	
N.R.	6	I.F.	8
N.S	6.1	V.R.	8
L.S.	6.6	J.B.	8.4
A.M.	6.8	S.B.	8.7
I.G.	6.8	R.S	9
M.C.	6.11	I.B.	9.11
J.L.	6.11	N.R.	9.11
M.S.	7.2	N.L.	10.9
G.G.	7.7	J.W.	11
V.S.	7.7	N.B.	11.8
V.G.	7.10	M-A.B.	11.8
		J.G.	12.5

Tableau 4.2 – Age des enfants (M =Moyenne ; SD =Ecart-type)

2.1.2 Matériel

Comme indiqué sur la Figure 4.5, le dispositif expérimental est identique à celui de l'Expérience 2, présentée au Chapitre 3.

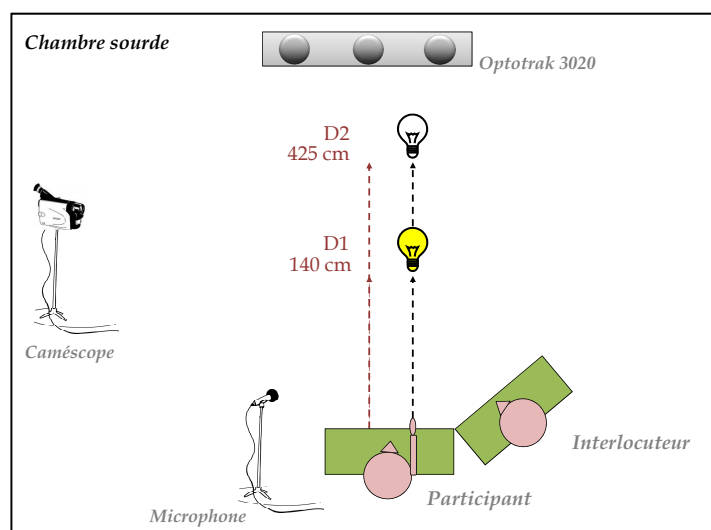


FIGURE 4.5 – Dispositif expérimental

Les enfants sont donc évalués individuellement en chambre sourde, et sont assis face à deux LED, placées à deux distances différentes d'eux :

- D1, située à 140 cm, dans l'espace **extra-personnel proche** ;
- D2, située à 425 cm, dans l'espace **extra-personnel lointain**.

2. Expérience 3

Ces deux LED ambigües, situées hors de portée, sont alignées l'une par rapport à l'autre et par rapport à l'axe œil-index de l'enfant, qui les ajuste lui-même sur une même ligne (i.e. "l'une dans l'autre"). Elles permettront de tester un encodage particulier de la distance, via les propriétés cinématiques du geste (l'angle du bras seul ne peut pas les distinguer) mais également via les propriétés phonétiques du pointage vocal. Le microphone AKG C1000S, relié à un enregistreur multipistes Korg D1200, ainsi que le système opto-électronique de capture de mouvements Optotrak 3020, vont permettre l'enregistrement des productions orales et gestuelles ; l'ensemble de la passation est également enregistré par caméscope numérique, afin de vérifier ultérieurement la validité des tâches effectuées et le déroulement de l'expérience (bien que l'expérimentateur effectue un contrôle en direct et intervient si nécessaire).

2.1.3 Procédure

La tâche des enfants, dont la main est initialement placée en position de repos en milieu de table, est de désigner la LED allumée à l'interlocuteur situé sur leur droite (ou leur gauche, selon leur préférence manuelle), en réponse à la question « C'est où ? », prononcée de manière neutre. La situation d'interaction, sous forme d'un jeu de questions/réponses est relativement claire et plutôt naturelle. Néanmoins l'interlocuteur s'efforce d'être le plus constant possible (au cours d'une même session, mais également entre différentes sessions) : tourné vers l'enfant lorsqu'il s'adresse à lui, il ne regarde en direction des cibles qu'une fois la réponse donnée.

Trois conditions expérimentales sont testées : deux situations unimodales (**Parole Seule**, **Geste Seul**) et une situation bimodale (**Parole+Geste**) :

- **Parole Seule** : En situation de parole seule, les enfants doivent nommer la LED allumée en utilisant l'adverbe locatif « là », inséré dans la phrase porteuse : « c'est [+ item] ». Il leur est demandé de ne pas utiliser leurs mains, qu'ils peuvent donc laisser en position de repos, sur la table, ou sur leurs genoux.
- **Geste Seul** : En situation de geste seul, ils doivent pointer la cible par un geste de l'index droit ou gauche.
- **Parole+Geste** : En situation de parole+geste, il leur est demandé de nommer et simultanément de pointer la LED allumée.

Ces trois conditions permettront de tester l'interaction entre le geste et la parole, par la comparaison des productions unimodales et bimodales.

Le paradigme utilisé est toujours un paradigme en blocs (2×3 Conditions, soit six blocs successifs). Le premier bloc, pour chaque condition, est précédé d'une courte session d'entraînement (2×4 essais, avec possibilité de recommencer l'entraînement si nécessaire). La moitié des enfants commencent par la condition Geste Seul et l'autre moitié par la condition Parole+Geste (la condition Parole Seule étant potentiellement la moins naturelle, et de fait, la plus difficile). Chaque essai débute par la question de l'interlocuteur, suivie de l'allumage de l'une des deux diodes, signal de départ pour l'enfant.

La durée totale de l'Expérience 3, au cours de laquelle les enfants sont confrontés à 60 essais (= 2 Distances $\times$ 3 Conditions $\times$ 10 Itérations), est légèrement supérieure à celle de l'Expérience 2, puisqu'elle dure une trentaine de minutes (soit une dizaine de minutes supplémentaires). En effet, afin de rassurer les enfants et pour qu'ils soient le plus à l'aise possible, nous avons pris le temps de discuter, de répondre à toutes leurs questions, mais également de leur présenter le matériel de façon détaillée (essai des capteurs et visualisation de leurs mouvements sur écran d'ordinateur, exemples d'applications concrètes d'un tel dispositif, telles que la réalisation d'images animées, etc.).

Le plan d'expérience est un plan intra-sujets, de manière à ce que chaque participant soit confronté à chaque modalité de chaque variable (dont le récapitulatif est disponible Tableau 4.3).

Variables Indépendantes	Modalités
Distance	140 cm - 425 cm
Condition	Parole Seule - Geste Seul - Parole+Geste

Tableau 4.3 – Récapitulatif des variables indépendantes

2.2 Acquisition des données

L'acquisition des données est strictement identique à celle de l'Expérience 1 (présentée au Chapitre 3, Section 1.2). Pour rappel, l'enregistrement du signal de parole et du bip de synchronisation est réalisé par un microphone AKG C1000S, relié à un enregistreur multipistes Korg D1200, respectivement en voies 1 (les niveaux sont réglés individuellement et restent invariables au cours d'une même session) et 2 ; l'enregistrement des mouvements manuels et articulatoires est réalisé par le système Optotrak 3020, permettant de suivre la trajectoire de diodes émettrices infra-rouge au cours du temps.

Sept diodes émettrices (en plus de celle fixée sur pied métallique et synchronisée au signal sonore, dans le but de repérer le début de chaque nouvel essai) sont donc mises en place, comme indiqué sur la Figure 4.6 ci-après :

- Trois diodes de référence placées dans un plan rigide, au niveau du front de l'enfant ¹⁶ ;
- Deux diodes placées au centre des contours des lèvres supérieure et inférieure (notées D_{LS} et D_{LI}) ;
- Une diode placée sur l'ongle de l'index droit ou gauche (notée D_I), collée sur une équerre métallique ;
- L'ajout d'une diode sur le dessus de la main, droite ou gauche, permettra, éventuellement, de compenser une perte trop importante de données, liée au masquage de la diode placée sur l'index.

¹⁶. Les trois diodes de référence n'ont pas été exploitées, n'étant pas nécessaires pour les mesures effectuées.

2. Expérience 3



FIGURE 4.6 – Position des diodes émettrices infra-rouge chez l'enfant (© C. Gonseth)

2.3 Mesures

Les mesures effectuées, acoustiques, articulatoires et cinématiques (résumées au Tableau 4.4 ci-après), sont identiques à celles de la première expérience, et le lecteur pourra donc trouver des informations supplémentaires au Chapitre 3, Section 1.3.

Les productions orales ont été segmentées puis enregistrées au format .wav, avant d'être "nettoyées". Le repérage des pics d'énergie nous a permis de détecter de façon automatique la voyelle [a], dont les frontières temporelles (i.e. début, milieu et fin) ont été annotées. Nous avons finalement mesuré, au milieu de chaque voyelle, les valeurs de fréquence fondamentale, d'intensité, de premier et de second formants. La trajectoire des diodes placées sur les lèvres a été segmentée et mise en correspondance avec les données acoustiques, afin de calculer le maximum d'ouverture des lèvres, pour chaque voyelle. La trajectoire de la diode placée sur l'index a également été segmentée, puis annotée (départ de la position de repos ; extension maximale ; départ de la position d'apex ; retour en position de repos), grâce à la détection préalable des deux pics de vitesse correspondant à l'aller et au retour du geste. Ceci nous a permis de calculer le pic de vitesse et l'amplitude de l'aller du geste (i.e. premier stroke), ainsi que la durée du plateau (i.e. tenue). La durée totale du pointage ne sera pas prise en compte dans cette étude, du fait d'un risque élevé (et confirmé) de masquage de la diode en fin de mouvement et/ou d'ajout de gestes parasites pendant la phase de retour (e.g. lorsque l'enfant se gratte, lorsqu'il ferme le poing ou qu'il "joue" avec la diode et la cache avec son majeur).

Variables Dépendantes	Mesures
Données acoustiques	F0 (Hz) - Intensité (dB) - F1 (Hz) - F2 (Hz)
Données articulatoires	Ouverture des lèvres (mm)
Données cinématiques	Durée plateau (s) - Pic de vitesse (mm/s) - Amplitude (mm)

Tableau 4.4 – Récapitulatif des variables dépendantes

2.4 Analyses statistiques et prédictions

Le lecteur pourra trouver le détail technique et théorique des analyses statistiques au Chapitre 3, Section 1.4.

Certains essais ont dû être exclus de l'analyse statistique : les oublis (d'item et/ou de pointage) mais également les données manquantes liées à des problèmes techniques (masquage des diodes émettrices infra-rouge notamment), soit 5% et 4.54% des données acoustiques (pour les groupes 1 et 2 respectivement), 5.25% et 4.8% des données articulatoires et 9.75% et 21.8% des données cinématiques. Par ailleurs, du fait de la taille de nos échantillons et de leur dispersion, nous avons également choisi d'exclure les outliers (i.e. valeurs supérieures ou égales à 2 écart-type), soit 3% et 3.2% des données acoustiques (pour les groupes 1 et 2 respectivement), 2.6% et 3.1% des données articulatoires et 3.3% et 3% des données cinématiques. Ce pourcentage d'essais exclus est relativement important et correspond aux données acoustiques et articulatoires d'un participant de chaque groupe, et aux données cinématiques d'un participant du premier groupe et de trois participants du second groupe, exclus de l'analyse statistique finale. Les analyses statistiques acoustiques et articulatoires portent donc sur 10 et 11 participants (dans les groupes 1 et 2 respectivement) et les analyses cinématiques portent quant à elles sur 10 et 9 participants (dans les groupes 1 et 2 respectivement).

Le test de Shapiro-Wilk ([61]) nous a permis de vérifier la normalité des données (une transformation logarithmique¹⁷ étant appliquée si besoin). Le test de Mauchly ([47]) et le correctif de Greenhouse-Geisser ([27]), relatifs à la sphéricité des données, n'ont pas été nécessaires, le $dl1$ étant égal à 1 pour chaque facteur étudié.

2.4.1 Analyse intra-sujets & analyse individuelle

Une première série d'analyses va consister à étudier l'effet des facteurs intra-sujets, pour chaque groupe d'âge. Pour ce faire, nous allons effectuer deux ANOVA (sur les médianes), pour plan à mesures répétées et deux facteurs intra-sujets :

Groupe 1 (6-7 ans) : $S_{10}[Distance_2 \times Condition_3]$
 Groupe 2 (8-12 ans) : $S_{11|9}[Distance_2 \times Condition_3]$

L'ANOVA à mesures répétées va nous permettre de tester deux effets principaux, celui de la Distance et celui de la Condition, ainsi qu'un effet d'interaction, $Condition * Distance$. Néanmoins, du fait du nombre restreint de participants dans les deux groupes (11 au maximum), la variabilité de l'échantillon (en termes de dispersion) est relativement importante ; elle représente donc un biais potentiel, au niveau de l'effet des facteurs et de leur interprétation. De fait, nous présenterons également les ANOVA individuelles, qui nous permettront d'évaluer au mieux les stratégies utilisées par chacun des participants.

17. $Y' = LN(Y) + \text{une constante}$, en cas de valeurs négatives.

2. Expérience 3

2.4.2 Analyse inter-sujets

Dans le cas où un effet serait observé pour au moins deux des trois groupes d'âge, une seconde série d'analyses nous permettra d'étudier l'impact du facteur *Groupe* inter-sujets (dont les modalités sont résumées Tableau 4.5), par le biais d'une ANOVA à mesures répétées, portant sur les médianes, pour plan mixte (i.e. un facteur inter-sujets et deux facteurs intra-sujets) : $Distance_2 \times Condition_3 < Groupe_3 >$

Variable Inter-sujets	Modalités	Nb Individus Acoustique	Nb Individus Articulatoire	Nb Individus Cinématique
<i>Groupe</i>	Groupe 1 [6-7 ans [	10	10	10
	Groupe 2 [8-12 ans[	11	11	9
	Groupe 3 [Adultes [	27	26	25

Tableau 4.5 – Variables inter-sujets et modalités associées. Les données du Groupe 3 sont celles obtenues dans l'Expérience 2 et sont disponibles au Chapitre 3, Section 2.5)

2.4.3 Prédictions

Une évolution du couplage geste/parole au cours du développement, en période scolaire, aurait pour conséquence des différences relatives au groupe d'âge ; de plus, en admettant que le geste soit un outil communicatif privilégié par les jeunes enfants, nous pouvons envisager que (1) l'encodage de la distance via les propriétés cinématiques du pointage soit plus précoce que l'encodage de la distance via les propriétés phonétiques du pointage ; (2) l'influence du geste sur la parole soit plus précoce que l'influence de la parole sur le geste.

Plus précisément, nous attendons un encodage cinématique de la distance, similaire à celui observé chez l'adulte, pour les enfants des groupes 1 et 2 (potentiellement plus marqué chez les enfants plus âgés), suivi de l'émergence de l'encodage phonétique aux alentours de 8 ans, qui résulte, à l'âge adulte, en un encodage de la distance via les deux modalités. De la même façon, nous attendons une influence du geste sur la parole, similaire à celle observée chez l'adulte, pour les enfants des groupes 1 et 2 (potentiellement plus marquée chez les enfants plus âgés), suivie de l'émergence de l'influence de la parole sur le geste aux alentours de 8 ans, qui résulte, à l'âge adulte, en une coopération inter-modalités.

2.5 Résultats de l'Expérience 3

Cette section présente un descriptif des effets observés pour chaque facteur, selon l'analyse considérée (i.e. intra-sujets, individuelle, inter-sujets si nécessaire). Les résultats du Groupe 3 y sont également rappelés à titre indicatif. Les implications théoriques qui en découlent seront discutées dans la section suivante (voir Section 2.6).

2.5.1 Propriétés articulatoires du pointage vocal

Ouverture des lèvres : Analyses intra-sujets et individuelle La figure 4.7 montre l'influence de la Condition et de la Distance sur l'ouverture des lèvres (*OL*), pour les groupes 1 et 2.

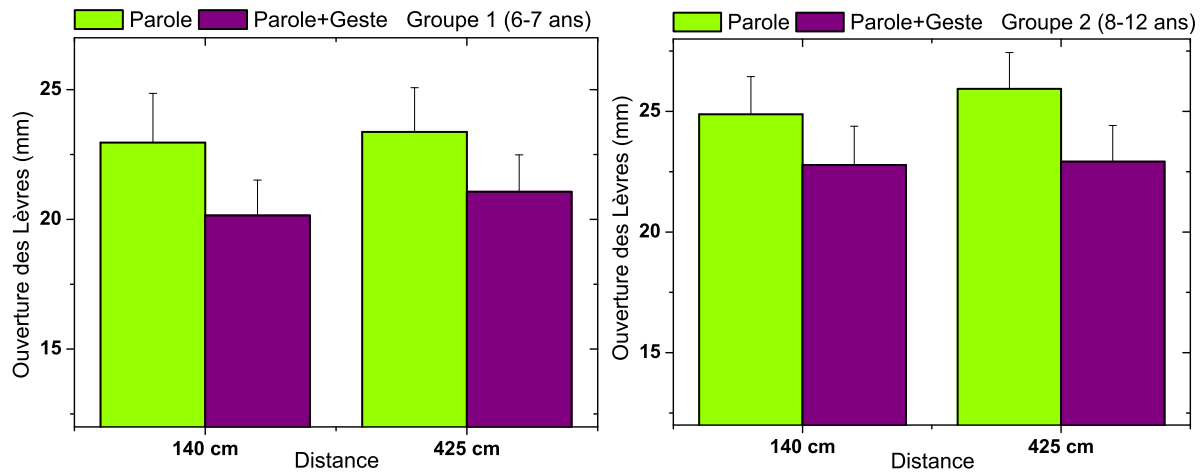


FIGURE 4.7 – *Ouverture des lèvres* (mm) en fonction de la Distance et de la Condition

Effets principaux Pour le Groupe 1, les résultats montrent un effet principal de la Condition ($F(1,9) = 15.77, p < .01$) : l'ouverture des lèvres augmente en situation unimodale. Au niveau individuel, ce pattern est observé pour six participants¹⁸, tandis qu'aucune différence n'est observée pour les quatre autres participants. En revanche, l'effet de la Distance n'est pas significatif ($F(1,9) = 2.06, p = 0.18, NS$), et cela pour neuf enfants (un seul participant présente une ouverture des lèvres significativement plus grande pour la distance la plus éloignée, bien que l'effet $D1 < D2$ soit proche du seuil de significativité, avec $p = .05$, pour deux enfants). Pour le Groupe 2, les résultats montrent un effet principal de la Condition et de la Distance. L'ouverture des lèvres est plus grande en situation unimodale ($F(1,10) = 12.27, p < .01$), et ce pour 10 participants (le dernier ne présente aucune différence significative), et augmente avec la distance du référent ($F(1,10) = 5.95, p < .05$), mais pour deux enfants seulement, les neuf autres ne montrant aucune variation.

Effets d'interaction L'effet d'interaction $C * D$ n'est pas significatif pour le Groupe 1 ($F(1,9) = 0.25, p = 0.63, NS$) : cette absence d'effet est observée pour neuf participants, un enfant montre en revanche un effet de la Distance plus important pour la condition bimodale. L'interaction est par contre quasi-significative pour le Groupe 2 ($F(1,10) = 4.92, p = .05$), i.e. l'effet de la Distance tend à être significatif pour la condition unimodale uniquement (toutefois, au niveau individuel, aucune interaction n'est observée).

18. Les résultats individuels présentés dans l'ensemble de la section sont significatifs au seuil $p < .05$.

2. Expérience 3

A retenir ! Le Tableau 4.6 présente un bref aperçu des principales différences observées en intra-sujets sur les données articulatoires.

Ouverture des lèvres			
	Groupe 1	Groupe 2	Rappel Groupe 3
Distance	ϕ	$D1 < D2$	$D1 < D2$
Condition	$Unimodale > Bimodale$	$Unimodale > Bimodale$	$Unimodale > Bimodale$
Interaction	ϕ	$Interaction (p = .05)$	$Interaction$

Tableau 4.6 – Principaux résultats intra-sujets (données articulatoires : ouverture des lèvres)

Ouverture des lèvres : Analyse inter-sujets Une analyse inter-sujets est nécessaire, l'effet de la Condition étant observé pour chaque groupe, celui de la Distance pour les groupes 2 et 3. L'interaction *Groupe * Distance* n'est pas significative : l'effet observé chez l'adulte (une ouverture plus grande pour la distance la plus éloignée) est donc comparable à celui observé chez les enfants de 8 ans et plus. En revanche, l'interaction *Groupe * Condition* est significative, et ce pour les groupes 1 vs 3 et 2 vs 3 : l'effet observé sur l'ouverture des lèvres (une ouverture plus grande pour la condition unimodale) dépend donc du groupe d'âge considéré. Toutefois, le déséquilibre inter-groupes (i.e. la taille de l'échantillon diffère selon le groupe considéré) ne permet pas, à notre connaissance, de conduire d'analyse *post-hoc* sur ces données. Une analyse au premier ordre, illustrée Figure 4.8, semble néanmoins indiquer que l'effet de la Condition soit plus important pour les deux groupes d'enfants que pour les adultes¹⁹.

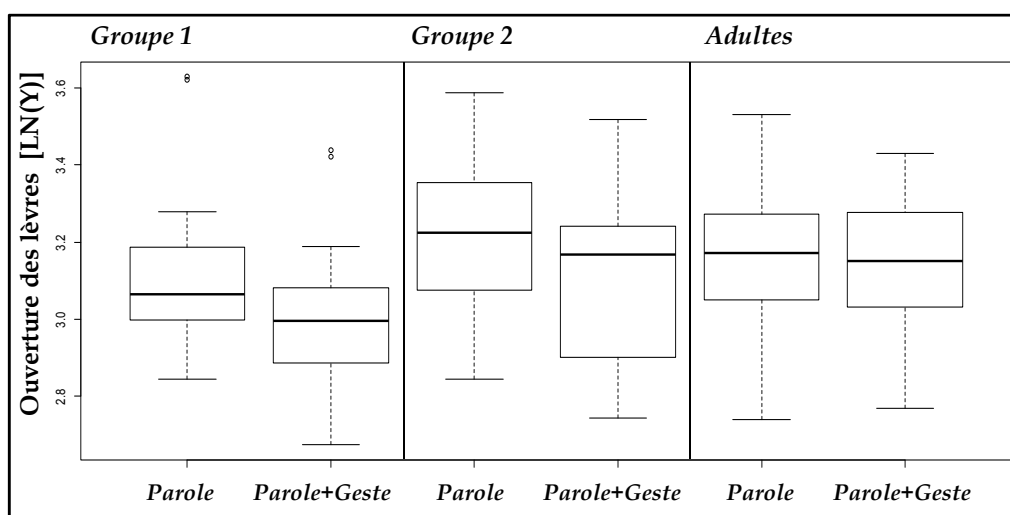


FIGURE 4.8 – *Ouverture des lèvres* (après transformation logarithmique) en fonction de la Condition et du Groupe d'âge

19. La *boîte à moustaches* offre un résumé visuel des données, basé sur le minimum, les trois quartiles Q1, Q2 (i.e. la médiane), Q3, et le maximum. Ici, la hauteur de la boîte représente la distance inter-quartile $Q3 - Q1$, et les moustaches sont de longueur $Q3 - Q1 \times 1.5$. Une valeur est considérée comme atypique (et indiquée par le symbole $\circ$) lorsqu'elle dépasse cette valeur, au-dessous du 1^{er} quartile ou au-dessus du 3^{ème} quartile.

2.5.2 Propriétés acoustiques du pointage vocal

Premier formant : Analyses intra-sujets et individuelle

Effets principaux Pour le Groupe 1, nous n'observons aucun effet significatif. En d'autres termes, les valeurs de premier formant ne sont influencées ni par la Condition ($F(1, 9) = 0.62$, $p = .45$, *NS*), ni par la Distance ($F(1, 9) = .64$, $p = 0.44$, *NS*). Au niveau individuel, l'absence d'effet de la Condition n'est observée que pour trois enfants (cinq enfants présentent des valeurs plus élevées en condition unimodale, deux enfants présentent des valeurs plus élevées en condition bimodale), tandis que l'absence d'effet de la Distance concerne huit enfants (deux seulement montrent des valeurs plus élevées pour la distance la plus éloignée).

Des résultats similaires sont observés pour le Groupe 2, pour lequel l'effet de la Condition ($F(1, 10) = 2.54$, $p = .14$, *NS*) et de la Distance ($F(1, 10) = 0.1$, $p = .76$, *NS*) ne sont pas significatifs. L'absence d'effet de la Condition est observée chez sept enfants (les quatre autres présentent des valeurs plus élevées en situation unimodale) ; l'absence d'effet de la Distance est observée chez 10 participants (seul un enfant montre des valeurs plus élevées pour la distance la plus éloignée).

Effets d'interaction De la même manière, l'interaction $C * D$ n'est significative pour aucun des groupes ($F(1, 9) = 0.15$, $p = .71$, *NS* pour le Groupe 1 ; $F(1, 10) = 4.42$, $p = .06$, *NS* pour le Groupe 2). Au niveau individuel, cette absence d'interaction est observée pour neuf enfants du Groupe 1 (un enfant montre un effet de la Distance plus important en condition unimodale) et pour les onze enfants du Groupe 2.

A retenir ! Le Tableau 4.7 présente un bref aperçu des principales différences observées en intra-sujets sur les données acoustiques (F1).

F1			
	Groupe 1	Groupe 2	Rappel Groupe 3
Condition	ϕ	ϕ	<i>Unimodale > Bimodale</i>

Tableau 4.7 – Principaux résultats intra-sujets (données acoustiques : F1)

Second formant : Analyses intra-sujets et individuelle La figure 4.9 montre l'influence de la Condition et de la Distance sur les valeurs de second formant, pour les groupes 1 et 2.

Effets principaux Nous n'observons pas d'effet significatif pour le premier groupe. La Condition n'influence pas les valeurs de F2 ($F(1, 9) = 3.7$, $p = .09$, *NS*), bien que seuls trois individus soient concernés ; la majorité des enfants montre en effet des valeurs plus élevées en condition bimodale, tandis qu'un seul enfant montre des valeurs plus élevées en condition unimodale. Le second formant n'est pas davantage influencé par la Distance ($F(1, 9) = 0.03$, $p = .86$, *NS*), ce qui caractérise les productions de neuf participants (un seul enfant montre des valeurs de F2 plus élevées pour la distance la plus éloignée).

2. Expérience 3

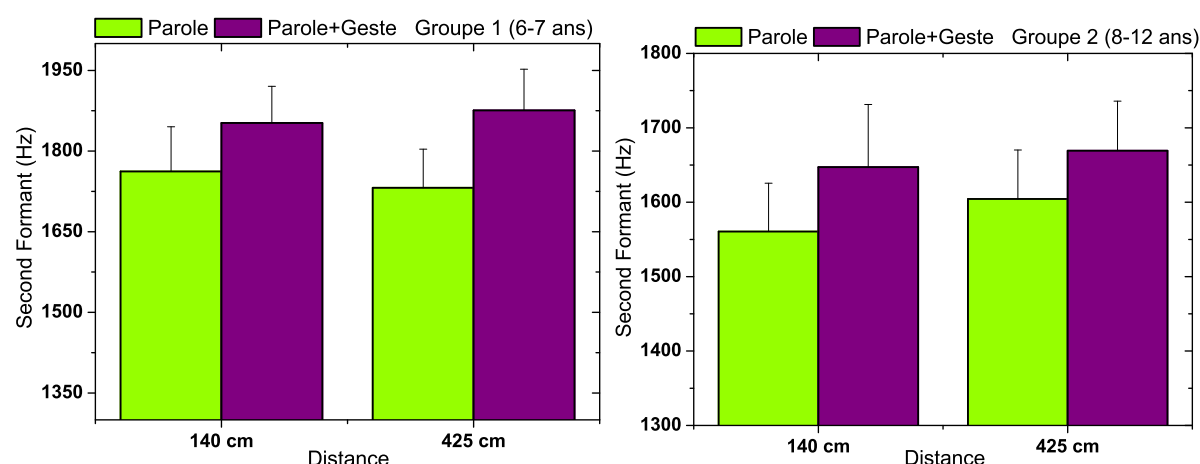


FIGURE 4.9 – F_2 (Hz) en fonction de la Distance et de la Condition

En revanche, pour le Groupe 2, nous observons un effet significatif de la Condition ($F(1, 10) = 9.67, p < .05$) : les valeurs de second formant sont plus élevées en situation bimodale, mais cet effet ne concerne que trois enfants (les huit autres ne présentent pas de différence significative, liée à l'utilisation d'une seule ou de deux modalités). L'effet de la Distance n'est en revanche pas significatif ($F(1, 10) = 1.29, p = .28, NS$), et cela pour 10 participants (seul un enfant montre des valeurs plus élevées pour la distance la plus éloignée).

Effets d'interaction L'interaction $C*D$ n'est significative pour aucun des groupes ($F(1, 9) = 2.04, p = .19, NS$ pour le Groupe 1 ; $F(1, 10) = 0.27, p = .61, NS$ pour le Groupe 2), et ceci pour chaque enfant.

A retenir ! Le Tableau 4.8 présente un bref aperçu des principales différences observées en intra-sujets sur les données acoustiques (F_2).

Second formant			
	Groupe 1	Groupe 2	Rappel Groupe 3
Condition	ϕ	$Unimodale < Bimodale$	ϕ

Tableau 4.8 – Principaux résultats (données acoustiques : F_2)

Fréquence fondamentale (F_0) : Analyses intra-sujets et individuelle

Effets principaux Les résultats ne montrent aucun effet significatif, et ce quel que soit le groupe d'âge (pour rappel, aucun effet significatif lié à la fréquence fondamentale n'était observé chez l'adulte). Pour le premier groupe, l'absence d'effet de la Condition ($F(1, 9) = 2.74, p = .13, NS$) est observée chez neuf participants (un seul enfant présente des variations de la fréquence fondamentale, plus élevée en condition unimodale), et l'absence d'effet de la Distance ($F(1, 9) = 1.39, p = .27, NS$) est observée chez sept participants (deux enfants montrent des

valeurs plus élevées pour la distance la plus éloignée, un enfant présente le pattern inverse, soit des valeurs plus élevées pour la distance la plus proche). Pour le second groupe, l'absence d'effet de la Condition ($F(1, 10) = 1.27, p = .28, NS$) concerne six enfants (deux participants montrent des valeurs plus élevées en condition bimodale, tandis que les trois derniers montrent des valeurs plus élevées en condition unimodale), et l'absence d'effet de la Distance ($F(1, 10) \approx 0, p = .98, NS$) est observée pour chacun des participants.

