La Langue Française Parlée Complétée: Production et Perception

Virginie Attina

INSTITUT NATIONAL POLYTECHNIQUE DE GRENOBLE

T H E S E

pour obtenir le grade de

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

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

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

présentée et soutenue publiquement par

Virginie Attina Dubesset

le 25 novembre 2005

La Langue française Parlée Complétée (LPC) : Production et Perception

———

Directeurs de thèse : Denis Beautemps, Marie-Agnès Cathiard, Jean-Luc Schwartz

———

JURY

P.Y. Coulon	Président
J.-F. P. Bonnot	Rapporteur
S. Gibet	Rapporteur
J. Leybaert	Examinateur
J.-L. Schwartz	Directeur de thèse
D. Beautemps	Co-directeur de thèse
M.-A. Cathiard	Co-directeur de thèse

INSTITUT NATIONAL POLYTECHNIQUE DE GRENOBLE

T H E S E

pour obtenir le grade de
DOCTEUR DE L'INPG, Spécialité : « Sciences Cognitives »

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

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

présentée et soutenue publiquement par

Virginie Attina Dubesset

le 25 novembre 2005

La Langue française Parlée Complétée (LPC) : Production et Perception

———

Thèse réalisée sous la direction de
Denis Beautemps, Marie-Agnès Cathiard
et Jean-Luc Schwartz

———

JURY

P.Y. Coulon	Président
J.-F. P. Bonnot	Rapporteur
S. Gibet	Rapporteur
J. Leybaert	Examinateur
J.-L. Schwartz	Directeur de thèse
D. Beautemps	Co-directeur de thèse
M.-A. Cathiard	Co-directeur de thèse

Remerciements

Mes remerciements s'adressent en premier lieu à Marie-Agnès Cathiard et à Denis Beautemps, mes « co-directeurs » de thèse, pour avoir encadré cette thèse. Merci de votre aide, de votre disponibilité et de vos encouragements tout au long de ces années. Merci de m'avoir formée et de m'avoir donné l'envie de continuer dans cette voie. Je te remercie, Denis, de m'avoir toujours laissé une grande liberté dans mon travail, me permettant ainsi d'acquérir l'autonomie nécessaire à tout travail de recherche. Merci pour ton honnêteté scientifique et tes enseignements en Matlab et en statistiques. Marie, je te remercie très sincèrement pour ta motivation, ton implication et ta confiance, sans lesquelles je ne serai pas arrivée au bout de cette thèse. Je te dois la grande majorité des connaissances en parole que j'ai acquises. Merci pour ton écoute attentive, nos discussions, ta relecture minutieuse, tes conseils et tes encouragements ; merci d'avoir toujours su trouver les bons mots quand j'en avais besoin. Merci simplement de m'avoir transmis le goût de la recherche ; tu resteras dans mon cœur ma directrice de thèse.

J'adresse mes plus sincères remerciements aux membres du jury pour avoir accepté d'évaluer ce travail. A M. Jean-François Bonnot et Mme Sylvie Gibet, pour l'honneur qu'ils me font d'être les rapporteurs de cette thèse. Un grand merci également à Mme Jacqueline Leybaert, pour votre appui à notre recherche, et à M. Pierre-Yves Coulon, pour avoir accepté de faire partie de ce jury.

Je remercie vivement Jean-Luc Schwartz d'avoir accepté de diriger officiellement cette thèse. Merci d'avoir toujours soutenu nos travaux et d'avoir su me garder une place dans un planning déjà bien chargé de directeur de labo, chercheur et directeur d'autres thèses. Un grand merci également pour ta relecture et tes propositions pertinentes pour améliorer ce manuscrit.

Merci à Pierre Escudier, directeur du laboratoire, au moment de mon entrée à l'ICP, mais également directeur du DEA Sciences cognitives que j'ai suivi avec beaucoup d'intérêt. Merci de m'avoir donné l'envie d'effectuer mes recherches à l'ICP.

Un remerciement particulier à Christian Abry, qui représente pour moi une grande source d'inspiration. Merci d'avoir partagé vos connaissances et votre savoir visionnaire et d'avoir accepté de relire des parties de ce manuscrit. Merci pour vos idées géniales et pour l'intérêt que vous avez porté à notre recherche. Merci de m'avoir ouvert l'esprit et de m'avoir donné l'envie d'explorer de nouveaux horizons.

Merci à Martine Marthouret, orthophoniste au CHU de Grenoble, pour ses conseils et ses discussions autour du LPC. Merci de m'avoir permis d'établir un réel contact avec la réalité de la surdité et de m'avoir initiée au codage LPC. Merci également à Anne-Marie Fluttaz, qui m'a appris les bases du codage. Merci à

l'association ALPC et localement à l'ADIDA, pour tout ce que vous faites pour promouvoir ce fantastique système.

Un grand merci à tous mes « sujets » : aux codeuses, Glawdys Brunel, Evelyne Huriez, Anne Magnin, Sabine Chevalier et Roseline Vannier, qui ont subi les contraintes de l'enregistrement et sans qui je n'aurais pu obtenir ces résultats. Merci à tous les enfants et adultes qui ont passé mes tests de perception, à Richard Nomballais, pour son aide précieuse pour trouver d'autres sujets décodeurs, mais aussi aux parents qui ont toujours manifesté un grand intérêt à l'égard de mes travaux.

Je tiens à remercier également Delphine Alloatti et Florence Bouaouni, qui ont, par leur stage, participé à l'avancement de cette thèse. Merci également pour vos encouragements. Merci à Matthias Odisio pour s'être intéressé au code LPC et s'être impliqué dans nos recherches.

Un grand merci à toutes les personnes du laboratoire, qui ont participé de près ou de loin à l'aboutissement de cette thèse. Merci à tous les chercheurs de l'ICP et d'ailleurs, aux rencontres de congrès, pour leur partage de la science et les nombreuses discussions. En particulier à Pierre Badin pour sa rigueur dans les démarches mais aussi pour sa disponibilité et sa patience à mon égard : merci de m'avoir permis de récolter un maximum de bibliographie ; à Pascal Perrier pour ses explications en traitement du signal et ses réponses à mes nombreuses questions ; à Xavier Pelorson pour tous ses conseils mais aussi pour son sens de l'humour et ses bonbons et à Willy Serniclaes, pour sa gentillesse et son aide en statistiques. A tous ceux qui m'ont aidé dans mes recherches bibliographiques : Marie, Christian A., Pierre, Virginie D., Jean-Luc, Pascal P., Aude, Pauline, Francesca, Chloé et bien d'autres. Merci à toute l'équipe administrative et technique. Je tiens à remercier en particulier, Alain Arnal et Christophe Savariaux pour leur expertise et leur présence technique à tous nos enregistrements ; Nino Medves pour sa disponibilité et sa présence, ses « dépannages » à toute heure et sa bonne humeur ; et Christian Bulfone pour m'avoir toujours éclairée lors de mes interrogations « informatiques », et pour ses conseils et ses apports théoriques qui m'ont permis de préciser les cours que je donne. Pour leur gentillesse et leur aide administrative, merci à Nadine Bioud, à Dominique Vuillet, à Marie-Thérèse Delfarguiel, à Razika Hammache, secrétaire du Master Sciences Cognitives, à Annie Caillat, à Marie-Jeanne Deshayes et à Isabelle Raffin. Un grand merci à Mme Gaude, qui a toujours su faciliter mes démarches administratives, qui a été une oreille attentive et une parole très réconfortante.

Merci à tous mes collègues du Département d'Informatique Pédagogique pour leurs encouragements, en particulier durant cette période critique qu'est la rédaction, et pour avoir partagé avec moi leur passion de l'enseignement. Un grand merci à tous les jeunes chercheurs, thésards et post-doctorants, mais aussi aux nombreux stagiaires, pour l'ambiance et la convivialité qu'ils ont su établir dans nos locaux. Merci en particulier à tous ceux qui ont partagé avec moi de bons moments, discussions et rigolades ; je sais que vous vous reconnaîtrez. Merci à tous mes amis, connaissances et proches pour tous leurs encouragements.

Un merci spécial pour Annemie qui m'a supportée dans beaucoup de sens du terme (et ce n'était pas toujours facile…), merci pour ton amitié, pour ton humour dépoilant, pour ton accent pas belge, pour ton écoute et tes encouragements incessants. Een dikke dank je wel voor het gedeelde dagdagelijkse lief en leed in "den bureau" en voor het gegeven vertrouwen.

Un grand merci à toute ma famille et belle-famille qui m'a toujours encouragée. Merci à mes parents, qui ont toujours cru en moi, pour leur confiance, leur fierté et leur amour. Grazie per tutto l'affetto che mi avete dato e che mi ha permesso di diventare quella che sono.

Enfin, c'est à toi, Pascal, que j'adresse mes derniers remerciements et que je dédie cette thèse. C'est bien à toi que je la dois pour m'avoir toujours encouragée et soutenue. Merci pour ton amour et ta patience, pour tes paroles réconfortantes et ton bonheur de me voir réussir. Merci d'avoir supporté les contraintes d'une fin de thèse laborieuse. Merci simplement d'avoir vécu cette thèse avec moi.

Cette thèse a bénéficié d'une bourse BDI du CNRS et d'un financement ATER à l'Université Stendhal.

Résumé

La *LPC* ou *Cued Speech* est un augment manuel qui permet au sourd de désambiguïser l'information phonologique visible sur le visage. L'efficacité de ce système pour l'acquisition de la phonologie de la langue est bien établie. Mais la production du code LPC n'avait jamais été étudiée, et nous l'avons fait par une technique de suivi des mouvements labiaux et manuels de quatre codeuses professionnelles. Notre résultat comportemental majeur est que le geste de la main – contre toute attente – *précède* le geste des lèvres. Cette anticipation donne un rôle inattendu à la parole visible : celui de venir désambiguïser le geste manuel, conçu au départ pour désambiguïser la parole… Notre hypothèse est que le système de Cornett a été recodé en termes neuralement compatibles pour le contrôle des gestes des voyelles et des consonnes dans la LPC et la parole. Ainsi le contrôle des *contacts* vocaliques manuels va se trouver en phase avec celui des contacts consonantiques visibles. Ce phasage est assez précis pour que, quelles que soient les variations de la durée de la production de la syllabe CV, l'aboutissement de la détente (*stroke*) du système main-bras se produise dans la phase de tenue de l'attaque consonantique. L'incorporation de la main et de la face dans un espace de contrôle neural commun peut être ainsi pleinement réalisée dans la LPC.

Mots-clés : code LPC, Cued Speech, surdité, production de parole, coordination gestes manuel et articulatoire, perception/intégration

Abstract

Cued Speech is a manual method that allows deaf people to disambiguate the phonological information through the visual channel. Its efficiency for phonological speech acquisition is well established, but a study of Cued Speech production is lacking. Therefore the aim of this work is to investigate the temporal organization of French Cued Speech production on four experienced cuers. Our main result is that the hand gesture anticipates the lip gesture. It is hypothesized that the found pattern of coordination results from the neural compatibility between movement control of consonants and vowels in Cued Speech and visible speech. Thus the vocalic manual contact control is in-phase with the consonantal contact control of visible speech. This phasing is maintained regardless the variability of the syllable duration. This way, in Cued Speech, hand and face are completely incorporated in a common neural control space.

Key-words: Cued Speech, deafness, speech production, manual and articulatory gestures coordination, speech perception/integration

Sommaire

INTRODUCTION GENERALE....1

PARTIE I LA LANGUE FRANÇAISE PARLEE COMPLETEE, UN SYSTEME GESTUEL DE COMMUNICATION COORDONNE AVEC LA PAROLE AUDIO-VISUELLE....3

CHAPITRE I. La vision pour percevoir la parole : « Quand les lèvres ont besoin d'un coup de main… » 5

I.1. La parole audio-visuelle 6
I.1.1. Le rôle de la vision 6
I.1.2. Les modèles d'intégration 9
I.2. La lecture labiale 12
I.2.1. Une compétence limitée 12
I.2.2. Les sosies labiaux....13
I.2.3. Les effets coarticulatoires 14
I.2.4. Autres facteurs....15
I.2.5. Implications chez les déficients auditifs 15
I.3. Le Cued Speech 16
I.3.1. Définition et principes de construction 16
I.3.2. Un système syllabique 18
I.3.3. Organisation du geste....19
I.4. Adaptation du CS au Français : La Langue française Parlée Complétée....21
I.5. Efficacité perceptive du code manuel....22
I.5.1. Réception de syllabes et de mots 22
I.5.2. Réception de parole conversationnelle et de phrases complexes 22
I.5.3. Influence du temps d'exposition aux clés....22
I.6. Représentations phonologiques et développement du langage....22
I.7. Intégration des informations manuelles et labiales....22
I.8. Que nous apprennent les technologies innovantes sur la coordination LPC-parole ?....22
I.8.1. L'Autocuer : un système avant-gardiste....22
I.8.2. Les « kinèmes » de Vilaclara : de la reconnaissance de parole pour les sourds 22
I.8.3. De l'utilité d'un système de reconnaissance de parole efficace pour transmettre des clés automatiques....22
I.8.4. Un générateur automatique de Cued Speech développé au MIT 22
I.8.4.1. Etudes de simulation : de l'importance de la ressemblance au Cued Speech manuel 22
I.8.4.2. Le codeur automatique de Cued Speech temps réel 22

CHAPITRE II. La parole, une structure co-articulée 22

II.1. La coarticulation 22
II.2. La production de la parole coarticulée....22
II.2.1. Le modèle de coarticulation d'Öhman 22
II.2.2. Syllabes et segments : les contrôles de la parole....22
II.3. Les modèles d'anticipation labiale....22
II.3.1. Trois modèles de production pour l'anticipation....22
II.3.2. Le Modèle d'Expansion du Mouvement (M.E.M.) 22
II.3.2.1. Le MEM de protrusion 22
II.3.2.2. Le MEM de constriction 22

CHAPITRE III. Gestes et parole 22

III.1. Une communication multimodale ... 22
III.1.1. Les emblèmes ... 22
III.1.2. Les gestes co-verbaux ... 22
III.1.3. Les fonctions du geste ... 22
III.2. Interdépendance du geste et de la parole ... 22
III.2.1. Les différentes phases du geste ... 22
III.2.2. Anticipation et synchronisation ... 22
III.2.2.1. Des règles de synchronisation 22
III.2.2.2. Timing du début de geste 22
III.2.2.3. Timing du stroke 22
III.2.2.4. Variabilité 22
III.2.3. Coordination du geste et de la parole ... 22
III.2.4. Quand gestes et parole sont étroitement liés ... 22
III.2.4.1. Formulette d'incantation 22
III.2.4.2. La gestualité dans les enfantines 22
CHAPITRE IV. Modélisation de la production de parole et de gestes 22
IV.1. Un modèle psycholinguistique de production de mots : le modèle « Speaking » ... 22
IV.1.1. La conceptualisation ... 22
IV.1.2. La sélection lexicale ... 22
IV.1.3. L'encodage phonologique ... 22
IV.1.4. L'encodage phonétique ... 22
IV.1.5. L'articulation ... 22
IV.1.6. La coopération temporelle de ces modules ... 22
IV.2. Un modèle pour les gestes et la parole : le Sketch Model ... 22
IV.2.1. La conception du sketch ... 22
IV.2.2. Le planificateur de gestes ... 22
IV.3. Coordination temporelle des deux modèles ... 22
PARTIE II PARTIE EXPÉRIMENTALE ... 22
CHAPITRE V. Etude pilote de la production de syllabes codées 22
V.1. La locutrice-codeuse : GB ... 22
V.2. Expérience 1 : étude des transitions manuelles avec configuration consonantique fixe ... 22
V.2.1. Corpus d'étude ... 22
V.2.2. Acquisition des données ... 22
V.2.2.1. Description du matériel 22
V.2.2.2. Traitement des données 22
V.2.2.3. Caractéristiques de production 22
V.2.3. Résultats ... 22
V.2.3.1. Séquences VCV 22
V.2.3.2. Séquences V-V 22
V.2.4. Vers un schéma de coordination « position-lèvres-son » ... 22
V.2.5. En résumé ... 22
V.3. Expérience 2 : formation des clés manuelles ... 22
V.3.1. Corpus d'étude ... 22
V.3.1.1. Changements de configurations consonantiques 22
V.3.1.2. Changements de configuration consonantique avec transitions manuelles 22
V.3.2. Acquisition des données ... 22
V.3.2.1. Description du matériel 22
V.3.2.2. Traitement des données 22
V.3.2.3. Caractéristiques de production 22

V.3.3. Résultats 22
V.3.3.1. Coordination manuelle et orofaciale à positions vocaliques manuelles fixes 22
V.3.3.2. Coordination manuelle et orofaciale avec changement de positions vocaliques 22
V.3.4. Vers un schéma de coordination « position-clé-lèvres-son » 22
V.3.5. En résumé 22
V.4. Conclusion sur les deux expériences 22
V.4.1. Comparaison normalisée 22
V.4.2. Un schéma de coordination temporelle main-lèvres-son 22
CHAPITRE VI. Expérience 3 : généralisation à d'autres locuteurs-codeurs sur des séquences syllabiques CV 22
VI.1. Méthode 22
VI.1.1. Sujets 22
VI.1.2. Corpus d'étude 22
VI.1.3. Acquisition des données et traitement 22
VI.2. Résultats 22
VI.2.1. Schéma temporel de coordination pour les trois codeuses 22
VI.2.1.1. Organisation temporelle main-lèvres-son 22
VI.2.1.2. Comparaison interlocuteur 22
VI.2.2. Evolution des relations LPC-parole dans le domaine syllabique 22
VI.2.2.1. Geste manuel 22
VI.2.2.2. Coordination main-lèvres-son dans le domaine syllabique : à la recherche des invariants 22
VI.2.3. Variance et variabilité régulière (*lawful variability* ou invariance) 22
VI.3. Conclusion et discussion générale sur la production de syllabes CV codées 22
VI.3.1. Schéma général de coordination temporelle LPC-parole 22
VI.3.2. L'anticipation, résultat de la compatibilité des contrôles LPC-parole 22
CHAPITRE VII. Etude de l'anticipation coarticulatoire chez la codeuse GB 22
VII.1. Expérience 4 : anticipation vocalique d'arrondissement et de hauteur au travers d'une pause prosodique 22
VII.1.1. Corpus 22
VII.1.2. Traitement des données 22
VII.1.3. Résultats 22
VII.1.3.1. Anticipation d'arrondissement 22
VII.1.3.2. Anticipation de hauteur 22
VII.1.4. Conclusion 22
VII.2. Expérience 5 : anticipation vocalique d'arrondissement au travers d'une séquence consonantique 22
VII.2.1. Corpus 22
VII.2.2. Traitement des données 22
VII.2.3. Résultats 22
VII.2.4. Conclusion 22
VII.3. Conclusion générale sur l'anticipation coarticulatoire 22
CHAPITRE VIII. Expérience 6 : perception de syllabes CV codées 22
VIII.1. Corpus 22
VIII.2. Etude préliminaire : coordination main-lèvres-son 22
VIII.2.1. Acquisition et traitement des données 22
VIII.2.2. Résultats et conclusion 22
VIII.3. Etude perceptive 22
VIII.3.1. Montage et passation du test 22

VIII.3.1.1. Création des stimuli 22
VIII.3.1.2. Sujets et procédurẽ 22
VIII.3.2. Résultats 22
VIII.3.3. Discussion 22
VIII.3.3.1. Le geste de la main est bien perçu en avance des lèvres 22
VIII.3.3.2. Sujets « précoces » versus « tardifs » 22
VIII.3.4. Conclusion 22
DISCUSSION GENERALE ET CONCLUSION 22
Vers un modèle de génération de parole codée 22
Vers un modèle d'intégration parole-LPC 22
Amélioration des technologies cognitives du handicap 22
La LPC : un codage phonologique incorporé 22
REFERENCES BIBLIOGRAPHIQUES 22
LISTE DES PUBLICATIONS 22
TABLE DES FIGURES 22
INDEX DES TABLEAUX 22
TABLE DES ANNEXES 22
ANNEXES 22

Introduction générale

« *Deux ou trois sens valent mieux qu'un* », tel pourrait être l'adage de la parole car contrairement à ce que l'on pourrait penser, elle est par nature multimodale et multisensorielle (Schwartz et al., 2002a). Tous les moyens sont bons *pour récupérer les intentions communicatives* de celui qui parle : l'audition, la vision, qui représente bien plus qu'un simple relais, mais aussi le toucher. Lorsque l'un de nos sens vient à nous faire défaut, de véritables stratégies palliatives se mettent en place. Ainsi, pour les sourds, la vision est primordiale : elle leur permet de réaliser une *lecture* de ce que dit le locuteur en décodant ses mouvements labiaux. C'est de cette manière que les sourds ont accès au message linguistique oral, même s'il reste malgré tout ambigu.