Effets d'interaction L'interaction $C*D$ n'est significative pour aucun des groupes ($F(1, 9) = 1.6, p = .24, NS$ pour le groupe 1 ; $F(1, 10) = 0.38, p = .55, NS$, pour le groupe 2). Ce résultat concerne l'ensemble des participants du premier groupe, et seuls deux enfants du second groupe montrent un effet significatif de la Condition uniquement pour la distance la plus éloignée.

Intensité : Analyses intra-sujets et individuelle La figure 4.10 montre l'influence de la Condition et de la Distance sur l'intensité du signal de parole, pour les groupes 1 et 2.

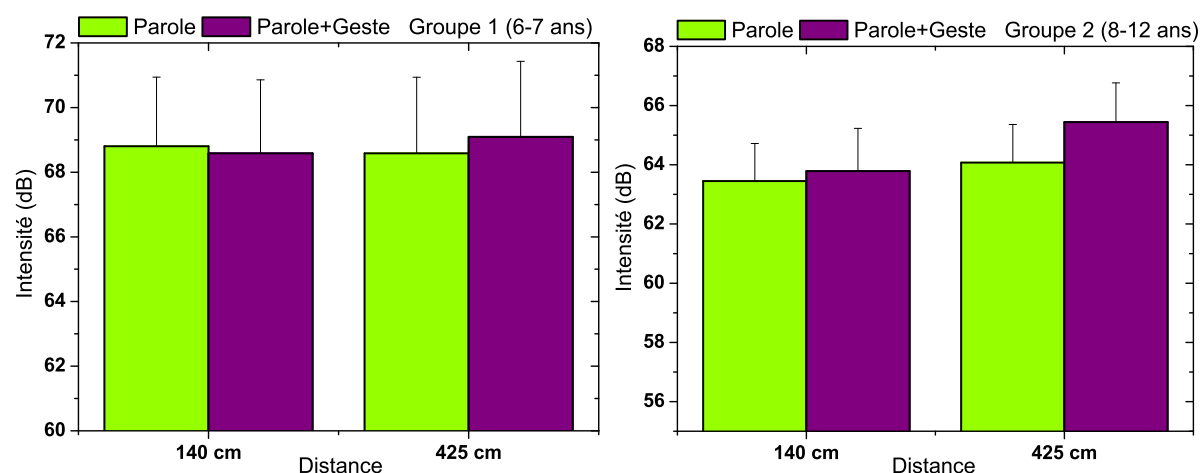


FIGURE 4.10 – *Intensité* (dB) en fonction de la Distance et de la Condition

Effets principaux Nous n'observons pas d'effet significatif pour le Groupe 1 : la Condition n'influence pas l'intensité du signal de parole ($F(1, 9) = 0.01, p = .91, NS$) mais ce résultat ne concerne, individuellement, que trois participants. Trois autres participants montrent une intensité plus élevée en condition bimodale, tandis que quatre participants présentent le pattern inverse, soit une intensité plus élevée en condition unimodale. De la même façon, la Distance n'a pas d'influence significative sur l'intensité ($F(1, 9) = 0.28, p = .61, NS$), mais cet effet concerne neuf participants (un seul enfant montre une intensité plus élevée pour la distance la plus éloignée).

Pour le Groupe 2, l'effet de la Condition n'est pas significatif ($F(1, 10) = 1.61, p = .23, NS$). Ce résultat est observé pour huit participants (un seul enfant présente des valeurs plus élevées en condition unimodale, deux en condition bimodale). En revanche, l'effet de la Distance est significatif ($F(1, 10) = 5.14, p < .05$) : l'intensité est plus élevée pour la distance la plus éloignée. Néanmoins, ce résultat n'est observé que pour trois participants, les huit autres ne montrent aucune variation d'intensité relative à la distance de la cible.

2. Expérience 3

Effets d'interaction L'interaction $C*D$ n'est significative pour aucun des groupes ($F(1, 9) = 0.82, p = .39, NS$ pour le Groupe 1 ; $F(1, 10) = 4.52, p = .06, NS$ pour le Groupe 2) : ceci vaut pour l'ensemble des participants du Groupe 1 tandis que deux enfants du Groupe 2 montrent un effet de la Condition plus important pour la distance la plus éloignée.

A retenir ! Le Tableau 4.9 présente un bref aperçu des principales différences observées en intra-sujets sur les données acoustiques (intensité).

Intensité			
	Groupe 1	Groupe 2	Rappel Groupe 3
Distance	ϕ	$D1 < D2$	$D1 < D2$

Tableau 4.9 – Principaux résultats (données acoustiques : intensité)

Intensité : Analyse inter-sujets Une analyse inter-sujets s'avère nécessaire puisque l'effet de la Distance (une intensité plus élevée pour la distance la plus éloignée) est observé dans les groupes 2 et 3. L'interaction $Groupe * Distance$ n'est pas significative, ce qui suggère que l'effet observé chez l'adulte soit comparable à celui observé chez l'enfant âgé de 8 ans et plus.

2.5.3 Propriétés cinématiques du pointage manuel

Amplitude : analyses intra-sujets et individuelle La figure 4.11 montre l'influence de la Condition et de la Distance sur l'amplitude du pointage manuel, pour les groupes 1 et 2.

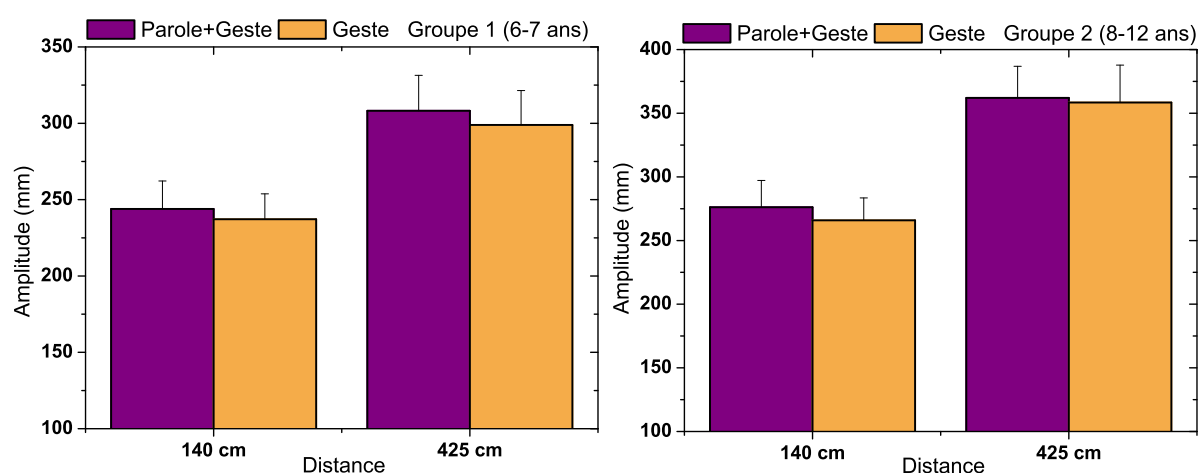


FIGURE 4.11 – *Amplitude* (mm) en fonction de la Distance et de la Condition

Effets principaux Pour le Groupe 1, l'effet de la Condition n'est pas significatif ($F(1, 9) = 3.2, p = .1, NS$) ; ce résultat s'observe pour sept participants (un enfant produit des pointages plus amples en condition unimodale, tandis que deux enfants présentent le pattern inverse, soit une amplitude plus importante en condition bimodale). L'effet de la Distance est en revanche significatif ($F(1, 9) = 8.34, p < .05$), et ce pour neuf participants : l'amplitude augmente pour la distance la plus éloignée (un seul enfant ne présente pas de différence significative). Des résultats similaires sont observés pour le Groupe 2 : l'amplitude ne varie pas en fonction de la Condition ($F(1, 8) = 0.57, p = .47, NS$), ce qui concerne, au niveau individuel, cinq enfants (trois participants montrent une amplitude plus élevée en condition bimodale et un participant en condition unimodale), mais l'effet de la Distance est significatif ($F(1, 8) = 28.38, p < .01$), l'amplitude augmente avec la distance de la cible, un résultat qui concerne l'ensemble des participants.

Effets d'interaction L'interaction $C*D$ n'est significative pour aucun des groupes ($F(1, 9) = 0.11, p = .75, NS$ pour le Groupe 1 ; $F(1, 8) = 0.43, p = .53, NS$ pour le Groupe 2). Un seul enfant du Groupe 1 présente un effet de la Distance plus important en condition bimodale, tandis que trois enfants (un du Groupe 1, deux du Groupe 2) montrent un effet de la Distance plus important en condition unimodale.

A retenir ! Le Tableau 4.10 présente un bref aperçu des principales différences observées en intra-sujets sur les données cinématiques (amplitude).

Amplitude			
	Groupe 1	Groupe 2	Rappel Adulte
Distance	$D1 < D2$	$D1 < D2$	$D1 < D2$

Tableau 4.10 – Principaux résultats (données cinématiques : amplitude)

Amplitude : Analyse inter-sujets Une analyse inter-sujets s'avère nécessaire, puisque l'effet de la Distance est observé dans les trois groupes. L'interaction $Groupe * Distance$ est significative pour les groupes 2 vs 3. Autrement dit, l'effet observé sur l'amplitude, soit des valeurs plus élevées pour la distance la plus éloignée, est fonction de l'âge des participants. L'analyse au premier ordre, illustrée Figure 4.12, suggère que l'effet de la Distance sur l'amplitude soit plus important chez les enfants âgés de 8 à 12 ans que chez les adultes.

2. Expérience 3

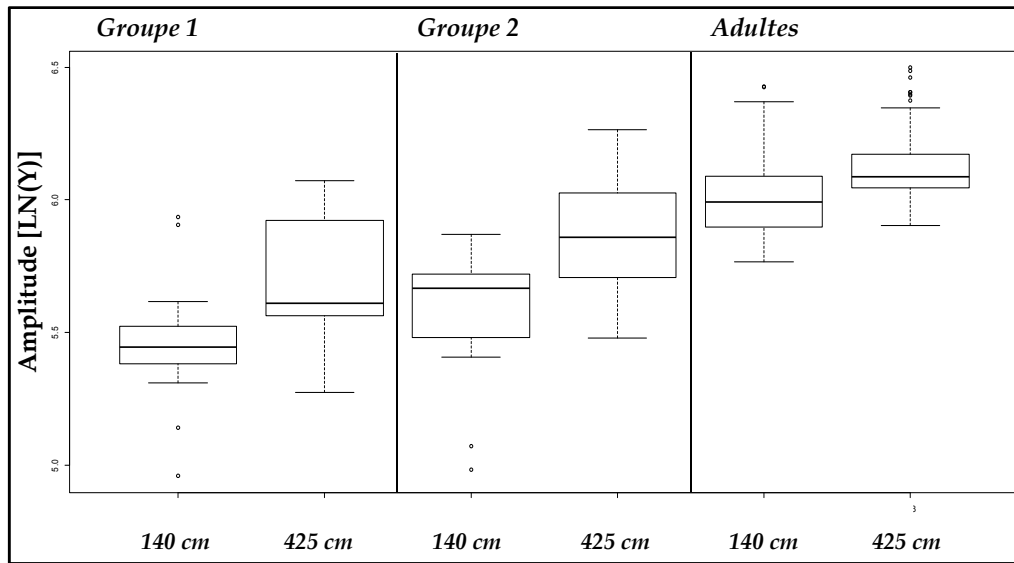


FIGURE 4.12 – *Amplitude* (valeurs représentées après transformation logarithmique) en fonction de la Distance et du Groupe

Durée du plateau : Analyses intra-sujets et individuelle La figure 4.13 montre l'influence de la Condition et de la Distance sur la durée du plateau, pour les groupes 1 et 2.

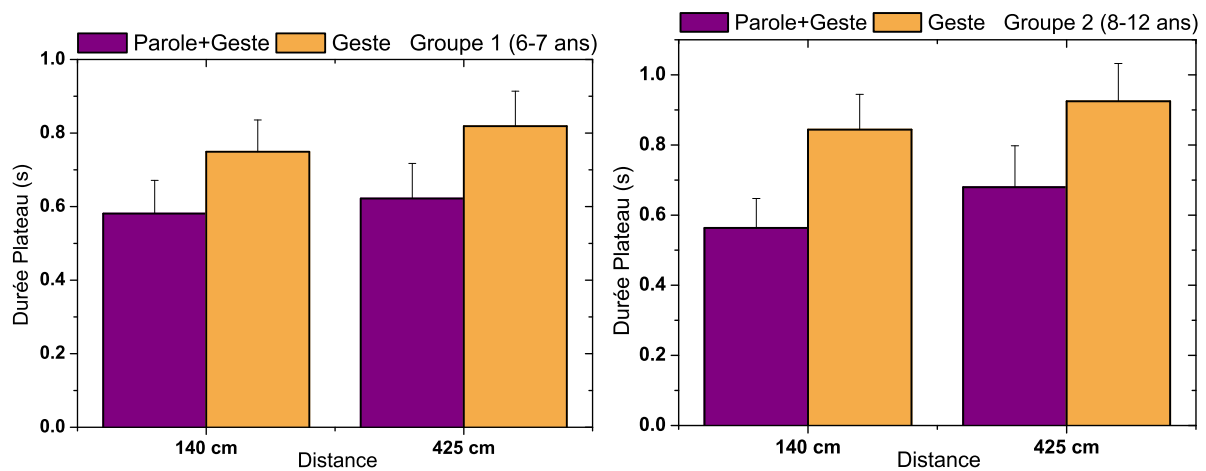


FIGURE 4.13 – *Durée Plateau* (s) en fonction de la Distance et de la Condition

Effets principaux Pour le premier groupe, l'effet de la Condition est significatif ($F(1, 9) = 13.08$, $p < .01$) : le pointage est maintenu plus longtemps en condition unimodale. Ce résultat peut s'observer chez quatre participants, les six autres ne montrent aucune variation. En revanche, l'effet de la Distance n'est pas significatif ($F(1, 9) = 2.39$, $p = 0.15$, *NS*), ce qui concerne neuf enfants, un seul montre une durée du plateau significativement plus longue pour la distance la plus éloignée.

Pour le second groupe, l'effet de la Condition est significatif ($F(1, 8) = 13.77, p < .01$) : la durée du plateau augmente en condition unimodale (ce qui est observé chez six enfants, trois participants en revanche ne présentent aucune différence). Par ailleurs, l'effet de la Distance est également significatif ($F(1, 8) = 6.39, p < .05$) : le pointage est maintenu plus longtemps lorsqu'il est dirigé vers la cible la plus éloignée. Toutefois, ceci caractérise seulement deux participants (les sept autres ne présentent aucune différence).

Effets d'interaction L'interaction $C*D$ n'est significative pour aucun des groupes ($F(1, 9) = 0.09, p = .77, NS$ pour le Groupe 1 ; $F(1, 8) = 0.36, p = .56, NS$ pour le Groupe 2). Deux enfants du Groupe 1 présentent un effet de la Condition plus important pour la distance la plus éloignée, tandis qu'un enfant du Groupe 2 présente un effet de la Condition plus important pour la distance la plus proche.

A retenir ! Le Tableau 4.11 présente un bref aperçu des principales différences observées en intra-sujets sur les données cinématiques (durée du plateau).

	Plateau		
	Groupe 1	Groupe 2	Rappel Adultes
Condition	<i>Unimodale > Bimodale</i>	<i>Unimodale > Bimodale</i>	<i>Unimodale > Bimodale</i>
Distance	ϕ	$D1 < D2$	$D1 < D2$

Tableau 4.11 – Principaux résultats (données cinématiques : durée du plateau)

Durée du plateau : Analyse inter-sujets Des analyses inter-sujets s'avèrent nécessaires, l'effet de la Condition étant observé pour chaque groupe, et celui de la Distance pour les groupes 2 et 3. Les deux interactions *Groupe*Condition* et *Groupe*Distance* sont significatives. L'effet de la Condition sur la durée du plateau diffère pour les groupes 1 vs 3 et 2 vs 3. Les analyses au premier ordre, illustrées Figure 4.14 et Figure 4.15, indiquent que ces deux effets sont plus importants chez l'enfant relativement à l'adulte. Toutefois une variabilité très importante peut être observée au niveau des productions enfantines ; elle est, en moyenne, environ 25× plus importante que celle des productions adultes.

2. Expérience 3

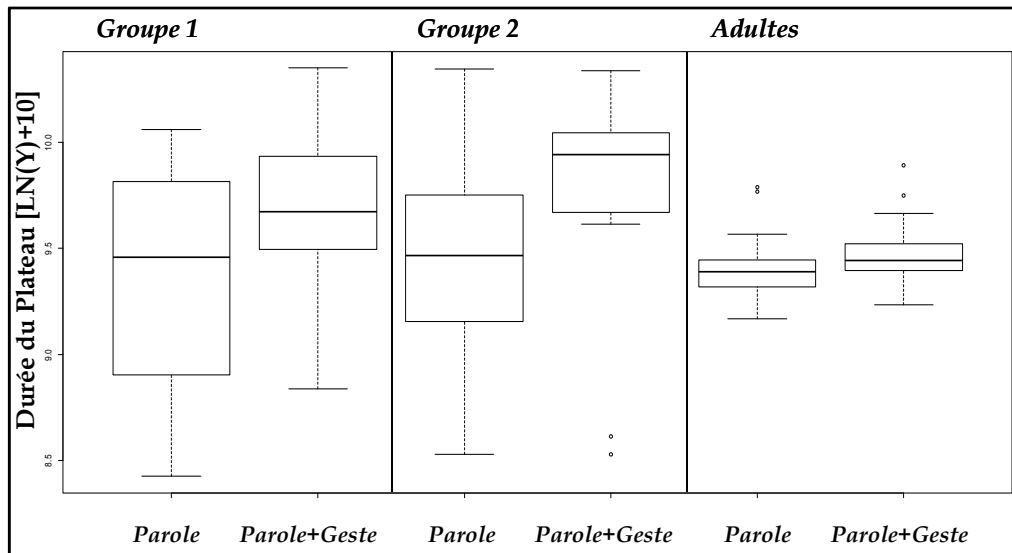


FIGURE 4.14 – *Durée du Plateau* (après transformation logarithmique) en fonction de la Condition et du Groupe.

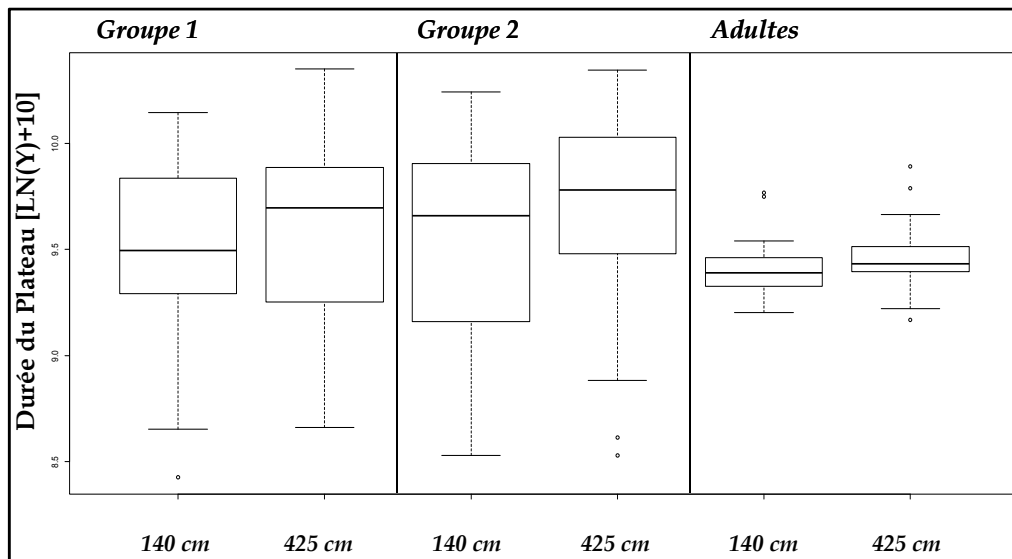


FIGURE 4.15 – *Durée du Plateau* (après transformation logarithmique) en fonction de la Distance et du Groupe.

Pic de vitesse : Analyses intra-sujets et individuelle La figure 4.16 montre l'influence de la Condition et de la Distance sur le pic de vitesse, pour les groupes 1 et 2.

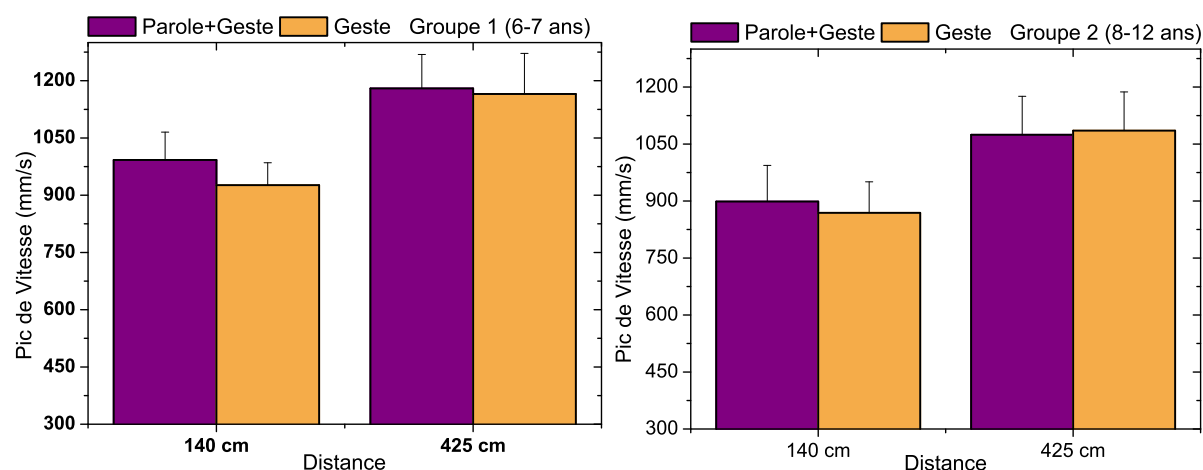


FIGURE 4.16 – *Pic de Vitesse* (mm/s) en fonction de la Distance et de la Condition

Effets principaux Pour le Groupe 1, l'effet de la Condition n'est pas significatif ($F(1, 9) = 0.42, p = .53, NS$), et ce pour six participants (trois participants présentent des valeurs plus élevées en condition bimodale, un participant en condition unimodale). En revanche, l'effet de la Distance est significatif ($F(1, 9) = 9.8, p < .05$), avec des valeurs plus élevées pour la distance la plus éloignée (cet effet est observé chez sept enfants, tandis que trois enfants ne montrent pas de différence).

Pour le Groupe 2, l'effet de la Condition n'est pas significatif ($F(1, 8) = 0.04, p = .84, NS$), et ce pour six participants (deux autres enfants montrent un pic de vitesse plus élevé en condition bimodale, et le dernier en condition unimodale). L'effet de la Distance est significatif ($F(1, 8) = 10.93, p < .05$) : le pic de vitesse augmente avec la distance de la cible, ce qui s'observe chez six participants (les trois autres ne montrent pas de différence).

Effets d'interaction L'interaction $C*D$ n'est significative pour aucun des groupes ($F(1, 9) = 1.79, p = .21, NS$ pour le Groupe 1 ; $F(1, 8) = 2.85, p = .13, NS$ pour le Groupe 2) et ceci pour l'ensemble des participants, groupes confondus.

A retenir ! Le Tableau 4.12 présente un bref aperçu des principales différences observées en intra-sujets sur les données cinématiques (pic de vitesse).

Vitesse			
	Groupe 1	Groupe 2	Rappel Adultes
Distance	$D1 < D2$	$D1 < D2$	$D1 < D2$
Condition	ϕ	ϕ	$Unimodale < Bimodale$

Tableau 4.12 – Principaux résultats (données cinématiques : pic de vitesse)

2. Expérience 3

Pic de vitesse : Analyse inter-sujets Une analyse inter-sujets s'avère nécessaire, puisque l'effet de la Distance (un pic de vitesse plus élevé pour la distance la plus éloignée) est observé pour chacun des groupes. L'interaction *Groupe * Distance* n'est pas significative, ce qui suggère que les performances des enfants, dès l'âge de 6 ans, sont comparables, sur ce point, à celles des adultes.

2.5.4 Synthèse illustrée des principaux résultats de l'Expérience 3

Les figures 4.17 et 4.18 proposent une synthèse des résultats obtenus pour l'effet de la Distance et celui de la Condition. Le "Pattern général" correspond aux résultats observés au niveau d'un groupe dans son ensemble (i.e. le Groupe 1, noté G1 ou le Groupe 2, noté G2), tandis que les "Autres patterns" correspondent aux résultats observés au niveau individuel. Le nombre d'enfants concernés par tel ou tel pattern est indiqué entre crochets. De plus, les patterns correspondant aux résultats observés chez l'adulte sont indiqués en rouge.

	DISTANCE	Pattern général		Autres patterns	
		G1[10]	G2[11]	G1[10]	G2[11]
VOIX	Ouverture des lèvres	D1=D2 [9]	D1<D2 [2]	D1<D2 [1]	D1=D2 [9]
	F1	D1=D2 [8]	D1=D2 [10]	D1<D2 [2]	D1<D2 [1]
	F2	D1=D2 [9]	D1=D2 [10]	D1<D2 [1]	D1<D2 [1]
	F0	D1=D2 [7]	D1=D2 [11]	D1<D2 [2]	[0]
				D1>D2 [1]	
	Intensité	D1=D2 [9]	D1<D2 [3]	D1<D2 [1]	D1=D2 [8]
GESTE		G1[10]	G2[9]	G1[10]	G2[9]
	Amplitude	D1<D2 [9]	D1<D2 [9]	D1=D2 [1]	[0]
	Durée du plateau	D1=D2 [9]	D1<D2 [2]	D1<D2 [1]	D1=D2 [7]
	Pic de vitesse	D1<D2 [7]	D1<D2 [6]	D1=D2 [3]	D1=D2 [3]

FIGURE 4.17 – Synthèse : Effet de la Distance (patterns adultes indiqués en rouge et en gras)

Les principaux résultats concernant l'effet de la Distance, illustrés Figure 4.17, peuvent être résumés en trois points :

1. L'encodage phonétique de la distance tel qu'observé chez l'adulte, une augmentation de l'ouverture des lèvres et de l'intensité du signal de parole pour les objets éloignés, n'est pas observé chez les enfants de 6-7 ans (pour la majorité d'entre eux) mais est observé chez l'enfant de 8-12 ans (pour une minorité d'entre eux).
2. L'encodage cinématique de la distance tel qu'observé chez l'adulte, une augmentation de l'amplitude, du pic de vitesse et de la durée du plateau pour les objets éloignés, n'est observé qu'en partie chez les enfants de 6-7 ans, qui présentent, en majorité, un effet sur

l'amplitude et le pic de vitesse; il est en revanche observé chez les enfants de 8-12 ans (l'effet sur l'amplitude et le pic de vitesse s'observe pour la majorité d'entre eux, l'effet sur la durée du plateau pour une minorité d'entre eux).

3. L'encodage cinématique de la distance semble disponible dans le répertoire de l'enfant avant l'encodage phonétique.

	CONDITION	Pattern général		Autres patterns	
		G1[10]	G2[11]	G1[10]	G2[11]
VOIX	Ouverture des lèvres	Uni>Bimodal [6]	Uni>Bimodal [10]	Uni=Bimodal [4]	Uni=Bimodal [1]
	F1	Uni=Bimodal [3]	Uni=Bimodal [7]	Uni>Bimodal [5] Uni<Bimodal [2]	Uni>Bimodal [4]
	F2	Uni=Bimodal [3]	Uni<Bimodal [3]	Uni<Bimodal [6] Uni>Bimodal [1]	Uni=Bimodal [8]
	F0	Uni=Bimodal [9]	Uni=Bimodal [6]	Uni>Bimodal [1]	Uni>Bimodal [3] Uni<Bimodal [2]
	Intensité	Uni=Bimodal [3]	Uni=Bimodal [8]	Uni>Bimodal [4] Uni<Bimodal [3]	Uni<Bimodal [2] Uni>Bimodal [1]
GESTE		G1[10]	G2[9]	G1[10]	G2[9]
	Amplitude	Uni=Bimodal [7]	Uni=Bimodal [5]	Uni<Bimodal [2] Uni>Bimodal [1]	Uni<Bimodal [3] Uni>Bimodal [2]
	Durée du plateau	Uni>Bimodal [4]	Uni>Bimodal [6]	Uni=Bimodal [6]	Uni=Bimodal [3]
	Pic de vitesse	Uni=Bimodal [6]	Uni=Bimodal [6]	Uni<Bimodal [3] Uni>Bimodal [1]	Uni<Bimodal [2] Uni>Bimodal [1]

FIGURE 4.18 – Synthèse : Effet de la Condition (patterns adultes indiqués en rouge et en gras)

Les principaux résultats concernant l'effet de la Condition, illustrés Figure 4.18, peuvent être résumés en trois points :

1. L'influence du geste sur la parole telle qu'observée chez l'adulte, une diminution de l'ouverture des lèvres et du premier formant en situation bimodale, n'est observée qu'en partie chez l'enfant : les enfants de 6-7 ans et ceux de 8-12 ans présentent un effet sur l'ouverture des lèvres mais non sur le premier formant. Par ailleurs, les enfants de 8-12 ans présentent une augmentation du second formant en situation bimodale.
2. L'influence de la parole sur le geste telle qu'observée chez l'adulte, une diminution de la durée du plateau associée à une augmentation de la vitesse en situation bimodale, n'est observée qu'en partie chez l'enfant : les enfants de 6-7 ans et les enfants de 8-12 ans présentent un effet sur la durée du plateau mais non sur le pic de vitesse.

2. Expérience 3

3. L'influence du geste sur l'ouverture des lèvres concerne la majorité des enfants de 6-7 ans et de 8-12 ans ; de plus, l'effet sur le premier formant est observé au niveau individuel pour près de la moitié des enfants, âges confondus. En revanche, l'influence de la parole sur la durée du plateau ne concerne qu'une minorité des enfants de 6-7 ans et la majorité des enfants de 8-12 ans ; de plus l'effet sur le pic de vitesse n'est observé au niveau individuel que pour une minorité d'enfants, âges confondus.

La figure 4.19 représente quant à elle le pourcentage d'enfants qui, au niveau individuel, respectent les patterns observés chez l'adulte (noté "Modèle théorique adulte"), en termes d'encodage moteur de la distance et de coopération inter-modalités, et ce au niveau du pointage vocal ("Voix") et du pointage manuel ("Geste").

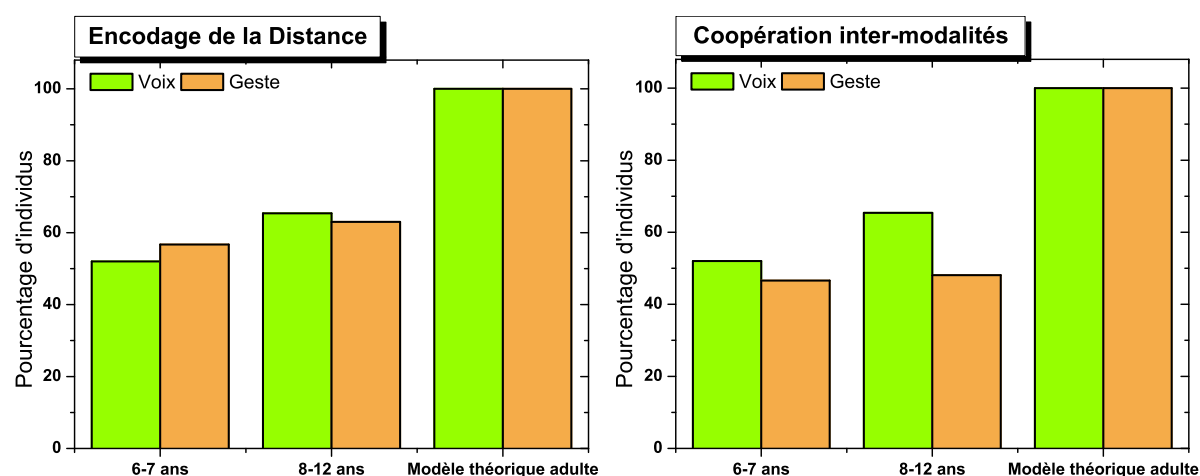


FIGURE 4.19 – Encodage de la distance & coopération inter-modalités chez l'enfant et l'adulte

Ces données peuvent être résumées en trois points :

1. La proportion d'enfants âgés de 8 à 12 ans, qui respectent les schémas adultes dans la modalité vocale, est plus élevée que celle des enfants de 6-7 ans.
2. La proportion d'enfants âgés de 8 à 12 ans, qui respectent les schémas adultes dans la modalité manuelle, est comparable à celle des enfants de 6-7 ans, et plus particulièrement concernant la coopération inter-modalités.
3. L'évolution du pointage vers un modèle adulte se fait progressivement ; cette progression semble plus marquée dans la modalité vocale.

2.6 Implications théoriques

Le lecteur pourra se rapporter aux figures 4.17, 4.18 et 4.19 pour une meilleure lecture des implications théoriques présentées ci-après.

L'objectif de notre Expérience 3 était de caractériser l'évolution des productions vocales et manuelles, et de leur interaction, au cours du développement, chez l'enfant d'âge scolaire.

Cette étude concernait plus particulièrement deux phénomènes observés chez l'adulte, à savoir (1) un encodage moteur de la distance, via les propriétés phonétiques et cinématiques du pointage, et (2) une interaction structurelle et bidirectionnelle, de nature coopérative, entre les systèmes de production de parole et de geste manuel ; deux phénomènes que nous envisageons comme des compétences tardives, élaborées et intégrées progressivement au répertoire communicatif des locuteurs. De fait, nous attendions des différences relatives au groupe d'âge (6-7 ans, 8-12 ans, adultes), avec notamment, chez le jeune enfant, une période "manuelle", caractérisée par un encodage cinématique de la distance ainsi qu'une influence du geste sur la parole, suivie d'une intégration progressive de nouvelles stratégies, aux alentours de 8 ans, caractérisée par l'émergence d'un encodage phonétique et d'une influence réciproque de la parole sur le geste.

2.6.1 Encodage moteur de la distance

Chez l'adulte, l'encodage moteur de la distance est manifeste au niveau de divers paramètres cinématiques (amplitude, pic de vitesse, durée totale, tenue du pointage manuel) et phonétiques (ouverture des lèvres, intensité du signal de parole). Ces différents paramètres augmentent avec la distance du référent, ce qui suggère l'utilisation d'une stratégie particulière pour désigner, de façon explicite, une cible éloignée, et ambiguë.

- Encodage cinématique de la distance

Au niveau du pointage manuel, on observe chez l'enfant âgé de 6-7 ans, un encodage de la distance au niveau de l'amplitude et du pic de vitesse. Ces deux paramètres augmentent avec la distance de la cible, un résultat plutôt robuste, puisqu'il concerne au niveau individuel une grande majorité d'enfants (respectivement neuf et sept enfants sur 10). La durée du plateau en revanche, ne varie pas en fonction de l'emplacement de l'objet (seul un enfant maintient ses pointages plus longtemps lorsque ceux-ci désignent un objet distant).

Chez l'enfant âgé de 8-12 ans, on observe un encodage de la distance via les trois paramètres testés : l'amplitude, le pic de vitesse et la durée du plateau augmentent avec la distance de la cible. Les résultats relatifs à l'amplitude et au pic de vitesse sont observés pour une grande majorité d'enfants (respectivement neuf et six enfants sur neuf), ce qui n'est pas le cas du résultat relatif à la durée du plateau, qui semble émerger à cette période. Il est en effet observé chez deux enfants seulement (âgés de 9.11 ans et 12.5 ans, situés dans la tranche d'âge supérieure de leur groupe), les neuf autres enfants ayant des résultats semblables à ceux des plus jeunes.

Par ailleurs, les analyses inter-groupes suggèrent un encodage de la distance via la vitesse du pointage relativement précoce, c'est-à-dire observé de manière indifférenciée pour les trois groupes d'âge. En revanche, l'effet observé sur l'amplitude ne diffère qu'entre les enfants de 8-12 ans et les adultes. En d'autres termes, l'effet observé chez les petits de 6-7 ans est comparable à celui observé chez les enfants plus âgés et chez les adultes ; en revanche, les enfants âgés de 8-12 ans montrent un effet plus important que les adultes. De la même façon, l'effet observé sur la durée du plateau (dans les Groupes 2 et 3) est significativement plus important pour les enfants de 8-12 ans que pour les adultes. De tels résultats (un effet plus important chez l'enfant relativement à l'adulte), pourraient traduire une période de transition, caractérisée par l'émergence puis l'exercice intensif des nouvelles stratégies dont l'enfant dispose.

2. Expérience 3

Néanmoins, ces résultats sont à interpréter avec précaution, le déséquilibre inter-groupes et la grande dispersion des productions chez l'enfant, pouvant en effet biaiser l'effet de nos différents facteurs.

- Encodage phonétique de la distance

Au niveau du pointage vocal, nous n'observons pas, chez l'enfant âgé de 6-7 ans, d'encodage de la distance. Cette absence d'encodage phonétique est observée pour neuf enfants sur 10 (tant au niveau des variations articulatoires que des variations d'intensité).

En revanche, chez l'enfant âgé de 8-12 ans, on observe un encodage de la distance comparable à celui observé chez l'adulte (i.e. absence d'interaction *Groupe * VI_{INTRA}*) : autrement dit, l'ouverture des lèvres et l'intensité du signal augmentent pour la distance la plus grande. Néanmoins, ce résultat ne concerne qu'une minorité d'enfants (deux enfants pour l'ouverture des lèvres, trois pour l'intensité) ; tous les autres enfants, sans exception, montrent des résultats similaires à ceux des enfants plus jeunes. L'encodage phonétique de la distance semble donc émerger aux alentours de 8-12 ans, plus tardivement que l'encodage cinématique.

Par ailleurs, l'absence d'effet de la Distance sur la fréquence fondamentale, le premier et le second formant, observée chez l'adulte, est également observée pour la majorité des enfants de 6-7 ans et de 8-12 ans. Concernant la fréquence fondamentale, seuls deux enfants de 6-7 ans montrent des valeurs plus élevées pour la distance la plus grande, et un enfant présente un résultat inverse, soit une fréquence fondamentale plus élevée pour la distance la plus petite. Concernant les variations du second formant, un seul enfant du premier groupe et un seul enfant du second groupe présentent des valeurs plus élevées pour la distance la plus éloignée. Potentiellement, l'acquisition de l'encodage phonétique de la distance nécessite de tester différentes stratégies (notamment des variations prosodiques, soit une accentuation particulière, ou une stratégie de protrusion plutôt que d'ouverture), avant de pouvoir atteindre les performances adultes. Concernant les variations du premier formant, deux enfants du premier groupe et un enfant du second groupe montrent des valeurs plus élevées pour la distance éloignée. Chez l'adulte, l'absence d'effet sur le premier formant était attribuée à un phénomène de compensation articulatoire (cf. Lindblom et collab. [44], Maeda [46] et Section 3 du Chapitre 3) ; il est de fait possible que certains enfants ne maîtrisent pas encore cette stratégie, tandis que les autres présentent une compensation articulatoire plus précoce.

Ces résultats, relatifs à l'encodage moteur de la distance, indiquent une évolution progressive des conduites, en direction des usages attestés chez l'adulte. L'encodage se met progressivement en place au cours du développement tardif de l'enfant. Visiblement, l'encodage cinématique émerge plus précocement (dès l'âge de 6 ans) que l'encodage phonétique (à partir de 8 ans). Ces résultats indiquent que l'encodage moteur de la distance via le pointage est un outil communicatif élaboré, intégré progressivement au répertoire communicatif du locuteur, et qui nécessite de fait l'apprentissage de divers mécanismes et stratégies de production. Il est envisageable qu'avant l'âge de 8 ans, l'enfant ne puisse maîtriser pleinement les stratégies émergentes dont il dispose ; celles-ci deviendront comparables à celles des adultes plus tardivement, après une période de transition. Ces données sont cohérentes avec le pattern évolutif observé d'une part, pour les performances motrices, et d'autre part, pour les performances verbales (voir les études présentées en Section 1.3, e.g. Badan et collab. [3], Colletta [15]).

2.6.2 Interaction voix/geste : coopération inter-modalités

L'interaction geste/parole chez l'adulte est de nature structurelle, bidirectionnelle, coopérative. En d'autres termes, les systèmes de production de parole et de gestes s'influencent mutuellement et leur contenu informatif est distribué en fonction de la, ou des modalités disponibles. Les productions bimodales, relativement aux productions unimodales, sont inhibées ; parallèlement, l'absence (contrainte) de l'une de ces modalités est compensée par un renforcement au niveau de l'autre modalité. Les propriétés acoustiques (F1) et articulatoires (ouverture des lèvres) du pointage vocal, ainsi que les propriétés cinématiques du pointage manuel (pic de vitesse, durée totale et tenue) sont ainsi renforcées lorsqu'une seule modalité est disponible pour transmettre l'information. On observe également, au niveau articulatoire, une interaction *Condition * Distance* ; celle-ci indique que l'encodage articulatoire est plus important en situation unimodale ; lorsque l'information visuelle, véhiculée en théorie par le geste manuel, est absente, elle est compensée par une ouverture des lèvres plus importante.

- Influence du geste sur la parole

L'influence du geste sur la parole (une variation articulatoire et acoustique) n'est observée qu'en partie chez les enfants de 6-7 ans et chez les enfants de 8-12 ans. La majorité des enfants, dans les deux groupes, montrent un effet sur l'ouverture des lèvres, soit une ouverture plus grande pour désigner l'objet éloigné. Par ailleurs, l'interaction *Condition * Distance* au niveau articulatoire émerge chez les enfants, à partir de 8-12 ans ($p = .05$). En revanche, l'effet sur le premier formant n'est significatif pour aucun des groupes. Néanmoins, cet effet est observé, au niveau individuel, pour cinq enfants du premier groupe et quatre enfants du second groupe. Ces données indiquent clairement l'émergence, dès 6 ans, d'une influence du geste sur les paramètres acoustiques du pointage vocal.

Nos résultats montrent également une influence du geste sur le second formant, qui est plus élevé en situation bimodale (des valeurs élevées de F2, soit un positionnement antérieur de la langue dans la bouche, sont cohérentes avec des valeurs basses de F1), effet par ailleurs inexistant chez l'adulte. Plus précisément, cet effet est observé pour six enfants du premier groupe (soit une majorité) sans être significatif au niveau global, tandis qu'il est observé au niveau du second groupe, mais ne concerne plus que trois enfants (les huit autres suivent le schéma adulte). Il semblerait que les enfants, dès l'âge de 6 ans, voire plus tôt, utilisent une stratégie coopérative précoce, basée sur des variations du second formant, avant d'évoluer progressivement vers une stratégie coopérative plus élaborée, de même type que chez adulte, i.e. liée à des variations du premier formant. A ce propos, Ménard [50] montre une évolution des affiliations formant/cavité, entre la naissance et l'âge adulte, qui pourrait expliquer cette transition. Son étude porte sur la comparaison de nourrissons, d'enfants âgés de 4 et 10 ans, d'adolescents de 16 ans et d'adultes de 21 ans. Relativement aux affiliations observées chez l'homme adulte (pour lequel le premier formant est associé à la cavité arrière et le second formant à la cavité avant²⁰), l'auteur observe chez le nourrisson, pour la voyelle [a], une antériorisation du corps de la langue, associée à une position plus élevée du dos de la langue : F1 est alors affilié à la cavité avant, F2 à la cavité arrière, et les lèvres apparaissent également plus ouvertes. Pour les enfants de 10 et 16 ans, si le corps de la langue est postériorisé, la position du dos de la langue en revanche reste inchangée.

20. Les cavités arrière/avant correspondent, dans l'étude de Ménard [50], aux cavités articulatoires créées par constriction en un point du conduit vocal (et non aux cavités physiologiques, pharyngale et buccale).

2. Expérience 3

Par ailleurs, l'absence d'effet observée chez l'adulte au niveau de la fréquence fondamentale et de l'intensité est également observée chez l'enfant, au niveau de chaque groupe. Néanmoins, la variabilité des productions individuelles révèle l'existence de multiples stratégies prosodiques, aussi bien chez les enfants de 6-7 ans (en particulier au niveau de l'intensité, où non moins de trois stratégies sont observées) que chez les enfants de 8-12 ans (en particulier au niveau de la fréquence fondamentale, où trois stratégies différentes sont également observées).

- Influence de la parole sur le geste

L'influence de la parole sur le geste, soit une augmentation de la tenue du pointage associée à une diminution du pic de vitesse, n'est observée qu'en partie chez les enfants de 6-7 ans et de 8-12 ans. L'effet relatif à la durée du plateau est observé dans les deux groupes, mais pour seulement quatre enfants du premier groupe (i.e. une minorité) et pour six enfants du second groupe (i.e. une majorité). L'effet relatif au pic de vitesse n'est quant à lui significatif pour aucun des groupes (et ce pour la majorité des enfants, âges confondus). De fait, l'influence de la parole sur le geste semble apparaître plus tardivement que celle du geste sur la parole. Par ailleurs, l'absence d'effet au niveau de l'amplitude chez l'adulte est observée pour la majorité des enfants (néanmoins, trois stratégies peuvent être identifiées au niveau individuel, et ce dans chaque groupe d'âge).

Les analyses inter-groupes, concernant l'effet de la Condition sur l'ouverture des lèvres et sur la durée du plateau, montrent que (1) les deux groupes d'enfants ne diffèrent pas l'un de l'autre, (2) chaque groupe d'enfants diffère du groupe adulte. L'effet de la Condition est en effet plus important chez l'enfant. De la même façon que précédemment, ce phénomène peut s'expliquer par une période de transition, caractérisée par l'apprentissage intensif de nouvelles stratégies (tout en gardant à l'esprit d'éventuels biais, liés au petit nombre de participants et à la dispersion des données). Par ailleurs, si la coopération entre deux modalités, qui intervient en situation bimodale, est une stratégie nouvelle, elle est par définition plus coûteuse et nécessite potentiellement une charge cognitive plus élevée. Si l'on considère qu'une charge cognitive importante induit une baisse de performances, soit une inhibition des productions, alors l'écart entre les productions unimodales et bimodales se creuse chez l'enfant (i.e. l'effet de la Condition est plus important chez l'enfant que chez l'adulte). Ceux-ci doivent en effet gérer (1) une charge cognitive élevée, liée à l'émergence de nouvelles stratégies et (2) une coopération inter-modalités (soit une "double inhibition" en situation bimodale).

Ces résultats indiquent une évolution progressive de l'interaction geste/parole, qui se met en place au cours du développement tardif de l'enfant et constitue en cela une stratégie communicative élaborée. L'influence du geste sur la parole semble précéder celle de la parole sur le geste, puisque la première est observée pour une majorité d'enfants âgés de 6-7 ans, tandis que la seconde n'en concerne qu'une minorité. D'autre part, le nombre de patterns observé pour l'effet de la Condition est sensiblement plus important que celui observé pour l'effet de la Distance ; il témoigne de la diversité des stratégies utilisées par les enfants à un âge donné, avant qu'une coopération mutuelle ne puisse être mise en place.

3 Conclusion

Les résultats de notre étude sont en faveur d'une évolution progressive du couplage geste et parole, au niveau du processus déictique. L'encodage moteur de la distance, ainsi que la coopération inter-modalités, se mettent en place au cours du développement tardif de l'enfant. De tels comportements nécessitent l'acquisition, par la pratique, de stratégies élaborées, intégrées progressivement au répertoire communicatif de l'enfant. On observe également un léger "avantage" du geste sur la parole chez les enfants les plus jeunes; l'information spatiale serait principalement portée par le geste avant d'être véhiculée via les deux modalités. Ainsi, l'encodage cinématique de la distance apparaît chez l'enfant de 6-7 ans tandis que l'encodage phonétique n'émerge qu'aux alentours de 8-12 ans; de la même façon, l'influence du geste sur la parole émerge plus précocement que l'influence de la parole sur le geste.

Néanmoins, il s'avère essentiel de poursuivre cette étude, afin d'augmenter (et d'équilibrer) le nombre de participants de chaque groupe. L'idéal serait de pouvoir évaluer les enfants par tranche d'âges minimales (soit d'un an), afin d'établir à long terme des profils développementaux précis et d'analyser finement la transition entre un modèle de production de gestes et de parole enfant et un modèle adulte. Par ailleurs, comme nous l'avons suggéré en Conclusion du Chapitre 3, ces premiers résultats pourraient être améliorés, notamment par le biais de procédures de normalisation sur nos données, permettant de réduire considérablement la variance de nos échantillons et de favoriser l'apparition d'effets plus robustes. Par exemple, certaines procédures de normalisation sur les données articulatoires et acoustiques permettraient de prendre en compte les différences anatomiques inter-locuteurs, et de limiter leurs répercussions au niveau de la fréquence fondamentale et des valeurs de formants (voir à ce propos les travaux de Nguyen et Espesser [51], présentés brièvement en Conclusion du Chapitre 3).

L'objectif général de ce travail de thèse étant d'évaluer précisément l'intégration geste/parole dans les mécanismes du langage, le dernier chapitre a pour but d'étudier un tel couplage via les mécanismes de perception du langage. Inspirés de l'hypothèse des systèmes intégrés, proposée par Kelly et collab. [39], et de la Théorie de la Perception pour le Contrôle de l'Action, proposée par Schwartz et collab. [60], nous avons mis au point un protocole expérimental basé sur le paradigme d'amorçage inter-modalités. Si geste et parole sont intégrés de manière générale dans les processus langagiers, produire un geste déictique pourrait bien faciliter la perception d'un terme déictique.

Annexe : Les tâches de conservation piagésiennes

Les épreuves piagésiennes de conservation physique et spatiale (Piaget [58]), dont les exemples les plus célèbres sont les tâches de conservation du liquide et du nombre (voir l'exemple illustré Figure 4.20) sont basées sur les trois étapes suivantes :

- Egalité initiale,
- Transformation de certaines caractéristiques de l'objet,
- Egalité finale.

L'enfant est dit conservant lorsqu'il reconnaît l'égalité finale, par jugement d'identité, de compensation ou de réversibilité.

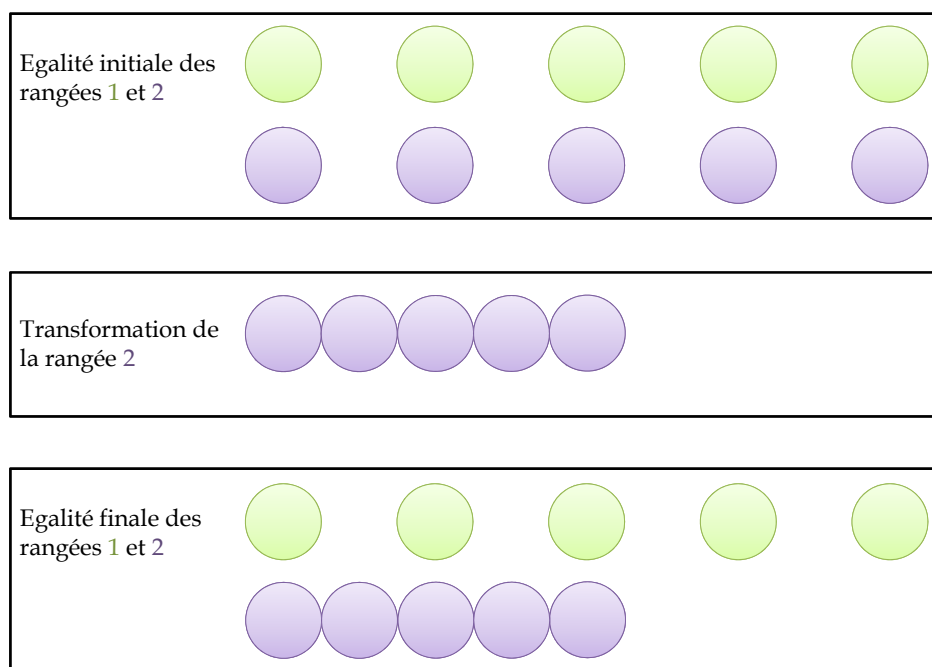


FIGURE 4.20 – Exemple de conservation spatiale : Les deux rangées sont-elles toujours identiques, en termes de nombre de jetons, **après** transformation ?

Bibliographie

- [1] Alibali, M. et S. Goldin-Meadow. 1993, «Gesture-speech mismatch and mechanisms of learning : What the hands reveal about a child's state of mind», *Cognitive Psychology*, vol. 25, p. 468–523.
- [2] Allen, G., K. Kirasic et R. Beard. 1989, «Children's expressions of spatial knowledge», *Journal of Experimental Child Psychology*, vol. 48, p. 114–130.
- [3] Badan, M., C. Hauert et P. Mounoud. 2000, «Sequential pointing in children and adults», *Journal of Experimental Child Psychology*, vol. 75, p. 43–69.
- [4] Bates, E., L. Benigni, I. Bretherton, L. Camaioni et V. Volterra. 1979, *The Emergence of Symbols : Cognition and Communication in Infancy*, New York : Academic Press.
- [5] Bates, E., D. Thal et V. Marchman. 1991, «Symbols and syntax : A Darwinian approach to language development», dans *Biological and Behavioral Determinants of Language Development*, édité par N. Krasnegor, D. Rumbaugh, R. Schiefelbusch et M. Studdert-Kennedy, Hillsdale, N.J. : Lawrence Erlbaum Associates, p. 29–65.
- [6] Batista, A. et M. Le Normand. 2010, «Gestes, verbalisations et combinaisons bimodales dans les productions d'enfants français âgés de 18 mois à 4 ans et demi», *Lidil (Multimodalité de la communication chez l'enfant)*, vol. 42, p. 55–75.
- [7] van Beinum, F., C. Clement et I. Van Den Dikkenberg-Pot. 2001, «Babbling and the lack of auditory speech perception : A matter of coordination?», *Developmental Science*, vol. 4(1), p. 61–70.
- [8] Bruner, J. 1983, *Child Talk*, Norton, New York.
- [9] Butcher, C. et S. Goldin-Meadow. 2000, «Gesture and the transition from one- to two-word speech : When hand and mouth come together», dans *Language and Gesture*, édité par D. McNeill, Cambridge University Press, New York, p. 235–257.
- [10] Butterworth, G. 1998, «What is special about pointing in babies?», dans *The development of sensory motor and cognitive capacities in early infancy, from perception to cognition*, vol. Simion, F. and Butterworth, G., Hove, England : Psychology Press/Erlbaum, p. 171–190.
- [11] Capone, N. et K. McGregor. 2004, «Gesture development : A review of clinical and research practices», *Journal of Speech, Language, and Hearing Research*, vol. 48, p. 173–186.
- [12] Caselli, M. 1990, «Communicative gestures and first words», dans *From Gesture to Language in Hearing and Deaf Children*, édité par V. Volterra et C. Erting, New York : Springer-Verlag, p. 56–67.
- [13] Caselli, M., V. Volterra et E. Pizzuto. 1984, «The relationship between vocal and gestural communication from the one-word to the two-word stage», dans *International Conference on Infant Studies (New York)*.

Bibliographie

- [14] Church, R. et S. Goldin-Meadow. 1986, «The mismatch between gesture and speech as an index of transitional knowledge», *Cognition*, vol. 23, p. 43–71.
- [15] Colletta, J. 2004, *Le développement de la parole chez l'enfant âgé de 6 à 11 ans : Corps, langage et cognition*, vol. 254, Hayen : Mardaga, Psychologie et Sciences Humaines.
- [16] Colletta, J. 2009, «Comparative analysis of children's narratives at different ages : A multimodal approach», *Gesture*, vol. 9(1), p. 61–97.
- [17] Connolly, K., K. Brown et E. Bassett. 1968, «Developmental changes in some components of a motor skill», *British Journal of Psychology*, vol. 59, p. 305–314.
- [18] Ejiri, K. 1998, «Synchronization between preverbal vocalizations and motor actions in early infancy : I. Pre-canonical babbling vocalizations synchronize with rhythmic body movements before the onset of canonical babbling», *Japanese Journal of Psychology*, vol. 68, p. 433–440.
- [19] Ejiri, K. et N. Masataka. 1999, «Synchronization between preverbal vocal behavior and motor action in early infancy : II. An acoustical examination of the functional significance of the synchronization», *Japanese Journal of Psychology*, vol. 69, p. 433–446.
- [20] Fogel, A. et T. E. Hannan. 1985, «Manual actions of nine- and fifteen-week-old human infants during face-to-face interaction with their mothers», *Child Development*, vol. 56, p. 1271–79.
- [21] Gesell, A. et F. Ilg. 1946, *Infant from 5 to 10*, New York : Harper & Row.
- [22] Gillis, S., K. Schauwers et P. Govaerts. 2002, «Babbling milestones and beyond : Early speech development in CI children», dans *Language Acquisition in Very Young Children With a Cochlear Implant*, édité par K. Schauwers, P. Govaerts et S. Gillis, University of Antwerp, p. 23–40.
- [23] Goldin-Meadow, S. et C. Butcher. 2003, «Pointing toward two-word speech in young children», dans *Pointing : Where Language, Culture, and Cognition Meet*, édité par S. Kita, Mahwah, NJ, US : Lawrence Erlbaum Associates Publishers, p. 85–107.
- [24] Goldin-Meadow, S., H. Nusbaum, P. Garber et R. Church. 1993, «Transitions in learning : Evidence for simultaneously activated strategies», *Journal of Experimental Psychology : Human Perception and Performance*, vol. 19(1), p. 92–107.
- [25] Goodwyn, S. et L. Acredolo. 1993, «Symbolic gesture versus word : Is there a modality advantage for onset of symbol use?», *Child Development*, vol. 64(3), p. 688–701.
- [26] Goodwyn, S. et L. Acredolo. 1998, «Encouraging symbolic gestures : A new perspective on the relationship between gesture and speech», dans *The Nature and Functions of Gesture in Children's Communications : New Directions for Child Development*, vol. 79, édité par J. Iverson et S. Goldin-Meadow, San Francisco : Jossey Bass, p. 61–73.
- [27] Greenhouse, S. et S. Geisser. 1959, «On methods in the analysis of profile data», *Psychometrika*, vol. 24, p. 95–112.