La Langue française Parlée Complétée (*Cued Speech* ou code LPC) a été inventée par le docteur Cornett (1967) comme une méthode de codage phonologique manuel pour désambiguïser les syllabes de la parole. Cette invention s'est révélée être un atout incontestable pour les sourds, permettant l'acquisition et le développement de compétences linguistiques, similaires à celles des enfants entendants. Nous disposons à l'heure actuelle de nombreuses données en faveur de cette méthode « artificielle ». Mais la problématique de l'intégration – aussi bien perceptive que productive – de ce code par les sourds n'est pas encore résolue. Comment procèdent les systèmes cognitifs pour encoder et décoder la parole visible (simultanément ?) sur les lèvres et la main ? Très peu de choses sont en fait connues sur la façon dont la LPC est produite par les codeurs, puisque jusqu'à présent aucune étude n'a réellement été menée en production. Cette thèse se propose donc d'étudier la coordination manuelle et orofaciale dans la production de ce système par des codeuses LPC expérimentées. Ces études nous permettront de mettre à jour les relations que la main entretient avec la parole, en production, mais aussi en perception. C'est vers une théorie de la *Perception pour le Contrôle de l'Action* (PACT, Schwartz et al., 2002b), développée à l'Institut de la Communication Parlée, faisant le lien dans la parole entre la perception, le contrôle de l'action et la phonologie, que nous nous orienterons comme cadre de travail (*framework*). Selon cette théorie, le sujet percevant forme des représentations sensori-motrices à partir de la récupération multimodale des gestes de parole de son interlocuteur et ces représentations contraignent son système phonologique et lui permettent de contrôler ses propres actions. C'est donc vers une hypothèse sur les contrôles de la coproduction parole-LPC que nos résultats nous mèneront. Plus spécifiquement, la nature spécifique des constituants de la syllabe en parole sera pour nous un véritable point d'ancrage, et nous permettra de

nous interroger en particulier sur la planification, la programmation et l'éxécution de la LPC chez les codeurs expérimentés et sur les mécanismes d'intégration de ce code chez les sourds LPC.

Dans cette thèse, nous présenterons dans un premier temps plus en détail le cadre général dans lequel s'insère cette étude : le processus de communication parlée. Nous ferons une brève revue sur la parole audio-visuelle, en insistant sur la modalité visuelle qui constitue l'accès principal à la parole pour les sourds afin de mieux comprendre l'intérêt du code LPC. Nous donnerons à la suite un état des connaissances sur cette méthode, en explorant son efficacité perceptive, ses répercussions pour le langage et l'intérêt qu'elle suscite au niveau des technologies du handicap (CHAPITRE I). Cette revue de la littérature nous permettra de rendre compte du fait qu'il n'existe à notre connaissance aucune étude consacrée à la production de ce code. Production et perception étant intrinsèquement liées dans les activités humaines, nous nous proposerons d'étudier la coordination de la main et de la parole en LPC. Nous donnerons un aperçu des connaissances sur la parole seule en nous focalisant particulièrement sur la complexité de cette activité *coarticulée* (CHAPITRE II) et nous présenterons également l'étude des gestes accompagnant naturellement la parole, les gestes *co-verbaux*, et la façon dont ils sont produits avec la parole (CHAPITRE III) ; cela nous permettra de présenter un modèle psycholinguistique pour la production de parole et de gestes, le modèle *Speaking* augmenté du *Sketch model* (CHAPITRE IV). Nous aborderons en seconde partie nos études expérimentales. La première expérience proposera un patron temporel de coordination main-lèvres-son, avec étude des positions et des configurations de main en LPC, que nous aurons mis en évidence chez une codeuse diplômée pour la production de syllabes de type Consonne-Voyelle, l'unité du code LPC (CHAPITRE V). L'étude de trois autres codeuses nous permettra de confirmer ce patron de coordination et de proposer des règles générales sur la façon dont est produit ce code (CHAPITRE VI). Nous aborderons également le cas particulier en parole de l'anticipation coarticulatoire en étudiant précisément comment la main se coordonne avec les lèvres dans un tel contexte (CHAPITRE VII). Nous proposerons alors une étude sur la perception de syllabes CV codées en LPC dans le cadre d'un paradigme de *gating* visuel (CHAPITRE VIII). Enfin, c'est selon trois perspectives différentes que nous discuterons de nos résultats et de leurs implications : la modélisation psycholinguistique de la production de la parole codée, la perception de ce code en insistant particulièrement sur les mécanismes d'intégration, et enfin l'intérêt de nos recherches pour le développement des technologies cognitives du handicap.

Partie I
la Langue française Parlée Complétée, un système gestuel de communication coordonné avec la parole audio-visuelle

« […] Our perception and understanding are influenced by a speaker's face and accompanying gestures, as well as by the actual sound of the speech »
Massaro, 2001

Nous présentons dans cette première partie le caractère complètement multimodal de la parole, afin de mieux appréhender le cadre général dans lequel s'inscrit la Langue française Parlée Complétée (LPC), c'est-à-dire un système « artificiel » gestuel en étroite relation avec la parole et qui va permettre de transmettre un message oral saisi dans son intégralité par la vision uniquement. La parole est par nature bimodale pour les entendants qui utilisent aussi bien l'audition et la vision pour la percevoir. Nous disposons d'une vaste littérature sur le sujet : la vision des gestes de parole interagit avec les sons que nous percevons durant la perception de parole. Nous n'en donnerons qu'un aperçu. Les mécanismes d'intégration de ces deux modalités ne sont pas encore exactement déterminés mais la richesse des données sur la parole audiovisuelle nous donne de nombreuses pistes. Pour les sourds, pour qui l'audition est déficiente, la lecture labiale constitue l'accès principal au langage parlé. Cependant, la lecture labiale seule est ambiguë (une même forme aux lèvres peut correspondre à plusieurs phonèmes). C'est dans ce contexte que le code LPC a été inventé, pour lever cette ambiguïté et transmettre un message linguistique précis. Si de nombreux travaux ont validé l'intérêt de ce code dans la perception de la parole, très peu de travaux se sont attelés à l'étude de sa production. Cette thèse présente une contribution à ce domaine en dévoilant l'organisation temporelle de ce code dans sa coproduction avec la parole. C'est pourquoi nous exposerons plus en détails comment la parole, cette structure coarticulée, est produite, en gardant toujours un lien avec sa perception. Puis nous présenterons une revue des travaux sur les gestes coverbaux qui nous donnera un aperçu des relations naturelles gestes-parole.

CHAPITRE I.

La vision pour percevoir la parole :

« Quand les lèvres ont besoin d’un coup de main… »

I.1. La parole audio-visuelle

En dehors des études très anciennes sur la lecture labiale utilisée par des sourds (depuis Bell, 1895) et de celles sur le rôle de la vision dans le bruit (depuis Sumby & Pollack, 1954), les recherches sur les traitements multimodaux – en particulier audiovisuels – de la parole se sont nettement développés ces vingt dernières années (tout particulièrement depuis Massaro, 1987a, et Dodd & Campbell, 1987). Depuis 1997, ce domaine de la perception audiovisuelle donne régulièrement lieu, tous les deux ans, aux rencontres *AVSP* « *Auditory Visual Speech Processing* ». Toutes ces recherches mettent en exergue la multisensorialité de la parole (pour une vision générale, voir en particulier l'intervention de J.-L. Schwartz durant les dernières *Journées d'Etude sur la Parole*, *2004*). Nous verrons, dans un premier temps, que la composante visuelle donnée par la lecture labiale joue un rôle important dans la perception de la parole. En effet, combinée à l'information auditive, elle constitue une aide non négligeable quand les conditions auditives sont dégradées (soit par du bruit soit par une surdité) ou simplement dans le cas de compréhension de textes difficiles. Nous évoquerons aussi l'illusion McGurk qui révèle que la composante visible de la parole ne peut être ignorée alors que le signal audio est pourtant parfaitement clair. Enfin, nous tenterons de comprendre la perception multimodale de la parole dans le cadre des modèles de fusion des flux audio-visuels proposés pour leur intégration.

I.1.1. Le rôle de la vision

Il est maintenant bien établi que la perception de la parole n'est pas seulement auditive mais complètement multimodale (Schwartz et al., 2002a, 2004). Pour les sourds, c'est la vision qui constitue l'entrée principale, via la lecture labiale, pour la perception de la parole (pour une revue, voir Bernstein et al., 1998). Nous verrons plus en détails les problèmes que cela implique (section I.2).

Mais la modalité visuelle est aussi naturellement utilisée par les bien-entendants, et est impliquée dans le module de perception de parole. Nous disposons d'une quantité de données témoignant du rôle important de la vision (pour une revue, voir Summerfield, 1979 ; Massaro, 1987a ; Dodd & Campbell, 1987 ; Cathiard, 1988/1989 ; Cathiard, 1994 ; Robert-Ribes, 1995 ; Schwartz et al., 2002a) et de l'existence de ces interactions audio-visuelles. Nous n'en donnerons qu'un bref aperçu.

Combinée à l'information auditive, la vue du visage du locuteur, en particulier les informations labiales, constitue une aide non négligeable quand les conditions de réception de parole sont dégradées par du bruit ambiant par exemple. De nombreuses données expérimentales ont en effet mis en évidence que la vue des informations faciales pouvait augmenter significativement le taux d'intelligibilité de la parole dans le bruit, et cela pour des entendants non entraînés à la lecture labiale (Sumby & Pollack, 1954 ;

Erber, 1969 ; MacLeod & Summerfield, 1987 ; et pour le français, Benoît et al., 1994). Sumby et Pollack (1954) ont mesuré le gain en intelligibilité apporté par la vision sur l'audio lors de la perception de la parole orale, en fonction d'un rapport signal-sur-bruit (RSB) variant de 0 à -30 dB, et du nombre de mots (bisyllabiques) à identifier. Ils montrent que le taux d'intelligibilité décroît quand le rapport signal sur bruit décroît et/ou que le nombre de stimuli augmente ; ce qui signifie que plus le signal acoustique est dégradé, plus il est difficile de percevoir correctement la parole. En ce qui concerne l'apport de la vision dans la condition audio-visuelle (par rapport à la condition audio seule), celui-ci s'avère être plus important quand le rapport signal-sur-bruit diminue. En effet, quand le signal acoustique est peu dégradé (RSB à 0 dB), la vue du locuteur n'améliore pas les scores d'intelligibilité (qui sont pratiquement à 100% en audio seul). En revanche, plus le signal acoustique est dégradé, plus la vision est utilisée : à un RSB de -30 dB, la vision dans la condition audio-visuelle permet d'identifier de 40% à 80% de mots en plus (selon le nombre de mots) qu'en condition audio seule. MacLeod et Summerfield (1987) estiment un gain moyen de 11 dB apporté par la lecture labiale : le sujet obtient le même taux moyen d'identification correcte que dans le cas où le signal acoustique est moins dégradé (de 11 dB). Pour le français, Benoît et al. (1994) teste le bénéfice de la lecture labiale (en comparant les scores en condition audio seule et en audio-visuelle) pour des séquences syllabiques sans sens de type [VCVCVz] avec V= [a, i, y] et C= [b, v, z, ʒ, ʁ, l] (par exemple, [ababaz]), dans cinq conditions de RSB (variant de -24 dB à 0 dB par pas de 6 dB). En condition audio, les scores d'identification correcte des syllabes passent de près de 90% (pour un RSB de 0 dB) à pratiquement 0% pour un RSB de -24 dB. En condition audio-visuelle en revanche, ces scores s'élèvent encore à 65%, ce qui est comparable à une condition purement visuelle. Ainsi la vision permet incontestablement d'améliorer l'intelligibilité de la parole quand les conditions auditives sont dégradées. En outre, il apparaît que les deux modalités sont « complémentaires ». Pour les consonnes, le lieu d'articulation est mieux récupéré par la vision que par l'audition (quand le signal est bruité) et inversement pour le mode (Summerfield, 1987). Pour les voyelles, on trouve également une certaine complémentarité. Benoît et al. (1994) l'avaient clairement observée pour la voyelle arrondie-protruse [y] en particulier : parmi les voyelles [a, i, y], [y] était la mieux identifiée en vision et la moins bien identifiée en audio. Dans une étude plus vaste et systématique, Robert-Ribes et al. (1998 ; voir aussi Robert-Ribes, 1995) ont confirmé cette tendance pour les voyelles du français ([i, e, ɛ, y, ø, œ, u, o, ɔ, a]) : en condition audio seule, le trait de hauteur est mieux perçu que le trait avant-arrière, qui est lui-même mieux perçu que l'arrondissement alors qu'en condition visuelle seule, c'est le trait d'arrondissement qui est le mieux perçu, suivi du trait de hauteur, le trait avant-arrière n'étant pratiquement pas identifié. En outre, il apparaît que l'intégration de ces deux modalités est *synergique* : les scores en audio-visuel sont toujours meilleurs qu'en audio

seul ou en visuel seul (Robert-Ribes et al., 1998 ; voir aussi, Schwartz et al., 2002a). Ces résultats amènent les auteurs à qualifier la perception audio-visuelle de *doublement optimale* : « Ainsi, il semble y avoir ce que l'on pourrait appeler une « double optimalité » de la perception audio-visuelle, aussi bien au niveau de l'information présente (*complémentarité*) que du traitement de cette information (*synergie*). » (Schwartz et al., 2002a, p. 146).

Il ne faut pas croire que la vision vient en relais de l'audition seulement quand celle-ci est perturbée par du bruit. La supériorité des informations audio-visuelles sur celles fournies par les deux modalités seules en est la preuve. Même dans des conditions où le signal auditif est clair, la vue du locuteur apporte des informations. Ceci a été clairement mis en évidence par Reisberg et al. (1987) pour des messages difficiles à comprendre et pourtant faciles à entendre (« easy to hear but hard to understand »). Dans le cadre de quatre expériences, ces auteurs ont testé la contribution de la vision en utilisant un paradigme de *shadowing* : les sujets devaient répéter ce qu'ils entendaient en condition audio seule et en condition audio-visuelle. Les stimuli étaient prononcés dans une langue étrangère (pour l'expérience 1, la langue testée était le français et les sujets étaient de langue maternelle anglaise, en apprentissage du français et pour l'expérience 2, la langue testée était l'allemand et les sujets de langue maternelle anglaise et en apprentissage de l'allemand), ou bien consistaient en des messages produits dans la langue maternelle des sujets mais prononcés avec un accent étranger (le locuteur avait un accent belge), ou encore les stimuli étaient des textes sémantiquement difficiles à comprendre (extraits de la *Critique de la Raison Pure* de Kant) énoncés par un locuteur de langue maternelle anglaise, sans accent. Les résultats montrent tous un bénéfice significatif de l'ajout de l'information labiale, de 4,2% en moyenne à 21,5%. Ainsi la vision permet de rehausser la perception du message linguistique même quand les conditions d'écoute ne sont pas dégradées (voir aussi, Davis & Kim, 1998, pour une autre tâche de *shadowing* en langue étrangère ; et plus récemment dans notre laboratoire, voir Beautemps et al., 2003, pour une tâche de *close shadowing* chez des sujets français, montrant un bénéfice de la vision sur des temps de réaction).

Finalement, la modalité visuelle a une influence irrépressible sur le son : c'est ce qui est démontré par le célèbre *effet McGurk* (McGurk & MacDonald, 1976 ; MacDonald & McGurk, 1978). Ces auteurs ont présentés à des sujets des stimuli audiovisuels conflictuels, pour lesquels le son ne correspond pas à l'image. La vidéo d'un locuteur articulant une syllabe était montrée aux sujets alors qu'ils entendaient distinctement une autre syllabe (le son était monté sur la vidéo de manière synchrone). Classiquement, les syllabes sont de type [ba] et [ga] combinées audiovisuellement : $[ba]_A$ + $[ba]_V$, $[ga]_A$ + $[ga]_V$, $[ba]_A$ + $[ga]_V$ et $[ga]_A$ + $[ba]_V$ (A correspond au son et V à la vidéo). Sans détailler tous les résultats, ce qui ressort fortement est le phénomène illusoire. Quand l'image de [ga] est présentée simultanément avec

le son de [ba], les sujets perçoivent majoritairement un [da], soit un percept qui ne correspond à aucun des deux stimuli mais à une sorte de compromis : il s'agit d'une illusion dite de *fusion* des deux informations. Dans la condition inverse, avec l'image de [ba] et le son de [ga], les sujets perçoivent un [bga], soit un percept incluant les deux consonnes : il s'agit d'une illusion dite de *combinaison*. Ces effets sont particulièrement forts car même lorsque les sujets connaissent la manipulation des stimuli, ou bien ont pour consigne de se concentrer sur ce qu'ils entendent, ils ne peuvent s'empêcher d'être leurrés (Liberman, 1982 ; Summerfield & McGrath, 1984). Cet *effet McGurk* a été largement observé et répliqué (pour une revue, voir Green, 1998 ; pour le français, Cathiard, 1994 ; Cathiard et al., 2001 ; Colin, 2001 ; Colin & Radeau, 2003). Ces expériences montrent que les informations auditives et visuelles sont toutes deux prises en compte et intégrées lors de la perception bimodale de la parole.

I.1.2. Les modèles d'intégration

Les résultats présentés précédemment révèlent l'existence d'interactions entre les modalités visuelle et auditive. Comment dans le processus de communication les informations issues de ces deux modalités sont-elles intégrées pour l'accès aux représentations ? De nombreux modèles ont été proposés pour rendre compte de cette intégration (entre autres le plus connu, le *Fuzzy-Logical Model of Perception* de Massaro, 1987b). Nous ne passerons pas en revue toutes les propositions, mais nous présenterons très brièvement quatre architectures, proposées par Schwartz et al., (1998, 2002a), adaptées et modifiées des cinq *métriques* de Summerfield (1987), et au carrefour des théories issues de la psychologie cognitive et du domaine du traitement de l'information traitant de la fusion de capteurs dans les processus de décision.