- [28] Hay, L. 1978, «Accuracy of children on an open-loop pointing task», *Perceptual and Motor Skills*, vol. 47, p. 1079–1082.
- [29] Hay, L. 1979, «Le mouvement dirigé vers un objectif visuel, chez l'adulte et chez l'enfant», *L'Année Psychologique*, vol. 79(2), p. 559–588.
- [30] Hay, L. 1979, «Spatial-temporal analysis of movements in children : Motor programs versus feedback in the development of reaching», *Journal of Motor Behavior*, vol. 11, p. 189–200.
- [31] Hazen, N., J. Lockman et H. J. Pick. 1978, «The development of children's representations of large-scale environments», *Child Development*, vol. 49, p. 623–636.
- [32] Holender, D. 1980, «Interference between a vocal and a manual response to the same stimulus», dans *Tutorials in motor behavior*, édité par G. Stelmach et J. Requin, Amsterdam : North-Holland, p. 421–431.
- [33] Iverson, J. 2010, «Developing language in a developing body : The relationship between motor development and language development», *Journal of Child Language*, vol. 37, p. 229–261.
- [34] Iverson, J., O. Capirci et M. Caselli. 1994, «From communication to language in two modalities», *Cognitive Development*, vol. 9, p. 23–43.
- [35] Iverson, J. et S. Goldin-Meadow. 2005, «Gesture paves the way for language development», *Psychological Science*, vol. 16, n° 5, p. 367–371.
- [36] Iverson, J. et E. Thelen. 1999, «Hand, mouth and brain : The dynamic emergence of speech and gesture», *Journal of Consciousness Studies*, vol. 6(11-12), p. 19–40.
- [37] Keller, J., P. Fleurance et B. Candau. 1987, «Ontogenèse d'une habileté motrice chez les enfants de 3 à 9 ans dans une tâche d'anticipation-coïncidence : L'attraper de balle», dans *Recherches en psychologie du sport*, édité par A. Vom Hofe et Simonet, Issy-les-Moulineaux, p. 86–97.
- [38] Kelly, S., S. Manning et S. Rodak. 2008, «Gesture gives a hand to language and learning : Perspectives from cognitive neuroscience, developmental psychology and education», *Language and Linguistics Compass*, vol. 2, p. 569–588.
- [39] Kelly, S., A. Ozyürek et E. Maris. 2010, «Two sides of the same coin : Speech and gesture mutually interact to enhance comprehension», *Psychological Science*, vol. 21(2), p. 260–267.
- [40] Kelso, J. 1995, *Dynamic Patterns : The Self-Organization of Brain and Behavior*, Cambridge, MA : MIT Press.
- [41] Kugler, P. et M. Turvey. 1987, *Information, Natural Law, and the Self-Assembly of Rhythmic Movement*, Hillsdale, NJ : Erlbaum.
- [42] Levelt, W., G. Richardson et W. L. Heij. 1985, «Pointing and voicing in deictic expressions», *Journal of Memory and Language*, vol. 24, p. 133–164.

Bibliographie

- [43] Lew, A. et G. Butterworth. 1997, «The development of hand-mouth coordination in 2- to 5-month-old infants : Similarities with reaching and grasping», *Infant Behavior and Development*, vol. 20, p. 59–69.
- [44] Lindblom, B., J. Lubker et T. Gay. 1979, «Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation», *Journal of Phonetics*, vol. 7, p. 147–161.
- [45] Locke, J., K. Bekken, L. McMinn-Larson et D. Wein. 1995, «Emergent control of manual and vocal-motor activity in relation to the development of speech», *Brain and Language*, vol. 51, p. 498–508.
- [46] Maeda, S. 1991, «On articulatory and acoustic variabilities», *Journal of Phonetics*, vol. 19, p. 321–331.
- [47] Mauchly, J. 1940, «Significance test for sphericity of a normal n -variate distribution», *The Annals of Mathematical Statistics*, vol. 11, p. 204–209.
- [48] McNeill, D. 1992, *Hand and Mind : What Gestures Reveal about Thought*, University of Chicago Press.
- [49] Meier, R. et R. Willerman. 1995, «Prelinguistic gesture in deaf and hearing infants», dans *Language, Gesture, and Space*, édité par K. Emmorey et J. Reilly, Hillsdale, NJ : Erlbaum, p. 391–409.
- [50] Ménard, L. 2002, *Production et perception des voyelles au cours de la croissance du conduit vocal : Variabilité, invariance et normalisation*, thèse de doctorat, Université Stendhal (Grenoble III).
- [51] Nguyen, N. et R. Espesser. 2004, «Méthodes et outils pour l'analyse acoustique des systèmes vocaliques», *Bulletin PFC (Phonologie du Français Contemporain)*, vol. 3, n° 3, p. 77–85. 2182 2182 Programme interdisciplinaire CNRS TCAN.
- [52] Oller, D. et R. Eilers. 1988, «The role of audition in infant babbling», *Child Development*, vol. 59, p. 441–466.
- [53] Özçalışkan, S. et S. Goldin-Meadow. 2005, «Gesture is at the cutting edge of early language development», *Cognition*, vol. 96(3), p. 101–113.
- [54] Pechmann, T. et W. Deutsch. 1982, «The development of verbal and nonverbal devices for reference», *Journal of Experimental Child Psychology*, vol. 34, p. 330–341.
- [55] Pellizzer, G. et C. Hauert. 1996, «Visuomanual aiming movements in 6- to 10-year-old children : Evidence for an asymmetric and asynchronous development of information processes», *Brain and Cognition*, vol. 30, p. 175–193.
- [56] Petitto, L., S. Holowka, L. Sergio et D. Ostry. 2001, «Language rhythms in baby hand movements : Hearing babies born to deaf parents babble silently with their hands», *Nature*, vol. 413, p. 35–36.

- [57] Petitto, L. et P. Marentette. 1991, «Babbling in the manual mode : Evidence for the ontogeny of language», *Science*, vol. 251, p. 1493–1496.
- [58] Piaget, J. 1936, *La naissance de l'intelligence chez l'Enfant*, Delachaux et Niestlé.
- [59] Ripoll, H., J. Keller et I. Olivier. 1994, «Le développement du comportement moteur de l'enfant : L'exemple des saisies et des interceptions de balle», *Enfance*, vol. 47 (2-3), p. 265–284.
- [60] Schwartz, J., A. Basirat, L. Ménard et M. Sato. 2012, «The Perception-for-Action-Control Theory (PACT) : A perceptuo-motor theory of speech perception», *Journal of Neurolinguistics*, vol. 25(5), p. 336–354.
- [61] Shapiro, S. 1965, «An analysis of variance test for normality (complete samples)», *Biometrika*, vol. 52, p. 591–611.
- [62] Thal, D. et S. Tobias. 1992, «Communicative gestures in children with delayed onset of oral expressive vocabulary», *Journal of Speech and Hearing Research*, vol. 35, p. 1281–1289.
- [63] Tourrette, C. 1989, «La méthode longitudinale en psychologie développementale», dans *La Méthode Expérimentale en Psychologie*, édité par J. Rossi, Paris : Dunod, p. 180–237.
- [64] Volterra, V., M. Caselli, O. Capirci et E. Pizzuto. 2005, «Gesture and the emergence and development of language», dans *Elizabeth Bates : A Festschrift*, édité par M. Tomasello et D. Slobin, Mahwah, N.J. : Lawrence Erlbaum Associates, p. 3–40.
- [65] Von Holst, E. 1973, *The Behavioural Physiology of Animals and Man : The Collected Papers of Erich Von Holst (Vol.1)*, London : Methuen.

Couplage voix/geste dans la perception du langage

L'objectif de ce dernier chapitre est d'évaluer l'intégration des représentations vocales et manuelles dans les mécanismes de perception du langage. Comme le suggèrent en effet Kelly et collab. [51], « *the integrated relationship may hold for language comprehension as well*¹ » ([51], p.266). Notre hypothèse, basée notamment sur les expériences réalisées précédemment en production, est que les modalités vocales et manuelles participent ensemble aux représentations langagières ; de fait, on devrait pouvoir montrer l'effet de cette co-structuration dans les mécanismes de perception du langage, en particulier via les interactions sensorimotrices mises en jeu au cours de ce processus.

La problématique des interactions sensorimotrices, et de leurs représentations cognitives sous-jacentes, dans les mécanismes de perception de la parole, alimente bon nombre de travaux récents et se trouve au cœur des théories actuelles sur la perception (Glenberg et Gallese [34], Hickok et collab. [49], Schwartz et collab. [93], Skipper et collab. [99]) et la production du langage (Guenther [44], Perrier et collab. [73]). Ces modèles de perception de parole proposent, à divers degrés, que les représentations cognitives de la parole soient co-construites et co-contrôlées au cours de l'acquisition et du développement du langage, par l'interaction entre les systèmes de production et de perception. Notre objectif est de mettre en évidence la participation du geste manuel dans ces boucles de perception/action qui semblent structurer le langage.

Nous présenterons dans ce chapitre certains éléments de la littérature en faveur (1) de l'intervention des représentations sensorimotrices dans certaines tâches de perception des unités sonores de parole, et (2) de l'intervention du système moteur dans le traitement des unités sémantiques du langage (écrit ou oral). Sur la base de ces travaux, nous proposerons l'hypothèse d'une intégration du geste manuel dans les boucles de perception/action mises en jeu dans la perception du langage, selon laquelle les représentations motrices manuelles pourraient intervenir dans les mécanismes de perception du langage, et notamment d'accès lexical.

Afin de tester cette hypothèse, nous avons élaboré un protocole expérimental, nous permettant d'évaluer l'impact des représentations motrices, associées au geste de pointage, sur les mécanismes de compréhension et de perception de la parole. Plus concrètement, l'Expérience 4 est basée sur le paradigme d'amorçage intermodal ; les participants auront pour tâche de pro-

1. Traduction proposée : L'intégration [geste/parole] pourrait également être à l'œuvre dans les mécanismes de compréhension du langage.

duire différents types de gestes (communicatifs : geste déictique, et non communicatifs : geste de saisie), avant de décider le plus rapidement et justement possible si l'item perçu est un mot de la langue française ou non. L'hypothèse est qu'une interaction entre représentations motrices et lexicales du pointage permettrait au geste déictique de faciliter l'accès au terme déictique.

1 Mécanismes de perception du langage

Percevoir et identifier correctement un son de parole nécessite le décodage d'informations phonologiques à partir d'entrées sensorielles particulièrement variables. Cette variabilité est notamment liée aux différences anatomiques entre locuteurs et aux stratégies articulatoires propres à chacun, mais elle est également imputable au phénomène de coarticulation : la production de la parole nécessite « *une combinaison de gestes articulatoires à débit temporel élevé et largement parallélisé* » (Grabski [41], p.4), transformée par la suite, de manière fortement non linéaire, en une séquence de sons ; de fait, la correspondance entre gestes et sons est très peu transparente. Pour ces raisons, la nature des objets de perception de la parole, ou unités de parole, fait l'objet d'un débat animé autour de la nature acoustique ou articulatoire des unités de traitement de la parole. Ce débat, qui opposait initialement les théories acoustico-auditives aux théories motrices, a été relancé par les développements récents de la recherche, notamment en neurophysiologie, qui renforcent l'idée d'un rôle important des interactions sensorimotrices dans le traitement du langage².

1.1 Théorie acoustico-auditive *versus* théorie motrice

Selon l'approche motrice (et notamment la théorie motrice de la perception de la parole, e.g. Liberman [58], Liberman et Mattingly [60]) et la théorie réaliste directe (e.g. Fowler [27]), la perception de la parole ne pourrait être expliquée par les mécanismes généraux de la perception auditive, mais serait une capacité innée, fondamentalement différente de la perception d'autres types de sons, spécifique à l'homme et qui reposerait sur des unités de parole articulatoires. Le lien entre phonèmes et gestes articulatoires étant plus direct que celui entre phonèmes et sons, percevoir la parole reviendrait à percevoir le geste plutôt que le son. Les invariants phonémiques seraient donc récupérés grâce à un appariement entre les gestes articulatoires du locuteur, d'une part, et les connaissances et représentations motrices activées par ce signal chez l'auditeur, d'autre part. Autrement dit, l'auditeur exploite ses connaissances procédurales afin d'extraire, d'un signal acoustique complexe, les unités fonctionnelles directement liées aux commandes motrices (Fowler et Rosenblum [28], Galantucci et collab. [30]). La perception et la production du langage seraient donc basées sur un même répertoire de primitives motrices, qui (1) permettraient au locuteur de générer les gestes oro-faciaux nécessaires à la production de parole et (2) seraient activées chez l'auditeur, par un mécanisme de résonance audio-motrice (i.e. boucle perception/action), afin d'en faciliter la perception. Les auteurs prennent notamment pour arguments le phénomène de perception catégorielle³ et celui de la variabilité des

2. Le lecteur pourra trouver le détail de différentes théories de la perception de la parole dans les thèses de Basirat [4], Fort [26] ou Grabski [41], ainsi que dans les travaux de Schwartz et collab. [93].

3. Le phénomène de perception catégorielle fait référence à une relation non linéaire entre la production et la perception du signal de parole. Selon la définition de Liberman et collab. [59], cités par Grabski [41], la perception catégorielle est la « *capacité à mieux discriminer à écart acoustique constant deux phonèmes situés entre*

1. Mécanismes de perception du langage

unités sonores, liée aux effets de contexte.

L'approche auditive, comme son nom l'indique, considère les unités de parole comme purement auditives. L'auditeur, afin d'interpréter le signal acoustique perçu, récupère les invariants phonémiques directement à partir du signal acoustique. De fait, les processus de décodage acoustico-phonétiques ne font, en théorie, pas appel à des mécanismes moteurs articulatoires, mais reposeraient exclusivement sur des mécanismes auditifs et d'apprentissage perceptifs, liés à des prédispositions et capacités perceptives générales. Les tenants de cette approche montrent par exemple que les phénomènes de perception catégorielle et de coarticulation ne nécessitent pas de traitement spécifique à la parole ou à l'homme (Diehl et collab. [19]).

D'autres données viennent réfuter l'approche motrice. L'existence de différentes stratégies et configurations articulatoires pour la réalisation d'un même phonème (e.g. Savariaux et collab. [90, 91], Mielke et collab. [70], cités par Grabski [41]) remet par exemple en cause l'idée selon laquelle la perception de la parole reposerait uniquement sur des représentations de nature motrice. Les deux approches, auditive et motrice, ont peu à peu été reconsidérées, notamment grâce à l'amélioration des techniques d'investigation, qui a permis d'apporter de nouveaux éléments relatifs à l'implication d'interactions sensorimotrices dans la perception des sons de parole.

1.2 Interactions sensorimotrices en jeu dans la perception de la parole

La production et la perception de la parole apparaissent désormais comme deux mécanismes étroitement liés (voir à ce sujet la revue proposée par Grabski [41] ou Coello et Bartolo [7]), si bien que Schwartz et collab. [94] émettent l'hypothèse d'un « *common language of perception and action* » ([94], p.9), i.e. un langage commun de la perception et de l'action, dans la communication parlée. Le système moteur serait impliqué dans l'observation et la reconnaissance de l'action en général (comme nous l'aborderons dans la sous-section suivante) et de la parole en particulier. Autrement dit, « *the links between articulatory and perceptual mechanisms look like or derive from action-perception links that are observed for a range of non-linguistic actions*⁴ » (Schwartz et collab. [94], p.14) ; phénomène illustré, notamment, dans les travaux de Aboitiz et collab. [1], Arbib [2], Hickok [47], Rizzolatti et Arbib [82], Scott et Johnsrude [95], Wilson et Iacoboni [110] ou Skipper et collab. [97]. Ce lien, entre les mécanismes de production et de perception, se traduit notamment par une activation du système moteur et le recrutement de structures motrices corticales lors de certaines tâches de perception de sons de parole (voir à ce propos la revue de Galantucci et collab. [30] ou de Coello et Bartolo [7]).

D'un point de vue comportemental tout d'abord, Cooper et Lauritsen [11] montrent par exemple qu'une tâche d'adaptation sélective, soit l'écoute prolongée de séquences de parole contenant un phonème particulier, suivie de la présentation de stimuli ambigus (e.g. Crouzet [12], Eimas et Corbit [21]), induit non seulement une modification de la perception des items mais également une modification de leur production ultérieure ; réciproquement l'articulation répétée d'un son

deux catégories phonologiques plutôt qu'à l'intérieur de celles-ci » ([41], p.5). En d'autres termes, si l'exécution articulatoire est progressive et continue, le changement de perception s'effectue quant à lui de manière discrète. Plusieurs configurations motrices vont donc correspondre à un seul et même événement perceptif.

4. Traduction proposée : Les liens qu'entretiennent les mécanismes articulatoires et perceptifs dériveraient de liens action/perception, observés pour un ensemble d'actions non linguistiques.

de parole affecte la perception d'items, présentés ultérieurement (Cooper et collab. [9, 10], Sato et collab. [89]). Dans le même registre, Porter et collab. ([75, 76]) montrent, par le biais d'une tâche de *close shadowing* (soit la répétition rapide d'items), que le temps nécessaire pour reproduire des sons de parole perçus est très court, de l'ordre de 240 ms. Selon les auteurs, ce résultat suggère que la sélection du geste articulatoire, correspondant à l'item perçu, a lieu durant la tâche de perception elle-même ; autrement dit, la perception d'un son de parole induirait la sélection du geste articulatoire nécessaire à la réalisation acoustique du signal (en l'occurrence, sa répétition). Ces données, entre autres, suggèrent que la perception d'un son de parole activerait de façon plus ou moins automatique les représentations motrices qui lui sont associées.

Certaines études en neurophysiologie apportent des éléments supplémentaires, en faveur du lien perception/action et du rôle joué par les représentations motrices dans les mécanismes de perception de la parole.

Le système de neurones miroirs (abordé en Introduction), initialement découvert chez le singe macaque rhésus (e.g. Rizzolatti et collab. [83]), est généralement cité pour rendre compte de ce phénomène de résonance sensorimotrice, puisqu'il constitue la preuve neuroanatomique que certaines zones cérébrales sont recrutées pour la perception d'une action aussi bien que pour la production de celle-ci. Plus récemment, l'équipe de Kohler et collab. [54] a mis en évidence un phénomène de résonance audio-visuo-motrice, illustré par la décharge, au niveau de l'aire F5 du singe (homologue cytoarchitectonique de l'aire de Broca, considérée comme l'un des centres moteurs de la parole), de neurones miroirs lorsque l'individu observe, effectue et entend le son produit par, une action spécifique. L'existence d'un tel système miroir chez l'homme (e.g. Arbib [2], Gentilucci et Corballis [32], Rizzolatti et Arbib [82]) a donné naissance à l'hypothèse d'un fonctionnement similaire pour le traitement de la parole.

A ce sujet, certains travaux ont montré que les aires cérébrales impliquées dans la planification et l'exécution de gestes vocaux, ainsi que le cortex somatosensoriel lié aux mouvements de la bouche, sont activés durant certaines tâches de perception auditive, visuelle et audio-visuelle de la parole (Pulvermüller et collab. [78], Sato et collab. [88], Skipper et collab. [98, 99], Wilson et Iacoboni [110], Wilson et collab. [111]). Sato et collab. [88] montrent par exemple une activation précoce du cortex moteur primaire de la langue lors de la perception audio-visuelle de la parole. Dans le même registre, Fadiga et collab. [23] réalisent une étude en TMS (Stimulation Magnétique Transcrânienne) et observent une augmentation des potentiels moteurs évoqués, enregistrés au niveau des muscles des lèvres et de la langue, durant l'écoute passive de sons de parole (après stimulation des zones correspondantes, dans le cortex moteur primaire gauche) ; par ailleurs, cette activation est d'autant plus forte que le son en question est un mot (relativement aux pseudo-mots). Ces résultats montrent bien que la perception du langage active les aires motrices oro-pharyngales et faciales (voir également les travaux de Roy et collab. [85], Sundara et collab. [103], Watkins et Paus [108] ou Watkins et collab. [109]). De plus, cette activation induite par la seule perception de sons de parole, semble nettement liée aux articulateurs recrutés lors de la production du son perçu (D'Ausilio et collab. [16], Fadiga et collab. [23], Roy et collab. [85]). Autrement dit, la facilitation motrice est d'autant plus importante qu'elle est enregistrée au niveau du muscle correspondant au muscle effectivement utilisé pour produire le son perçu (e.g. la facilitation motrice est plus importante pour les aires associées aux mouvements de la langue lorsque le mot perçu contient des consonnes mobilisant prioritairement la langue, telles qu'un [l]). Cette "spécificité articulatoire" a par la suite été confirmée

1. Mécanismes de perception du langage

par le biais de l'Imagerie par Résonance Magnétique fonctionnelle (e.g. Pulvermüller et collab. [78], Skipper et collab. [99]).

Néanmoins, « *the fact that merely listening to speech sounds activates very early the neural motor system suggests a close connection between speech perception and the motor system, but does not mean per se that the motor activity is necessarily constitutive of speech perception*⁵ » (Coello et Bartolo [7], p.325-326). En effet, plusieurs études montrent qu'une perturbation des mécanismes langagiers (consécutive à des lésions cérébrales au niveau du cortex moteur associé au langage ou à des problèmes congénitaux) n'a pas d'impact sur la perception de la parole (e.g. Damasio [15], Goodglass [40], Lenneberg [57], cités par Coello et Bartolo [7]). Dans le même registre, une anesthésie temporaire des régions motrices de l'hémisphère gauche, suite à l'injection, intra-carotide, d'un barbiturique à courte durée d'action (i.e. test de Wada [106]), n'induit pas de moins bonnes performances dans certaines tâches de perception de parole, telles que la distinction phonémique ou discrimination de parole (Boatman et collab. [5], Hickok et collab. [50]). Par ailleurs, les enfants (e.g. Eimas et collab. [22]) et certains animaux (e.g. Kuhl et Miller [55]), sont capables de discriminer des sons qu'ils ne peuvent pourtant pas (encore, pour les premiers) produire.

L'ensemble de ces données a donc largement contribué à la formulation de nouvelles théories de perception de la parole (e.g. Hickok et collab. [49], Schwartz et collab. [93]), en repensant notamment la nature des représentations cognitives qu'elle implique. S'il semble aujourd'hui assez clair que le système moteur module le processus de perception de la parole, le "degré" avec lequel il intervient reste un sujet largement controversé : l'activation du système moteur est-elle automatique, comme le suggèrent Glenberg et Gallese [34], ou est-elle recrutée uniquement dans des cas particuliers, lorsque l'information est par exemple ambiguë ou incomplète, comme le proposent Hickok et collab. [49] ?

1.3 Vers une approche sensorimotrice des unités sonores de parole

Suite aux observations citées précédemment, un compromis entre une approche radicale purement auditive, et une approche, tout aussi radicale, purement motrice, de la perception de la parole, a été envisagé (Hickok et collab. [49], Remez [80, 81], Schwartz et collab. [93]) : la parole est alors considérée comme un objet multimodal, impliquant la mise en action de représentations sensorielles et motrices. Selon Sato [87] par exemple, « *les représentations de parole seraient intrinsèquement sensorimotrices, ni purs produits sensoriels, ni purs objets moteurs inférés, mais des percepts multimodaux régularisés par l'action* » ([87], Résumé).

C'est dans ce contexte que la Théorie de la Perception pour le Contrôle de l'Action (*Perception-for-Action-Control Theory*, i.e. PACT, Schwartz et collab. [92, 93]) a vu le jour. Elle considère la parole dans une perspective perception/action, puisqu'il semblerait que l'action mette en forme la perception (« *action shapes perception* », Schwartz et collab. [93], p.2) autant que la perception met en forme l'action (« *perception shapes action* », Schwartz et collab. [93], p.4). Dans cette optique, la production du langage oral est influencée par les gabarits auditifs et pho-

5. Traduction proposée : L'activation précoce du système moteur lors de l'écoute de sons de parole suggère une connexion étroite entre la perception de la parole et le système moteur, mais ne signifie pas pour autant que l'activité motrice soit nécessairement constitutive de la perception de la parole.

nologiques d'une langue donnée et structurée par les prédictions et représentations perceptives ; réciproquement, la perception de la parole dépend des contraintes motrices, i.e. articulatoires, et de l'existence de connaissances procédurales motrices, partagées par le locuteur et l'auditeur. Une co-structuration précoce entre les systèmes sensoriel et moteur de la parole (i.e. lors de l'acquisition de la parole, par le biais de l'imitation et de l'apprentissage sensorimoteur) permettrait la mise en place de cartes sensorimotrices, qui, à leur tour, permettraient (1) aux processus de décodage acoustico-phonétiques d'exploiter les capacités du système auditif, et (2) à l'interprétation phonétique finale des entrées sensorielles d'être en partie contrainte par les connaissances motrices procédurales de l'auditeur, par le biais de mécanismes de simulation motrice interne. A l'heure actuelle, le rôle de telles interactions sensorimotrices dans la perception du langage, en situation d'écoute normale, n'est pas clairement établi ; Schwartz et collab. [94] suggèrent néanmoins qu'un tel phénomène de résonance audio-motrice faciliterait l'échange conversationnel, par le biais d'un dispositif perceptuo-moteur commun aux interlocuteurs et favorisant l'émergence de comportements adaptatifs convergents.

1.4 Interactions sensorimotrices dans le traitement des unités sémantiques

D'autres études portent également sur les interactions sensorimotrices, mais à un autre niveau de traitement du langage, celui des unités sémantiques. Elles mettent en évidence l'intervention du système moteur en général (i.e. étendu à tout type de mouvements et au non aux seuls gestes articulatoires) dans le traitement du langage.

Afin d'évaluer le lien entre système moteur et traitement du langage, certaines études analysent l'effet de la production d'une action sur le traitement simultané d'une information linguistique ; en d'autres termes, elles cherchent à évaluer les interférences entre la production d'une action et la perception du langage (e.g. Bartolo et collab. [3], Glover et Dixon [36], Glover et collab. [37]). C'est le cas notamment de Gentilucci et collab. [31], qui analysent les gestes d'atteinte et de saisie vers un objet, sur lequel est inscrit une information linguistique. Leurs résultats montrent que le traitement de l'information linguistique est automatique et irrépressible, mais surtout que ce traitement affecte les performances motrices, en termes de planification de la trajectoire (liée au mouvement d'atteinte) ou de la saisie. Par ailleurs, leurs résultats suggèrent que ces interactions apparaissent uniquement lorsque le contenu linguistique est relatif à des propriétés pertinentes pour l'action (i.e. dimension spatiale, dynamique).

On peut également constater une réciprocité de ces interférences : l'activité motrice réalisée au cours d'une tâche langagière affecte le traitement du langage. Les performances de décision lexicale ou de compréhension de phrases sont ainsi modulées par la réalisation simultanée d'une action manuelle (e.g. Glenberg et Kaschak [35], Witt et collab. [112]). Rueschemeyer et collab. [86] montrent par exemple que des mouvements rotatifs de la main droite vont induire de meilleures performances (en termes de rapidité, de pourcentage d'erreurs) lorsque le mot à identifier est relatif à ce type d'action ; Zwaan et Taylor [113] montrent quant à eux que la lecture et la compréhension d'une phrase donnée sont plus rapides lorsque les participants produisent dans le même temps une action reliée au contenu de cette phrase.

Dans l'ensemble, ces données sont en faveur d'une interaction bidirectionnelle entre langage et action, et suggèrent donc que le geste manuel intervient dans les boucles perception/action en jeu dans le traitement du langage. Elles donnent lieu à l'élaboration de théories et modèles du

1. Mécanismes de perception du langage

traitement du langage, intégrant à des degrés divers le geste manuel : soit celui-ci joue un rôle de médiateur dans la perception du langage (e.g. Glenberg et Gallese [34]), soit sa représentation est simplement liée à celle des unités de parole, par la co-structuration de ces deux modalités (e.g. Mahon et Caramazza [62]).

L'étude menée par Flöel et collab. [25] apporte certains éléments neurophysiologiques à cette question. Les auteurs s'intéressent à l'effet de la perception et de la production de parole sur l'excitabilité corticale de la représentation motrice de la main. Selon les auteurs, l'existence d'un lien phylogénétique entre les systèmes manuel et langagier suggère que « *listening to "gestures" that compose spoken language should activate an extended articulatory/manual action-perception network*⁶ » ([25], p.704). En d'autres termes, la perception de la parole impliquerait de décoder l'ensemble des gestes communicatifs, et nécessiterait donc l'activation de gestes langagiers dans la modalité articuloire, certes, mais également manuelle. Leur étude, basée sur l'utilisation de la TMS, montre que les représentations motrices manuelles sont activées de manière équivalente dans les deux hémisphères lors de tâches linguistiques (production/perception de phrases ou de phonèmes), ce qui n'est pas le cas lors de tâches visuo-spatiales (de type Kimura, [52]) ou auditives, mais non linguistiques (telles que l'écoute de bruit). Selon les auteurs, ces résultats sont en faveur d'une évolution du langage à partir d'un système communicatif gestuel, au sein d'un réseau d'action/perception global et bilatéral. Ce phénomène de pré-excitation du cortex moteur de la main par le biais de la parole, avait déjà été établi en production par Tokimura et collab. [105], mais, à notre connaissance, l'étude de Flöel et collab. [25] est l'une des premières à mettre en évidence un lien direct entre perception du langage et système moteur manuel. Ainsi, la perception de la parole n'active pas seulement les muscles directement impliqués dans la production de la parole (i.e. oro-faciaux, cf. Fadiga et collab. [23]) mais également les muscles de la main.

Dans la lignée des modèles de traitement du geste et de la parole, présentés au Chapitre 1, Kelly et collab. [51] proposent l'*Integrated system hypothesis* (i.e. hypothèse des systèmes intégrés). Cette hypothèse propose que geste et parole interagissent de façon mutuelle et obligatoire (au sens d'irrépressible), non seulement dans les mécanismes de production (comme en témoignent les travaux de Kita et Ozyürek [53] ou de McNeill [67]), mais également dans les mécanismes de compréhension et de perception du langage. L'intérêt de cette étude est que les auteurs s'intéressent aux gestes iconiques, des gestes communicatifs donc, et proposent deux expériences, basées sur le principe de l'amorçage (défini en Section 2.1.4), où les participants regardent tout d'abord une vidéo, dite amorce, d'une action (e.g. action d'émincer des légumes) puis sont confrontés à une vidéo dite cible, contenant un geste iconique ainsi qu'une information verbale.

6. Traduction proposée : Ecouter les "gestes" qui composent les langues orales devrait activer un réseau étendu d'action-perception, au niveau articuloire et manuel.

L’amorce et la cible peuvent entretenir différentes relations :

- En situation contrôle (*baseline*), le geste et la parole présentés dans la cible sont tous deux reliés à l’action présentée dans l’amorce (e.g. action d’émincer des légumes ; ÉMINCER + « émincer »).
- En situation *filler* (non analysée), le geste et la parole présentés dans la cible sont incongruents, relativement à l’amorce (e.g. action d’émincer des légumes ; TORDRE + « tordre »).
- En situation d’incongruence partielle, quatre possibilités sont proposées : le geste cible est relié à l’amorce mais (1) la parole est faiblement incongruente (ÉMINCER + « couper »), ou (2) la parole est fortement incongruente (ÉMINCER + « tordre ») ; la parole est reliée à l’amorce mais (3) le geste est faiblement incongruent (COUPER + « émincer »), ou (4) le geste est fortement incongruent (TORDRE + « émincer »).

Dans la première expérience, la tâche des participants est de décider le plus rapidement et justement possible si la cible contient une information, peu importe la modalité, relative à l’amorce ; dans la seconde expérience, ils doivent décider si l’information verbale, contenue dans la cible, est relative à l’amorce.

Les auteurs, qui souhaitent montrer une influence mutuelle (1) et obligatoire (2) entre voix et geste, font les prédictions suivantes :

1. Lors de l’Expérience 1, les gestes incongruents devraient interagir avec le traitement de la parole congruente, ce qui résulterait en de moins bonnes performances (temps de réaction plus longs pour répondre que oui, une partie de l’information contenue dans la cible est relative à l’amorce) ; réciproquement la parole incongruente devrait interagir avec le traitement du geste congruent.
2. Lors de l’Expérience 2, le geste devrait intervenir, i.e. interférer, dans le traitement de la parole, bien qu’il soit demandé aux participants de se focaliser sur l’information verbale uniquement.

Les résultats de la première expérience montrent des temps de réaction plus rapides et moins d’erreurs en condition contrôle, relativement aux conditions partiellement incongruentes, ainsi qu’une absence d’interaction entre le degré d’incongruence et la modalité utilisée. Les auteurs observent également un résultat plutôt inattendu, puisque de manière générale les temps de réponse sont plus rapides lorsque l’information reliée est l’information verbale. Ils observent également que le taux d’erreurs augmente en situation de forte incongruence, relativement à la situation de faible incongruence. Les résultats de leur seconde expérience montrent que seul le taux d’erreurs augmente avec le degré d’incongruence (i.e. distance sémantique entre information gestuelle et verbale). Dans l’ensemble, ces résultats sont en faveur d’une interaction mutuelle, obligatoire et hiérarchisée entre le geste et la parole. Selon les auteurs, « *this tight relationship may reflect the basic multimodal architecture of the human brain, which may be designed to optimally process and integrate information from across modalities*⁷ » ([51], p.266). Geste et parole sont ainsi combinés afin de ne former qu’un seul signal, hétérogène ; geste et parole « *are simply two sides of the same coin : language*⁸ » (Kelly et collab. [51], p.266).

7. Traduction proposée : Cette relation étroite peut refléter l’architecture multimodale du cerveau humain, conçu pour traiter et intégrer au mieux l’information en provenance de diverses modalités.

8. Traduction proposée : [Geste et parole] ne sont autres que les deux revers d’une même médaille, le langage.

2. Expérience 4

L'ensemble de ces données est de toute première importance puisqu'elles suggèrent (1) que le mécanisme de perception/action s'applique aux gestes en général (dont les gestes langagiers, articulatoires et manuels) et non aux gestes articulatoires *en particulier*, et (2) que geste et parole forment un système intégré dans le processus de perception du langage. On peut alors envisager un lien direct entre représentations motrices et linguistiques : les représentations motrices manuelles interviendraient dans la perception du langage oral et l'accès aux représentations lexicales.

2 Expérience 4

« *Arm movement itself can be used as a linguistic signal*⁹ » Gentilucci et collab. ([33], p.944)

Dans un cadre théorique où la production et la perception du langage impliqueraient des représentations sensorimotrices de nature multimodale, notre hypothèse est que la production d'un geste manuel communicatif pourrait activer de façon automatique les représentations lexicales qui lui sont associées. La tâche proposée aux participants est une tâche de décision lexicale, qui consiste à déterminer le plus rapidement possible et sans erreur si l'item perçu est un mot de la langue française ou non. Cette expérience est basée sur le principe de l'amorçage (voir Section 2.1.4), c'est-à-dire que chaque item sera précédé d'une amorce, pouvant ou non faciliter son traitement. Plus précisément, nous souhaitons montrer que la production d'un geste déictique (relativement à l'absence de geste ou à la production d'un geste non communicatif) faciliterait le traitement de termes déictiques (relativement à celui de noms ou de pseudo-mots) : la représentation motrice du pointage en activerait la représentation lexicale. Parallèlement, nous souhaitons tester différents degrés de congruence, entre l'encodage spatial manuel d'une part, et l'encodage spatial lexical d'autre part. Plus concrètement, le geste de pointage (i.e. encodage spatial manuel) pourra être dirigé vers une cible proche ou plus éloignée, et de la même façon le déictique perçu (i.e. encodage spatial lexical) pourra être un déictique proximal ou distal. L'hypothèse envisagée est que l'effet d'amorçage dépendrait du degré de congruence de l'encodage spatial ; il serait donc plus important en situation congruente (déictique proximal amorcé par un pointage vers la cible proche ; déictique distal amorcé par un pointage vers la cible distante), relativement à la situation incongruente (déictique proximal amorcé par un pointage vers la cible distante et inversement).

2.1 Méthodologie

2.1.1 Participants

Trente-trois participants (dont 10 hommes et 23 femmes), âgés de 18 à 34 ans (Moyenne (M) = 23.9, Ecart-type (SD) = 3.4) ont participé bénévolement à cette étude. Tous étaient de langue maternelle française, sans déficit visuel ou auditif non corrigé, ni trouble moteur particulier, et droitiers. Recrutés majoritairement par diffusion d'e-mails, les participants étaient pour la plupart étudiants à l'Université de Grenoble.

9. Traduction proposée : Le mouvement du bras *per se* peut être interprété comme signal linguistique.

2.1.2 Items expérimentaux

Deux adverbes locatifs du français, *là-bas* (i.e. déictique distal) et *ici* (déictique proximal), ont été choisis comme items "cibles".

Nous avons par la suite utilisé la base de données Lexique 2 (New et collab. [72]), afin de choisir deux items de remplissage appariés aux cibles, tout en tenant compte des critères de sélection suivants, pouvant influencer le processus de reconnaissance de mots (voir Encadré, présenté en fin de sous-partie) : structure syllabique, fréquence lexicale dans le langage oral (base de données Freqfilm2, de New et collab. [72]), point d'unicité et densité du voisinage phonologique (estimée par l'outil générateur de voisins phonologiques de Lexique Toolbox, base de données Lexique 3.71, de New et collab. [71]). Sur cette base, deux noms masculins, du français également, *salut* et *ami*, ont été sélectionnés.

Enfin, quatre pseudo-mots ont été créés, [laga], [ini], [saʁy], [ati], "correspondant" chacun à l'un des items précédents, en termes de nombre de voisins phonologiques et point d'unicité (i.e. *nonword point*¹⁰ pour les pseudo-mots), structure syllabique et premier phonème (seule la consonne médiane diffère, comme par exemple pour la paire [saly] vs [saʁy]). Les items ont été appariés selon le premier phonème, puisque le début d'un mot, relativement à sa fin, semble jouer un rôle important dans le processus de reconnaissance de mots ; l'accès au lexique peut en effet se déclencher dès les 200 premières millisecondes, soit avant la fin, du signal acoustique (e.g. Marslen-Wilson [63], Marslen-Wilson et Zwitserlood [65], Spinelli et collab. [101], Warren et Marslen-Wilson [107]), l'information est ensuite décodée en temps réel et de manière continue (e.g. Dahan et collab. [14]), ce qui nous permet de moduler nos choix lexicaux au cours du déroulement temporel du signal (e.g. Warren et Marslen-Wilson [107]).

Notons que la proportion de cibles, relativement à l'ensemble des items, n'est que de 25%, ceci dans le but de limiter le développement de stratégies de réponse par les participants (Hamburger et Slowiaczek [45], cités par Fort [26]).

Récapitulatif des items expérimentaux Le Tableau 5.1 propose un récapitulatif des différents items et de leurs critères de sélection.

Il s'est avéré difficile de sélectionner des items de remplissage respectant chacun de nos critères. C'est le cas notamment pour la cible *ici*, ayant une fréquence d'occurrence particulièrement élevée (i.e. 2411.21 occurrences par million, ou opm). Toutefois, on peut envisager qu'au-delà d'une certaine fréquence, l'effet de fréquence d'occurrence n'intervienne plus (sachant qu'il n'y a pas de barrière précise entre fréquent et très fréquent, *ici* ne serait pas plus facilement reconnu que *ami* par exemple), ou que, de manière plus générale, l'effet de familiarité "domine". Cependant, les informations de familiarité, relatives à nos items, ne sont, à notre connaissance, pas disponibles (items introuvables dans les bases de données de fréquence subjective de Desrochers et Bergeron [18], Lachaud [56] ou encore de Bonin et collab. [6]). Enfin, on pourrait envisager que l'effet de fréquence soit compensé par le nombre de voisins phonologiques, qui, dans notre corpus, est plus faible pour l'item le plus fréquent. Les analyses statistiques ultérieures permettront en partie d'évaluer l'impact de ces différentes fréquences d'occurrence, notamment par la comparaison de l'effet d'amorçage obtenu pour l'item très fréquent avec celui obtenu pour les items fréquents.

10. Le point d'unicité pour un pseudo-mot désigne « le moment (en partant de son début) à partir duquel l'item est forcément un pseudo-mot » (Fort [26], p.58).

2. Expérience 4

Cibles	Fréquence lexicale (langage oral)	Densité du voisinage phonologique	PU	Pseudo-mots associés
Ici [isi]	2411.21	15	3	[ini] (15, 3)
Là-bas [laba]	263.15	17	4	[laga] (17, 4)

Items de remplissage	Fréquence lexicale (langage oral)	Densité du voisinage phonologique	PU	Pseudo-mots associés
Ami [ami]	360.90	30	3	[ati] (37, 3)
Salut [saly]	277.35	18	4	[savy] (1,4)

Tableau 5.1 – Critères de sélection pour chacun des items expérimentaux. La fréquence lexicale, dans le langage oral, est indiquée en occurrences par million (opm); la densité du voisinage phonologique des pseudo-mots, ainsi que leur *nonword point*, sont indiqués entre parenthèses

Enregistrement des items L'enregistrement des items a été réalisé en chambre sourde, à l'aide d'un microphone, connecté à un enregistreur numérique Marantz PMD 670, et ce dans le but de récupérer fichiers audio sous format .wav. Nous avons demandé à deux locuteurs naïfs (un homme et une femme), de langue maternelle française, de prononcer cinq fois la liste d'items (présentée dans cinq ordres différents), sur un ton neutre et constant. Deux versions de chaque item, pour chacun des locuteurs, ont finalement été retenues. Chacun de ces fichiers a ensuite été "nettoyé" sous Audacity¹¹ : un soin tout particulier a été apporté au repérage du début de l'item, de manière à ce que tout élément perturbateur, pouvant parasiter et/ou indiquer la reconnaissance de l'item (tel que des clics, des aspirations) soit supprimé. L'amplitude des sons a également été normalisée (niveau maximum fixé à -3dB) afin que le niveau sonore soit identique pour chaque item.

De cette façon, nous avons obtenu un corpus de 32 items ($8 \text{ items} \times 2 \text{ versions} \times 2 \text{ locuteurs}$), validé ultérieurement par deux autres individus naïfs, dont la tâche était simplement d'identifier l'item perçu. L'avantage d'un tel corpus est de (1) pouvoir présenter ultérieurement des items variés aux participants et éviter l'apparition de phénomènes d'habituation; (2) vérifier que les résultats obtenus ne sont pas liés aux propriétés idiosyncrasiques du locuteur mais bien à l'identité de l'item perçu.

11. audacity.sourceforge.net

CRITÈRES DE SÉLECTION

- FRÉQUENCE LEXICALE DANS LE LANGAGE ORAL

Les effets de fréquence d'occurrence et de familiarité^a sont deux effets robustes observés dans la reconnaissance des mots écrits ou parlés. Les mots de haute fréquence et/ou familiers, sont reconnus plus rapidement et correctement que les mots de basse fréquence^b, et/ou non familiers, et ce que les mots soient présentés en modalité visuelle (e.g. Connine et collab. [8], Ferrand [24]) ou auditive (e.g. Meunier et collab. [69]). Les quatre mots sélectionnés pour la présente étude sont considérés comme familiers et très fréquents dans le langage oral, puisque leur fréquence lexicale est supérieure à 100 occurrences par million (opm).

- POINT D'UNICITÉ

Le point d'unicité (PU), ou point de reconnaissance, correspond au phonème (en partant de la gauche, soit du début du mot), à partir duquel le mot peut être identifié sans ambiguïté et devient l'unique candidat possible (Fort [26], Goldinger [39], Marslen-Wilson [64]). En d'autres termes, il s'agit du moment de la réalisation acoustique du mot où celui-ci va être correctement identifié^c. Ce facteur peut moduler les effets de fréquence (Goldinger [39], Marslen-Wilson [64]), de supériorité du mot (Frauenfelder et collab. [29]) et est corrélé aux temps de réponse observés dans différentes tâches (Frauenfelder et collab. [29]). Chaque mot sélectionné dans notre étude possède un PU final, situé sur le dernier phonème, comme c'est d'ailleurs généralement le cas pour les mots courts.

- DENSITÉ DU VOISINAGE PHONOLOGIQUE

La densité du voisinage phonologique d'un mot correspond au nombre de ses voisins phonologiques, soit les mots pouvant être dérivés du mot original par substitution, addition ou délétion d'un phonème (Fort [26], Luce et Pisoni [61]). Les voisins phonologiques d'un mot donné n'en diffèrent donc que par un seul phonème^d. Les mots possédant peu de voisins phonologiques et/ou des voisins plus rares sont reconnus plus rapidement que les mots possédant beaucoup de voisins et/ou des voisins plus fréquents (Luce et Pisoni [61]).

^a. La fréquence est la mesure objective (fondée statistiquement) du nombre de fois qu'un individu a rencontré un mot ; la familiarité dépend quant à elle de l'expérience (ou "degré d'exposition") propre à chacun. On parle de fréquences objective et subjective (e.g. Spinelli et Ferrand [100]). Les mots fréquents sont en général familiers, tandis que les mots rares peuvent être familiers pour une catégorie d'individus donnée (e.g. le mot *formant* est rare mais familier pour les phonéticiens).

^b. Haute fréquence > 10 occurrences par million (opm) ; Basse fréquence < 10 opm.

^c. Le point d'unicité du mot *polochon* se réalise à l'entrée dans le système du phonème [ʃ] car à partir de ce point, aucune référence lexicale stockée en mémoire autre que *polochon* ne correspond à cet enchaînement phonologique.

^d. Le mot *salut* compte parmi ses voisins phonologiques les mots *talus* (substitution de phonème), *saluer* (addition de phonème) et *sale* (délétion de phonème).

2. Expérience 4

2.1.3 Matériel

Les participants sont testés individuellement en chambre sourde et sont assis à gauche de l'expérimentateur, face à deux diodes lumineuses situées à 140 et 425 cm. Ces deux diodes sont alignées l'une par rapport à l'autre et par rapport à l'axe œil-index du participant (la configuration est identique à celle de l'Expérience 2, présentée au Chapitre 3, Section 2).

Un clavier d'ordinateur, permettant aux participants de produire leurs réponses (par un appui clavier sur les touches *flèche droite* et *flèche gauche*), ainsi qu'un casque AKG K58, sont à disposition des participants. Ce dernier est utilisé pour la diffusion des items, le volume sonore étant adapté à chaque participant (c'est-à-dire réglé en début de passation) et invariable au cours d'une même session. Par ailleurs, un objet en bois cylindrique, d'une hauteur d'environ 6 cm et d'un diamètre d'environ 3 cm, nécessaire à la réalisation de l'une des tâches, est placé devant le participant (voir Section Procédure 2.1.4). L'expérimentateur dispose quant à lui d'un écran d'ordinateur, non visible par le participant, dont l'interface est contrôlée par le logiciel Presentation¹² (qui permet notamment l'enregistrement d'un fichier .log, soit un historique séquentiel des événements présentés). Il est ainsi en mesure de gérer les sessions d'entraînement, et les pauses, à l'aide d'un pavé numérique. Enfin, l'ensemble de la passation est enregistré via un caméscope numérique, ce qui permet de vérifier éventuellement la validité des tâches effectuées, bien que l'expérimentateur effectue ce contrôle en direct.

2.1.4 Procédure

Les participants doivent décider *le plus rapidement et justement possible* si l'item perçu est un mot de la langue française ou non : il s'agit là de la célèbre tâche de décision lexicale (e.g. Ferrand [24]), avec classification : les participants répondent « oui » - l'item est un mot français - ou « non » (réponse binaire, à l'inverse de la procédure *Go-No go*), par un appui clavier de la main gauche (*flèche droite* = « oui », *flèche gauche* = « non » pour la moitié des participants, et inversement pour l'autre moitié).

Le paradigme utilisé ici s'inspire de la technique de l'amorçage¹³, qui repose sur le principe suivant : le traitement d'un stimulus "amorce"¹⁴ influencerait (en termes de facilitation ou d'inhibition) le traitement d'un stimulus "cible"¹⁵, présenté ultérieurement. La tâche demandée aux participants est le plus souvent effectuée sur le second stimulus (e.g. lors d'une tâche de décision lexicale, le participant doit décider si la **cible** est un mot français ou non).

L'effet d'amorçage est mesuré en comparant les performances obtenues dans une condition dite reliée, où l'amorce et la cible entretiennent une relation de nature sémantique, phonologique ou orthographique¹⁶, à celles obtenues dans une condition dite non reliée, où elles n'entretiennent aucune relation particulière¹⁷. L'hypothèse générale est que la condition reliée devrait induire

12. <http://www.neurobs.com>

13. Grosjean et Frauenfelder [43] proposent une revue des différents paradigmes utilisés pour étudier le processus de reconnaissance de mots.

14. Les amorces seront notées par la suite en lettres minuscules et italiques.

15. Les cibles seront notées par la suite en lettres capitales.

16. Exemple de relation (1) sémantique : *poulet* - COCHON ; (2) phonologique : *rope* - ROBE ; (3) orthographique : *lone* - LONG.

17. Exemple d'items non reliés : *veau* - PULL.

de meilleures performances : percevoir l'amorce permettrait en effet d'activer un certain nombre de représentations lexicales, dont celle de la cible. De fait, lors de la tâche à proprement parler, la cible est activée, au sens de disponible en mémoire, et devrait donc être traitée plus rapidement. À l'inverse, en condition non reliée, aucune activation résiduelle, liée à la présentation de l'amorce, ne peut faciliter la reconnaissance de la cible.

On distingue classiquement le paradigme d'amorçage **intra**modal, qui consiste à présenter l'amorce et la cible dans une seule et même modalité (dans ce cas, l'effet d'amorçage peut être lié à l'intervention de mécanismes spécifiques à la modalité en question, i.e. pré-lexicaux) du paradigme d'amorçage **inter**modal, qui consiste à présenter l'amorce et la cible dans deux modalités sensorielles distinctes (e.g. amorce visuelle et cible auditive ; dans ce cas, l'effet d'amorçage implique un décodage amodal des deux sources d'information, i.e. lexical).

Nous proposons ici un paradigme d'amorçage intermodal "revisité", dans le sens où l'amorce ne sera pas perçue, mais produite, par le participant. Trois conditions d'amorçage sont ainsi testées (voir Figure 5.1 ci-après) :

- Condition **reliée**¹⁸ : **Pointage**

Le participant doit désigner de l'index droit la diode allumée, jusqu'à la diffusion de l'item et effectuer ensuite son jugement de lexicalité de la main gauche.

- Condition **non reliée** : **Saisie**

Le participant doit saisir l'objet cylindrique (décrit précédemment) dans sa main droite, jusqu'à la diffusion de l'item et effectuer ensuite son jugement de lexicalité de la main gauche.

- Condition **neutre** : **Contrôle**

Le participant doit effectuer son jugement de lexicalité de la main gauche (aucune amorce gestuelle).

Le paradigme utilisé est un paradigme en blocs : l'expérience se déroule donc en six parties successives (2×3 conditions), l'ordre des blocs étant aléatoire, i.e. différent selon participant¹⁹). Chaque nouveau bloc est précédé d'une courte session d'entraînement (quatre essais, avec possibilité de recommencer l'entraînement si nécessaire). La durée de l'expérience est d'une vingtaine de minutes (comprenant les consignes ainsi qu'un court debriefing), pendant lesquelles les participants sont confrontés à 288 essais ($3 \text{ Conditions} \times 8 \text{ Items} \times 12 \text{ Itérations}$).

Le plan d'expérience est un plan mixte : le participant est confronté aux différentes modalités des variables Item, Condition et Distance (cette dernière ne concerne que la condition de Pointage) mais à une seule modalité de la variable Locuteur (un participant sur deux perçoit les items prononcés par le locuteur 1, l'autre ceux prononcés par le locuteur 2). Le Tableau 5.2, présenté ci-après, propose un récapitulatif de nos variables indépendantes, intra-sujets et inter-sujets.

18. La nature de la condition (reliée, non reliée, neutre) est fonction des cibles, soit les adverbes déictiques.

19. Une seule contrainte est imposée pour la génération de l'ordre des blocs : deux blocs de Pointage ne peuvent pas se succéder, afin que le geste ne devienne pas "automatique" et que les participants restent dans une optique d'interaction.

2. Expérience 4

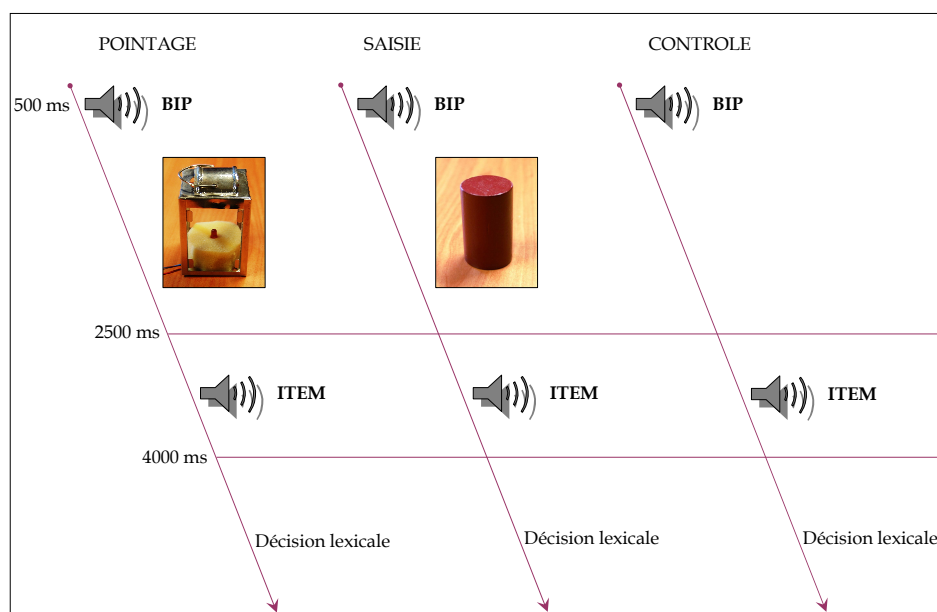


FIGURE 5.1 – Déroulement temporel d'un essai type pour chaque condition. Chaque essai débute par un signal sonore, signal de départ pour le participant

Variables Intra-sujets	Modalités
<i>Item</i>	
<i>I1</i>	Cibles <i>ici, là-bas</i>
<i>I2</i>	Remplissage <i>ami, salut</i>
<i>I3</i>	Pseudo-mots [ini] [laga] [ati] [saby]
<i>Condition</i>	
<i>C1</i>	Pointage
<i>C2</i>	Saisie
<i>C3</i>	Contrôle
<i>Distance</i>	
<i>D1</i>	140 cm
<i>D2</i>	425 cm
Variables Inter-sujets	Modalités
<i>Locuteur</i>	
<i>L1</i>	Locuteur 1 ♂
<i>L2</i>	Locuteur 2 ♀

Tableau 5.2 – Récapitulatif des variables indépendantes

2.2 Mesures

Le pourcentage d'erreurs ainsi que les temps de réponse de choix (*TRC*), mesurés uniquement sur les réponses correctes, ont été calculés pour chacun des participants. Ces mesures ne concernent que 29 des 33 participants ; trois participants ont dû être exclus pour non respect des consignes, notamment des gestes de saisie très rapides voire absents, ainsi qu'un participant,

ayant systématiquement répondu « non » pour la cible *là-bas*, classée ainsi en pseudo-mot.

Ces deux variables dépendantes (% d'erreurs, TRC) sont classiquement mesurées dans les tâches de décision lexicale (e.g. Donders [20], Reinwein [79]). Plus précisément, le TRC , supposé refléter les mécanismes mentaux impliqués dans une tâche donnée (e.g. Donders [20], Posner [77]), est calculé en millisecondes de la façon suivante : $TRC = T_{[APPUI\ CLAVIER]} - T_{[DÉBUT\ ITEM]}$.

Notons qu'à la différence des temps de réponse *Go-No go*, qui nécessitent trois étapes de traitement (i.e. perception, discrimination, production de la réponse), les TRC impliquent une étape supplémentaire, avant l'étape finale, la sélection de la réponse. Les TRC aberrants (i.e. inférieurs à 100 ms = T_{INF} , ou supérieurs à 2000 ms = T_{SUP} , selon Tessier [104]) ont été supprimés : 0.28%²⁰ de l'ensemble des données sont classés en T_{INF} (soit une moyenne, négligeable, de 0.82 T_{INF} par participant) ; 1.2%²¹ des données sont classés en T_{SUP} . Ce pourcentage est plus élevé que le précédent, notamment parce qu'un participant (exclu de l'analyse) totalise à lui seul près de 15% de T_{SUP} .

Par ailleurs, le pourcentage d'erreurs est relativement faible, d'en moyenne 0.9% par participant.

2.3 Analyses statistiques & prédictions

Le lecteur pourra se référer au Chapitre 3, Section 1.4, pour plus de précisions concernant le détail technique et théorique des analyses statistiques.

Les analyses statistiques portent ici sur 28 participants. Seuls les TRC seront analysés ; le pourcentage d'erreurs étant négligeable, aucune analyse statistique n'a été conduite sur cette mesure.

La normalité ainsi que la sphéricité des données ont été testées (respectivement par les tests de Shapiro-Wilk [96] et de Mauchly [66]) et corrigées si nécessaire (respectivement par une transformation par la racine carrée, $Y' = \sqrt{Y}$, la distribution initiale suivant plutôt une loi de Poisson, et le correctif de Greenhouse-Geisser [42]).

2.3.1 Analyse principale : Temps de réponse selon l'Item et de la Condition

Une première série d'analyses consistera à étudier l'effet du Locuteur L , de la Condition C et de l'Item I sur les performances (i.e. TRC).

L'ANOVA à mesures répétées pour plan mixte, $S_{14} < L_2 > \times C_3 \times I_3$, porte sur les médianes. Elle nous permettra de tester trois effets principaux (L ; C ; I) ainsi que les effets d'interaction correspondants ($L * C$; $L * I$; $L * C * I$; $C * I$). Par ailleurs, si l'ANOVA permet de mettre en évidence une différence significative entre les modalités d'un facteur, elle ne permet pas de déterminer précisément quelles modalités diffèrent l'une de l'autre. Les tests de comparaisons *post-hoc* s'avèrent donc nécessaires lorsqu'un effet significatif est observé pour les facteurs à trois modalités ou plus (ici la Condition et l'Item).

20. On observe 0.28% de temps inférieurs à 100 ms, répartis de la façon suivante : 22.2% en condition Saisie, 63% en condition Pointage et 14.8% en condition Contrôle.

21. On observe 1.2% de temps supérieurs à 2000 ms, répartis de la façon suivante : 50% en condition Saisie, 36% en condition Pointage et 19% en condition Contrôle.

2. Expérience 4

Notre hypothèse principale porte sur l'interaction $C * I$: le traitement des cibles (déictiques) devrait être facilité par la production du geste de pointage. Un tel résultat traduirait un effet d'amorçage, soit une facilitation, en condition reliée : *pointage* - DÉICTIQUE, relativement aux conditions non reliées : *pointage* - NOM, *pointage* - PSEUDO-MOT. Par ailleurs, cet effet d'amorçage ne devrait s'observer qu'en condition Pointage et non en l'absence de geste amorce (condition Contrôle) ou lorsque le geste n'est pas communicatif (condition Saisie). En d'autres termes, nous souhaitons mettre en évidence un effet d'interaction entre représentations motrices et représentations lexicales.