Schwartz et al. (1998, 2002a ; voir aussi Robert-Ribes, 1995) proposent quatre architectures possibles pour la modélisation du processus de fusion des informations auditives et visuelles (voir Figure 1) :

- Le modèle à identification directe (ID) : dans ce modèle, il n'y a pas d'interaction intermédiaire entre les deux flux d'information avant l'accès au code linguistique : il n'y a donc pas de représentation commune aux deux modalités entre l'entrée et le percept identifié. L'étape de classification se fait donc directement sur l'entrée bimodale.
- Le modèle à identification séparée (IS) : il y a deux processus de reconnaissance qui opèrent séparément sur chacune des entrées auditive et visuelle, puis il y a un processus de fusion tardive des deux représentations phonétiques. Notons que cette architecture permet de modéliser le McGurk tel qu'il a pu être expliqué par MacDonald et McGurk (1978 ; voir aussi Summerfield, 1987) par leur hypothèse VPAM (*Vision: Place, Audition: Manner*), qui prédit que l'on récupère

l'information de lieu de la consonne sur les lèvres et l'information de mode par l'audition. Ce modèle IS est aussi à la base du Fuzzy-Logical Model of Perception (FLMP, Massaro, 1987b).

– Le modèle de recodage dans la modalité dominante (RD) : l'audition est supposée être la modalité dominante dans la parole. Dans cette perspective, l'entrée visuelle est recodée sous un format compatible avec celui des représentations auditives, la fonction de transfert du conduit vocal, et la fusion se fait de manière précoce, c'est-à-dire avant l'accès au code.

– Le modèle de recodage dans la modalité motrice (RM) : l'intégration des données bimodales est également précoce dans ce modèle, mais le recodage commun des deux modalités se fait dans la modalité motrice, soit dans un format commun *amodal* représentant la *cause* commune du son et de l'image.

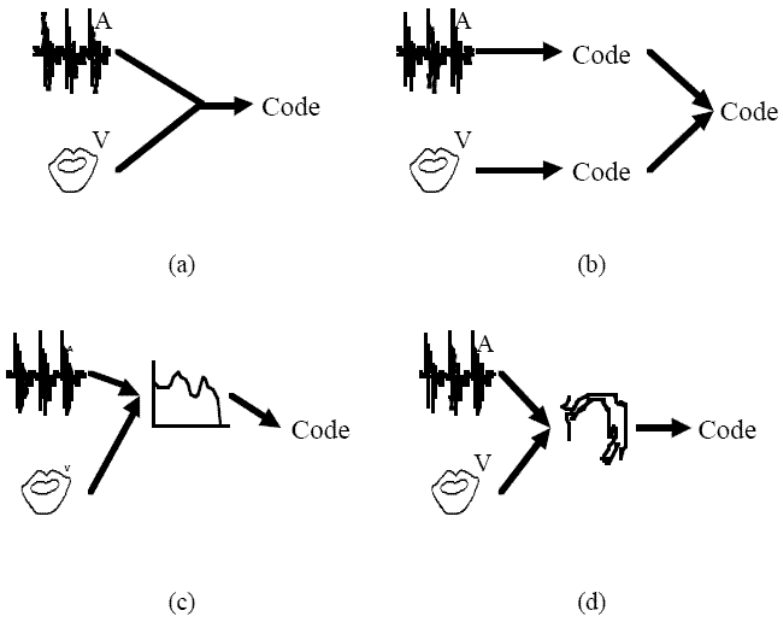


Figure 1. Architectures de base pour la fusion audiovisuelle proposées par Schwartz et al. (1998) : (a) identification directe, (b) identification séparée, (c) recodage dans la modalité dominante et (d) recodage dans la modalité motrice. Figure tirée de Schwartz et al., 2002a.

Les auteurs proposent une taxinomie de ces modèles afin de choisir le modèle adéquat, celui qui rend le mieux compte des données ou celui le plus efficace, selon qu'on se place dans une perspective de modélisation des processus cognitifs en jeu dans la perception audio-visuelle (sciences de la cognition) ou bien de reconnaissance de parole (sciences de l'ingénieur) (voir Figure 2). Trois questions peuvent être posées afin de se guider vers l'une ou l'autre de ces architectures : y a-t-il une représentation commune aux deux modalités ? Si oui, est-elle précoce ou tardive, par rapport à l'accès au code ? Enfin, s'il y a une représentation commune précoce, quel est son format ? Pour de plus amples détails sur cette taxinomie, avec exemples de modèles et adéquations aux données expérimentales, voir

Schwartz et al. (1998, 2002a ; voir aussi, Robert-Ribes, 1995). En comparant les données expérimentales très riches dans le domaine de la perception audio-visuelle de la parole et les prédictions des différents modèles, ces auteurs proposent pour l'intégration de la parole bimodale une architecture qui serait plutôt de type RM (Robert-Ribes, 1995).

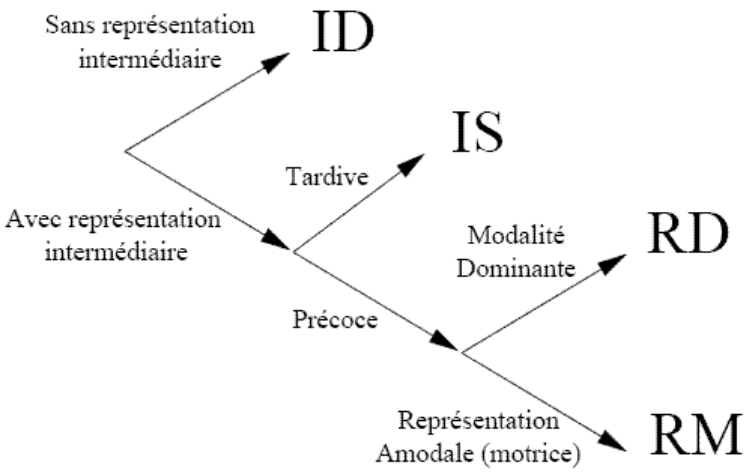


Figure 2. Taxinomie des modèles de fusion. Figure tirée de Schwartz et al., 2002a.

Schwartz et al. (2002b ; pour une version moins étendue en français, voir Schwartz, 2001) proposent un cadre théorique général, la Perception pour le Contrôle de l'Action (PACT) faisant le lien dans la parole entre la perception, le contrôle de l'action et la phonologie. Selon cette théorie, l'auditeur forme des représentations sensori-motrices à partir de la récupération multimodale des gestes de parole de son interlocuteur (c'est-à-dire les composantes de base à contrôler, soit le timing et les cibles) ; ces représentations contraignent son système phonologique et lui permettent de contrôler ses propres actions. Dans ce cadre, les représentations sont intrinsèquement sensori-motrices : « These representations, aiming at recovering and specifying speech controls, are neither pure sensorial patterns, nor pure inferred motor objects, but multimodal percepts constrained by action regularities, or gestures shaped by multimodal perception. » (p. 271).

Dans ce cadre, ils proposent le T*iming-Target model* comme architecture possible, rendant compte des interactions sensori-motrices en partant du fait que la parole est dynamique, audiovisuelle et motrice (Schwartz et al., 1992 ; voir également la proposition d'un modèle *RM dynamique* dans Robert-Ribes, 1995, pp. 202-211). Ce modèle, élaboré à partir du modèle *phasic/tonic* de Chistovich (1980) pour le traitement auditif de la parole, est composé d'un module audio-visuel d'analyse des trajectoires qui permet de récupérer les cibles (module tonique) et d'un module audio-visuel qui permet de récupérer les événements temporels auditifs et visuels, soit le timing (module phasique). Le timing et les cibles donnent la base de la représentation commune perceptuo-motrice où les entrées auditives et visuelles

peuvent être projetées, comparées et fusionnées. Les interactions audio-visuelles s'effectuent séparément dans chacun de ces deux modules, qui sont considérés comme indépendants, mais la fusion dans l'espace commun est précoce (avant la catégorisation) et peut donc tenir compte de rendez-vous temporels entre cibles. Ainsi les contrôles de la parole sont récupérés et ils permettent de former des représentations complètement sensori-motrices.

I.2. La lecture labiale

« Lipreading is a skill that can be developed to useful levels only after one knows the spoken language. It is not effective for learning a language »

Cornett, 1988

Nous revenons maintenant à un cadre plus limité, celui de la lecture labiale, processus directement en jeu dans la perception de la parole par les malentendants. Ses limites nous permettront de comprendre l'utilité d'un système complémentaire, pouvant utiliser la main pour lever l'ambiguïté inhérente à la lecture sur les lèvres, tel que la Langue française Parlée Complétée.

I.2.1. Une compétence limitée

La lecture labiale est une méthode permettant de percevoir la parole en regardant uniquement les mouvements des lèvres et le visage de son interlocuteur (pour une revue, voir Summerfield, 1983).

Utiliser les modalités visuelle et auditive s'avère donc être d'une grande efficacité pour percevoir la parole. Les personnes sourdes qui ne peuvent se fier uniquement à leurs restes auditifs pour communiquer font un grand usage de l'information visuelle. Ils sont « dépendants » de la lecture labiale pour accéder à la langue et percevoir la parole (Erber, 1972 ; Bernstein & Auer, 2003).

Mais savons-nous tous lire sur les lèvres ? La réponse est « oui » sans aucun doute. Lire sur les lèvres est une faculté présente chez tout le monde, sans besoin préalable d'apprentissage. Par contre les compétences d'un individu à l'autre sont très variables : dans une étude de MacLeod et Summerfield (1990), le score de lecture labiale de mots dans des phrases variait de 20% à 70% selon le sujet. Il n'y a pas vraiment de différence démontrée entre entendants et sourds (Owens & Blazek, 1985 ; MacLeod & Summerfield, 1990), même si ceux qui ont les meilleures performances dans le domaine se retrouvent souvent parmi les sourds (Bernstein et al. 1998). Que sommes-nous capables exactement de percevoir ? En moyenne, la lecture labiale seule permet d'appréhender 40 à 60 % des phonèmes d'une langue donnée (Montgomery & Jackson, 1983 ; Owens & Blazek, 1985), et seulement 10 à 30 % des mots (Nicholls & Ling, 1982 ; Bernstein et al., 2000). Certes les compétences à lire sur les lèvres

peuvent être plus ou moins améliorées par l'entraînement (Walden et al., 1977 ; Walden et al., 1981), cependant les meilleures performances en lecture labiale n'atteignent généralement pas la perfection.

I.2.2. Les sosies labiaux

D'où vient donc cette limite de perception ? Directement de l'entrée visuelle, c'est-à-dire de ce qui est fourni par la production de la parole. La parole est en effet « visible » essentiellement au niveau des lèvres à la sortie du conduit vocal. Ce sont donc les différences de formes labiales qui nous permettent de distinguer les phonèmes entre eux. Mais chaque phonème n'est pas caractérisé par une forme labiale unique ; certains phonèmes ont une apparence aux lèvres tellement similaire que la distinction aux lèvres en est rendue très difficile. Ces groupes de phonèmes visuellement identiques et considérés comme un percept unique sont appelés « sosies labiaux » ou « visèmes » (Fisher, 1968). Les phonèmes confondus visuellement entre eux appartiennent au même groupe et sont donc bien distincts des phonèmes de groupes différents. Ainsi les phonèmes /p/, /b/ et /m/ qui sont tous les trois caractérisés par une occlusion bilabiale bien visible sont jugés similaires si l'on s'appuie uniquement sur la vision pour les distinguer et en l'absence de tout contexte : le trait de mode qui les distingue acoustiquement (le voisement pour /b/ et la nasalité pour /m/) ne permet pas de les distinguer visuellement.

Comment se regroupent les sosies labiaux ? C'est à cette question que de nombreuses études, menées principalement sur l'anglais (de Woodward & Barber, 1960 à Owens & Blazek, 1985), ont tenté de répondre en effectuant des tests de perception des voyelles et des consonnes. Pour le Français, Gentil (1981) a mis en évidence quatre groupes de confusions visuelles pour les voyelles en testant des enfants et adolescents sourds : [i], [o-u-y-ɔ-ɔ̃], [ã] et [ɑ-a-ɛ-e-œ̃-ɛ̃]. Il apparaît que pour identifier les voyelles nous nous appuyons sur les caractéristiques articulatoires visibles, la dimension d'arrondissement et la dimension d'aperture ou séparation verticale des lèvres, apparaissant comme les caractéristiques les plus saillantes pour juger de l'identité d'une voyelle (Jackson et al., 1976 ; Montgomery & Jackson, 1983).

Pour les consonnes du français, Gentil (1981) met en évidence trois groupes prédominants de sosies labiaux : les bilabiales [p-b-m], les labiodentales [f-v] et les consonnes [ʃ-ʒ] caractérisées par une protrusion labiale avec éversion. Chacun de ces groupes est caractérisé par une activité articulatoire directement visible au niveau des lèvres, à l'inverse des autres consonnes [k-g-ɲ-ʀ-s-z-t-d-n] dont les changements articulatoires se produisent plus à l'intérieur du conduit vocal. Les regroupements en

différentes classes de visèmes font donc apparaître le lieu d'articulation de la consonne comme distinction visible.

I.2.3. Les effets coarticulatoires

Il apparaît cependant que les regroupements en visèmes des différents phonèmes de la langue, aussi bien pour les consonnes que pour les voyelles, ne semblent pas se résumer à cela. La lecture de la littérature à ce sujet fait ressortir une forte variabilité qui met en contradiction les différentes études (pour une comparaison, voir Owens & Blazek, 1985 ; Jackson, 1988 ; Cathiard, 1988/89). Cette variabilité s'explique principalement par les effets coarticulatoires en jeu dans la production de la parole. Produire de la parole ne revient pas à enchaîner une suite de phonèmes indépendants les uns à la suite des autres, mais chaque son a une influence sur son voisin tant au niveau acoustique qu'au niveau articulatoire. La coarticulation implique tous les articulateurs en jeu dans la production de la parole, mais c'est à la sortie du conduit vocal, au niveau des lèvres que les configurations sont visuellement différentes selon le contexte avoisinant, modifiant ainsi notre perception des différents phonèmes. C'est ce que Gentil (1981) a observé pour le français. Les consonnes associées à la voyelle [a] dans des logatomes de type consonne-voyelle (CV) étaient plus facilement reconnues que lorsqu'elles étaient associées à la voyelle [u]. La production de la voyelle [a] à la suite de la consonne implique une grande aperture aux lèvres qui facilite la distinction visuelle de la consonne et de la voyelle (Erber et al., 1979). En revanche, le geste de protrusion (c'est-à-dire l'avancée des lèvres) et d'arrondissement pour produire le [u], a une forte tendance à masquer la consonne précédente ; les lèvres anticipent ce mouvement de protrusion durant la production de la consonne et de ce fait « masquent » les traits articulatoires caractéristiques de la consonne et la rendent très difficile à reconnaître. De nombreuses études ont mis en évidence l'influence du contexte vocalique sur l'intelligibilité de la consonne (pour l'anglais, Benguérel & Pichora-Fuller, 1982 ; Owens & Blazek, 1985 ; pour le français, Benoît et al., 1994). Il en va de même pour la voyelle ; en effet, il apparaît également que le contexte consonantique peut influencer la perception de la voyelle. Montgomery et al. (1987) ont montré pour des logatomes de type CVC, que l'identification des voyelles en contexte consonantique neutre, du point de vue de la labialité (c'est-à-dire les consonnes postérieures comme [g] par exemple), était meilleure que celles en contexte labialisé (par exemple [p] qui implique une fermeture bilabiale). Benoît et al. (1994) montrent pour le français que ce sont surtout les consonnes impliquant une forte protrusion des lèvres (comme [ʒ]) qui affectent l'identification des voyelles (voir aussi, Cathiard, 1988/89 ; Tseva & Cathiard, 1990).

I.2.4. Autres facteurs

D'autres variables ont également leur rôle à jouer. Les différences entre les études ont été expliquées par les différences de stimuli qui impliquaient des effets coarticulatoires différents, par les différences dans la méthodologie employée notamment pour les critères statistiques utilisés pour déterminer les classes de visèmes à partir des matrices de confusions, mais aussi par une forte variabilité interlocuteurs concernant la production de la parole et la formation des différentes formes labiales (Owens & Blazek, 1985 ; Jackson, 1988 ; pour une revue, voir Cathiard, 1988/89). Gentil (1981) a pointé la différence entre les deux locuteurs utilisés dans son expérience et leur influence dans la tâche de reconnaissance des différentes consonnes et voyelles par les sujets sourds. Sur un plus grand nombre de locuteurs, il a été démontré que les groupes de sosies labiaux vocaliques et consonantiques pouvaient varier suivant la personne qui prononçait les phonèmes et en fonction de la facilité que les sujets avaient à lire sur les lèvres du locuteur (Lesner & Kricos, 1981 ; Kricos & Lesner, 1982 ; Yakel et al., 2000).

I.2.5. Implications chez les déficients auditifs

Il va de soi que dans ces conditions, percevoir la parole par la lecture labiale seule est une activité difficile car pleine d'ambiguïtés et de surcroît tributaire du locuteur *et* du labiolecteur. En clair, cela signifie que les personnes sourdes n'ont qu'une perception partielle de la parole par le biais de la lecture labiale. Elles peuvent parfois reconstituer le message oral par suppléance mentale à l'aide des indices contextuels. Mais en général cette charge cognitive importante ne leur permet pas de suivre un discours aisément. Par ailleurs, ce déficit constitue un problème majeur pour les enfants sourds prélinguaux qui n'ont que ce moyen pour apprendre la langue parlée. A partir de la lecture labiale qui ne fournit qu'une partie des contrastes phonologiques de la langue (par exemple, des informations sur le lieu d'articulation), il a été montré que ces enfants acquièrent et développent des représentations phonologiques, mais que celles-ci sont inexactes et sous-spécifiées (Dodd, 1976, 1987 ; Alegria et al., 1992 ; pour une revue, voir Colin, 2004). Chez les entendants, ces codes phonologiques dérivés en partie de la lecture labiale sont impliqués dans de nombreuses activités cognitives au cours du développement : perception et production de parole, jugement de rimes, tâches de mémorisation impliquant la mémoire à court terme et apprentissage de la lecture et de l'écriture (Alegria et al., 1992). Les enfants sourds ont recours également à des codes phonologiques mais démontrent d'importants retards liés à l'inexactitude de ces représentations (Leybaert, 1996). En effet, ces enfants présentent en moyenne un grand retard dans la compréhension de la lecture et plus généralement ont un niveau de langage beaucoup plus faible que celui d'enfants entendants du même âge. La transmission du langage oral de manière complète apparaît donc comme une nécessité.

I.3. Le Cued Speech

« I developed Cued Speech as a result of my decision that the spoken language must be made clear to the deaf child through vision if he is to learn it rapidly and well »
Cornett, 1988.

« A sender is able to transmit the cues in real-time synchronously with speech, thus conveying a visual analog of the syllabic-phonemic-rythmic patterns of spoken language »
Nicholls & Ling, 1982.