Il est également possible que l'on observe un effet principal de l'Item sur le *TRC*, avec de meilleures performances globales pour les mots (indépendamment du type de mots, déictiques ou noms) que pour les pseudo-mots. Un tel résultat serait lié à l'effet de supériorité du mot (Cutler et collab. [13], Frauenfelder et collab. [29]) ou de supériorité de détection des stimulations lexicales (i.e. « *les stimulations avec des caractéristiques lexicales (mots) sont mieux détectées que les stimulations sans (pseudo-mots et sons complexes)* », Perrin [74], p.17).

Néanmoins, selon Rossi [84], « *il semble incontestable que la supériorité du mot sur le non-mot tient plus à sa prononciabilité qu'à sa signification, de sorte qu'avec des non-mots prononçables et des mots de fréquence moyenne on obtient des performances identiques* » ([84], p.492) ; il s'agit là de l'effet de supériorité de détection de la parole (i.e. « *les stimulations avec des caractéristiques phonologiques (mots et pseudo-mots) sont mieux détectées que les stimulations sans (sons complexes)* », Perrin [74], p.17).

Un effet principal de la Condition, avec de meilleures performances globales en condition Contrôle, est également envisageable. Cette tâche devrait être plus facile à réaliser que les tâches impliquant la production d'un geste manuel. Elle ne nécessite à priori « que » les étapes liées au *TRC* (i.e. perception, discrimination, production de la réponse, sélection de la réponse) alors que les deux autres impliquent une tâche supplémentaire, soit une charge cognitive plus élevée (en termes notamment de coordination, de concentration, d'attention auditive et visuelle).

Enfin, nous n'attendons aucun effet impliquant la variable Locuteur (L ; $L * C$; $L * I$; $L * I * C$). Les effets principaux et d'interaction observés ne devraient en aucun cas être liés au locuteur, mais bien à l'identité de l'item perçu.

2.3.2 Analyse complémentaire : Temps de réponse selon l'Item et la Distance

Une seconde série d'analyse portera plus précisément sur les résultats obtenus en condition Pointage, afin d'étudier l'effet de la Distance D et de l'Item I' sur les performances. La variable I , renommée I' , compte à présent non plus trois (déictiques, noms, pseudo-mots) mais quatre modalités, les déictiques étant scindés en deux catégories : proximaux et distaux.

L'ANOVA à mesures répétées pour plan mixte, $S_{14} < L_2 > D_2 \times I'_4$, porte également sur les médianes et nous permettra de tester trois effets principaux (L ; D ; I') ainsi que les effets d'interaction associés ($L * D$; $L * I'$; $L * D * I'$; $D * I'$). Des analyses *post hoc* permettront d'évaluer les effets liés à l'Item de façon plus précise.

Notre hypothèse principale porte ici sur l'interaction $D * I'$: le traitement des déictiques proximaux devrait être facilité par la production d'un geste de pointage vers la diode proche. Réciproquement, le traitement des déictiques distaux devrait être facilité par la production d'un geste de pointage vers la diode éloignée. Un tel résultat traduirait un effet d'amorçage, soit une facilitation, en condition reliée (i.e. congruence entre l'encodage spatial manuel et l'encodage spatial lexical) : *pointage proche* - PROXIMAL (et réciproquement *pointage éloigné* - DISTAL), relativement aux conditions non reliées (i.e. incongruence entre les deux types d'encodage) : *pointage proche* - DISTAL (et réciproquement *pointage éloigné* - PROXIMAL), *pointage* - NOM, *pointage* - PSEUDO-MOT. En d'autres termes, nous souhaitons mettre en évidence un effet d'interaction entre les représentations motrices et les représentations lexicales liées à l'encodage de la distance.

De nouveau, il est possible que l'on observe un effet principal de l'Item (*TRC* plus rapides pour les mots que pour les pseudo-mots), mais nous n'attendons pas d'effet principal de la Distance, ni d'effets, principal et d'interaction, liés à la variable Locuteur.

2.4 Résultats de l'Expérience 4

Cette section présente un descriptif des effets observés pour chaque facteur considéré. Les implications théoriques qui en découlent seront discutées dans la section 2.5.

2.4.1 Analyse principale : Temps de réponse selon l'Item et la Condition

L'effet principal de la Condition n'est pas significatif ($F(2, 52) = 2.37, p = 0.10, NS$), bien que l'analyse au premier ordre, illustrée Figure 5.2²², indique une tendance à répondre plus rapidement en Condition Contrôle.

En revanche, l'effet principal de l'Item, illustré Figure 5.3 ci-après, est significatif ($F(2, 52) = 8.58, p < .01$). Plus précisément, les comparaisons *post hoc* montrent que les mots, de manière générale (i.e. adverbes et noms), sont traités plus rapidement que les pseudo-mots ($p < .05$).

22. La *boîte à moustaches* offre un résumé visuel des données, basé sur le minimum, les trois quartiles Q1, Q2 (i.e. la médiane), Q3, et le maximum. Ici, la hauteur de la boîte représente la distance inter-quartile $Q3 - Q1$, et les moustaches sont de longueur $Q3 - Q1 \times 1.5$. Une donnée est considérée atypique lorsqu'elle dépasse cette valeur, au-dessous du 1^{er} quartile ou au-dessus du 3^{ème} quartile. Ces valeurs sont indiquées par le symbole $\circ$.

2. Expérience 4

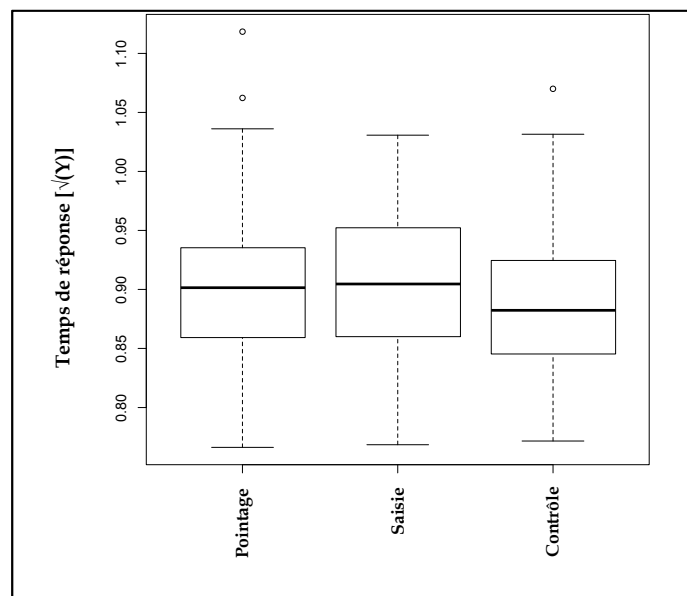


FIGURE 5.2 – Temps de réponse, après transformation par la racine carrée, en fonction de la Condition

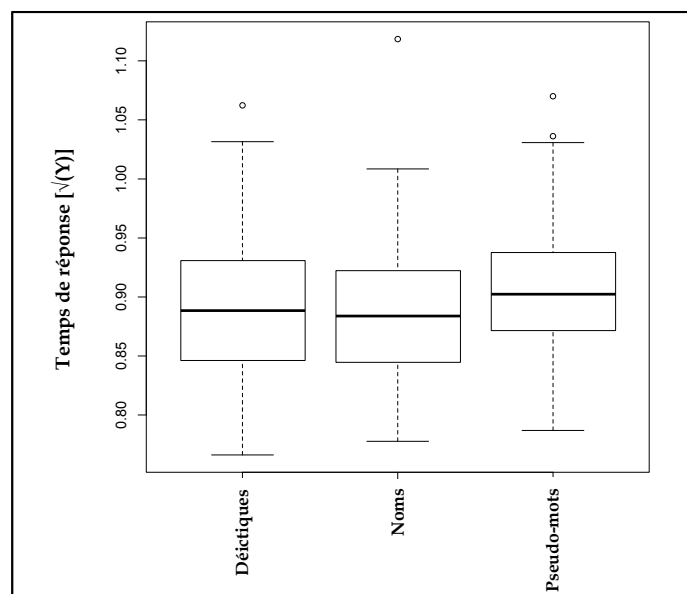


FIGURE 5.3 – Temps de réponse, après transformation par la racine carrée, en fonction de l'Item

D'autre part, l'interaction $C * I$ est notable (voir Figure 5.4); elle n'est pas robuste, statistiquement parlant, mais p atteint la valeur critique de .05 ($F(4, 104) = 2.42$, $p = 0.05$). Les analyses *post hoc* n'indiquent aucune différence inter-items en condition Contrôle (soit la condition neutre, de référence); en revanche, on observe (1) de meilleures performances (i.e. TRC plus rapides) pour les cibles déictiques en condition Pointage ($p < .05$) et (2) de meilleures

performances pour les items de remplissage en condition Saisie ($p < .05$).

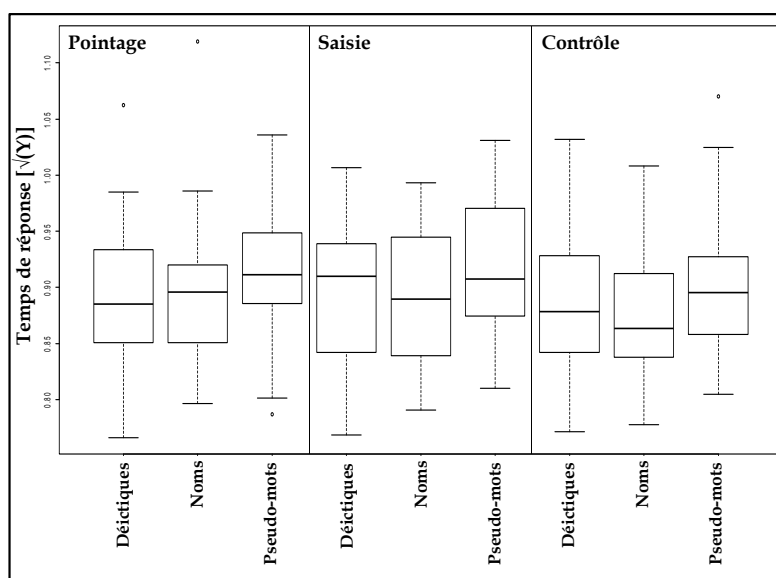


FIGURE 5.4 – Temps de réponse, après transformation par la racine carrée, en fonction de la Condition et de l'Item

Enfin, comme attendu, la variable Locuteur n'induit aucun effet significatif (effet principal : $F(1, 26) = 0.47$, $p = 0.62$, NS ; effets d'interaction : $F_L * I(2, 52) = 0.83$, $p = .44$, NS ; $F_L * C(2, 52) = 0.47$, $p = .62$, NS ; $F_L * C * I(4, 104) = 1.67$, $p = .16$, NS), ce qui suggère que les effets sont imputables à l'identité de l'item perçu plutôt qu'aux caractéristiques propres des locuteurs.

2.4.2 Analyse complémentaire : Temps de réponse selon l'Item et la Distance

L'effet principal de la Distance, illustré Figure 5.5, est significatif ($F(1, 26) = 15.42$, $p < .01$) : de manière générale, les items sont reconnus plus rapidement lorsque les participants désignent la diode la plus proche.

Les résultats montrent également un effet significatif de l'Item ($F(3, 78) = 6.25$, $p < .01$), illustré Figure 5.6 ci-après. Les analyses *post hoc* indiquent que seuls les déictiques proximaux et les noms sont reconnus plus rapidement que les pseudo-mots ($p < .05$) ; autrement dit, les déictiques distaux et les pseudo-mots sont traités de manière équivalente. Les déictiques proximaux, à la différence des noms, sont également traités plus rapidement que les déictiques distaux ($p < .05$).

L'effet d'interaction $D * I'$ n'est pas significatif ($F(3, 78) = 1.04$, $p = .38$, NS). L'analyse au premier ordre, illustrée Figure 5.7 ci-après, semble pourtant montrer que les déictiques proximaux, relativement à l'ensemble des autres items, tendent à être reconnus plus rapidement en situation de pointage proche. L'absence d'interaction pourrait être liée à une dispersion trop importante des données mais également au nombre limité d'items proximaux et distaux, relativement aux noms et pseudo-mots.

2. Expérience 4

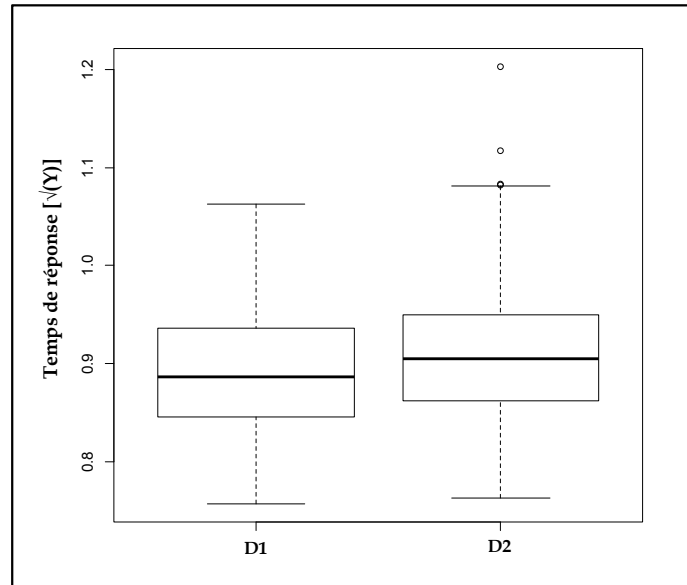


FIGURE 5.5 – Temps de réponse, après normalisation par la racine carrée, en fonction de la Distance

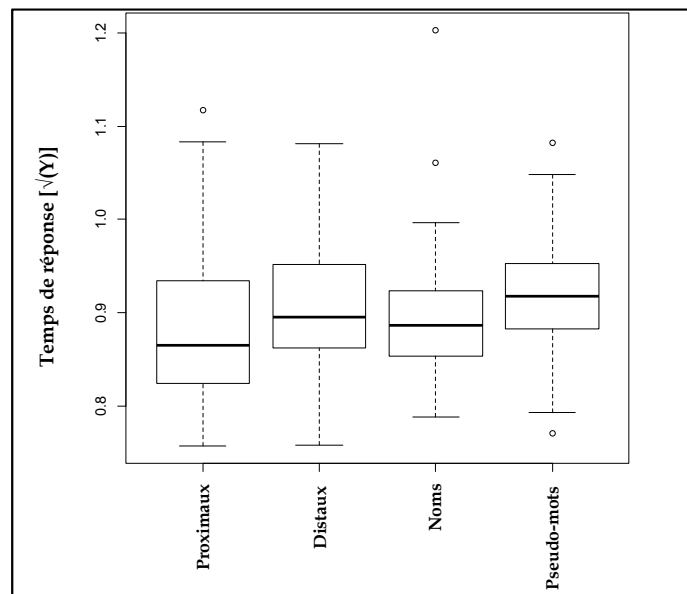


FIGURE 5.6 – Temps de réponse (normalisés) en fonction de l'Item

Enfin, comme attendu, nous n'observons pas d'effet significatif lié à la variable Locuteur (effet principal : $F(1, 26) = 1.32, p = .26, NS$; effets d'interaction : $F_{L*I'}(3, 78) = 1.22, p = .31, NS$; $F_{L*D}(3, 78) = 0.28, p = .60, NS$; $F_{L*I'*D}(3, 78) = 0.48, p = .70, NS$), les résultats obtenus ne sont donc pas liés aux caractéristiques propres du locuteur mais bien à l'identité de l'item perçu (i.e. déictiques proximaux et distaux, noms, pseudo-mots).

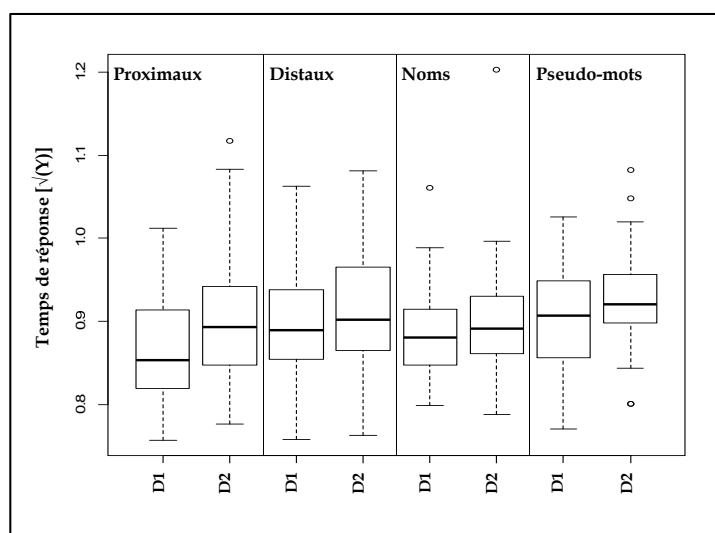


FIGURE 5.7 – Temps de réponse, après transformation par la racine carrée, en fonction de la Distance et de l’Item

2.5 Implications théoriques

2.5.1 Représentations motrices et lexicales en jeu dans la perception de la deixis

L’objectif de notre analyse principale était d’évaluer l’intégration geste/parole dans les mécanismes de perception du langage. Notre hypothèse était que le processus de perception de la parole reposait sur la mise en action d’unités sensorimotrices multimodales, le système moteur en général (et manuel en particulier) étant recruté dans le traitement du langage. Nous avons souhaité montrer que les gestes manuels communicatifs et la parole participaient ensemble aux représentations langagières, par nature multimodale. Plus concrètement, la production d’un geste de pointage pourrait activer de manière automatique les représentations lexicales qui lui sont associées ; en d’autres termes, un accès direct, à partir des représentations motrices, au lexique mental. Nous attendions de fait une interaction entre représentations déictiques motrices et représentations déictiques linguistiques.

De manière générale, les résultats discutés ci-après sont liés à la nature de l’item perçu plutôt qu’aux propriétés idiosyncrasiques du locuteur, comme l’indique l’absence d’effet impliquant cette variable.

Par ailleurs, nos résultats ne montrent pas d’effet principal de la Condition. L’analyse au premier ordre suggère un léger avantage de la condition Contrôle, relativement aux conditions de Pointage et de Saisie, mais néanmoins les temps de réponse ne sont pas retardés dans ces deux conditions, qui nécessitent pourtant la production d’un geste manuel, et de fait, une charge cognitive plus élevée.

D’autre part, les mots, indépendamment de leur nature (adverbe déictique ou nom), sont traités, de manière générale, plus rapidement que les pseudo-mots. Ce résultat est à priori en faveur d’un effet de supériorité de détection des stimulations lexicales, mais l’effet d’interaction, entre l’Item et la Condition, indique une facilitation du traitement lexical uniquement en condition

2. Expérience 4

de Pointage et de Saisie. Autrement dit, en condition Contrôle (soit la condition de référence), les mots ne sont pas traités plus rapidement que les pseudo-mots. Cet effet d'interaction suggère que l'effet de supériorité du mot est lié à la prononciabilité, plutôt qu'au caractère lexical, de l'item perçu (cf. Rossi [84]). De fait, nos résultats sont en faveur d'un effet de supériorité de détection de la parole plus que de stimulations lexicales (en situation Contrôle). Cet effet de supériorité de détection des stimulations lexicales interviendrait lorsque la tâche demandée aux participants mobilise une plus grande partie des ressources cognitives (i.e. en situation de Pointage et de Saisie).

L'effet d'interaction entre l'Item et la Condition suggère également un meilleur traitement des déictiques en situation de Pointage. Lorsque le participant effectue un geste de pointage, il reconnaît plus rapidement les déictiques, relativement aux pseudo-mots, ce qui n'est pas le cas des noms (les noms et les pseudo-mots sont reconnus de façon semblable, tout comme en condition de référence). Ces résultats suggèrent que la production du geste de pointage, soit l'activation de représentations déictiques motrices, active de façon automatique les représentations lexicales qui lui sont associées, et ce par un mécanisme d'activation résiduelle, qui permet aux participants d'identifier plus rapidement les adverbes déictiques. En condition de Saisie en revanche, c'est le traitement des noms qui est avantagé. Lorsque le participant effectue un geste de saisie, il reconnaît plus rapidement les noms, relativement aux pseudo-mots, ce qui n'est pas le cas des déictiques. Cet effet peut potentiellement être lié à un mécanisme d'inhibition : le geste non relié (i.e. de saisie), inhiberait la reconnaissance des déictiques, de la même façon qu'un geste de pointage inhiberait la reconnaissance des noms. Il s'agit là d'un effet de type Stroop [102], soit l'interférence produite par une information non pertinente (en l'occurrence le geste de saisie) lors de l'exécution d'une tâche cognitive (ici la reconnaissance de déictiques) ; la difficulté ainsi induite se traduit alors par un ralentissement des temps de réponse.

Les données de l'analyse principale suggèrent donc une activation automatique des représentations déictiques linguistiques, à partir de l'activation des représentations motrices du pointage. Elles sont ainsi en faveur d'une intégration geste/parole dans le processus de perception du langage, une interaction entre le système manuel et le système vocal. La perception du langage, contrainte, au niveau phonologique, par la production des gestes articulatoires, serait également contrainte, au niveau lexical, par la production des gestes manuels. La perception du langage serait donc basée sur la mise en action d'unités perceptivo-motrices multimodales, au niveau phonologique et lexical.

2.5.2 Représentations motrices et lexicales liées à l'encodage de la distance

L'objectif de l'analyse complémentaire était d'étudier l'interaction entre représentations motrices et lexicales, liées à l'encodage de la distance, qui est un processus déictique lexico-moteur. Notre hypothèse était que le geste manuel activerait les représentations lexicales qui lui sont associées de manière **spécifique** ; en d'autres termes, le geste de pointage activerait les représentations lexicales associées au pointage, mais un pointage proche activerait préférentiellement les termes proximaux, tandis qu'un pointage éloigné activerait davantage les termes distaux.

Notons tout d'abord que l'ensemble des résultats de cette seconde analyse sont liés à la nature de l'item perçu plutôt qu'aux caractéristiques propres aux productions des locuteurs (i.e.

absence d'effet significatif impliquant cette variable).

D'autre part, les résultats indiquent qu'indépendamment de l'item perçu, les temps de réponse sont plus rapides lorsque le participant désigne l'objet proche, relativement à l'objet éloigné. Cet effet, inattendu, pourrait être expliqué par la disposition des cibles et notamment leur ambiguïté spatiale. Il serait potentiellement plus difficile pour le participant de désigner la diode éloignée et ambiguë, puisque son objectif (indiquer un objet particulier à quelqu'un) est alors imprécis, au sens de moins concret, plus difficilement interprétable par l'interlocuteur. La réalisation d'un geste spécifique pour désigner un objet éloigné et ambigu pourrait donc induire une diminution des ressources cognitives allouées au jugement de lexicalité. Dans une certaine mesure, ces résultats suggèrent que les participants sont bien en situation d'interaction communicative : ils considèrent le point de vue de leur interlocuteur et s'efforcent de lui communiquer l'information de la manière la plus adéquate possible (le pointage n'est donc pas "automatique" mais reflète bien un acte communicatif, désigner un objet *pour* quelqu'un).

On observe par ailleurs, en situation de pointage donc, un effet principal de l'Item : les déictiques proximaux et les noms sont reconnus plus rapidement que les pseudo-mots ; les déictiques proximaux sont également reconnus plus rapidement que les déictiques distaux. Le déictique proximal *ici*, malgré une fréquence d'occurrence plus élevée que celle des mots (respectivement 2411.21 opm contre, en moyenne, 319.12 opm), n'est pas traité plus rapidement que ces derniers. Plusieurs interprétations sont possibles : (1) au-delà d'une certaine fréquence, l'effet de fréquence d'occurrence n'intervient plus, de fait il n'existe pas de bénéfice pour les items très fréquents relativement aux items fréquents ; (2) l'effet de familiarité est potentiellement dominant et l'on peut penser qu'*ici* est sensiblement identique, en termes de fréquence subjective, aux items *ami* et *salut* ; (3) le nombre de voisins phonologiques, plus faible pour *ici* relativement à *ami* et *salut* (respectivement, 15 contre 30 et 18), module l'effet de fréquence. Ce résultat suggère donc un avantage du déictique proximal, relativement aux pseudo-mots, en situation de pointage. La production d'un geste de pointage active préférentiellement les représentations lexicales proximales et inhibe le traitement des représentations lexicales distales. Rappelons-nous à ce propos qu'un participant a dû être exclu de l'analyse statistique finale, parce qu'il avait systématiquement classé l'item *là-bas* en tant que pseudo-mot. Alors si la majorité des participants a hésité avant d'accepter cet item comme mot de la langue française, possiblement parce que la consigne n'indiquait pas clairement qu'un item composé pouvait être considéré comme une seule unité lexicale, les temps de réactions ont nécessairement été retardés. Néanmoins, si le traitement des déictiques distaux est plus lent que celui des déictiques proximaux, il est en revanche équivalent à celui des noms.

D'autre part, l'interaction entre la Distance et l'Item n'est pas significative. L'analyse au premier ordre semble pourtant montrer que les déictiques proximaux, relativement à l'ensemble des autres items, tendent à être reconnus plus rapidement en situation de pointage proche, relativement à la situation de pointage éloigné. L'absence de significativité peut-être liée à la dispersion de nos données mais également au faible pourcentage d'adverbes déictiques : l'utilisation de la variable I' (c'est-à-dire la variable Item, à quatre modalités) induit une proportion plus faible d'items reliés que la variable I (i.e. la variable Item, à trois modalités) : le pourcentage de cible, initialement de 25%, est finalement de 12.5%.

De manière générale, les résultats de l'analyse complémentaire suggèrent que le geste manuel active les représentations lexicales qui lui sont associées de manière générique plutôt que spéci-

3. Conclusion

fique ; en d'autres termes, les représentations motrices associées à un espace perceptif plus ou moins lointain ne semblent pas strictement associées aux représentations lexicales proximales et distales. Néanmoins, la dispersion des données, la faible proportion d'items cibles, mais également une difficulté potentiellement accrue en situation de pointage éloigné, peuvent expliquer certains de nos résultats. Il serait donc pertinent d'augmenter le nombre de participants et le pourcentage d'items reliés afin de diminuer la variance et de favoriser l'émergence d'effets plus robustes. Il serait également intéressant de conduire une expérience similaire, mais tenant compte de la difficulté induite par un positionnement ambigu des objets : s'affranchir de cette difficulté, notamment en testant deux objets éloignés l'un de l'autre mais non alignés, pourrait favoriser l'apparition d'un effet d'amorçage en situation de pointage éloigné.

3 Conclusion

L'objectif de ce dernier chapitre était de tester l'intégration des représentations vocales et manuelles dans les mécanismes de perception du langage, par le biais d'une procédure d'amorçage intermodal. Nos résultats sont en faveur d'une facilitation du traitement de la parole en général, et des termes déictiques en particulier, par l'activation des représentations motrices du geste de pointage manuel. Nous avons ainsi pu démontrer que la production d'un geste communicatif active plus particulièrement les représentations lexicales directement associées à ce geste. En d'autres termes, la perception de la parole pourrait être basée sur l'activation de représentations sensorimotrices multimodales, i.e. dans les modalités vocales et manuelles. Ces données suggèrent que geste et parole ne forment qu'un seul et même système communicatif, multimodal, les représentations déictiques motrices et lexicales étant étroitement liées.

Néanmoins, nos résultats pourraient être améliorés ; dans un premier temps, par le recrutement d'un nombre plus important de participants, et dans un second temps, par une amélioration du protocole expérimental. Le débat quant au niveau de traitement du langage auquel les représentations motrices interviennent reste en effet ouvert. Si nos données suggèrent que les représentations motrices du pointage interviennent au niveau lexical, l'élaboration d'un nouveau protocole permettrait de confirmer ce phénomène. On pourrait envisager de comparer un geste de pointage vers un objet situé dans l'espace péripersonnel, c'est-à-dire à portée de main, à un geste d'atteinte et de saisie vers ce même objet. De cette façon, la comparaison serait effectuée entre deux gestes similaires en termes de configuration motrice (e.g. direction, extension) mais porteurs de fonctions distinctes (montrer/saisir) et de nature différente (communicative/non communicative). Un effet d'amorçage équivalent pour ces deux types de geste serait en faveur d'une intervention des représentations motrices de bas niveau, tandis qu'un effet d'amorçage supérieur pour le geste de pointage traduirait une intervention au niveau lexical.

Par ailleurs, la question du rôle exact joué par de telles représentations motrices est à l'heure actuelle vivement controversée (voir à ce sujet la revue proposée par Coello et Bartolo [7]). Glenberg et Gallese [34] proposent par exemple le modèle ABL (i.e. *Action-Based Language*, modèle du Langage Basé sur l'Action) qui postule un rôle fonctionnel et automatique du système moteur dans l'acquisition, la compréhension et la production du langage. Nos connaissances motrices nous permettraient d'apprendre puis de comprendre le langage (la maîtrise de certains compor-

tements serait par exemple directement liée à l'apprentissage de nouveaux mots chez l'enfant et à une compréhension optimale du langage chez l'adulte). « *We tend to do what we say, and we tend to say [...] what we do*²³ » ([34], p.911) ; en d'autres termes l'action et le langage seraient intrinsèquement et fonctionnellement liés.

Schwartz et collab. [93] suggèrent également, pour le traitement des unités sonores, une intervention particulière des représentations motrices au cours du développement du langage : si le rôle des interactions sensorimotrices n'est pas clairement établi en situation d'écoute "normale", on peut en revanche envisager que celles-ci structurent nos apprentissages.