I.3.1. Définition et principes de construction

Dans les années 60, le docteur R. Orin Cornett, vice-président du Gallaudet College, conscient des difficultés que rencontraient les enfants sourds à travers le monde dans leur rapport à la lecture et de manière générale au langage, a décidé de mettre en place un système permettant à l'enfant sourd d'apprendre le langage parlé au quotidien. En 1967 (plus précisément durant une partie des années 1965 et 1966), le docteur R. Orin Cornett a mis au point pour l'anglais-américain une méthode permettant de transmettre visuellement la totalité du message oral à un rythme naturel de parole, le « Cued Speech » (Cornett, 1967). Conçu comme un compromis entre une communication *purement oraliste* qui ne permet pas d'acquérir un bon niveau de langage du fait des ambiguïtés inhérentes à la lecture labiale et une communication *gestuelle* qui éloigne trop la personne sourde de la langue utilisée couramment, le Cued Speech est une méthode qui utilise la main en complément de la parole visuelle. La personne quand elle parle utilise en même temps une suite de clés manuelles, le dos de la main en face de l'interlocuteur, pour désambiguïser les formes aux lèvres qui sont identiques.

Ces clés manuelles ont deux composantes : la forme de la main (ou configuration des doigts) utilisée pour coder les phonèmes consonantiques et la position de la main près du visage (dans une des deux hémifaces) pour les phonèmes vocaliques (voir les clés du Cued Speech Figure 3, adaptée de Cornett, 1967). Comme nous pouvons le voir sur la Figure 3, il y a huit configurations de doigts pour les consonnes et quatre positions de main pour les voyelles. Afin de limiter le nombre de clés, les clés manuelles n'identifient pas des phonèmes de manière individuelle mais regroupent, chacune, des phonèmes bien différents aux lèvres. Les phonèmes visuellement similaires sont codés par des clés différentes (c'est le cas par exemple pour /p/, /b/ et /m/ qui sont codés respectivement par les configurations n°1, n°4 et n°5). C'est l'*intersection* entre l'information manuelle (une clé identifiant un

sous-groupe de phonèmes possibles) et l'information labiale (un visème) qui donne un percept unique de ce qui est prononcé.

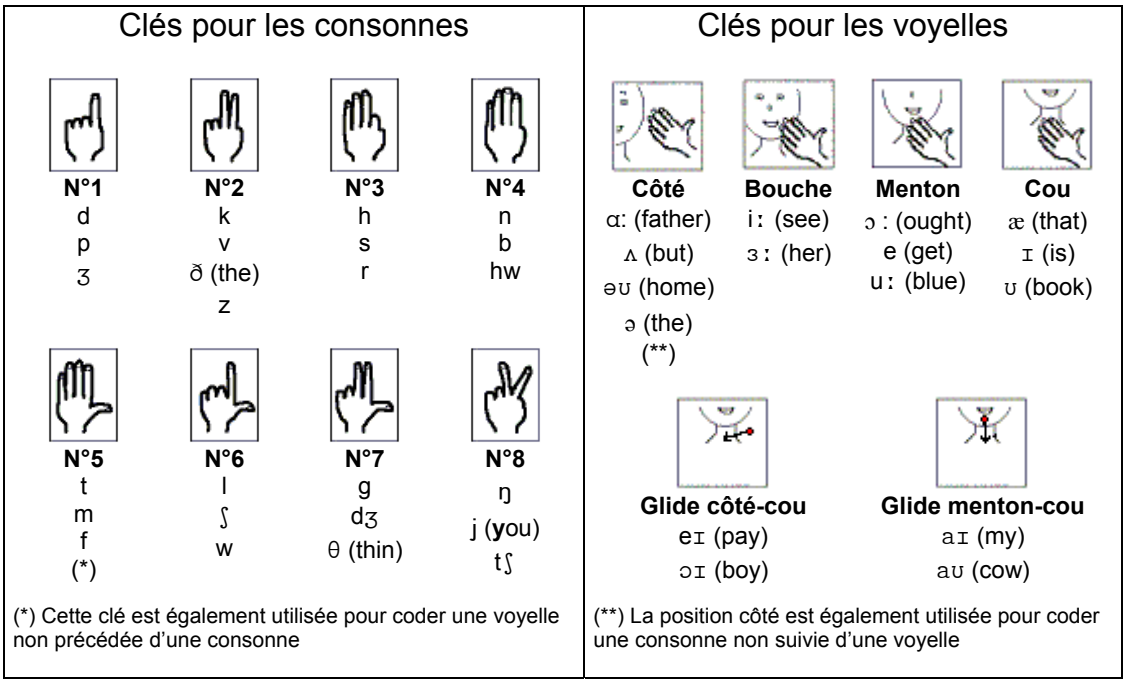


Figure 3. Clés manuelles du Cued Speech conçu par Cornett pour l'anglais-américain

Comment Cornett a-t-il procédé pour constituer son système ? Physicien de formation, Cornett a pensé le système en termes mathématiques. Il lui apparaissait évident que toute l'information phonologique ne pouvait pas être transmise par la main uniquement, à un rythme normal d'élocution. Celle-ci devait être transmise de manière complémentaire par les lèvres et par la main. De plus, le codage syllabique (le fait qu'un mouvement de main pouvait transmettre une information sur la consonne et sur la voyelle en même temps) lui garantissait la rapidité d'exécution du système. Son but était donc double : maximiser le contraste visuel entre les différents phonèmes et suivre un principe d'économie d'énergie afin que la personne utilisant ce système (le codeur) ne se fatigue trop vite (Cornett, 1982). Il s'est aidé principalement des visèmes établis par Woodward et Barber (1960) pour regrouper les différents phonèmes consonantiques de l'anglais en groupes de phonèmes visuellement contrastés au maximum. A l'aide des tables de fréquence de Denes (1963), il a réparti ces groupes de phonèmes en groupes de clés manuelles de façon à faciliter les enchaînements de mouvements de main et de doigts pour coder les combinaisons de consonnes les plus fréquentes de la langue. Il a par exemple attribué la configuration n°5 (main grande ouverte), qui est la plus facile à exécuter, au groupe de phonèmes le plus fréquent de la langue. Pour les voyelles, il a réparti au sein de chaque position des voyelles ouvertes, étirées et arrondies qui sont donc bien différentes aux lèvres. Comme les diphtongues sont codées par le glissement de la main entre les deux positions vocaliques correspondantes, Cornett a choisi les positions de façon à faciliter ces mouvements.

Ainsi les clés manuelles associées aux formes labiales fournissent l'information nécessaire pour permettre d'identifier précisément le phonème à partir de ce qui est visible sur les lèvres.

I.3.2. Un système syllabique

Afin de permettre une transmission des clés à un débit de parole naturel, la syllabe CV (consonne-voyelle) est l'unité de ce système. Ainsi la parole est codée en suites syllabiques CV, c'est-à-dire resyllabifiée. Pour chaque syllabe CV, la consonne C est codée par la configuration des doigts correspondante en pointant la position du visage qui code la voyelle V (voir exemple pour [pi] sur la Figure 4). Une consonne isolée est codée par la forme de main correspondante sur le « côté », position dite « neutre ». Une voyelle isolée est codée avec la configuration digitale n°5 (main ouverte, voir Figure 3) qui est une configuration dite « neutre » pointant la position adéquate. Dans le cas particulier de deux clés identiques à la suite (par exemple « ft » dans le mot anglais left), un léger mouvement d'avant-arrière de la main est effectué de manière à distinguer les deux phonèmes (voir Figure 4).

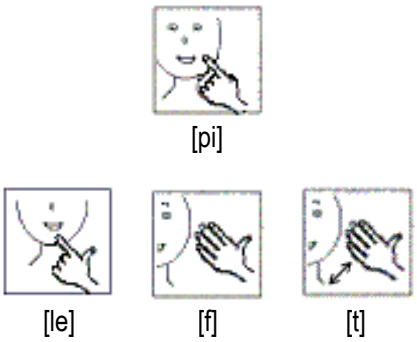


Figure 4. En haut : codage de la syllabe [pi] (la configuration n°1 est utilisée pour coder la consonne [p] en pointant la « bouche », position codant la voyelle [i], voir Figure 3) ; en bas, codage du mot « left » impliquant la même clé à la suite.

Avec l'expérience, Cornett a également défini quelques mouvements particuliers permettant de transmettre des informations prosodiques. Par exemple, l'inclinaison de la main par rapport à l'horizontale donne une indication sur l'intonation de la phrase (une inclinaison de 90° indique une forte accentuation alors qu'une inclinaison presque horizontale rend compte d'une faible accentuation du locuteur). Nous ne détaillerons pas la liste de ces mouvements particuliers car ils n'ont pas été maintenus dans l'adaptation du Cued Speech au Français. Le lecteur intéressé pourra se référer à quelques indications données par le Docteur Cornett (1982, 1994), aux recommandations de la NCSA, l'association nationale du Cued Speech aux Etats-Unis (1994) ainsi qu'au manuel « Gaining Cued Speech proficiency » de Walter Beaupré (1997) accessible sur le Web.

I.3.3. Organisation du geste

« Speech accompanied by these synchronized cues is referred to as "cued speech" »
Cornett, 1967

Que sait-on sur l'organisation de ce geste en relation avec ce qu'il code, la parole ? Jusqu'à présent, il n'y a pas eu d'étude entièrement consacrée à cette organisation temporelle et c'est le travail que nous présenterons dans cette thèse. Nous pouvons tout de même trouver quelques recommandations de la part de l'auteur du système sur le timing du codage. Cornett insiste à plusieurs reprises sur l'importance de la synchronie des mouvements manuels avec le son produit correspondant. Il définit le Cued Speech comme « […] a time-locked system; that is, the cues must synchronize with the spoken sounds. Every cue is essentially a hand movement that is timed relative to the sound » (Cornett, 1994). Il indique plus précisément pour la syllabe CV que ce sont les transitions manuelles d'une position à l'autre et les changements de configurations consonantiques qui indiquent comment la main et le son doivent se synchroniser. « The time of arrival of the hand at a given location indicates the instant at which the next sound is to begin. The time the hand reaches a new configuration likewise indicates accurately when the associated sound begins. », (Cornett, 1994). Cela signifie clairement que dès que la configuration manuelle est formée, on commence à produire la consonne acoustique correspondante. De la même façon, dès que la position de main est atteinte, on commence à produire la voyelle. Ainsi, la consonne et la voyelle sont complètement synchronisées avec leur correspondant LPC, soit la configuration et la position de main.

Dans le cas de deux consonnes précédant une voyelle, Cornett indique une synchronie des mouvements labiaux avec les mouvements manuels mais un retard de l'émission du son afin de garantir une prononciation « naturelle » de la syllabe. « When two consonants precede a vowel, as in the word *steep*, the first consonant is cued in the base position and the hand moved quickly to the vowel position while the second consonant cue is formed, in synchronization with the lip movements. The lips should assume the position for the first consonant as it is cued, but one should not begin making the sound until the hand is approaching the position in which the contiguous consonant and the following vowel are to be cued. This makes it possible to pronounce the syllable naturally. » (Cornett, 1967, p. 9). En clair, dans le cas du mot *steep*, Cornett préconise de retarder le début de la friction du [s] (par rapport au geste labial), afin d'éviter une production acoustique saccadée du mot (du type [s#ti:p] en insérant une pause entre [s] et [t]).

La synchronie des clés manuelles à la parole est aussi un point fondamental (« synchronization of cues with speech movements is essential ») pour Walter Beaupré (1997) qui a rédigé un manuel contenant des conseils aux codeurs pour améliorer leur codage, et plus précisément la précision, la vitesse, la fluidité et l'expression. Il pense que la désynchronisation des deux gestes peut amener de la confusion chez le sourd qui décode. De plus, il déduit de son raisonnement que si les clés arrivaient après les éléments de parole, cela obligerait la personne à ralentir son rythme de parole ; ce qui n'est parfois pas nécessaire et non désiré. Dans les recommandations de la NCSA (1994) nous pouvons également retrouver cette nécessité de synchronie qui permettrait au sourd décodeur de voir tout ce qui est prononcé. Ils recommandent de faire un mouvement particulier pour chaque phonème produit, qui n'est en fait pas forcément visible sur les lèvres : « the execution of each cue must include a discernible movement or event that clearly indicates the time at which the key articulatory action takes place ». Par exemple, ils préconisent le fait de toucher la position (quand c'est possible) avec la main de façon à avoir soi-même un retour (feedback) sensoriel sur la synchronie entre le geste manuel et le geste articulatoire. Ainsi le moment où la partie du corps est touchée donne une indication au sourd décodeur sur le timing de la parole et donc sur la durée des différents gestes articulatoires. Cette synchronie des deux gestes se justifierait donc en termes d'efficacité perceptive pour la personne qui décode. Il est important de noter que ce timing particulier de synchronisation du mouvement manuel à la parole est une recommandation pour les personnes qui codent le Cued Speech américain. Ce genre de règle n'existe pas à notre connaissance pour le français. De plus, il n'a jamais été mesuré que dans la pratique de la production du Cued Speech ce timing soit respecté par les codeurs.

L'équipe de Duchnowski au MIT (1998a, 1998b, 2000) a cependant observé sur des films (sans réaliser de mesures) le timing des gestes manuels en relation avec la parole produite par des locuteurs-codeurs de Cued Speech américain. Ils ont remarqué que « […] human cuers often begin to form cues well before producing audible sound » (Duchnowski et al., 2000, p.491). Par ailleurs des études de simulations menées par Bratakos et al. (1998) ont exploré les effets sur la réception des mots du retard de la clé sur le son. Leurs résultats sont indiscutables : plus la clé est montrée avec du retard, plus les performances en réception correcte baissent (voir section I.8.4.1). Duchnowski et son équipe se sont servi de ces observations pour implémenter un système de codage automatique de clés du Cued Speech comme système d'aide à la lecture labiale pour les personnes sourdes (voir section I.8.4.2). Afin d'améliorer leur système, les auteurs ont avancé de 100 ms l'affichage des clés par rapport au signal acoustique correspondant. Leur système adopte donc un pattern de synchronisation entre la main et le son un peu différent de celui recommandé par Cornett : la clé, c'est-à-dire

l'information sur la consonne et sur la voyelle est donnée 100 ms en avance par rapport au début acoustique de la syllabe.

Le geste manuel semble donc suivre une organisation temporelle spécifique extrêmement reliée à la parole. Les informations, que nous avons retirées des différents auteurs cités dans cette section ne nous permettent pas de décrire un patron temporel clair de l'organisation du geste de la main en rapport avec la parole visible sur les lèvres et le son produit. Nous verrons par la suite comment nous nous proposons d'étudier cette organisation temporelle afin de déterminer des règles de production fiables.

I.4. Adaptation du CS au Français : La Langue française Parlée Complétée

En 1993, le Cued Speech avait été adapté à 56 langues et dialectes à travers le monde (dans les premiers temps, toutes les variantes de l'anglais, l'espagnol, le français, l'allemand, puis l'italien, le russe, le tchèque, l'arabe, le thaï, le cantonais, etc.) par Cornett avec l'aide de personnes expertes natives du pays concerné (Cornett, 1994). Il était important que les mêmes critères que ceux utilisés pour la construction du Cued Speech (maximisation du contraste visuel et économie d'effort) soient maintenus. Cependant, les données sur les différentes langues, telles que les fréquences statistiques des différents phonèmes, n'étaient pas toujours accessibles. C'est donc le critère de compatibilité entre les langues qui a été principalement pris en considération pour les différentes adaptations. Ainsi les consonnes communes sont toutes réalisées par les mêmes clés digitales. Cela permettait de plus d'assurer la possibilité d'un bilinguisme (voire plurilinguisme) pour les enfants sourds bénéficiant de codage. Ainsi on retrouve en grande partie les mêmes clés que celles du Cued Speech. Dans les différentes adaptations, on retrouve les huit configurations de main et de deux à cinq positions selon le nombre des phonèmes vocaliques de la langue concernée.

L'adaptation du Cued Speech américain au français a été introduite en France en 1977 sous le nom de « Langage Complété Cornett » (LCC) qui a donné plus tard le « Langage Parlé Complété » ou LPC, nom sous lequel il est majoritairement connu en France. Récemment le LPC a été renommé en « Langue française Parlée Complétée » afin d'insister sur le fait que ce système était entièrement basé sur la langue française et ne constituait pas une langue à part entière. C'est le nom que nous retiendrons pour la suite : la Langue française Parlée Complétée ou la LPC. Depuis son apparition en France, la LPC s'est largement répandue dans le pays et connaît un développement important auprès de familles touchées par la surdité, de professionnels et de centres spécialisés. Plusieurs associations faisant la promotion de ce système se sont créées. Parmi elles, l'A.L.P.C. (l'Association pour la

promotion et le développement de la Langue française Parlée), qui se trouve être l'association nationale de la LPC, est en charge de former et délivrer le diplôme professionnel de codeur. Pour l'attribution de ce diplôme, le jury (constitué de professionnels de la surdité et de la LPC et de sourds décodeurs) évalue la précision des codes manuels et la fluidité et la rapidité des mouvements. Bien que l'apprentissage des clés puisse être très rapide (quelques heures), plusieurs mois d'entraînement sont nécessaires pour atteindre un rythme de codage plus élevé.

En ce qui concerne les codes, en français, il y a huit configurations de main pour les consonnes et cinq positions nécessaires pour coder tous les phonèmes vocaliques, la position pommette étant la position supplémentaire par rapport aux clés du Cued Speech. Il n'y a en revanche pas de règles particulières pour des diphtongues, inexistantes en français standard. La Figure 5 indique la répartition des phonèmes du français au sein des différentes clés. Le principe de fonctionnement reste le même que pour le Cued Speech : le codage est syllabique avec comme unité la syllabe CV. Les règles spécifiques pour les voyelles non précédées de consonne (V) et pour les consonnes non suivies de voyelle (C ou C_n) sont maintenues : comme pour le Cued Speech, la position « côté » est également une position neutre utilisée pour coder les consonnes isolées (souvent le cas des groupes consonantiques où la consonne n'est pas suivie d'une voyelle ou en finale de mot précédant un schwa) et la configuration n°5, avec la main grande ouverte, est aussi la configuration neutre utilisée pour coder les voyelles isolées (non précédées d'une consonne) (voir exemple de mot codé impliquant différentes structures syllabiques Figure 6). Les mouvements impliqués dans la production de ce code sont donc la transition manuelle d'une position à une autre, le changement de configuration consonantique et dans certains cas où le code se répète (comme c'est le cas par exemple pour « papa »), un léger mouvement d'avant arrière de l'ensemble avant-bras - main.

Il est important d'insister sur le fait que ce système permet de transmettre tous les contrastes phonologiques de la langue sans ambiguïté. La personne, qui l'utilise pour s'adresser au sourd, code tout ce qu'elle prononce : ainsi les variantes phonologiques de prononciation (ex : pour le mot « lait », certains vont le prononcer [le] et coder le mot à la position « cou » alors que d'autres le prononceront [lɛ] et le coderont à la position « menton ») et les liaisons entre les mots d'une même phrase (voir Figure 7) vont être transmis. Le codage LPC s'appuie donc bien sur la langue orale (c'est-à-dire sur la réalisation phonétique) et non pas sur l'orthographe du mot.

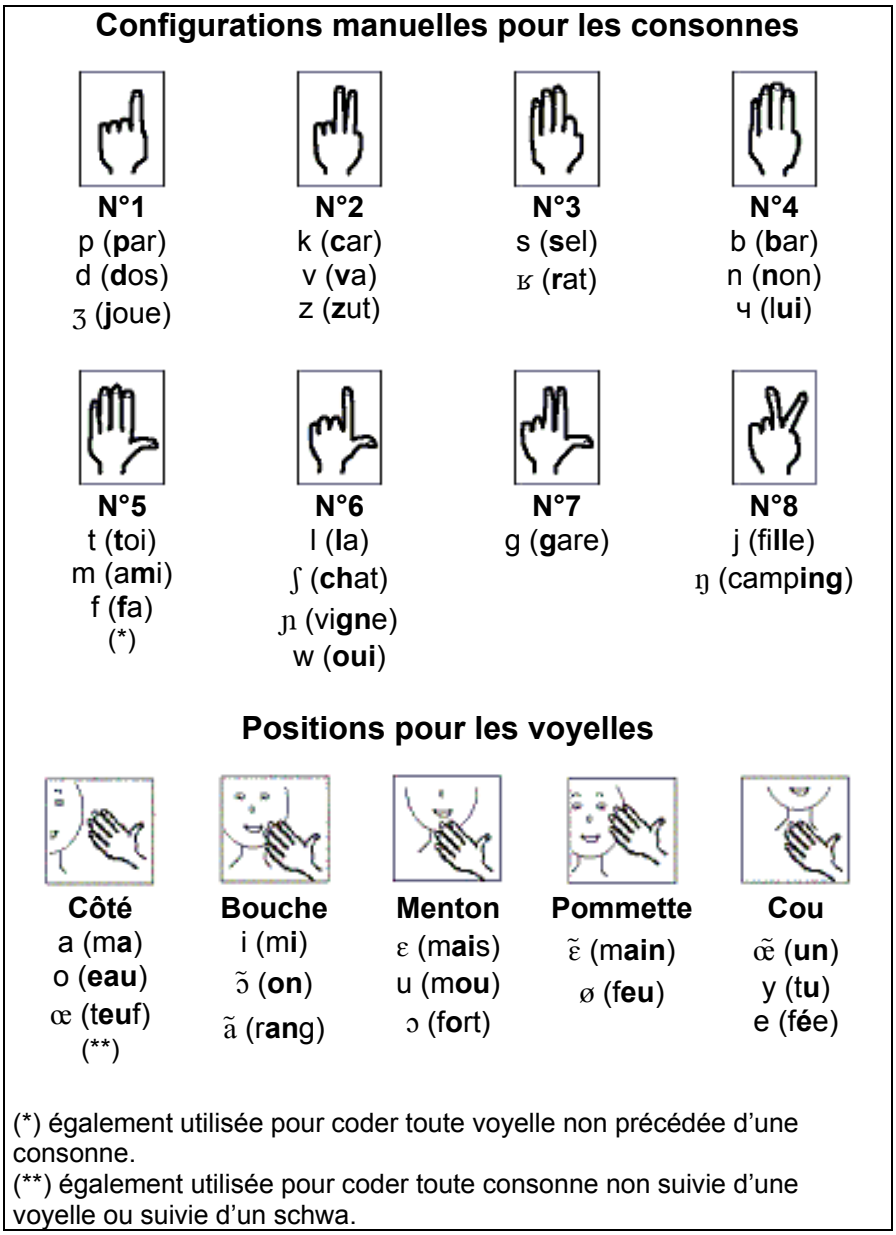


Figure 5. Clés manuelles de la LPC pour le français. Adaptation du Cued Speech.

Figure 6. Codage manuel du mot « structure » en LPC suivant la décomposition syllabique C.C.CV.C.CV.C

Figure 7. Exemple d'expression codée avec et sans liaison, « mon avion ». A gauche, le [n] de liaison est codé par la configuration appropriée avec la main en position « côté » pour la voyelle [a] associée à cette consonne.

I.5. Efficacité perceptive du code manuel

Nous avons vu que la lecture labiale à elle seule ne permettait pas d'identifier de manière exacte la parole. Le Cued Speech avait été inventé pour justement pallier cette difficulté. Dès les débuts de son utilisation, de nombreux acteurs (parents, sourds utilisateurs, professionnels de la surdité) ont témoigné de son efficacité pour la réception de la parole. La multiplication de son utilisation à travers le monde atteste également de son utilité pour les sourds. Qu'est-ce qui est perçu exactement dans la parole codée ? De manière quantitative, des études ont testé expérimentalement l'apport des clés manuelles à la lecture labiale. Globalement, l'ajout des clés améliore significativement la réception des syllabes (Nicholls & Ling, 1982), des mots (Ling & Clarke, 1975 ; Clarke & Ling, 1976 ; Nicholls & Ling, 1982 ; Gregory, 1987) et des phrases (Uchanski et al., 1994 ; Bratakos et al. 1998 ; Duchnowski et al., 2000).

I.5.1. Réception de syllabes et de mots

Nicholls et Ling (1982) ont testé la réception de syllabes et de mots-clés insérés dans des phrases chez 18 enfants sourds (de 9 à 17 ans) qui avaient eu au moins quatre ans d'exposition au Cued Speech. Les syllabes utilisées étaient du type consonne-voyelle (CV) ou voyelle-consonne (VC) et combinaient 28 consonnes avec les voyelles [a, i, u], donnant ainsi 84 stimuli différents. La réception de ces syllabes était testée sous de nombreuses conditions : en audio seul (A), en lecture labiale seule (L), en audio + lecture labiale (AL), avec les clés seules (C), en audio + clés (AC), en lecture labiale + clés (LC) et en audio + lecture labiale + clés (ALC). Les résultats montrent un bénéfice incontestable des clés du Cued Speech. Les auteurs obtiennent en effet une réception exacte des syllabes de plus de 80% pour les conditions LC et ALC alors que les résultats en lecture labiale (ainsi qu'en condition AL, C et AC) ne dépassent pas les 30-40%. Pour les conditions L, AL, LC et ALC, les résultats sont meilleurs en contexte vocalique [a] et [i] qu'en contexte [u], contexte vocalique qui pénalise fortement l'identification des consonnes (ainsi que nous l'avons déjà évoqué en section I.2.3).

Nicholls et Ling (1982) ont également testé la réception de mots-clés (monosyllabiques) insérés à la fin de phrases simples dont le contexte sémantique était soit fortement prédictible (*High-Predictability*, HP) soit très peu (*Low-Predictability*, LP). Les phrases HP donnent des indices sémantiques contextuels permettant d'identifier le mot-clé (exemple : « Go to sleep in your bed. »). En revanche les phrases LP ne donnent aucun de ces indices sur le mot (exemple : « Where is my book ? »). Les auteurs obtiennent des résultats frôlant la perfection (95%) quand les clés sont ajoutées (conditions LC et ALC) et cela quel que soit le niveau contextuel de la phrase. Il est à noter que le score de 95% est atteint en condition LC – sans audition – ce qui signifie que la parole codée peut donc être perçue de manière

exacte par la modalité visuelle uniquement. Ce résultat met l'emphase sur le rôle des codeurs qui interviennent en classe pour retransmettre aux élèves décodeurs le discours du professeur, et de manière plus générale l'utilité des personnes qui codent ce qui a été dit par une autre personne même sans émettre de son. En effet, dans ce cas là, les codeurs subvocalisent : ils articulent bien mais n'émettent pas de son.

I.5.2. Réception de parole conversationnelle et de phrases complexes

Nous avons rapporté l'efficacité du Cued Speech pour des mots insérés dans des phrases très simples. Est-ce que la réception de parole codée est toujours aussi efficace quand le niveau de langue est plus élevé ? En clair, le code permet-il aux sourds décodeurs de suivre une conversation normale ?

Uchanski et al. (1994) ont testé quatre sujets sourds adultes (de 18 à 27 ans), qui avaient bénéficié auparavant d'au moins huit ans de code de manière intensive (à la maison et à l'école), sur des phrases enregistrées sur cassette vidéo par deux locutrices-codeuses de Cued Speech. Le test comprenait trois types de phrases tirées de la liste CID qui contient essentiellement de la parole conversationnelle, de la liste Clarke qui contient également de la parole conversationnelle mais avec un vocabulaire adapté pour des enfants de 7-8 ans et donc avec un fort contexte prédictible et enfin de la liste Harvard qui contient des phrases d'un niveau beaucoup plus difficile avec un contexte interne très peu prédictible. Chaque phrase contenait un certain nombre de mots clés que le sujet devait identifier par écrit. Le test était passé dans deux conditions : en lecture labiale seule et en lecture labiale avec clés manuelles du Cued Speech en complément. Les résultats en moyenne montrent un bénéfice notable de l'ajout des clés à la lecture labiale. En effet, les quatre sujets ont obtenu en lecture labiale seule des scores moyens d'identification correcte des mots de 62% pour les listes CID, 45% pour les listes Clarke et 25% pour les listes Harvard. Avec l'ajout des clés du Cued Speech, les sujets obtiennent des scores moyens de 97% pour la parole conversationnelle et de 84% pour les listes Harvard qui étaient beaucoup plus difficiles. Il apparaît donc que le complément d'information apporté par le Cued Speech permet aux personnes sourdes de percevoir la parole « de tous les jours » de manière aussi efficace que les personnes entendantes. Bien que les clés ne leur permettent pas d'atteindre un si bon niveau pour des phrases beaucoup plus difficiles, on note cependant une large amélioration (de 25% à 84%).

Ces résultats sont également confirmés par d'autres études. Bratakos et al. (1998) ont testé six adultes sourds (19-27 ans) qui avaient bénéficié d'au moins 12 ans de code. En réception de phrases complexes à contexte interne très peu prédictible, les résultats obtenus sont similaires à ceux de Uchanski et al. (1994) : seulement 30% de réception correcte en lecture labiale seule et 84% en lecture

labiale augmentée de clés du Cued Speech. Duchnowski et al. (2000) ont testé cinq adultes (19-24 ans) avec un minimum de 14 ans d'exposition au Cued Speech sur un matériel de parole similaire. La lecture labiale seule ne permet d'identifier que 35% des mots-clés constituant les phrases alors qu'avec l'ajout des clés du Cued Speech les résultats atteignent 91%.

I.5.3. Influence du temps d'exposition aux clés

Les premières études sur la perception de parole codée avaient montré un bénéfice moins avantageux des clés. Ling et Clarke (1975) avaient testé 12 enfants sourds (7-12 ans) qui n'avaient bénéficié que d'une seule année de codage et surtout à l'école. Les stimuli testés consistaient en des mots reliés dans des expressions (ex : « boy and girl ») et des phrases simples composées de quatre mots (ex : « she has five books »). Les résultats en lecture labiale seule et en lecture labiale avec clés étaient de seulement 9% pour les phrases entièrement identifiées. Un léger bénéfice des clés était obtenu pour les mots dans les expressions (52% de réception correcte en lecture labiale + clés contre 35% en lecture labiale seule). En comptabilisant tous les mots correctement identifiés (y compris ceux des phrases incomplètes), les résultats restent assez bas : 68% en lecture labiale + clés contre 50% en lecture labiale seule. Clarke et Ling (1976) ont testé l'année suivante huit enfants de cette étude, qui avaient donc une année supplémentaire d'exposition au code. Les résultats pour les phrases correctement identifiés sont de 23% en lecture labiale seule contre plus de 62% en lecture labiale avec clés du Cued Speech. L'ajout des clés est, cette fois-ci, plus avantageux pour les enfants. Le temps d'exposition au Cued Speech semble donc avoir un rôle important dans son efficacité.

L'entraînement intensif au décodage sur une courte période ne semble en revanche pas avoir d'effet, du moins dans l'immédiat. C'est ce qu'a montré Gregory (1987) dans son étude sur la réception de mots isolés. Il a testé un public assez hétérogène de sujets sourds âgés de 11 à 68 ans et qui avaient bénéficié de deux à neuf ans de Cued Speech. Le matériel testé consistait en des listes composées de mots isolés (50 mots par liste) codés ou non. Les sujets ont été pré-testés et post-testés. Entre les deux phases, ils ont bénéficié de 18 heures de code intensif. Les résultats ne montrent aucune différence entre les deux phases : en moyenne, les sujets identifient correctement les mots à 43% en lecture labiale seule et à plus de 76% avec l'ajout des clés. Le Cued Speech a donc un effet favorable sur la réception des mots mais cet effet ne semble pas amélioré par l'entraînement à court terme.

Pour le français, les résultats montrent également un bénéfice des clés de la LPC. Périer et al. (1987, 1990) ont testé des sujets sourds, répartis selon leur degré d'exposition à la LPC, sur la réception du langage parlé. L'un des groupes (groupe « école ») était constitué de 24 enfants (de 8 à 14 ans) qui avaient reçu du code LPC uniquement à l'école (durant les heures de cours et pendant les séances

d'orthophonie). L'autre groupe (groupe « maison ») était constitué de 11 enfants (de 5 à 10 ans) qui avaient bénéficié de code à la maison de manière intensive et à l'école. Le test consistait en 78 phrases simples présentées avec ou sans code. Les sujets devaient choisir une image (sur quatre) illustrant la phrase qui venait d'être prononcée. Les phrases étaient réparties en trois niveaux de difficulté en fonction de la similarité des phrases sur les lèvres : le niveau facile (tous les mots étaient visuellement différents), le niveau moyen (un mot de chaque phrase dans les quatre choix était similaire) et le niveau difficile (les quatre choix possibles étaient identiques aux lèvres et aucun indice contextuel n'était disponible). Les résultats montrent que l'ajout des clés améliore significativement les performances des sujets : en moyenne toutes phrases confondues, les sujets du groupe « école » ont une performance de 60% en lecture labiale seule contre 71% en lecture labiale + clés et les sujets du groupe « maison » ont une performance de 66% en lecture labiale seule contre 86% avec l'ajout des clés. Les clés de la LPC sont plus exploitées pour les phrases difficiles, et cela surtout pour les enfants du groupe « maison » (de 39% en lecture labiale à 72% avec LPC). De manière générale, il n'y a pas de différence entre les deux groupes en ce qui concerne la lecture labiale seule ; en revanche, le groupe « maison » a de meilleurs résultats que le groupe « école » quand les clés sont ajoutées.

Alegria et al. (1999) ont trouvé des résultats similaires chez 31 sujets sourds prélinguaux qui étaient répartis en deux groupes selon leur âge de début d'exposition au code LPC. Le groupe « LPC précoces » était constitué de sept sujets (âgés de 8 à 12 ans) qui avaient eu une exposition au code à la maison avant l'âge de 2 ans et durant au moins six ans. Le groupe « LPC tardifs » était constitué de 24 sujets (de 11 à 19 ans) qui avaient reçu du code pendant au moins trois ans et seulement à partir de l'âge de 5 à 9 ans. Les stimuli étaient des mots ou pseudo-mots (mots phonologiquement possibles mais qui n'existent pas en français) bisyllabiques de quatre phonèmes (CV-CV, VC-CV, V-CVC et V-CCV), présentés avec ou sans code. Les auteurs ont obtenu une amélioration significative des performances de tous les sujets avec l'ajout des clés manuelles (de 40% en lecture labiale seule à plus de 70% avec LPC). Globalement, les sujets « LPC précoces » avaient de meilleures performances que les sujets « LPC tardifs » surtout pour l'identification des pseudo-mots, ce qui suggère que l'âge de début d'exposition à la LPC est un facteur déterminant (Leybaert, 1998). C'est ce que démontrent aussi de nombreuses études qui mettent en évidence des différences entre enfants sourds ayant eu du code LPC précocement (avant l'âge de 3 ans) et intensivement et enfants sourds ayant également bénéficié de code mais plus tardivement (voir section suivante I.6).

I.6. Représentations phonologiques et développement du langage

Le complément d'information apporté par les clés du Cued Speech (et de son équivalent français, la LPC) permet donc de percevoir visuellement la parole orale de manière exacte. Bien que ce code ait été inventé à l'origine pour améliorer la réception de parole, il se révèle comme un outil loin d'être négligeable dans le développement du langage des enfants sourds bénéficiant de cette méthode, surtout pour ces enfants, autrement appelés enfants « biberonnés » au LPC, qui ont bénéficié d'une exposition très précoce aux clés (Leybaert et al., 1998 ; pour une revue, voir Leybaert & Alegria, 2003 ; Colin, 2004).

Les enfants sourds exposés à la LPC saisissent toute la langue parlée de manière exacte avec ses caractéristiques phonologiques, grammaticales et syntaxiques. Les « petits mots » (mots de fonctions comme les articles par exemple) et les finales de mots sont également rendus visibles permettant le développement correct des connaissances morphologiques. Hage et al. (1990, 1991) ont testé la capacité d'enfants sourds à déterminer le genre des mots. Le genre est assez souvent marqué par la morphologie finale du mot. Certaines finales de mots indiquent en effet un genre féminin (par exemple les finales : cigar<u>ette</u>, tart<u>ine</u>) alors que d'autres un genre plutôt masculin (par exemple : mant<u>eau</u>, lap<u>in</u>). Certaines morphologies finales en revanche ne marquent pas le genre du mot (par exemple : p<u>oire</u>, v<u>erre</u>). Les résultats montrent que les enfants sourds qui ont bénéficié très tôt de LPC ont des performances similaires à celles d'enfants entendants en ce qui concerne la détermination du genre des mots. Ce n'est généralement pas le cas pour les enfants sourds éduqués oralement car souvent, par le biais de la lecture labiale seule, ils ont des difficultés à percevoir les articles et les fins de mots qui peuvent donner un indice sur le genre.

Ainsi, contrairement aux autres enfants sourds, pour qui les représentations phonologiques sont sous-spécifiées du fait des ambiguïtés inhérentes à la lecture labiale, l'exposition précoce et intensive au code LPC va permettre à ces enfants, par le biais de la vision uniquement, de développer des représentations phonologiques exactes des mots. Quand ils ont été « baignés » dans un flot linguistique correctement spécifié dès leur plus jeune âge (avant l'âge de 3 ans), les enfants sourds ont en effet une sensibilité à la rime proche de celle des enfants entendants et différente de celle des autres sourds (sourds exposés plus tardivement au code à l'école, sourds éduqués oralement ou sourds pratiquant la langue des signes ; Charlier & Leybaert, 2000). Ceci montre que bien que dépourvus d'audition, ils ont accès à des représentations phonologiques leur donnant de bonnes capacités de comparaison et de manipulation des unités segmentales et/ou syllabiques. De plus, contrairement aux autres enfants sourds, les enfants LPC précoces ne s'appuient pas sur l'orthographe

des mots pour juger de leur similarité phonologique (ils ne vont pas faire d'erreurs pour juger de la rime des mots « tasse » et « glace » par exemple). Dans son étude longitudinale, Colin (2004) a d'ailleurs montré que l'acquisition des représentations phonologiques des mots permettant d'effectuer la tâche de jugement de rimes précède l'apprentissage de la lecture chez ces enfants, de la même manière que chez les enfants entendants. Il est à noter que ces habiletés à juger de la similarité sonore des mots ont également été démontrées pour des enfants sourds bénéficiant de l'équivalent anglais du code LPC, le Cued Speech (LaSasso et al., 2003). Une autre preuve de l'existence de ces représentations phonologiques est mise en évidence par des expériences testant la mémoire à court terme. Dans une tâche de mémorisation et de rappel d'une liste de mots, les entendants ont en général plus de difficultés à retenir les mots quand ils riment et quand ils sont plus longs : ce sont les effets dits de similarité phonologique et de longueur des mots. Ces effets ont également été mis en évidence chez les enfants LPC précoces (Leybaert & Charlier, 1996 ; Leybaert et al., 1998).