Néanmoins, certains auteurs, tels que Hickok [48], Mahon et Caramazza [62] ou Meteyard et collab. [68] considèrent que le système moteur ne joue pas de rôle direct au niveau du traitement du langage mais que tous deux seraient structurellement liés. Une hypothèse possible est que ce lien entre les systèmes moteur et langagier serait naturellement établi au niveau cérébral par un mécanisme d'apprentissage hebbien (Hebb [46], cité par Coello et Bartolo [7]), sans aucune contribution des aires perceptives et motrices au niveau des processus sémantiques d'accès lexical (Del Giudice et collab. [17], cités par Coello et Bartolo [7]). Cette hypothèse se base sur le fait que les mots d'action (généralement des verbes) sont initialement appris puis utilisés dans des contextes d'exécution et/ou d'observation de l'action qu'ils décrivent (Goldfield [38], cité par Coello et Bartolo [7]). En d'autres termes, l'apparition d'un mot et celle de son référent moteur sont bien souvent temporellement liées. Sachant que le mécanisme d'apprentissage associatif hebbien postule qu'une activité synchrone de neurones permette d'établir des regroupements entre ces neurones, les aires du langage et les représentations sensorimotrices seraient liées du fait de la proximité temporelle entre ces deux types d'événements (voir également les travaux de Mahon et Caramazza [62], qui décrivent ce type d'activation en cascade de deux niveaux de traitement qualitativement différents mais reliés au niveau cérébral par un circuit neuronal commun).

De nombreuses études peuvent encore être envisagées afin d'éclaircir ce point, et notamment celle du décours temporel des interférences entre l'action et le langage, de l'influence du type de tâches (e.g. identifier, classer, ou interpréter des unités de parole pourrait moduler de diverses façons ces interférences), mais encore celle des réseaux de neurones impliqués dans les représentations sensori-motrices et langagières. « *All these issues represent original and fascinating challenges in the field of cognitive neurosciences and would obviously viewed as important avenues for the research on action and language in the future.*²⁴ » (Coello et Bartolo [7], p.338).

23. Traduction proposée : On tend à faire ce que l'on dit, et à dire ce que l'on fait.

24. Traduction proposée : De telles perspectives sont autant de défis originaux et fascinants dans le champ des neurosciences cognitives et représentent à coup sûr des pistes de recherche essentielles pour les recherches futures sur l'action et le langage.

Bibliographie

- [1] Aboitiz, F., V. Garcia, C. Bosman et E. Brunetti. 2006, «Cortical memory mechanisms and language origins», *Brain and Language*, vol. 98(1), p. 40–56.
- [2] Arbib, M. A. 2005, «From monkey-like action recognition to human language : An evolutionary framework for neurolinguistics», *Behavioral and Brain Sciences*, vol. 28, p. 105–167.
- [3] Bartolo, A., A. Weisbecker et Y. Coello. 2007, «Linguistic and spatial information for action», *Behavioural Brain Research*, vol. 184, p. 19–30.
- [4] Basirat, A. 2010, *Émergence des représentations perceptives de la parole : Des transformations verbales sensorielles à des éléments de modélisation computationnelle*, thèse de doctorat, Institut National Polytechnique de Grenoble.
- [5] Boatman, D., J. Hart, R. Lesser, N. Honeycutt, N. Anderson, D. Miglioretti et B. Gordon. 1998, «Right hemisphere speech perception revealed by amobarbital injection and electrical interference», *Neurology*, vol. 51, p. 458–464.
- [6] Bonin, P., A. Méot, L. Aubert, N. Malardier, P. Niedenthal et M. C. Capelle-Toczek. 2003, «Normes de concrétude de valeur d'imagerie, de fréquence subjective et de valence émotionnelle pour 866 mots», *L'Année Psychologique*, vol. 103(4), p. 655–694.
- [7] Coello, Y. et A. Bartolo. 2012, «Contribution of the action system to language perception and comprehension : Evidence and controversies», dans *Language and Action in Cognitive Neuroscience*, édité par Y. Coello et A. Bartolo, London : Psychology Press, p. 321–342.
- [8] Connine, C., J. Mullennix, E. Shernoff et J. Yelen. 1990, «Word familiarity and frequency in auditory and visual word recognition», *Journal of Experimental Psychology : Learning, Memory, & Cognition*, vol. 16, p. 1084–1096.
- [9] Cooper, W., D. Billings et R. Cole. 1976, «Articulatory effects on speech perception : A second report», *Journal of Phonetics*, vol. 4, p. 219–232.
- [10] Cooper, W., S. Blumstein et G. Nigro. 1975, «Articulatory effects on speech perception : A preliminary report», *Journal of Phonetics*, vol. 3, p. 87–98.
- [11] Cooper, W. et M. Lauritsen. 1974, «Feature processing in the perception and production of speech», *Nature*, vol. 252, p. 121–123.
- [12] Crouzet, O. 2000, *Segmentation de la parole en mots et régularités phonotactiques : Effets phonologiques, probabilistes ou lexicaux ?*, thèse de doctorat, Université Paris 5 - René Descartes.
- [13] Cutler, A., J. Mehler, D. Norris et J. Segui. 1987, «Phoneme identification and the lexicon», *Cognitive Psychology*, vol. 19, p. 141–177.
- [14] Dahan, D., J. Magnuson et M. Tanenhaus. 2001, «Time course of frequency effects in spoken-word recognition : Evidence from eye movements», *Cognitive Psychology*, vol. 42, p. 317–367.

- [15] Damasio, A. 1992, «Aphasia», *New England Journal of Medicine*, vol. 326, p. 531–539.
- [16] D’Ausilio, A., F. Pulvermüller, P. Salmas, I. Bufalari, C. Begliomini et L. Fadiga. 2009, «The motor somatotopy of speech perception», *Current Biology*, vol. 19, p. 381–385.
- [17] Del Giudice, M., V. Manera et C. Keysers. 2009, «Programmed to learn? The ontogeny of mirror neurons», *Developmental Science*, vol. 12, p. 350–363.
- [18] Desrochers, A. et M. Bergeron. 2000, «Valeurs de fréquence subjective et d’imagerie pour un échantillon de 1,916 substantifs de la langue française», *Canadian Journal of Experimental Psychology*, vol. 54(4), p. 274–325.
- [19] Diehl, R., A. Lotto et L. Holt. 2004, «Speech perception», *Annual Review of Psychology*, vol. 74, p. 431–461.
- [20] Donders, F. 1868, «La vitesse des actes psychiques», *Archives Néerlandaises des Sciences Exactes et Naturelles*, vol. 3, p. 296–317.
- [21] Eimas, P. et J. Corbit. 1973, «Selective adaptation of linguistic feature detectors», *Cognitive Psychology*, vol. 4, p. 99–109.
- [22] Eimas, P., E. Siqueland, P. Jusczyk et J. Vigorito. 1971, «Speech perception in infants», *Science*, vol. 171(968), p. 303–306.
- [23] Fadiga, L., L. Craighero, G. Buccino et G. Rizzolatti. 2002, «Speech listening specifically modulates the excitability of tongue muscles : A TMS study», *European Journal of Neuroscience*, vol. 15, p. 399–402.
- [24] Ferrand, L. 2001, *Cognition et lecture : Processus de base de la reconnaissance de mots écrits chez l’adulte*, Bruxelles : DeBoeck Université.
- [25] Flöel, A., T. Ellger, C. Breitenstein et S. Knecht. 2003, «Language perception activates the hand motor cortex : Implications for motor theories of speech perception», *European Journal of Neuroscience*, vol. 18(3), p. 704–8.
- [26] Fort, M. 2011, *L’accès au lexique dans la perception audiovisuelle et visuelle de la parole*, thèse de doctorat, Université de Grenoble.
- [27] Fowler, C. 1986, «An event approach to the study of speech perception from a direct-realist perspective», *Journal of Phonetics*, vol. 14, p. 3–28.
- [28] Fowler, C. et L. Rosenblum. 1991, «The perception of phonetic gestures», dans *Modularity and the Motor Theory of Speech Perception*, édité par I. Mattingly et K. Studdert-Kennedy, Lawrence, NJ : Erlbaum, p. 33–60.
- [29] Frauenfelder, U., J. Segui et T. Dijkstra. 1990, «Lexical effects in phonemic processing : Facilitatory or inhibitory», *Journal of Experimental Psychology : Human Perception and Performance*, vol. 16, p. 77–91.
- [30] Galantucci, B., C. Fowler et M. Turvey. 2006, «The motor theory of speech perception reviewed», *Psychonomic Bulletin & Review*, vol. 13, p. 361–377.

Bibliographie

- [31] Gentilucci, M., F. Benuzzi, L. Bertolani, E. Daprati et M. Gangitano. 2000, «Language and motor control», *Experimental Brain Research*, vol. 133, p. 468–490.
- [32] Gentilucci, M. et M. Corballis. 2006, «From manual gesture to speech : A gradual transition», *Neurosciences and Behavioral Reviews*, vol. 30, p. 949–960.
- [33] Gentilucci, M., R. Dalla Volta et C. Gianelli. 2008, «When the hands speak.», *Journal of Physiology*, vol. 102, n° 1-3, p. 21–30, ISSN 0928-4257.
- [34] Glenberg, A. et V. Gallese. 2012, «Action-based language : A theory of language acquisition, comprehension, and production», *Cortex*, vol. 48(7), p. 905–922.
- [35] Glenberg, A. et M. Kaschak. 2002, «Grounding language in action», *Psychonomic Bulletin & Review*, vol. 9, p. 558–565.
- [36] Glover, S. et P. Dixon. 2002, «Semantics affect the planning but not control of grasping», *Experimental Brain Research*, vol. 146, p. 383–387.
- [37] Glover, S., D. Rosenbaum, J. Graham et P. Dixon. 2004, «Grasping the meaning of words», *Experimental Brain Research*, vol. 154, p. 103–108.
- [38] Goldfield, B. 2000, «Nouns before verbs in comprehension vs. production : The view from pragmatics», *Journal of Child Language*, vol. 27, p. 501–520.
- [39] Goldinger, S. 1996, «Auditory lexical decision», *Language & Cognitive Processes*, vol. 11, p. 559–568.
- [40] Goodglass, H. 1993, *Understanding Aphasia*, San Diego, CA : Academic Press.
- [41] Grabski, K. 2012, *Les cartes sensorimotrices de la parole : Corrélats neurocognitifs et couplage fonctionnel des systèmes de perception et de production des voyelles du Français*, thèse de doctorat, Université de Grenoble.
- [42] Greenhouse, S. et S. Geisser. 1959, «On methods in the analysis of profile data», *Psychometrika*, vol. 24, p. 95–112.
- [43] Grosjean, F. et U. Frauenfelder. 1996, «A guide to spoken word recognition paradigms : Introduction», *Language and Cognitive Processes*, vol. 11, p. 553–558.
- [44] Guenther, F. 2006, «Cortical interactions underlying the production of speech sounds», *Journal of Communication Disorders*, vol. 39, p. 350–365.
- [45] Hamburger, M. et L. Slowiczek. 1996, «Phonological priming reflects lexical competition», *Psychonomic Bulletin & Review*, vol. 3, p. 520–525.
- [46] Hebb, D. 1949, *The Organization of Behavior : A Neuropsychological Theory*, New York : Wiley.
- [47] Hickok, D., G. & Poeppel. 2007, «The cortical organization of speech processing», *Nature Reviews Neuroscience*, vol. 8, p. 393–402.

- [48] Hickok, G. 2010, «The role of mirror neurons in speech perception and action word semantics», *Language and Cognitive Processes*, vol. 25, p. 749–776.
- [49] Hickok, G., J. Houde et F. Rong. 2011, «Sensorimotor integration in speech processing : Computational basis and neural organization», *Neuron*, vol. 69(3), p. 407–422.
- [50] Hickok, G., K. Okada, W. Barr, J. Pa, C. Rogalsky, K. Donnelly, L. Barde et A. Grant. 2008, «Bilateral capacity for speech sound processing in auditory comprehension : Evidence from Wada procedures», *Brain and Language*, vol. 107(3), p. 179–184.
- [51] Kelly, S., A. Ozyürek et E. Maris. 2010, «Two sides of the same coin : Speech and gesture mutually interact to enhance comprehension», *Psychological Science*, vol. 21(2), p. 260–267.
- [52] Kimura, D. 1963, «Right temporal-lobe damage : Perception of unfamiliar stimuli after damage», *Archives of Neurology*, vol. 8(3), p. 264–271.
- [53] Kita, S. et A. Ozyürek. 2003, «What does cross-linguistic variation in semantic coordination of speech and gesture reveal ? : Evidence for an interface representation of spatial thinking and speaking», *Journal of Memory and Language*, vol. 48, n° 1, p. 16–32.
- [54] Kohler, E., C. Keysers, A. Umiltà, L. Fogassi, V. Gallese et G. Rizzolatti. 2002, «Hearing sounds, understanding actions : Action representation in mirror neurons», *Science*, vol. 297, p. 846–848.
- [55] Kuhl, P. et J. Miller. 1975, «Speech perception by the chinchilla : Voiced-voiceless distinction in alveolar plosive consonants», *Science*, vol. 190, p. 69–72.
- [56] Lachaud, C. 2007, «CHACQFAM : Une base de données renseignant l’âge d’acquisition estimé et la familiarité pour 1225 mots monosyllabiques et bisyllabiques du Français», *L’Année Psychologique*, vol. 107(1), p. 39–63.
- [57] Lenneberg, E. 1962, «Understanding language without ability to speak : A case report», *Journal of Abnormal and Social Psychology*, vol. 65, p. 419–425.
- [58] Liberman, A. 1996, *Speech : A special code*, The MIT Press.
- [59] Liberman, A., K. Harris, H. Hoffman et B. Griffith. 1957, «The discrimination of speech sounds within and across phoneme boundaries», *Journal of Experimental Psychology*, vol. 54, p. 358–368.
- [60] Liberman, A. et I. Mattingly. 1985, «The motor theory of speech perception revised», *Cognition*, vol. 21, p. 1–36.
- [61] Luce, P. et D. Pisoni. 1998, «Recognizing spoken words : The neighborhood activation model», *Ear and Hearing*, vol. 19, p. 1–36.
- [62] Mahon, B. et A. Caramazza. 2008, «A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content», *Journal of Physiology*, vol. 102, p. 59–70.

- [63] Marslen-Wilson, W. 1987, «Functional parallelism in spoken word-recognition», *Cognition*, vol. 25, p. 71–102.
- [64] Marslen-Wilson, W. 1990, «Activation, competition, and frequency in lexical access», dans *Cognitive Models of Speech Processing : Psycholinguistics and Computational Perspectives*, édité par G. Altmann, Cambridge, MA : The MIT Press, p. 148–172.
- [65] Marslen-Wilson, W. et P. Zwitserlood. 1989, «Accessing spoken words : The importance of word onsets», *Journal of Experimental Psychology : Human Perception and Performance*, vol. 15, p. 576–585.
- [66] Mauchly, J. 1940, «Significance test for sphericity of a normal n -variate distribution», *The Annals of Mathematical Statistics*, vol. 11, p. 204–209.
- [67] McNeill, D. 2010, «Gesture», dans *Cambridge Encyclopedia of the Language Sciences*, édité par P. Hogan, Cambridge, England : Cambridge University Press.
- [68] Meteyard, L., S. Cuadrado, B. Bahrami et G. Vigliocco. 2012, «Coming of age : A review of embodiment and the neuroscience of semantics», *Cortex*, vol. 48(7), p. 788–804.
- [69] Meunier, F., M. Hoen et J. Segui. 2008, «L'effet de fréquence en modalité auditive», *Parole*, vol. 43, p. 241–262.
- [70] Mielke, J., A. Baker et D. Archangeli. 2010, «Variability and homogeneity in American English /r/ allophony and /s/ retraction», dans *Variation, Detail, and Representation : LabPhon 10*, Berlin : Mouton de Gruyter, p. 699–719.
- [71] New, B., C. Pallier, M. Brysbaert et L. Ferrand. 2004, «Lexique 2 : A new French lexical database», *Behavior Research Methods, Instruments & Computers : A Journal of the Psychonomic Society Inc.*, vol. 36, p. 516–524.
- [72] New, B., C. Pallier, L. Ferrand et R. Matos. 2001, «Une base de données lexicales du français contemporain sur internet : Lexique», *L'Année Psychologique*, vol. 101, p. 447–462.
- [73] Perrier, P., L. Ma et Y. Payan. 2005, «Modeling the production of vcv sequences via the inversion of the biomechanical of the tongue», dans *Proceedings of Interspeech 2005*, p. 1041–1044.
- [74] Perrin, F. 2011, «*Les zombies comprennent-ils ce qu'on leur dit ? Les mécanismes inconscients de la perception de la parole*», habilitation à diriger des recherches, Université Claude Bernard Lyon 1.
- [75] Porter, R. et F. Castellanos. 1980, «Speech-production measures of speech perception : Rapid shadowing of VCV syllables», *Journal of Acoustical Society of America*, vol. 67, p. 1349–1356.
- [76] Porter, R. et J. Lubker. 1980, «Rapid reproduction of vowel-vowel sequences : Evidence for a fast and direct acoustic motoric linkage in speech», *Journal of Speech and Hearing Research*, vol. 23, p. 593–602.

- [77] Posner, M. 1986, *Chronometric explorations of the mind*, New York Oxford University Press.
- [78] Pulvermüller, F., M. Huss, F. Kherif, F. Moscoso del Prado Martin, O. Hauk et Y. Shtyrov. 2006, «Motor cortex maps articulatory features of speech sounds», dans *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, p. 7865–7870.
- [79] Reinwein, I. 1996, «Le temps est-il un indicateur fiable de la difficulté de lecture du texte / de son traitement cognitif par le lecteur?», *Revue Québécoise de Linguistique*, vol. 25, p. 145–162.
- [80] Remez, R. 1996, «Critique : Auditory form and gestural topology in the perception of speech», *Journal of the Acoustical Society of America*, vol. 99, p. 1695–1698.
- [81] Remez, R. 2005, «The perceptual organization of speech», dans *The Handbook of Speech Perception*, édité par D. Pisoni et R. Remez, Oxford : Blackwell, p. 28–50.
- [82] Rizzolatti, G. et M. A. Arbib. 1998, «Language within our grasp», *Trends in Neurosciences*, vol. 21, p. p. 188–194.
- [83] Rizzolatti, G., L. Fadiga, V. Gallese et L. Fogassi. 1996, «Premotor cortex and the recognition of motor actions», *Cognitive Brain Research*, vol. 3, p. p. 131–141.
- [84] Rossi, J. 1977, «La supériorité de l'identification des mots par rapport aux non-mots», *L'année psychologique*, vol. 77, n° 2, p. 475–495, ISSN 0003-5033.
- [85] Roy, A., L. Craighero, M. Fabbri-Destro et L. Fadiga. 2008, «Phonological and lexical motor facilitation during speech listening : A transcranial magnetic stimulation study», *Journal of Physiology*, vol. 102, p. 101–105.
- [86] Rueschemeyer, S., O. Lindemann, D. van Rooij, W. van Dam et H. Bekkering. 2010, «Effects of intentional motor actions on embodied language processing», *Experimental Psychology*, vol. 57, p. 260–266.
- [87] Sato, M. 2004, *Représentations verbales multistables en mémoire de travail : Vers une perception active des unités de parole*, thèse de doctorat, Institut National Polytechnique de Grenoble.
- [88] Sato, M., G. Buccino, M. Gentilucci et L. Cattaneo. 2010, «On the tip of the tongue : Modulation of the primary motor cortex during audiovisual speech perception», *Speech Communication*, vol. 52(6), p. 533–541.
- [89] Sato, M., K. Grabski, A. Glenberg, A. Brisebois, A. Basirat, L. Ménard et L. Cattaneo. 2011, «Articulatory bias in speech categorization : Evidence from use-induced motor plasticity», *Cortex*, vol. 47(8), p. 1001–1003.
- [90] Savariaux, C., P. Perrier et J. Orliaguet. 1995, «Compensation strategies for the perturbation of the rounded vowel [u] using a lip tube : A study of the control space in speech production», *Journal of the Acoustical Society of America*, vol. 98, p. 2428–2842.

Bibliographie

- [91] Savariaux, C., P. Perrier, J. Orliaguet et J. Schwartz. 1999, «Compensation for the perturbation of French [u] using a lip tube : II. Perceptual analysis», *Journal of the Acoustical Society of America*, vol. 106(1), p. 381–393.
- [92] Schwartz, J., C. Abry, L. Boë et M. Cathiard. 2002, «Phonology in a Theory of Perception-for-Action-Control», dans *Phonetics, Phonology, and Cognition*, édité par J. Durand et B. Laks, Oxford University Press, p. 244–280.
- [93] Schwartz, J., A. Basirat, L. Ménard et M. Sato. 2012, «The Perception-for-Action-Control Theory (PACT) : A perceptuo-motor theory of speech perception», *Journal of Neurolinguistics*, vol. 25(5), p. 336–354.
- [94] Schwartz, J., M. Sato et L. Fadiga. 2008, «The common language of speech perception and action : A neurocognitive perspective», *Perception*, vol. 13, p. 9–22.
- [95] Scott, S. et I. Johnsrude. 2003, «The neuroanatomical and functional organization of speech perception», *Trends in Neurosciences*, vol. 26, p. 100–107.
- [96] Shapiro, S. 1965, «An analysis of variance test for normality (complete samples)», *Biometrika*, vol. 52, p. 591–611.
- [97] Skipper, J., S. Goldin-Meadow, H. Nusbaum et S. Small. 2007, «Speech-associated gestures, Broca’s area, and the human mirror system», *Brain and Language*, vol. 101(3), p. 260–277.
- [98] Skipper, J., H. Nusbaum et S. Small. 2005, «Listening to talking faces : Motor cortical activation during speech perception», *NeuroImage*, vol. 25, p. 76–89.
- [99] Skipper, J., V. Van Wassenhove, H. Nusbaum et S. Small. 2007, «Hearing lips and seeing voices : How cortical areas supporting speech production mediate audiovisual speech perception», *Cerebral Cortex*, vol. 17, p. 2387–2399.
- [100] Spinelli, E. et L. Ferrand. 2005, *Psychologie du langage : L’écrit et le parlé, du signal à la signification*, Paris : Armand Colin.
- [101] Spinelli, E., J. Segui et M. Radeau. 2001, «Phonological priming in spoken word recognition with bisyllabic targets», *Language and Cognitive Processes*, vol. 16, p. 367–392.
- [102] Stroop, J. 1935, «Studies of interference in serial verbal reactions», *Journal of Experimental Psychology*, vol. 18, p. 643–662.
- [103] Sundara, M., A. Namasivayam et R. Chen. 2001, «Observation-execution matching system for speech : A magnetic stimulation study», *NeuroReport*, vol. 12, p. 1341–1344.
- [104] Tessier, C. 2006, *Le traitement des homographes en temps réel : Interaction du degré de polarité et du SOA*, mémoire de maîtrise, Université du Québec à Montréal.
- [105] Tokimura, H., Y. Tokimura, A. Oliviero, T. Asakura et J. Rothwell. 1996, «Speech-induced changes in corticospinal excitability», *Annals of Neurology*, vol. 40(4), p. 628–634.

- [106] Wada, J. 1949, «A new method for the determination of the side of cerebral speech dominance : A preliminary report on the intracarotid injection of sodium amytal in man», *Medicine and Biology*, vol. 4, p. 221–222.
- [107] Warren, P. et W. Marslen-Wilson. 1988, «Cues to lexical choice : Discriminating place and voice», *Perception & Psychophysics*, vol. 43, p. 21–30.
- [108] Watkins, K. et T. Paus. 2004, «Modulation of motor excitability during speech perception : The role of Broca’s area», *Journal of Cognitive Neuroscience*, vol. 16, p. 978–987.
- [109] Watkins, K., A. Strafella et T. Paus. 2003, «Seeing and hearing speech excites the motor system involved in speech production», *Neuropsychologia*, vol. 41, p. 989–994.
- [110] Wilson, S. et M. Iacoboni. 2006, «Neural responses to non-native phonemes varying in producibility : Evidence for the sensorimotor nature of speech perception», *NeuroImage*, vol. 33, p. 316–325.
- [111] Wilson, S., A. Saygin, M. Sereno et M. Iacoboni. 2004, «Listening to speech activates motor areas involved in speech production», *Nature Neuroscience*, vol. 7, p. 701–702.
- [112] Witt, J., D. Kemmerer, S. Linkenauer et J. Culham. 2010, «A functional role for motor simulation in naming tools», *Psychological Science*, vol. 21, p. 1215–1219.
- [113] Zwaan, R. et L. Taylor. 2006, «Seeing, acting, understanding : motor resonance in language comprehension», *Journal of Experimental Psychology : General*, vol. 135, p. 1–11.

Conclusion générale

« *Quelquefois, hasarder des réponses est seulement une manière d'éclaircir pour soi-même des questions* » Alessandro Baricco (*L'âme de Hegel et les vaches du Wisconsin*)

Ce travail de thèse, dont l'objectif était de comprendre et de caractériser au mieux les liens qui unissent le geste manuel, la parole et le langage, s'inscrit dans une littérature abondante et pluridisciplinaire. La question des relations geste/parole est en effet au cœur de débats passionnés, relatifs à la nature même du langage humain, de sa structure, mais également de son développement ontogénétique et phylogénétique.

L'hypothèse générale que nous avons souhaité tester est celle d'un couplage entre geste et parole, au sein d'un seul et même système communicatif, par nature multimodal ; notre objectif était donc de montrer que le langage pouvait être considéré comme un système vocal autant que gestuel, reposant sur l'utilisation conjointe de deux modalités.

Nous avons choisi d'évaluer le lien entre gestes manuels, geste vocaux, et structures linguistiques associées, au sein d'un processus communicatif particulier, la deixis spatiale, et par le biais d'une approche expérimentale contrôlée. L'étude du lien geste/parole s'est déclinée selon trois axes de recherche, (1) dans les mécanismes de production du langage (présenté au Chapitre 3) ; (2) au cours du développement tardif du langage (présenté au Chapitre 4) ; (3) dans les mécanismes de perception du langage (présenté au Chapitre 5).

Au terme de ce travail de thèse, nous proposons un bref rappel des principaux résultats observés, à partir desquels nous envisagerons une perspective plus large, consacrée à la phylogenèse de la multimodalité de la communication langagière humaine.

4 Synthèse des principaux résultats

4.1 Axe de recherche 1 : Lien geste/parole dans les mécanismes de production du langage

Le premier objectif de cette étude (Expériences 1 & 2, présentées au Chapitre 3) était d'évaluer la nature de l'interaction geste/parole. Notre postulat de base étant que geste et parole sont deux modalités d'un système unifié, l'hypothèse relative à cette question était celle d'une génération commune des productions vocales et manuelles, c'est-à-dire au même niveau de traitement (i.e. hypothèse linguistique de génération du geste, proposée notamment par Kita et Ozyürek [13] et Hostetter et Alibali [11], inspirés des travaux de McNeill et Duncan [15]). Dans cette optique, nous avons souhaité mettre en évidence une interaction particulière entre ces deux systèmes de production, par le biais d'une tâche de production de pointage unimodal et bimodal, vers des cibles situées à différentes distances.

Les résultats de cette étude sont en faveur d'un couplage particulier entre le geste et la parole, qui apparaissent comme deux modalités langagières à part entière, étroitement liées par une interaction structurelle de bas niveau, bidirectionnelle et complémentaire. Ainsi, les productions vocales et manuelles sont structurées relativement les unes aux autres et non indépendamment les unes des autres. En d'autres termes, la quantité d'information véhiculée par le geste dépend de celle véhiculée par la parole, et vice versa ; plus précisément, l'absence d'une modalité est compensée par un renforcement des productions dans l'autre modalité, alors qu'inversement, l'information est répartie sur l'ensemble des modalités si toutes deux sont disponibles pour transmettre le message souhaité.

Ces données peuvent contribuer à enrichir la littérature actuelle consacrée aux modèles de production conjointe de geste et de parole, en y apportant des informations relatives au processus déictique. Comme le souligne Roustan [24], « *la plupart de ces modèles ont été conçus (et testés) pour les gestes iconiques, or ces modèles prétendent à l'explication de la coordination des deux modalités pour les "gesticulations" en général* » ([24], p.135). De fait, les modèles actuels ne sont pas relatifs à un type de geste en particulier, mais il est pourtant envisageable que chaque type de geste entretienne une relation particulière à la parole qu'il accompagne (Roustan [24] montre par exemple des patrons de coordination temporelle différents selon que le geste considéré est un geste de battement ou de pointage). Dans cette optique, nos résultats pourraient être en partie expliqués par le modèle d'Interface proposé par Kita et Ozyürek [13], qui suggère un échange d'informations bidirectionnel de haut niveau (c'est-à-dire entre les Générateurs d'Actions et de Messages), permettant aux gestes d'être influencés par la structure de l'énoncé verbal. Nos résultats suggèrent que l'énoncé verbal puisse également être structuré relativement au geste, phénomène que nous expliquons par un échange bidirectionnel de plus bas niveau (c'est-à-dire entre les centres moteurs, soit le Contrôle moteur et le Formulateur).

Comme nous l'avons évoqué précédemment dans le Chapitre 3, ces résultats pourraient être améliorés, notamment par l'application de procédures de normalisation et l'augmentation du nombre de participants, qui favoriseraient l'apparition d'effets plus robustes. Ce travail pourrait par ailleurs être complété par l'évaluation de l'aspect temporel des productions unimodales et bimodales. Cette thématique a été abordée dans le cadre d'un stage ingénieur, ayant donné lieu à des analyses temporelles préliminaires (Chadœuf [3]). A plus long terme, l'ensemble de ces données pourrait permettre l'élaboration d'un modèle de production plus complet, caractérisant la nature structurelle et temporelle du couplage geste/parole.