Chez les enfants entendants, les représentations phonologiques sont impliquées dans les activités de lecture et d'écriture. Il s'ensuit que ces codes phonologiques exacts ont également de remarquables répercussions chez les enfants sourds LPC précoces. Ces enfants atteignent en effet un niveau de lecture et d'orthographe comparable à celui d'enfants entendants du même âge (Leybaert, 1996, 2000). Dans des tâches d'orthographe sur des mots nouveaux, ils font le même type d'erreurs que les enfants entendants : les mots sont mal orthographiés mais restent corrects phonologiquement (exemple « sitron » à la place de « citron », « trin » à la place de « train »). A l'inverse, les enfants LPC plus tardifs (à l'école) font des erreurs sur l'orthographe sans respecter la phonologie du mot (exemple « copat » à la place de « copain »). Ces enfants ont en fait plus de mal à tirer profit des relations phonèmes-graphèmes (relations entre formes phonologiques et patrons orthographiques) car leurs représentations phonologiques ne sont pas totalement exactes ; il s'ensuit qu'ils sont beaucoup moins autonomes dans la génération d'écriture (en particulier, pour écrire des mots qu'ils n'ont jamais rencontrés auparavant).

L'exposition précoce aux codes manuels en complément à la lecture labiale permet donc aux enfants sourds de mettre en place des représentations phonologiques exactes de la langue à partir de la modalité visuelle uniquement et de développer correctement les différents mécanismes cognitifs pour une acquisition de la langue comparable à celle d'enfants entendants.

Quelle est donc la nature de ces représentations phonologiques formées à partir de la parole codée ? Fleetwood et Metzger (1998) proposent que les représentations phonologiques de base soient constituées de composantes purement visuelles, soit la forme labiale associée à la forme de main et la

position de main. Ces auteurs proposent même pour l'anglais-américain de complètement dissocier ces composantes de la parole. Ils inventent ainsi le terme de « cuem », pour remplacer le terme de Cued Speech, qu'ils définissent ainsi (p. 29) : « Cuem refers to an articulatory system that employs non-manual signals (NMS) found on the mouth and the handshapes and hand placements of Cued Speech to provide visibly discrete symbols that represent phonemic (and tonemic) values. Neither the production or the reception of acoustic information nor of speech is entailed in the meaning of the term *cuem*. ». Cette dissociation de la parole et du *cuem* viendrait du fait que le son émis en parole n'est pas forcément nécessaire pour percevoir ce code de manière efficace (Nicholls & Ling, 1982). Ainsi dans leur proposition, le *cuem* serait un autre médium pour transmettre le langage, par le canal visuel uniquement, tout comme la parole qui utiliserait la voie auditive. En clair, c'est un autre système de codage qui est proposé, qui ne se base pas sur les composantes articulatoires de la parole, ce qui expliquerait que les sourds utilisant cette méthode peuvent parfois ne pas bien oraliser.

Dans une étude sur la mémoire de travail, Leybaert et Lechat (2001) ont testé l'existence d'une boucle phonologique basée sur ces composantes phonologiques visuelles du code LPC – soit la forme labiale, la forme de main et la position de main – chez des jeunes sourds bénéficiant de LPC de manière plus ou moins intensive, en testant les effets de rimes, de similarité des formes labiales et de similarité des positions de main. Les résultats montrent que les sujets ont de moins bonnes performances de rappel pour les listes de mots qui sont similaires du point de vue des rimes que pour les listes phonologiquement différentes, ce qui confirme le fait qu'ils ont bien accès à des représentations phonologiques non basées sur le son et qui sont encodées en mémoire. De plus, les sujets sont également moins performants quand les listes sont similaires du point de vue des formes labiales et des positions de main (bien que ces listes ne soient pas rimantes). Les auteurs concluent donc que les formes labiales et la position de main sont codées en mémoire sous une forme phonologique visuelle, non dérivée des caractéristiques articulatoires de la parole (« We do not think that our deaf participants processed their perception of C[ued] S[peech] in a code derived from articulatory features of speech », p. 960). Sur ce point, les auteurs sont donc en accord avec les propositions de Fleetwood et Metzger (1998).

Cependant, cette position semble s'être nuancée au cours du temps. Dans un autre article en effet, Leybaert (Leybaert & Van Reybroeck, 2004) fait remarquer que : « L'enfant sourd exposé au LPC peut avoir un programme moteur entièrement précis, même si le résultat de son articulation n'est pas intelligible. Il s'agit bien de la précision des commandes articulatoires, liées à la qualité des représentations phonologiques, et non du résultat apparent de ces commandes, souvent altéré lorsque la boucle audio-phonatoire est déficiente ou absente » (p. 203). Elle conclut ainsi : « Il semble donc

que l'exposition précoce et intensive au LPC conduise au développement de représentations phonologiques et de programmes moteurs précis, pouvant servir de support à des jugements de rime adéquats, et ce, même avant d'avoir été mis en contact avec le code écrit » (p. 206). Ainsi, il semblerait bien que les représentations phonologiques développées par les sourds exposés au LPC ne soient pas si indépendantes de la parole. On peut en effet raisonnablement supposer que certaines commandes articulatoires soient récupérables à travers les formes labiales vues puisque celles-ci sont bien le résultat de gestes articulatoires.

I.7. Intégration des informations manuelles et labiales

Il est évident que la forme labiale, la forme de main et la position de main ont un rôle particulier dans la perception du code LPC, quelle que soit la nature du format dans lequel elles sont encodées. En atteste la remarquable amélioration des performances en lecture labiale quand le code manuel est ajouté. La question qui se pose maintenant est de savoir comment ces informations sont intégrées par le sourd décodeur pour l'identification d'un percept unique. Nous disposons à l'heure actuelle d'assez peu d'études ayant posé cette question.

Alegria et al. (1999), dans une étude sur l'identification des mots et pseudo-mots perçus par des sourds LPC précoces et LPC tardifs (voir plus haut), ont étudié cette question en analysant pour les pseudo-mots, les erreurs liées aux caractéristiques de la structure particulière du code LPC. Ils se sont intéressés particulièrement aux erreurs liées à la structure CV du code, les *CS errors*, pouvant donner des percepts phonémiques supplémentaires dans le cas de suites syllabiques non-CV (VC-CV, V-CVC et V-CCV), et aux erreurs liées à des substitutions phonémiques au sein de chaque clé. L'analyse de ces erreurs devrait en effet donner des indices sur la façon dont les clés sont traitées en rapport avec la lecture labiale. Nous rappelons que la syllabe CV est l'unité de base du système LPC : elle est codée à la fois par la forme de main pour la consonne et par la position pour la voyelle. La parole en LPC est resyllabisée en suites de syllabes CV ; dans la construction de ce code, une voyelle isolée V est codée à la position adéquate en utilisant la forme de main dite « neutre » (main ouverte, configuration 5 sur la Figure 5), de même, les consonnes isolées (C ou une suite consonantique C_n) sont codées avec la configuration de main adéquate positionnée à la position « côté ». Ainsi, le nombre de clés ne va pas forcément correspondre au nombre de syllabes CV réellement articulées (il faut deux clés pour coder la structure CVCV, mais il en faut également deux pour coder la structure CCV). Dès lors, il se peut que le sourd perçoive des segments qui n'ont pas été prononcés (par exemple une voyelle dans C[V]CV). De plus, nous rappelons que chaque clé code un sous-ensemble de phonèmes, qui sont clairement distincts aux lèvres. Les substitutions au sein d'une même clé signifieraient que les

sourds traiteraient l'information manuelle sans intégrer l'information labiale. Les résultats montrent d'une part que les erreurs de substitutions au sein d'une même clé se produisent surtout pour les consonnes (c'est-à-dire pour les configurations de main) plutôt que pour les voyelles, avec une tendance à être plus importantes pour les LPC précoces. D'autre part, en ce qui concerne le nombre de syllabes identifiées par rapport au nombre réellement articulé, il apparaît que le code LPC aide à déterminer le nombre de syllabes exact quand les syllabes produites sont de type CV seulement. Pour les autres types de structures, les sujets ont en effet tendance à interpréter le nombre supplémentaire de clés comme des syllabes supplémentaires. En particulier, ils se pourraient qu'ils décodent la consonne isolée C codée sur le « côté » comme une syllabe [Cə] contenant un schwa (par exemple, [bəli] à la place de [bli] ; rappelons que le schwa est codé sur le côté en LPC). Les auteurs font donc l'hypothèse que ces erreurs apparaissent dans le cas où la visibilité labiale des segments est insuffisante. Les erreurs analysées tendraient à montrer que les clés manuelles peuvent parfois être interprétées indépendamment de la lecture labiale.

Dans une étude plus récente, Alegria et Lechat (2005) ont testé cette intégration main-lèvres dans le cadre d'un paradigme McGurk, en présentant à 20 sujets sourds, répartis en LPC précoce (exposition au code avant l'âge de 2 ans ; âge moyen : 9 ans) et LPC tardif (âge moyen : 11 ans et 8 mois), des informations labiales et manuelles concordantes et discordantes. Ce conflit main-lèvres devrait en effet révéler la façon dont les deux informations se combinent et leur importance respective. Les auteurs ont également manipulé la saillance des informations labiales liée aux effets de coarticulation : les consonnes très visibles aux lèvres, comme les consonnes protruses, vont réduire l'intelligibilité des voyelles et les voyelles arrondies vont réduire la perception des consonnes (voir section I.2.3). Les stimuli testés étaient des syllabes CV (avec C= [ʃ, ʒ, s, z, f, v, k, ʁ] et V= [a, ã, o, ɔ̃]) produites avec et sans code et présentés sans son dans trois conditions : en lecture labiale seule, en lecture labiale + clés correctes et en lecture labiale + clés discordantes. Les voyelles testées étaient les voyelles ouvertes [a, ã] présentées en contexte favorable (avec [s, z]) ou défavorable ([ʃ, ʒ]). La condition discordante consistait à associer la position « menton » (correspondant à [ɛ, u, ɔ] ; pour un rappel voir Figure 5 p. 22) au [a] (codée sur le « côté ») et la position « cou » (correspondant à [e, y, œ̃]) au [ã] (codée à la position « bouche »). Les consonnes testées étaient les labiodentales [f, v] et les consonnes postérieures [k, ʁ], beaucoup moins visibles aux lèvres, présentées en contexte favorable [a, ã] et en contexte défavorable [o, ɔ̃]. La condition discordante consistait à associer la configuration 1 (correspondant à [p, d, ʒ]) à [v, k] et la configuration 4 (correspondant à [b, n, ɥ]) à [f, ʁ]. Ainsi les cas

de discordance mettent en conflit une forme labiale qui ne correspond à aucune des formes labiales des phonèmes codés par la main. Les sujets devaient regarder les stimuli filmés et donner leur réponse par écrit sur la consonne ou sur la voyelle. Les auteurs retrouvent les résultats classiques : dans la condition concordante, l'ajout des clés améliore significativement les scores d'identification par rapport à la lecture labiale seule et de manière plus importante pour les LPC précoces. La saillance des informations labiales a également un effet pour les LPC précoces : l'utilisation des informations manuelles est plus importante quand les consonnes sont difficiles à lire sur les lèvres (en contexte vocalique défavorable et selon le lieu d'articulation de la consonne), il en va de même pour les voyelles avec cependant un effet beaucoup moins marqué. L'analyse des erreurs montre également une différence entre les deux groupes de sujets. Les LPC tardifs font en général beaucoup plus d'erreurs que les LPC précoces. De plus, les erreurs des LPC précoces sont davantage liées à la structure du code, les *CS errors*. Dans la condition discordante, les sourds (surtout les LPC précoces) vont choisir les consonnes les moins visibles aux lèvres qui représenteront une sorte de compromis entre ce qu'ils voient sur la main et ce qu'ils voient aux lèvres. Ainsi les LPC précoces vont mieux exploiter les informations manuelles qui vont être intégrées avec les informations labiales selon le degré de saillance de ces dernières.

Selon les auteurs, l'intégration main-lèvres semble donc suivre des principes similaires à ceux observés en perception de parole audio-visuelle. Deux sortes de modèles de traitement de parole pouvant rendre compte de l'intégration main-lèvres sont proposés (Alegria et al., 1992, 1999 ; Alegria & Lechat, 2005). Le premier est un modèle hiérarchique, dans lequel l'information de lecture labiale, qui serait première, fournirait le corps de l'information phonologique et l'information manuelle, plus tardive et optionnelle, permettrait de résoudre les ambiguïtés restantes. Le second modèle repose sur une véritable intégration des informations manuelles et labiales, les deux informations ayant un poids équivalent : « the Lip-reading/Cues compound would produce a unique amodal phonemic percept conceptually similar to Summerfield's 'common metric' [1987] which integrates auditory and lip-reading information to generate a vocal tract filter function. » (p. 468). Ainsi le code LPC serait dans le premier cas conçu comme un indice « artificiel » dans une approche de type résolution de problèmes, alors que dans le second cas, il serait conçu comme une des entrées (au même titre que la lecture labiale) d'un système de traitement automatique de la parole. Les auteurs proposent (1992) que cette modélisation dépende de l'âge et du degré d'exposition au code : « Subjects early exposed to C[ued] S[peech] could process it phonemically because they have developed normal phonemic representations of speech. Subjects exposed to CS later can be limited to use it as an artificial signal in a problem solving way. » (p. 128).

Afin de mieux comprendre comment se réalise l'intégration des deux informations labiale et manuelle, nous nous proposons d'étudier précisément les relations de ces deux systèmes et l'organisation de leur coordination. Nous défendons l'idée que percevoir de la parole, c'est d'une certaine manière récupérer les gestes articulatoires de celui qui a produit cette parole. Ainsi nous faisons l'hypothèse que quand le sourd perçoit de la LPC, il va récupérer le code par la vision mais plus spécifiquement la manière dont a été produit ce code. Pour comprendre le mécanisme d'intégration LPC-parole, il faut donc comprendre comment le code est produit par les codeurs. Hormis les quelques indications données pour la pratique du code, il n'y a pas d'étude consacrée à la production du code LPC, à savoir comment la main et la parole se coordonnent naturellement dans la pratique quotidienne de codeurs professionnels. Nous pouvons trouver néanmoins quelques indices sur cette coordination dans le domaine des technologies, où les clés de ce code manuel ont été intégrées à des systèmes de synthèse.

I.8. Que nous apprennent les technologies innovantes sur la coordination LPC-parole ?

Nous passerons en revue dans cette section les différents systèmes de codage automatique qui ont été développés. Nous nous limiterons aux systèmes se rapprochant du principe du Cued Speech naturel, c'est-à-dire aux systèmes de compléments à la lecture labiale. De ce fait, nous n'aborderons pas les systèmes d'aide pour les sourds qui font une transcription phonétique à partir du langage parlé, tels que le projet Vidvox (Destombes, 1982), et qui donc se suffisent à eux-mêmes.

Les systèmes que nous présenterons ont une optique d'utilisation en temps réel dans les situations quotidiennes et sont basés sur un système phonétique de reconnaissance automatique (*Automatic Speech Recognition*) de parole. L'idée est de compléter la parole d'un interlocuteur ne connaissant pas les clés manuelles par l'ajout d'indices visuels. Ils nécessitent donc en plus de la présence d'une personne qui parle un système informatique de traitement important. Pour le système développé au MIT (le seul système actuellement en fonctionnement), nous présenterons les différentes étapes qui ont mené à son élaboration ainsi que les facteurs pris en compte pour l'amélioration de ce système.

I.8.1. L'Autocuer : un système avant-gardiste

Conscient des effets prometteurs liés à l'utilisation du Cued Speech au quotidien, Cornett commença en 1969 à travailler sur un système informatisé portatif qui donnerait à la personne sourde l'équivalent des clés du Cued Speech manuel. En collaboration dès 1971 avec Beadles du Research Trinagle Institute à Durham en Californie du nord, ils mirent au point des lunettes générant des images correspondant à un équivalent des clés du Cued Speech, l'Autocuer (Cornett, 1982 ; Cornett, 1988).

Ces lunettes sont reliées par un câble à un système informatique porté à la ceinture. Les lunettes sont munies de diodes lumineuses (deux groupes de 28 petites lampes formant sept segments lumineux) qui projettent, grâce à une courbure particulière du verre extérieur, des images virtuelles qui paraissent être dans l'air à environ un mètre en face de la personne qui les porte (le sourd connaissant les clés du Cued Speech et ayant eu un entraînement à l'utilisation de ces lunettes). Les images font apparaître des segments lumineux qui peuvent prendre par combinaison 9 formes différentes, désignées pour coder les consonnes à quatre emplacements différents, pour coder les voyelles. Bien que ne correspondant pas aux groupes de clés du Cued Speech, ces codes respectent le fait que les phonèmes similaires aux lèvres sont placés dans différents groupes. En orientant sa tête, la personne sourde peut voir ces images virtuelles à côté du visage de son interlocuteur. Le son produit par l'interlocuteur est capté par un microphone unidirectionnel placé sur les branches des lunettes. Le signal acoustique est ensuite transmis à un microprocesseur qui est en charge de faire l'analyse de parole et de commander l'allumage approprié des diodes afin de faire apparaître les images virtuelles « codant » les éléments de parole qui ont été reconnus. Il est important de souligner que le traitement complet du signal de parole requiert 150 ms. De ce fait, les images clés sont projetées 150 à 200 ms **après** l'émission du son de l'interlocuteur. D'après le docteur Cornett, « cela rend en fait cette indication plus facile à lire que si elle était fournie en synchronisme absolu avec la parole. Nous avons beaucoup de chance, et nous avons vérifié ce point par simulation. » (Cornett, 1982, p.34). L'Autocuer a fait l'objet d'entraînement et de tests durant plusieurs années auprès de 103 étudiants sourds de l'Université Gallaudet. Les premiers résultats indiquaient un taux de 76,8% de perception correcte de mots sélectionnés de manière aléatoire sur une liste définie de 500 mots. Les auteurs ont par la suite apporté quelques améliorations au système en remplaçant par une clé neutre la moitié des erreurs faites par le module de reconnaissance de parole. Les scores moyens s'élevaient alors à 82% pour des mots isolés. Selon Bratakos et al. (1998), le bénéfice des clés est peu conséquent si l'on admet que 63% des mots étaient correctement identifiés en lecture labiale seule (notons qu'il est tout de même de 30% d'après le calcul du gain [(LPC-LL)/LL]). Les derniers tests effectués sur l'Autocuer révèlent des résultats corrects de seulement 60 à 67% sur des mots isolés (d'après Uchanski et al., 1994 ; Bratakos et al. 1998). Le problème majeur viendrait du module de reconnaissance de phonèmes qui laisse passer plus d'un tiers des phonèmes sans les reconnaître (il y aurait en fait une reconnaissance correcte des phonèmes de 54% seulement). Ce système ne semble donc pas adapté à une communication quotidienne. L'utilisation de l'Autocuer nécessite de plus une importante période d'entraînement (au moins 40 heures). Par ailleurs, des tests effectués sur des mots inclus dans des phrases courtes ne montrent aucun bénéfice des clés complémentaires. Une période d'au moins 8

mois d'exposition à ce système serait nécessaire pour décoder des phrases à un rythme naturel de parole (d'après Bratakos et al. 1998). Il semblerait que par la suite l'Autocuer n'ait jamais été utilisé.

I.8.2. Les « kinèmes » de Vilaclara : de la reconnaissance de parole pour les sourds

Georges Vilaclara (1988) a tenté de mettre au point un système qui transmettrait l'information de parole aux sourds profonds à travers une autre modalité sensorielle que l'audition. Il s'agit d'un système de reconnaissance des phonèmes du français qui produit des « kinèmes », soient des éléments qui viendraient en complément de la lecture labiale. Six groupes de kinèmes consonantiques : [p, t, k], [b, d, g], [m, n, ŋ], [f, s, ʃ], [v, z, ʒ] et [l, r] et cinq groupes de kinèmes vocaliques ont été définis pour le français. Il est à noter que le regroupement de ces kinèmes ressemble aux groupes de kinèmes de l'AKA (l'Alphabet des Kinèmes Assistés) de Walter Wouts (1982), qui se trouve être, de la même manière que le Cued Speech, un complément manuel à la lecture labiale mais basé sur des caractéristiques phonétiques des sons. Dans le système de Vilaclara, la reconnaissance des phonèmes se fait par apprentissage sur une base de données contenant des signaux acoustiques de parole segmentée en phonèmes. Chaque segment a des paramètres qui sont caractérisés par un schéma statistique. Tous les attributs de chaque unité sont stockés dans une base de connaissance. Un système expert est chargé de trouver pour chaque unité à reconnaître le pattern statistique connu qui correspond le mieux. La décision est alors donnée par inférence. Les résultats sur une cinquantaine de phrases contenant au total 450 consonnes et 603 voyelles produites par un seul locuteur montrent un taux d'identification des kinèmes par le système de reconnaissance de 79% (d'après Uchanski et al., 1994). Aucune information n'a été donnée concernant le nombre de phonèmes que le système ne détecte pas (*deletions*) et sur le nombre de phonèmes que le système rajoute par erreur (*insertions*). Aucun test pour déterminer l'utilité de ce code n'a été à ce jour effectué ; il semblerait en fait que le système n'ait jamais été utilisé.

I.8.3. De l'utilité d'un système de reconnaissance de parole efficace pour transmettre des clés automatiques

Uchanski et al. (1994) ont étudié les performances des systèmes de reconnaissance de la parole existants dans les systèmes de production de clés automatiques, afin de voir si l'utilisation de tels systèmes pouvaient réellement apporter un bénéfice aux sourds usagers.

Afin d'évaluer les possibilités de reconnaissance dont étaient capables ces systèmes, ils ont tout d'abord comparé les performances entre le système développé au MIT (*Massachusetts Institute of Technology*, Cambridge) et des experts entraînés à la lecture de spectrogrammes. Les résultats des

experts représenteraient le potentiel des futures technologies. Ils ont de la même façon que Vilaclara (1988) défini des groupes de nouvelles clés (CBG). Les résultats corrects dans l'identification de ces groupes indiquent, comme attendus, une moins bonne performance pour le système de reconnaissance automatique (en moyenne 67,9% à 69,2% dépendant de la base d'apprentissage du système) que pour les experts humains. Il est à noter que ces résultats sont également moins bons que le système développé par Vilaclara.

Les auteurs ont par la suite effectué des analyses théoriques afin d'estimer le bénéfice éventuel de clés automatiques sur l'identification des consonnes et des voyelles. Pour leurs prédictions, ils ont utilisé un modèle d'intégration audio-visuelle, le « Post-Labelling Model » (Braida, 1991). A partir de données sur les confusions faites sur les consonnes et les voyelles en condition audio seule et visuelle seule, ce modèle peut prédire des confusions audiovisuelles des consonnes et des voyelles. Les auteurs se sont servis de données de la littérature audiovisuelle pour les confusions des consonnes (Erber, 1972) et les confusions des voyelles (Wozniack & Jackson, 1979 ; Hack & Erber, 1982 ; Montgomery & Jackson, 1983). Les auteurs ont fait ces estimations pour différents systèmes de reconnaissance : les systèmes de reconnaissance de parole du MIT, le système de Vilaclara, un système de reconnaissance « idéale » identifiant des clés parfaitement codées, et des experts humains. Ils ont également évalué les différents groupes de clés : « Phone » pour lequel une clé correspond à un segment de parole, « MC », les clés du Cued Speech, « AC », groupes définis pour l'Autocuer, « CBG », groupes définis par les auteurs et enfin les kinèmes de Vilaclara. Les auteurs ont donc pu par prédiction estimer le pourcentage d'identification correcte des phonèmes reconnus et codés par ces systèmes chez des sourds en condition visuelle, audio-visuelle et audio. Les résultats montrent qu'avec une reconnaissance de parole parfaite (« idéale ») les scores prédits pour MC, AC et CBG atteignent plus de 90% d'exactitude en condition visuelle et audiovisuelle. « This indicates that these three cue groups were well chosen as supplements to speechreading, given that the cues could be produced accurately » (Uchanski et al., 1994, p. 31). Il semblerait donc qu'en tant que compléments à la lecture labiale, ces groupements de clés ainsi constitués soient efficaces. De manière générale, les résultats indiquent qu'une réception exacte de 75 à 90% pour les phonèmes peut être envisagée avec de tels systèmes (Uchanski et al., 1994 ; Braida et al. 1995). Afin de généraliser leurs résultats à la réception de parole continue, les auteurs ont estimé qu'une précision correcte de 70-80% des phonèmes pour un système de reconnaissance était nécessaire pour permettre une réception correcte de phrases. Les analyses prédictives faites par ces auteurs confirment donc l'utilité de systèmes de codage automatique basés sur la reconnaissance de parole pour des sourds usagers.

En conclusion de leur étude, les auteurs avancent l'idée que, pour un système produisant des clés automatiquement, ce n'est pas seulement la performance du système de reconnaissance qui constituerait le point fondamental, mais le plus important serait plutôt la manière dont les compléments de synthèse sont transmis : « the cues derived by the recognizers must be presented so that they can be perceived accurately and integrated well with speechreading » (p. 36). L'utilisation d'un tel système par un sourd devrait être facile et ne devrait pas demander plusieurs heures d'entraînement (ce que nécessite l'utilisation de l'Autocuer). Ils proposent alors de garder les groupes de clés du Cued Speech manuel qui se sont révélées efficaces à maintes reprises dans la réception de parole. Par contre, ils rejettent l'idée de prendre des formes « symboliques » telles que celles utilisées dans l'Autocuer (groupes de LED formant des dessins géométriques).

I.8.4. Un générateur automatique de Cued Speech développé au MIT

L'équipe de Duchnowski (Duchnowski et al., 2000) au MIT a mis au point un système donnant au sourd décodeur l'image du locuteur avec des images de clés (photos de main) venant en superposition et correspondant aux codes du Cued Speech. Afin de construire le système le plus optimal, ils ont d'abord effectué une série de simulations.

I.8.4.1. Etudes de simulation : de l'importance de la ressemblance au Cued Speech manuel

Ces premières études menées par Bratakos et al. (1998) visaient à évaluer l'importance des différences qui existent entre un locuteur-codeur humain et un système automatique de codage dans une optique d'utilisation de systèmes de reconnaissance automatique de parole en temps réel. Ils ont évalué les effets liés à l'apparence des clés de synthèse par rapport aux clés manuelles, les effets des erreurs généralement commises par un système de reconnaissance automatique de parole et enfin les effets du retard de l'affichage des clés par rapport au son.

Les auteurs ont utilisé la vidéo d'une locutrice-codeuse expérimentée en production de Cued Speech pour l'anglais-américain prononçant des phrases avec (100 mots par minute) ou sans code (140 mots par minute). Comme pour l'étude de Uchanski et al. (1994), trois types de phrases étaient utilisés : des phrases tirées de la liste CID qui consiste en de la parole conversationnelle, des phrases tirées de la liste Clarke à fort contexte et des phrases tirées de la liste Harvard, assez complexes et à très faible contexte. Ces phrases contenaient chacune cinq mots-clés (quatre mots monosyllabiques et un mot bisyllabique). Le signal acoustique de parole était pour chacune des phrases segmenté et étiqueté (début et fin de chaque phonème) par des experts.

Pour les clés, les auteurs ont capturé, à partir des phrases codées, les huit configurations consonantiques du Cued Speech. Ils ont ensuite superposé ces photos aux positions appropriées sur l'image de la locutrice pour les phrases non codées. Il est à noter que ces photos de clés sont par nature discrètes c'est-à-dire qu'il n'y a pas de mouvement des doigts ni de transition manuelle. La clé (c'est-à-dire la configuration consonantique à la position appropriée près du visage) est en fait affichée au début de la consonne acoustique et est maintenue jusqu'à la fin de la voyelle dans le cas d'une syllabe CV ou bien jusqu'à la fin de la consonne dans le cas d'une consonne isolée. Pour le choix des clés, les auteurs ont développé une grammaire qu'ils ont implémentée en automates à états finis à partir des combinaisons possibles du Cued Speech manuel (c'est-à-dire CV, VC, C, CC et CVC). Cette machine peut ainsi à partir des étiquettes phonétiques dérivées de l'acoustique spécifier une séquence appropriée de paires configuration-position.

Les auteurs ont testé la réception de mots dans trois conditions : en lecture labiale seule, en lecture labiale (phrases non codées) + clés manuelles (phrases codées) et en lecture labiale + clés de synthèse (phrases non codées + clés superposées). Les clés de synthèse ont été obtenues de trois manières différentes : des clés totalement correctes, dérivées de la transcription phonétique manuelle (PSC, *perfect synthetic cue*), des clés dérivées d'un système de reconnaissance de parole indépendant du contexte (qui utilise un modèle différent pour chaque phonème) et des clés dérivées d'un système de reconnaissance de parole dépendant du contexte (qui utilise différents modèles pour le même phonème selon le phonème qui est à venir). Dans tous les cas le son n'était pas ajouté.

Le test s'est déroulé en deux phases, séparées de deux mois. Les sujets testés étaient des jeunes adultes sourds (de 19 à 27 ans) qui avaient eu au moins 12 ans d'exposition au Cued Speech. Avant le test, les sujets étaient familiarisés, avec les différentes conditions, sur les phrases des listes CID et Clarke. Pour cet entraînement, les réponses étaient données par écrit et ils avaient après chaque réponse une correction. Le test proprement dit s'est effectué sur les phrases Harvard sans correction : chaque sujet a été testé sur au moins 40 phrases dans chaque condition (200 mots clés).

Les résultats moyens obtenus tous sujets confondus indiquent une performance correcte de 30% en lecture labiale seule et de 84% en lecture labiale + clés manuelles, soit un bénéfice conséquent des codes manuels (Bratakos et al., 1998). Il est à noter que ces résultats sont similaires à ceux obtenus par Uchanski et al. (1994) sur des phrases de même type. En ce qui concerne les clés de synthèse, pour les clés de synthèse PSC, les résultats s'élèvent en moyenne à 77%. Les auteurs expliquent cette moins bonne performance par le fait que, pour les clés de synthèse, la main n'est pas articulée. Ils ne connaissent pas réellement l'importance de ce facteur mais ont pu déduire, d'après les témoignages

des sujets qui ont passé les tests, que le fait qu'il n'y avait pas de transition entre les positions était gênant. De plus, les images de main de synthèse étaient plus petites en taille qu'une main réelle. Le timing des clés a également un rôle à jouer : pour les clés de synthèse, le timing est fixé par la segmentation du son. On peut facilement imaginer que, dans la réalité, les choses ne sont pas aussi rigides. Par ailleurs, ces résultats peuvent également s'expliquer en partie par la vitesse plus rapide des phrases codées avec les clés de synthèse (140 mots par minute) que celles codées manuellement (100 mots par minute). Il est à noter cependant que l'entraînement avec ces clés de synthèse peut permettre d'améliorer les performances en réception (les scores pour la condition PSC étaient meilleurs durant la phase 2). Pour les clés dérivées du système de reconnaissance dépendant du contexte et du locuteur, les résultats s'élèvent à 70%.

Les auteurs voulaient également voir l'impact des erreurs faites par le système de reconnaissance sur la réception des mots. Le système de reconnaissance de parole n'étant pas parfait, il lui arrive de mal reconnaître des phonèmes (*substitution*) c'est-à-dire qu'il remplace un phonème par un autre, ou encore de laisser passer certains phonèmes sans les reconnaître (*deletion*) et enfin de reconnaître des phonèmes qui n'ont en fait pas été produits (*insertion*). Ce genre d'erreurs peut être parfois commis par un locuteur-codeur, surtout dans le cas de codage rapide différé de ce qui est dit à un rythme très soutenu. Dans leurs simulations, les auteurs ont testé les effets de clés erronées introduites à un taux de 10% et 20%. De manière générale, les performances des sujets diminuent quand le nombre d'erreurs augmente. Les insertions qui correspondent à quelque chose de nouveau (non visible sur les lèvres) ont un effet néfaste sur la réception des mots.

Les auteurs voulaient enfin évaluer l'impact du retard de l'affichage des clés inhérent aux systèmes de reconnaissance (pour l'Autocuer, ce temps de traitement s'élevait à 150-200 ms). L'identité d'une clé ne peut en effet être connue qu'une fois que le phonème aura été prononcé par le locuteur et qu'il aura été reconnu par le système. De ce fait, l'affichage des clés se fera toujours après les mouvements labiaux. Lors de l'évaluation de l'Autocuer, Cornett prétendait que ce retard n'affectait pas la réception de la parole codée (voir section I.8.1). Or nous avons vu plus haut (section I.3.3) l'importance de la synchronisation des clés manuelles avec la parole, et dans tous les cas les effets néfastes du retard de la clé sur les segments de parole. Les auteurs ont donc testé les effets d'un retard d'affichage de clés de 33 ms, 99 ms et 165 ms par rapport au son. Les résultats montrent que plus le retard est grand, plus les performances diminuent. De manière générale, un retard de 33 ms (c'est-à-dire que la clé est affichée 33 ms après le début acoustique de la consonne correspondant) n'affecte pas la performance des sujets. En revanche, pour les autres retards, combinés à un certain nombre d'erreurs, les scores diminuent de manière très significative. Si l'on considère qu'une syllabe CV dure en moyenne 165 à

200 ms pour un rythme normal de parole, et que le système de reconnaissance doit attendre d'avoir la syllabe complète pour pouvoir déterminer la clé correspondante, on comprend que la clé va être affichée bien après le début de la syllabe prononcée (durée de la CV + durée du temps de traitement). Ceci entraînera sans aucun doute une performance en réception fortement diminuée.

Il ressort donc de ces études de simulations qu'un système de codage automatique à partir de système de reconnaissance de parole en temps réel est faisable et utile pour les personnes sourdes. Les résultats de cette étude confirment de plus les analyses théoriques de Uchanski et al. (1994). Afin d'assurer un minimum d'erreurs, les auteurs préconisent un système de reconnaissance dépendant du locuteur plutôt qu'un système multi locuteur. Par ailleurs, l'impression de mouvement des clés semble importante pour la perception des sujets : tout rapprochement au système naturel manuel serait un bénéfice pour la réception par les personnes utilisatrices. De la même façon, le timing des clés est important : les auteurs ont mis en évidence la nécessité de la synchronie des clés avec le son. Dans tous les cas, les clés doivent être présentées avec un minimum de retard.

I.8.4.2. Le codeur automatique de Cued Speech temps réel

I.8.4.2.1 Composition du système

A la suite de ces simulations, Duchnowski et al. (1998a, 1998b, 2000) ont développé un système de codage automatique de clés du Cued Speech qui fonctionne en temps réel à partir d'un système de reconnaissance automatique de parole. Dans ce système, un locuteur est filmé en train de parler sans coder. Dans une autre salle, l'image du locuteur avec les clés de synthèse suivant les règles du Cued Speech est affichée sur un écran.

Deux ordinateurs sont chargés du traitement. Le PC1 (un Pentium Pro Class) s'occupe de numériser le signal acoustique (échantillonnage à 10 KHz) qui est capté par un microphone monodirectionnel porté par le locuteur afin de le pré-traiter (amplification des hautes fréquences, paramétrisation du signal en vecteurs de paramètres). Il est également en charge de la numérisation de la vidéo (à 30 images par seconde), du stockage de l'image capturée du locuteur durant le temps du traitement et de l'affichage des clés. Il est relié au deuxième ordinateur (PC2) par un réseau local de type Ethernet.

Le PC2 (une station Alpha DEC) s'occupe de la reconnaissance phonétique : il est muni d'un système de reconnaissance automatique de parole dépendant du locuteur. Les programmes de reconnaissance sont basés sur le logiciel HTK du laboratoire Entropic utilisant des modèles de Markov cachés à trois états. Le décodeur Viterbi de HTK a été modifié par les auteurs afin d'avoir une recherche plus rapide et de produire une séquence phonétique continue. Trois types de modèles sont utilisés : C1 sont des

modèles indépendants du contexte (46 modèles) qui atteignent une précision de reconnaissance de 65% en temps réel ; C2 sont des modèles dépendants du contexte à droite (2116 modèles) qui ont une performance temps réel de 66 à 70% ; enfin C3 sont des modèles dépendants du contexte à gauche et à droite (plus de 3800 modèles) qui nécessitent un temps de traitement plus long mais qui atteignent en contrepartie un taux de précision de 74%. Le PC2 transcrit donc automatiquement le son en séquences phonétiques et à partir de là identifie une séquence de clés correspondantes suivant les règles du Cued Speech grâce à une grammaire à états finis (voir Bratakos et al., 1998).

Le système fonctionne de la façon suivante. Le locuteur est filmé en train de parler. Chaque image est capturée et transmise au PC1 qui les garde en mémoire le temps du traitement (pendant deux secondes ; il est à noter que ce temps est largement suffisant). Pendant ce temps, le son est traité par le PC1 et envoyé au PC2 pour la reconnaissance. Le PC2 identifie les phonèmes et les clés correspondant à chaque image. Une fois l'identification effectuée, le PC2 envoie au PC1 l'information sur la clé (la configuration consonantique et la position) ainsi que les informations sur le timing d'affichage (début et fin d'affichage de clé). Enfin (à la fin des deux secondes), le PC1 affiche sur un moniteur TV l'image finale du locuteur avec la clé superposée.

On comprend donc que, dans les faits, ce système ne fait pas exactement du temps réel : la vidéo du locuteur avec les clés superposées est en fait affichée deux secondes après ce qui se produit dans la réalité (sorte de playback). Les auteurs ont choisi cette astuce pour éviter le problème du retard lié au temps de traitement, qui avait été démontré comme très néfaste dans la réception des mots (Bratakos et al., 1998). Ainsi, le fait de stocker les images en mémoire peut permettre d'afficher les clés de synthèse sur le visage du locuteur suivant un timing se rapprochant du codage manuel. De plus, il est à noter que l'affichage de l'image recomposée est continu ; tout se passe comme si la vidéo était différée de deux secondes.

I.8.4.2.2 Affichage des clés de synthèse

Pour l'affichage des clés, les auteurs ont testé différentes visualisations suggérant un mouvement de transition lisse de la main entre les positions avec différentes stratégies de timing des clés. Ils ont mis au point différentes versions du système (Duchnowski et al., 2000) :

– L'affichage « discrete » est le type d'affichage utilisé durant les études de simulation : il n'y a pas de mouvement simulé, le changement de clé est brutal en début de consonne. La clé est maintenue durant toute la durée de la syllabe.

– L'affichage « smooth » fait une approximation de mouvement de transition. L'image de la main est déplacée à vitesse uniforme suivant une trajectoire linéaire calculée par interpolation entre les deux positions. Il n'y a pas de pause (maintien de la clé) aux positions.