Le second objectif de cette étude était d'évaluer l'encodage de la distance, une fonction déictique particulière, via les propriétés lexicales, phonétiques (i.e. acoustiques et articulatoires) et cinématiques du pointage. Deux aspects de cet encodage ont été considérés. Nous avons tout d'abord souhaité montrer que cet encodage était issu d'un phénomène moteur, de bas niveau, dans les deux modalités : désigner un objet éloigné impliquerait un geste plus ample, que celui-ci soit un geste vocal ou manuel ; par ailleurs, nous avons souhaiter relier cet encodage moteur à l'encodage phonologique de la distance (i.e. distinction proximal/distal) et déterminer ainsi dans quelle mesure le système moteur, vocal et manuel, pouvait être impliqué dans la construction de nos représentations linguistiques spatiales.

4. Synthèse des principaux résultats

Nos résultats montrent un effet de la distance de l'objet pointé sur les propriétés du pointage manuel et vocal. Plus particulièrement, nous avons mis en évidence des variations phonétiques et cinématiques en fonction de la distance de la cible, soit une augmentation de divers paramètres, et ce indépendamment de la nature lexicale du terme utilisé en parole ou du type de voyelle produit. En d'autres termes, nous utilisons une stratégie particulière pour désigner de manière explicite un objet particulier, ce qui nous permet de faire la distinction entre deux objets situés dans l'espace péripersonnel et extrapersonnel, mais également entre deux objets ambigus, situés dans un même espace perceptif. Il semble également que le point de vue de notre interlocuteur influence nos productions, puisque cet encodage de la distance, au niveau articulatoire, n'est observé que lorsque nous disposons uniquement de la modalité vocale (message alors moins "visible" pour l'interlocuteur). L'encodage de la distance, qui apparaît donc comme une fonction motrice de bas niveau, pourrait également expliquer la distinction phonologique observée entre les termes proximaux et distaux, dans diverses langues du monde. En effet, les patrons phonétiques, cinématiques et phonologiques d'encodage sont cohérents et pourraient donc être corrélés.

Ces résultats pourraient être approfondis par le biais d'études cross-linguistiques. Il est en effet possible que l'encodage observé dans notre expérience ne soit pas purement moteur mais plutôt le "simple" reflet du patron phonologique de la langue maternelle des individus. Notre étude ne permet pas d'éclaircir ce point puisque l'encodage lexical du français correspond à l'encodage phonétique/cinématique observé dans notre expérience. Il apparaît donc pertinent de tester ce protocole auprès d'individus de langues maternelles variées, respectant différents patrons d'encodage phonologique.

4.2 Axe de recherche 2 : Lien geste/parole au cours du développement tardif du langage

L'objectif de notre seconde étude, présentée au Chapitre 4, était de caractériser l'évolution de la relation geste/parole au cours du développement tardif de l'enfant, c'est-à-dire chez l'enfant d'âge scolaire, afin de déterminer si les performances observées chez l'adulte (en termes d'interaction geste/parole et d'encodage moteur de la distance) pouvaient être considérées comme des compétences précoces, i.e. des caractéristiques intrinsèques du processus déictique, ou des compétences plus tardives, i.e. des fonctions cognitives élaborées.

Nos résultats sont en faveur d'une évolution progressive du couplage geste/parole. L'encodage moteur de la distance, ainsi que la coopération inter-modalités, se mettent en place au cours du développement tardif de l'enfant et nécessitent donc l'acquisition, par la pratique, de stratégies élaborées, intégrées progressivement au répertoire communicatif de l'enfant. Nos données montrent également une certaine asymétrie en terme d'utilisation des deux modalités chez les enfants les plus jeunes, qui semblent favoriser la modalité manuelle. L'information spatiale serait donc principalement portée par le geste, avant d'être véhiculée via les deux modalités (l'encodage cinématique de la distance apparaît plus tôt que l'encodage phonétique, tout comme l'influence du geste sur la parole émerge avant l'influence de la parole sur le geste).

Notre objectif initial était de pouvoir évaluer les enfants par tranche d'âge minimal, afin d'élaborer un "agenda développemental" du pointage linguistique, dans les modalités vocale et manuelle. Il s'avère donc capital de poursuivre cette étude afin d'obtenir un plus large panel d'observations. Nos données pourraient ainsi, à long terme, permettre de faire le lien entre les performances des tout-petits (au cours des deux premières années de vie) et celles des adultes. De cette façon, on pourrait envisager l'élaboration d'un modèle de production conjointe de geste et de parole, de la petite enfance à l'âge adulte, intégrant des données relatives à la coordination structurelle et temporelle des signaux. Dans cette optique, un nouveau protocole a été élaboré au Gipsa-Lab et a donné lieu à un stage de Master Recherche, consacré à l'évaluation des aspects temporels du pointage vocal et manuel, chez l'adulte et l'enfant (Clarke [4], Vilain et collab. [28]).

4.3 Axe de recherche 3 : Lien geste/parole dans les mécanismes de perception du langage

L'objectif de notre dernière étude, présentée au Chapitre 5, était d'évaluer le couplage geste/parole dans les mécanismes de perception du langage, qui semblent être sous-tendus par des boucles de perception/action et l'activation de représentations de nature sensorimotrices. Notre hypothèse étant celle d'une connexion étroite entre geste et parole, nous avons souhaité montrer que la production d'un geste manuel pouvait activer les représentations lexicales qui lui sont associées. Pour ce faire, la tâche proposée aux participants était une tâche de décision lexicale (décider le plus rapidement et justement possible si l'item perçu est un mot de la langue française), basée sur le principe de l'amorçage (chaque item perçu est précédé d'une amorce pouvant ou non faciliter son traitement). Plus précisément, nous avons souhaité montrer que la production d'un geste déictique (relativement à l'absence de geste ou à la production d'un geste non communicatif) facilitait le traitement de termes déictiques (relativement à celui de noms ou de pseudo-mots), un argument supplémentaire en faveur d'un système multimodal unifié.

Les résultats de cette étude suggèrent que l'activation de représentations motrices manuelles induit une activation des représentations linguistiques qui lui sont associées. Dans cette mesure, ils sont en faveur d'une intégration geste/parole dans le processus de perception du langage, sous la forme d'une interaction de nature sémantique. Par ailleurs, le geste manuel active les représentations lexicales de manière générique plutôt que spécifique, puisque les représentations motrices associées à l'espace perceptif proche *versus* lointain ne sont pas strictement associées aux représentations lexicales proximales *versus* distales.

Néanmoins, comme évoqué dans le Chapitre 5, les résultats relatifs aux représentations lexicales proximales et distales pourraient être liés à des biais méthodologiques (notamment dûs à l'utilisation d'un protocole complexe et un nombre relativement faible de cibles lexicales reliées). L'élaboration d'un nouveau protocole pourrait donc être envisagée afin de confirmer, ou non, ces observations. Par ailleurs, la comparaison de deux gestes de même structure mais porteurs d'une fonction distincte et de nature différente (tel qu'un geste de pointage et un mouvement d'atteinte et de saisie vers un objet situé à portée de main) permettrait d'évaluer plus précisément le niveau de traitement du langage auquel les représentations motrices interviennent.

5. Perspective : Phylogénèse de la multimodalité

Enfin, il est souhaitable que se développent des études sur la co-structuration des unités sensorimotrices au cours du développement. Quelques rares études (e.g. Depaolis et collab. [7]) commencent à montrer que la structuration de l'espace phonétique perceptif de l'enfant se fait à travers le crible de ses capacités motrices, et l'on pourrait penser qu'il en va de même pour la perception des unités sémantiques.

5 Perspective : Phylogénèse de la multimodalité

Nos différentes études ayant apporté des arguments pour une intégration forte du geste et de la parole dans le langage humain, se pose logiquement la question de la place de cette intégration au cours de l'émergence de la communication langagière dans l'espèce humaine. Comme évoqué dans l'Introduction de ce manuscrit, la question de la nature du langage humain, de son émergence et de son évolution, est de toute première importance ; elle anime la communauté scientifique dans son ensemble et suscite un nombre considérable de travaux, issus de domaines divers et variés. Plus particulièrement, l'étude comparative des comportements des primates humains et non humains présente l'intérêt de mettre en évidence certaines continuités/discontinuités entre le système communicatif de nos proches cousins et certains traits de notre langage. Elle présente ainsi un cadre idéal pour évaluer différentes hypothèses relatives à l'origine du langage (dont une revue est disponible notamment dans les travaux de Fitch [8]).

Dans ce cadre là, de nombreuses études, synthétisées dans les travaux de Tomasello et Zuberbuehler [27] ou Arbib et collab. [1], portent sur les vocalisations (e.g. Candiotti et collab. [2], Cäsar et collab. [5], Cäsar et Zuberbuehler [6], Seyfarth et collab. [25], Wallez et Vauclair [29]) ou les gestes manuels (e.g. Meguerditchian [16], Meguerditchian et collab. [17], Meguerditchian et Vauclair [18], Meunier et collab. [19]), produits par les singes ou grands singes dans des contextes variés, avec notamment l'objectif d'établir une typologie des productions dans chacune des modalités, afin de pouvoir les comparer aux productions humaines.

Toutefois, comme nous l'avons démontré dans ce travail de thèse, le langage humain apparaît comme un processus par nature multimodal, et repose donc sur l'utilisation conjointe de diverses modalités. Afin de comprendre l'établissement phylogénétique d'un tel système, il s'avère alors crucial d'évaluer non pas les productions vocales et gestuelles indépendamment les unes des autres, mais bien relativement les unes aux autres.

Cet aspect multimodal des conduites reste, à notre connaissance, encore largement inexploré. Slocombe et collab. [26] ont à ce sujet établi un classement de diverses recherches empiriques, relatives aux comportements communicatifs spontanés, observés chez différentes espèces de primates non humains. Plus précisément, leur étude porte sur des articles scientifiques, rédigés en anglais, approuvés par comité de lecture, sur la période 1960-2008. Les auteurs ont effectué une recherche à partir de mots-clé, tels que *monkey*, *call*, *vocalization*, *communication*, *gesture*^{*}, dans trois bases de données couramment utilisées, *Web of Science*²⁵, *ScienceDirect*²⁶ et *Pri-*

25. <http://wokinfo.com>

26. <http://www.sciencedirect.com>

*mateLit*²⁷. Leur corpus final, comptant non moins de 553 articles, constitue un ensemble varié, bien que non exhaustif, de recherches relatives aux comportements manuels, oro-faciaux et/ou vocaux.

Les auteurs observent que sur l'ensemble de leur corpus, seuls 5% d'articles concernent l'aspect multimodal des conduites ! En revanche, 64% sont consacrés aux vocalisations, 22% aux expressions faciales et seulement 9% aux productions manuelles. Outre cette asymétrie en termes de quantité d'informations disponibles sur chaque modalité, les auteurs révèlent des différences considérables entre les différents travaux. Par exemple, la communication vocale est davantage étudiée chez le singe alors que la plupart des études relatives à la communication manuelle portent sur les grands singes. Les approches méthodologiques utilisées diffèrent également ; alors que certaines études consacrées aux vocalisations sont réalisées dans des contextes relativement stressants pour l'animal (le paradigme du *playback* peut par exemple être utilisé pour générer des situations d'attaque, de menace), que ce soit en milieu naturel ou en captivité, la plupart des études consacrées aux gestes manuels sont en revanche réalisées dans des contextes d'interaction sociale, de jeu, et le plus souvent en captivité. Les auteurs questionnent alors la validité des approches multimodales, basées sur la mise en relation de données unimodales pourtant incomparables au vu des discontinuités quantitatives et méthodologiques qu'elles présentent. Selon eux, « *we are unlikely to understand multimodal signals fully by studying the component signals independently*²⁸ » (Slocombe et collab. [26], p.922).

Concernant plus précisément les 28 articles qui portent effectivement sur l'aspect multimodal des conduites (et non sur le regroupement d'observations initialement indépendantes), neuf sont consacrés aux expressions faciales, vocalisations et gestes, et seuls cinq s'intéressent aux comportements vocaux et manuels. On trouve également des données relatives au lien entre expressions faciales et gestes, ou expressions faciales et vocalisations (e.g. Ghazanfar et Logothetis [9], Ghazanfar et Schroeder [10], Parr [20], Partan [21]).

La plupart de ces études ont pour objectif d'établir l'aspect multimodal des conduites et montrent qu'à l'instar de la communication langagière humaine, la communication des primates est basée sur une grande variété de comportements, provenant de diverses modalités. Leavens et collab. [14] et Hostetter et collab. [12] ont par exemple montré une modulation des productions chez le grand singe, relative aux comportements de leur interlocuteur ; plus précisément, les chimpanzés, qui disposent d'un répertoire riche, composé de signaux visuels, gestuels, tactiles et auditifs leur permettant de communiquer avec l'homme, ont la possibilité de combiner ou non certains signaux, en fonction de l'état attentionnel de l'interlocuteur. Pollick et de Waal [22] ont quant à eux observé les comportements spontanés de chimpanzés et de bonobos, pour comparer les productions vocales et manuelles de ces deux espèces, notamment en termes de flexibilité, et pour étudier l'effet de signaux multimodaux sur le comportement du destinataire. D'autres études, telles que celle réalisée par Partan [21], montrent que certaines combinaisons d'expressions faciales et de vocalisations sont fortement contraintes (i.e. combinaisons systématiques) tandis que d'autres associations sont plus flexibles.

27. <http://primatelit.library.wisc.edu/>

28. Traduction proposée : Il est peu probable d'espérer comprendre l'aspect multimodal des signaux en évaluant leurs différents composants de manière indépendante.

5. Perspective : Phylogénèse de la multimodalité

Malgré un nombre restreint d'études concernant la multimodalité des comportements chez le primate non humain, Slocombe et collab. [26] suggèrent que le « *language may have evolved through an integrated combination of vocal, gestural and facial communication, rather than a unimodal system*²⁹ » ([26], p.923).

Néanmoins, ce type d'études, bien qu'essentiel, est pour l'heure insuffisant ; en effet, si les données ainsi obtenues indiquent clairement que la communication des primates est multimodale, elles ne permettent pas, en revanche, d'établir une typologie précise des productions unimodales et bimodales, observées dans un même contexte. En d'autres termes, elles fournissent des informations qualitatives mais n'évaluent pas l'intégration, au sens strict du terme, de différentes modalités.

Dans ce cadre là, un projet de recherche pourrait être envisagé, permettant à la fois d'évaluer l'aspect multimodal des conduites mais également de proposer une classification précise des signaux multimodaux.

Plusieurs perspectives de travail peuvent alors être envisagées :

- Observation de comportements spontanés en captivité (à la manière de Pollick et de Waal [22]) et en milieu naturel (à la manière de Roberts et collab. [23]), afin de déterminer dans quelle mesure les données récoltées en captivité sont dues à des artefacts liés à l'environnement social, humain, des animaux ; élicitation de comportements en captivité (à la manière d'Hostetter et collab. [12] ou de Leavens et collab. [14]) ;
- Observation des comportements chez le singe et le grand singe, afin d'obtenir des données sur des primates plus ou moins éloignés phylogénétiquement de l'espèce humaine et d'établir un continuum phylogénétique ;
- Classification des signaux communicatifs et intentionnels "isolés" : (1) vocalisations, en termes sémantique (e.g. cris d'alarme) et acoustique ; (2) comportements oro-faciaux ; (3) comportements manuels, en termes sémantique (e.g. quémante, pointage, menace, salut) et physique/morphologique (voir à ce propos les travaux de Roberts et collab. [23], qui proposent une classification en termes de trajectoire, d'orientation, durant les phases de préparation et de stroke du geste) ;
- Classification des combinaisons de ces différents types de signaux ;
- Comparaison des signaux "isolés" *versus* "combinés", afin d'évaluer les modulations des comportements (en termes acoustiques et physiques/ morphologiques), en fonction des modalités impliquées.

Un tel projet pourrait s'inscrire dans la littérature actuelle relative à l'origine du langage, en évaluant les précurseurs phylogénétiques de la multimodalité de la communication langagière humaine. Nous espérons que notre travail de thèse puisse constituer un argument supplémentaire pour le développement de ce type d'études et fournir certaines pistes de recherche relatives, par exemple, à l'interaction entre gestes vocaux et manuels chez le primate non humain.

29. Traduction proposée : Il est possible que le langage ait évolué à partir d'une combinaison intégrée de signaux vocaux, gestuels et faciaux, plutôt qu'à partir d'un système unimodal.

Bibliographie

- [1] Arbib, M., K. Liebal et S. Pika. 2008, «Primate vocalization, gesture, and the evolution of human language», *Current Anthropology*, vol. 59, p. 1053–1076.
- [2] Candiotti, A., K. Zuberbuehler et A. Lemasson. 2012, «Convergence and divergence in Diana monkey vocalizations», *Biology Letters*, vol. 8(3), p. 382–385.
- [3] Chadœuf, A. 2010, «Analyse de la production gestuelle d'agents en interaction», Mémoire de stage d'ingénieur 2A, Université de Grenoble.
- [4] Clarke, J. 2011, «Etude du développement de la synchronisation temporelle entre la parole et les gestes de pointage chez les enfants», mémoire de master, Université de Grenoble.
- [5] Cäsar, C., R. Byrne, R. Young et K. Zuberbuehler. 2012, «The alarm call system of wild black-fronted titi monkeys, *Callicebus nigrifrons*», *Behavioral Ecology and Sociobiology*, vol. 66(5), p. 653–667.
- [6] Cäsar, C. et K. Zuberbuehler. 2012, «Referential alarm calling behaviour in New World primates», *Current Zoology*, vol. 58(5), p. 680–697.
- [7] Depaolis, R., M. Vihman et T. Keren-Portnoy. 2011, «Do production patterns influence the processing of speech in prelinguistic infants?», *Infant Behavior and Development*, vol. 34(4), p. 590–601.
- [8] Fitch, W. 2005, «The evolution of language : A comparative review», *Biology and Philosophy*, vol. 20, p. 193–230.
- [9] Ghazanfar, A. et N. Logothetis. 2003, «Neuroperception : Facial expressions linked to monkey calls», *Nature*, vol. 423, p. 937–938.
- [10] Ghazanfar, A. et C. Schroeder. 2006, «Is neocortex essentially multisensory?», *Trends in Cognitive Science*, vol. 10, p. 278–285.
- [11] Hostetter, A. et M. Alibali. 2008, «Visible embodiment : Gestures as Simulated Action», *Psychonomic Bulletin & Review*, vol. 15, n° 3, p. 495–514.
- [12] Hostetter, A., M. Cantero et W. Hopkins. 2001, «Differential use of vocal and gestural communication by chimpanzees (*Pan troglodytes*) in response to the attentional status of a human (*Homo sapiens*)», *Journal of Comparative Psychology*, vol. 115(4), p. 337–43.
- [13] Kita, S. et A. Ozyürek. 2003, «What does cross-linguistic variation in semantic coordination of speech and gesture reveal? : Evidence for an interface representation of spatial thinking and speaking», *Journal of Memory and Language*, vol. 48, n° 1, p. 16–32.
- [14] Leavens, D., J. Russell et W. Hopkins. 2010, «Multimodal communication by captive chimpanzees (*Pan troglodytes*)», *Animal Cognition*, vol. 13, p. 33–40.
- [15] McNeill, D. et S. Duncan. 2000, «Growth points in thinking-for-speaking», dans *Language and Gesture*, Cambridge University Press, p. 141–161.

Bibliographie

- [16] Meguerditchian, A. 2009, *Latéralité et communication gestuelle chez le babouin et le chimpanzé : A la recherche des précurseurs du langage*, thèse de doctorat, Université Aix-Marseille.
- [17] Meguerditchian, A., S. Molesti et J. Vauclair. 2011, «Right-handedness predominance in 162 baboons for gestural communication : Consistency across time and groups», *Behavioral Neuroscience*, vol. 125(4), p. 653–660.
- [18] Meguerditchian, A. et J. Vauclair. 2006, «Baboons communicate with their right hand», *Behavioural Brain Research*, vol. 171, p. 170–174.
- [19] Meunier, H., J. Prieur et J. Vauclair. 2012, «Olive baboons communicate intentionally by pointing», *Animal Cognition*, vol. In press.
- [20] Parr, L. 2004, «Perceptual biases for multimodal cues in chimpanzee (*Pan troglodytes*) affect recognition», *Animal Cognition*, vol. 7, p. 171–178.
- [21] Partan, S. 2002, «Single and multichannel signal composition : Facial expressions and vocalizations of rhesus macaques (*Macaca mulatta*)», *Behaviour*, vol. 139, p. 993–1027.
- [22] Pollick, A. et F. de Waal. 2007, «Ape gestures and language evolution», *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104(19), p. 8184–8189.
- [23] Roberts, A., S. Vick, S. Roberts, H. Buchanan-Smith et K. Zuberbuehler. 2012, «A structure-based repertoire of manual gestures in wild chimpanzees : Statistical analyses of a graded communication system», *Evolution and Human Behavior*, vol. 33(5), p. 578–589.
- [24] Roustan, B. 2012, *Étude de la coordination gestes manuels/parole dans le cadre de la désignation*, thèse de doctorat, Université de Grenoble.
- [25] Seyfarth, R., D. Cheney et P. Marler. 1980, «Monkey responses to three different alarm calls : Evidence of predator classification and semantic communication», *Science*, vol. 210(4471), p. 801–803.
- [26] Slocombe, K., B. Wallerby et K. Liebal. 2011, «The language void : The need for multimodality in primate communication research», *Animal Behaviour*, vol. 81(5), p. 919–924.
- [27] Tomasello, M. et K. . Zuberbuehler. 2002, «Primate vocal and gestural communication», dans *The Cognitive Animal : Empirical and Theoretical Perspectives on Animal Cognition*, édité par M. Bekoff, C. Allen et G. Burghardt, Cambridge, Massachusetts : MIT Press, p. 293–299.
- [28] Vilain, A., C. Vilain et J. Clarke. 2012, «Vocal-manual coordination in child vs adult in two deictic tasks», dans *Papier présenté à Emergence of Language Abilities (ELA 2012)*, Lyon.
- [29] Wallez, C. et J. Vauclair. 2012, «First evidence of population-level oro-facial asymmetries during the production of distress calls by macaque (*Macaca mulatta*) and baboon (*Papio anubis*) infants», *Behavioural Brain Research*, vol. 234, p. 69–75.

Résumé/Abstract

Multimodalité de la communication langagière humaine : Interaction geste/parole et encodage de la distance dans le pointage

Désigner un objet à l'attention d'un interlocuteur est l'un des processus fondamentaux de la communication langagière. Ce processus est généralement réalisé à travers l'usage combiné de productions vocales et manuelles. Le but de ce travail est de comprendre et de caractériser les interactions entre la parole et le geste manuel dans des tâches de deixis spatiale, afin de déterminer de quelle façon l'information communicative est portée par chacun de ces deux systèmes, et de tester les principaux modèles de traitement du geste et de la parole.

La première partie de l'étude concerne la production du pointage vocal et manuel. L'originalité de ce travail consiste à rechercher des paramètres d'encodage de la distance dans les propriétés acoustiques, articulatoires et cinématiques du pointage multimodal, et à montrer que ces différentes caractéristiques sont corrélées les unes aux autres, et sous-tendues par un même comportement moteur de base : désigner un objet distant implique de produire des gestes plus larges, qu'ils soient vocaux ou manuels. Ce patron moteur peut alors être rapporté à l'encodage phonologique de la distance dans les langues du monde. Le paradigme expérimental utilisé dans cette étude contraste des pointages bimodaux vs unimodaux vocaux vs unimodaux manuels, et la comparaison de ces différentes conditions révèle par ailleurs une coopération bidirectionnelle entre les modalités vocales et manuelles, qui se répartissent la charge informative lorsqu'elles sont utilisées ensemble.

La deuxième partie de l'étude explore le développement du pointage multimodal. Les propriétés du pointage multimodal sont analysées chez des enfants de 6 à 12 ans, dans une tâche expérimentale comparable à celle des adultes. Cette deuxième expérience atteste d'un développement progressif du couplage geste/parole pour la deixis spatiale. Elle révèle que la distance est encodée préférentiellement dans les gestes manuels des enfants, par rapport à leurs gestes vocaux (surtout chez les jeunes enfants). Elle montre également que l'usage coopératif des gestes manuels et de la parole est déjà à l'œuvre chez les enfants, avec cependant plus d'influence du geste sur la parole que le patron inverse.

Le troisième volet de ce travail considère les interactions sensorimotrices dans la perception de la deixis spatiale. Une étude expérimentale basée sur un principe d'amorçage intermodal montre que le geste manuel joue un rôle dans le mécanisme de perception/production des unités sémantiques du langage. Ces données sont mises en parallèle avec les études sur la nature sensorimotrice des représentations dans le traitement des unités sonores du langage.

L'ensemble de ces résultats plaide pour une représentation fortement intégrée des gestes manuels et de la parole dans le cerveau langagier humain, et ce même à un stade relativement précoce du développement. Ils montrent également que l'encodage de la distance est un indice robuste, présent dans tous les aspects du pointage multimodal.

Mots-clés : Multimodalité, Deixis, Encodage de la distance, Production/Perception, Développement

Multimodality of linguistic communication : Gesture/speech interaction and distance encoding in pointing tasks

Designating an object for the benefit of another person is one of the most basic processes in linguistic communication. It is most of the time performed through the combined use of vocal and manual productions. The goal of this work is to understand and characterize the interactions between speech and manual gesture during pointing tasks, in order to determine how much linguistic information is carried by each of these two systems, and eventually to test the main models of speech and gesture production.

The first part of the study is about the production of vocal and manual pointing. The original aspect of this work is to look for distance encoding parameters in the lexical, acoustic, articulatory and kinematic properties of multimodal pointing, and to show that these different characteristics can be related with each other, and underlain by a similar basic motor behaviour : designating a distant object induces larger gestures, be they vocal or manual. This motor pattern can be related with the phonological pattern that is used for distance encoding in the world's languages. The experimental design that is used in this study contrasts bimodal vs. vocal monomodal vs. monomodal manual pointings, and a comparison between these conditions reveals that the vocal and manual modalities act in bidirectional cooperation for deixis, sharing the informational load when used together.

The second part of the study explores the development of multimodal pointing. The properties of multimodal pointing are assessed in 6-12 year-old children, in an experimental task similar to that of the adults. This second experiment attests a progressive evolution of speech/gesture interactions in the development of spatial deixis. It reveals that distance is preferentially encoded in manual gestures in children, rather than in vocal gestures (and especially so in younger children). It also shows that the cooperative use of speech and manual gesture in deixis is already at play in children, though with more influence of gesture on speech than the reversed pattern.

The third part of the study looks at sensorimotor interactions in the perception of spatial deixis. This experimental study, based on an intermodal priming paradigm, reveals that manual gesture plays a role in the production/perception mechanism associated with the semantic processing of language. These results can be related with those of studies on the sensorimotor nature of representations in the processing of linguistic sound units.

Altogether, these studies provide strong evidence for an integrated representation of speech and manual gestures in the human linguistic brain, even at a relatively early age in its development. They also show that distance encoding is a robust feature, which is also present in all aspects of multimodal pointing.

Keywords : Multimodality, Deixis, Distance encoding, Production/Perception, Development

Multimodalité de la communication langagière humaine : Interaction geste/parole et encodage de la distance dans le pointage

L'objectif de cette thèse est de caractériser les interactions entre parole, geste manuel et langage, via l'étude expérimentale de la deixis spatiale.

Une première étude, sur la production du pointage, nous a permis d'évaluer l'encodage de la distance via les propriétés phonétiques et cinématiques du pointage multimodal. Nos résultats montrent que ces paramètres sont corrélés et sous-tendus par un même comportement moteur (désigner un objet distant implique de produire des gestes, vocaux et manuels, plus larges), qui peut être rapporté à l'encodage phonologique de la distance dans les langues du monde. Nos résultats révèlent par ailleurs une coopération bidirectionnelle entre les modalités vocales et manuelles, qui se répartissent la charge informative lorsqu'elles sont combinées.

Une seconde étude, sur le développement tardif du pointage multimodal entre 6 et 12 ans, nous a permis d'attester d'une évolution progressive du couplage geste/parole : à cette période, la distance est encodée préférentiellement dans les gestes manuels, relativement aux gestes vocaux, et l'usage coopératif des gestes et de la parole est déjà à l'œuvre, avec néanmoins plus d'influence du geste sur la parole que l'inverse.

Une dernière étude, sur les interactions sensorimotrices dans la perception de la deixis spatiale, a montré que le geste manuel joue un rôle dans le mécanisme de perception/production des unités sémantiques du langage. Ces données sont mises en parallèle avec les études sur la nature des représentations mises en jeu dans le traitement des unités sonores du langage.

Mots-clés : Multimodalité, Deixis, Encodage de la distance, Production/Perception, Développement

Multimodality of linguistic communication : Gesture/speech interaction and distance encoding in pointing tasks

The aim of this thesis is to characterize speech, manual gesture and language interactions during pointing tasks.

The first part of the study is about the production of pointing. The original aspect of this work is to look for distance encoding in the lexical, phonetic and kinematic properties of multimodal pointing. Our results show that these parameters can be related with each other, and underlain by a similar motor behaviour (designating a distant object induces larger gestures, be they vocal or manual), that can be related with the phonological pattern that is used for distance encoding in the world's languages. Our results also reveal that the vocal and manual modalities act in bidirectional cooperation for deixis, sharing the informational load when used together.

The second part of the study explores the development of multimodal pointing in 6-12 year-old children and attests a progressive evolution of speech/gesture interactions. It reveals that distance is preferentially encoded in manual gestures, rather than in vocal gestures. It also shows that the cooperative use of speech and gesture in deixis is already at play in children, though with more influence of gesture on speech than the reversed pattern.

The third part of the study looks at sensorimotor interactions in the perception of spatial deixis. It reveals that manual gesture plays a role in the production/perception mechanism associated with the semantic processing of language. These results can be related with those of studies on the nature of representations in the processing of linguistic sound units.

Keywords : Multimodality, Deixis, Distance encoding, Production/Perception, Development