– L'affichage « dynamic » utilise des règles heuristiques qui permettent de rendre le mouvement plus naturel en répartissant le temps d'affichage de la clé entre le temps passé en position cible et celui nécessaire pour effectuer la transition vers la cible. Ils attribuent une durée de 150 ms pour une transition et 100 ms de tenue de la clé en position cible. Pour les transitions, le mouvement est de type « smooth ». Comme précédemment, la clé consonantique est changée à la fin de la transition, en début de consonne.

– L'affichage « synchronous » est le résultat de l'observation de vidéos de locuteurs-codeurs humains. A partir de ces vidéos, les auteurs ont pu constater que les codeurs humains semblaient souvent former les clés avant de produire le son. Pour simuler ce comportement, ils ont ajusté le moment d'affichage de la clé de façon à ce qu'elle apparaisse 100 ms avant le début acoustique de la consonne (déterminé par le système de reconnaissance). Dans cette version, la configuration consonantique de la main change au milieu de la transition manuelle (et non plus à la fin). Pour l'affichage de synthèse, les auteurs ont décidé d'allouer un minimum de 200 ms au mouvement d'une position à l'autre pour les voyelles de la diphtongue. Le changement de configuration consonantique (en configuration neutre) se fait au moment où la main a parcouru 75% de la transition. Enfin la transition de la dernière position (deuxième voyelle) vers la clé suivante se fait en 150 ms.

I.8.4.2.3 Evaluation

Cinq sujets sourds (19-24 ans) très expérimentés en Cued Speech (au moins 14 ans d'exposition, exposition actuelle de 1 à 4 heures par jour) ont été testés en condition de lecture labiale seule (*Speachreading Alone*, SA), de lecture labiale + clés manuelles (*Manual Cued Speech*, MCS) et de lecture labiale + clés de synthèse (*Automatic Cued Speech*, ACS). Dans toutes les conditions, l'audio a été enlevé.

Les phrases de test étaient lues et codées (ou non) par trois locutrices connaissant le Cued Speech. Trois tests de phrases ont été utilisés : les phrases CUNY à fort contexte constituées de 12 listes de phrases (102 mots-clés par liste) ; les phrases IEEE, beaucoup plus difficiles et à très faibles indices contextuels, sont regroupées en listes de 10 phrases (50 mots-clés par liste) et des phrases similaires aux phrases IEEE créées par les auteurs.

Le locuteur est assis dans une chambre sourde tandis que les sujets regardent la vidéo dans une autre pièce. Trois locuteurs se sont prêtés à l'expérience. Le test, durant au total environ 3 à 4 heures, est séparé en sessions de quatre ou cinq listes (une liste en SA, une liste en MCS et deux ou trois listes en ACS). Quatre expériences ont été effectuées afin de tester un maximum de conditions : variation du système de reconnaissance (*C1, C2 et C3*), variation de la version d'affichage du système (*smooth, discrete, dynamic, synchronous*), variation des phrases tests (*CUNY, IEEE, IEEE-like*) et variation du locuteur (*T1, T2, T3*) (voir les résultats obtenus dans le Tableau 1). Au début de chaque expérience, les sujets avaient une courte session d'entraînement.

Les résultats moyens tous sujets et expériences confondus montrent une performance de 35% en lecture labiale seule pour la réception des mots. Les résultats obtenus en lecture labiale + clés manuelles (MCS) démontrent un bénéfice conséquent de l'ajout des clés, dans la mesure où le taux de bonne réception des mots codés s'élève à 91%. Ce dernier résultat confirme les résultats précédents obtenus pour la réception de matériel codé (voir section I.5). Ces résultats s'avèrent dans tous les cas significativement meilleurs que ceux obtenus en lecture labiale seule et en lecture labiale + clés de synthèse. Pour les clés de synthèse, à titre indicatif, un résultat moyen de 52% a été obtenu. Cette valeur est à relativiser dans la mesure où la performance des sujets dépend de la précision du système de reconnaissance et de l'affichage des clés. Or, pour ce calcul moyen, toutes les conditions ont été mélangées. Par rapport aux performances en lecture labiale seule, l'apport de ces clés de synthèse n'est en fait significatif que dans les expériences III et IV, c'est-à-dire avec le système de reconnaissance le plus performant C3. Les meilleurs résultats sont obtenus avec le système C3 et l'affichage « synchronous » : ils s'élèvent en moyenne à 66% et constituent un bénéfice sur la lecture labiale de 57% (calculé par rapport au bénéfice du cued speech manuel sur la lecture labiale) pour les trois sujets confondus (voir Tableau 1). Par ailleurs, alors qu'il n'y a pas de différence significative entre les versions d'affichage « discrete » et « dynamic », la version « synchronous » apparaît comme bien meilleure que la version « dynamic » (dans l'expérience IV). Les performances obtenues avec ce système sont de plus renforcées par les témoignages positifs des sujets sur la version « synchronous ».

Expérience	Système automatique reco	affichage	Sujet	Locuteur	Matériel	Score des mots-clés SA	MCS	ACS	Bénéfice (%)
I	C1	smooth	S1	T1	CUNY	70,1	98,4	73,8	13,1
				T2		54,6	93,9	61	16,3
			S2	T3	IEEE	28	89,2	43,8	25,8
II	C2	smooth	S3	T3	IEEE	15,5	87,8	26,9	15,8
III	C3	discrete	S3	T3	IEEE-like	24,5	91	34,7	15,3
		dynamic						43,3	28,3
		discrete	S4		IEEE	29,5	92	47	28
		dynamic						48,7	30,7
		discrete	S5		IEEE	22,5	92	51	41
		dynamic						49	38,1
IV	C3	dynamic	S3	T3	IEEE-like	21,7	92,7	48,8	38,2
		synchronous						63,2	58,5
		dynamic	S4			42,4	86	53,6	25,7
		synchronous						67,2	56,9
		dynamic	S5			40,7	90	52,4	23,7
		synchronous						68,4	56,2

Tableau 1. Détail des scores d'identification obtenus dans les quatre expériences selon les différentes conditions (voir texte). Le calcul du bénéfice est donné par : 100(score_ACS - score_SA)/(score_MCS - score_SA). Tableau tiré de Duchnowski et al., 2000, p. 493.

Les résultats montrent donc que le meilleur système de reconnaissance C3 améliore les scores de réception par rapport à la lecture labiale seule mais il ressort que l'apparence de la main de synthèse ainsi que son timing d'apparition en relation avec la parole s'avère de plus grande importance : « The benefits provided by these cues strongly depend on articulation of hand movements and on precise synchronization of the actions of the hands and the face » (Braida et al., 1997, p. 3133). Il apparaît donc que, plus la main de synthèse se rapproche (dans son apparence et dans son mouvement) d'une main naturelle, plus la réception en est améliorée. Outre certaines astuces trouvées pour améliorer la discrimination des différentes configurations de main (les auteurs dans leur système ont coloré certaines formes de main difficiles à distinguer, Duchnowski et al., 1998b), le mimétisme du codage naturel est donc un point central. Duchnowski et al. (2000) ont tenté de se rapprocher du codage naturel en mettant en place des règles heuristiques de coordination. Cependant, ces règles sont basées sur le phasage des clés manuelles avec le son ; aucun lien n'est proposé avec la parole qui est visible sur les lèvres par les sourds décodeurs. Or comme nous le verrons dans le chapitre suivant, la parole est par nature coarticulée : ce qui implique le fait que les gestes articulatoires n'ont pas une correspondance directe avec les unités sonores. Les mouvements labiaux par exemple, peuvent, dans certains cas, être clairement en avance sur le son.

CHAPITRE II.

La parole, une structure co-articulée

II.1. La coarticulation

Le signal de parole est constitué d'une succession d'unités différentes. Cependant, contrairement à ce qu'on pourrait croire, ces unités ne sont pas indépendantes les unes des autres mais s'influencent mutuellement : c'est le phénomène de **coarticulation** (pour une revue, voir Bonnot, 1990a) ; « Sound segments are highly sensitive to context and show considerable influence from neighbouring segments. Such contextual effects are described as being the result of overlapping articulation or coarticulation » (Hardcastle & Hewlett, 1999, p. 1). En effet, quand on produit de la parole, on ne produit pas des segments individuels les uns après les autres : la parole n'est pas de l'épellation. Au contraire, la parole est produite par les gestes des différents articulateurs du conduit vocal (larynx, langue, lèvres, mâchoire, vélum) qui se chevauchent en partie au cours du temps car ils subissent des influences diverses : « Coarticulation is the superposition of multiple influences on the movement of an articulator. » (Perkell & Matthies, 1992, p. 2911). En clair, il n'y a pas de correspondance directe entre une entrée phonémique discrète et la sortie articulatoire continue.

Ainsi, chaque unité a une influence sur les unités voisines tant au niveau acoustique qu'au niveau articulatoire ; de cette manière, un segment de parole aura une configuration articulatoire différente selon les sons avoisinants et pourra être décrit différemment au niveau acoustique selon le contexte avoisinant. Si on compare, par exemple, la production articulatoire d'un [pu] à celle d'un [pi], on peut s'apercevoir facilement que durant la production du [p], les lèvres, bien que fermées, ne sont pas positionnées de la même manière : pour le [pu], les lèvres anticipent le mouvement de protrusion pour produire le [u] durant la consonne (les lèvres sont avancées) alors que pour le [pi], cette avancée des lèvres est remplacée par un étirement caractéristique de la production du [i]. Dans ce cas, les lèvres subissent l'influence du contexte ; on parle alors de *coproduction* (*temporal overlap*, Fowler, 1980 ; Fowler & Saltzman, 1993). Pour ce même exemple, la position de la langue au moment de la réalisation du [p] sera en arrière et haute dans [pu] et en avant et haute dans [pi] pour la réalisation de la voyelle suivante.

Perkell (1990, Perkell & Matthies, 1992) rajoute à sa définition de la coarticulation que les influences que subissent les mouvements des articulateurs peuvent venir du contexte (*acoustic-phonetic context*), comme dans le cas ci-dessus, mais aussi « from interactions with other articulators » (p. 2911). C'est ce que Fowler et Saltzman (1993) définissent sous le terme de *coordination* : il s'agit de contraintes de coordination (*coordinative constraints*), de dépendances, établies entre différentes articulateurs pour faire un « geste phonétique » (*phonetic gesture*, Liberman & Mattingly, 1985 ; voir aussi Browman & Goldstein, 1990) c'est-à-dire un geste articulatoire linguistique coordonné : « Essentially, phonetic

gestures are linguistically significant actions of structures of the vocal tract. […]. Physically, they are […] coordinated movements of the vocal tract that achieve a phonetically significant goal. » (Fowler & Saltzman, 1993, p. 172). Ainsi, pour produire un [p], c'est-à-dire faire un geste de fermeture bilabiale, le locuteur doit établir une coordination spatiale, un lien fonctionnel, entre la mâchoire, la lèvre inférieure et la lèvre supérieure : ces différents articulateurs sont, dans ce cas, coarticulés. Ainsi, même pour un phonème isolé qui ne subirait pas l'influence du contexte, il peut donc y avoir également coarticulation entre les différents articulateurs ; cette coarticulation est la conséquence de contraintes biomécaniques.

Le signal de parole est donc variable ; les entités abstraites, que sont les phonèmes, ne peuvent pas être décrites de manière unique au niveau acoustique et articulatoire car dans une séquence parlée, les signaux correspondants ne sont pas synchrones. Cette asynchronie est par ailleurs mise en évidence dans la théorie quantique de la parole proposée par Stevens (1989). Dans cette théorie, la relation entre les gestes articulatoires et les caractéristiques acoustiques des phonèmes est hautement non-linéaire ; dans la relation entre les paramètres acoustiques (par exemple, la fréquence relative des formants, les changements de fréquence fondamentale, etc.) et les paramètres articulatoires (la position et l'état des articulateurs du conduit vocal), il y a des régions où des grandes variations de paramètres articulatoires peuvent entraîner un effet acoustique minime, alors que dans d'autres régions, des petites variations articulatoires peuvent entraîner une très grande variation acoustique. De cette manière, il est difficile de segmenter le signal de parole en une suite unique de segments successifs. Les conséquences acoustiques de la coarticulation peuvent se manifester par exemple par des valeurs différentes de transitions de formants, des durées acoustiques plus longues (voir entre autres, Recasens, 1999, pour une technique d'analyse acoustique de la coarticulation). Au niveau articulatoire, les mouvements se chevauchent clairement : le geste de protrusion des lèvres, pour produire un [y] par exemple, peut anticiper le début acoustique de la voyelle et peut recouvrir ainsi plusieurs consonnes la précédant. De la même façon, bien que non visible directement, la position abaissée du vélum dans la production d'une consonne nasale en français peut persister durant la production d'une voyelle orale qui suit (Rossato et al., 2003). Ainsi, le phénomène de coarticulation peut être décrit de différentes manières selon l'articulateur principal en jeu dans les effets coarticulatoires (tous les articulateurs ne sont pas impliqués de la même façon selon les segments de parole produits. Pour une revue, voir Farnetani, 1997 ; Hardcastle & Hewlett, 1999). De la même manière, le recrutement des muscles responsables des gestes coarticulatoires diffère selon les segments produits. Il est à noter que la coarticulation labiale a été largement étudiée du fait de son accessibilité et de son observation directe (Farnetani, 1999). Par ailleurs, il a été montré que la

coarticulation, et plus précisément son étendue spatio-temporelle, pouvaient être modifiées par un certain nombre de facteurs, comme, par exemple, les caractéristiques suprasegmentales de la parole, telles que l'intonation et les frontières prosodiques, le rythme et le style de parole ; ce qui accentue d'autant la variabilité et la complexité de ce phénomène.

Au niveau temporel, la coarticulation peut se manifester de deux façons : elle peut avoir des effets d'*anticipation* (*anticipatory coarticulation*) ou des effets de *persévération* (*carryover coarticulation*). Dans certains cas en effet, les gestes articulatoires se chevauchent tellement que la réalisation articulatoire d'un phonème peut avoir une influence sur plusieurs segments précédents ou plusieurs segments suivants. L'anticipation coarticulatoire (« de droite à gauche ») est celle qui a été la plus étudiée, notamment au niveau des gestes labiaux qui sont directement observables (pour une revue, voir Cathiard, 1994 ; Farnetani, 1999). Dans ce cas-là, l'articulation d'un son subit l'influence d'un segment qui suit dans la chaîne parlée. C'est ce qui se produit pour la syllabe [tu] ; durant la production du [t], les lèvres sont déjà en position pour former l'arrondissement du [u]. La persévération (« de gauche à droite ») se manifeste par le fait que l'articulation d'un son subit l'influence d'un segment précédent. On peut observer ce phénomène dans l'articulation de la syllabe [ʃi] par exemple ; la protrusion-éversion labiale caractéristique du [ʃ] persiste durant l'articulation du [i].

Nous nous intéresserons plus particulièrement au versant anticipatoire de la coarticulation labiale. Plus concrètement au niveau temporel (voir une revue des travaux dans Cathiard, 1994), Benguérel et Cowan (1974) ont montré pour six locuteurs français que le geste de protrusion de la lèvre supérieure peut commencer jusqu'à six consonnes avant la production du [y] dans des séquences de type [iC_ny] (C_n représentant un nombre variable de consonnes), insérées dans des phrases du type « une sin**istre stru**cture » [istʁstʁy]. Ces résultats ont été retrouvés par Abry et Lallouache (1991) pour des séquences en français incluant cinq consonnes intervocaliques comme [ikstsky] dans « Ces deux Sixte sculptèrent » : le geste de protrusion labiale peut dans certains cas commencer dès la fin de la réalisation acoustique du [i] (voire durant sa production). Ils ont cependant mis en évidence une certaine variabilité intra-locuteur en montrant que le profil de mouvement de la protrusion de la lèvre supérieure pouvait varier (voir leur figure 2, p. 224, dans laquelle ils montrent trois exemples de profils différents) : l'arrondissement de la voyelle ne va pas systématiquement se rétro-propager à toutes les consonnes précédentes. Dans une étude sur la perception de l'anticipation du trait d'arrondissement, Cathiard (1994) a étudié les décours temporels du geste de protrusion de la lèvre supérieure et de la constriction, obtenus dans des transitions [i#y] (# représentant une pause prosodique silencieuse plus ou moins longue) produites par un locuteur français. Ces transitions étaient insérées dans des phrases

porteuses du type « *tu dis : UHI ise ?* ». Les deux conditions de pause s'élevaient en moyenne à160 ms pour les pauses courtes et 460 ms pour les pauses longues. Les résultats décrivent pour la transition [i#y] une forte anticipation du geste d'arrondissement des lèvres, avec une anticipation plus précoce de l'aire intérolabiale par rapport à la protrusion (de 40 ms). Le mouvement de constriction débute en moyenne 160 ms avant le début acoustique du [y] en condition de petite pause (soit dès la fin acoustique du [i]) et peut anticiper jusqu'à 240 ms en condition de longue pause. En résumé, le geste labial d'arrondissement vocalique anticipe clairement sur le début acoustique de la voyelle ; nous exposerons plus tard les différentes propositions de modélisation de ce phénomène (section II.3). Cathiard (1994) a également analysé l'anticipation de hauteur dans des transitions [i#a] (la durée de la pause # s'élevant en moyenne à 160 ms), incluse dans la phrase porteuse « T'as dit : AHA ase ? ». Pour cette transition, l'aperture des lèvres augmente, conjointement à un abaissement de la mandibule. L'étude de l'évolution temporelle de l'aire aux lèvres et d'un marqueur cutané placé sur le menton (donnant le mouvement propre de la mâchoire) met en évidence une anticipation sur le début du son des mouvements d'aperture des lèvres et d'abaissement de la mâchoire, ses deux mouvements étant synchrones. Plus précisément, le geste d'aperture débute en moyenne 180 ms avant le début acoustique du [a] (soit avant même la fin acoustique du [i]).

Les effets de coarticulation sont visibles sur la face mais sont-ils exploités par le sujet percevant ? En clair, l'anticipation labiale a-t-elle une efficacité perceptive ? Nous avons à ce propos déjà insisté sur les effets coarticulatoires dans la lecture labiale et la difficulté qui en résulte à établir des groupes de sosies labiaux réguliers (voir section I.2.3). Il semble donc bien que les effets de la coarticulation influencent la perception de la parole. Dans certains cas, la vision a un clair avantage sur l'audition pour ce qui est de percevoir l'anticipation (pour une revue détaillée, voir Cathiard, 1994). Cet auteur a testé en perception le décours temporel des transitions [i#y] (dont les caractéristiques articulatoires avaient été analysées), et montre, pour 25 sujets français, que l'anticipation d'arrondissement des lèvres, qui est récupérée par la vision seulement (du fait de la pause silencieuse), leur permet d'identifier correctement (à 95%) la voyelle arrondie jusqu'à 160 ms avant son début acoustique (la frontière à 50% sur les courbes d'identification pouvant se placer jusqu'à 210 ms avant le son). En outre, il apparaît que le début des courbes d'identification correspond d'une manière remarquable avec le pic d'accélération du geste de constriction. De la même manière, la dimension de hauteur peut être identifiée en vision seule en moyenne 160 ms avant le son, le démarrage des identifications étant également lié au pic d'accélération du mouvement labio-mandibulaire.

De nombreuses études se sont attachées à trouver les « règles » de ce phénomène : les effets de coarticulation sont-ils le résultat d'un contrôle à un haut niveau central ou bien découlent-ils simplement

d'une interaction entre les différentes contraintes biomécaniques des articulateurs en jeu ? C'est l'opposition classique entre une coarticulation « programmée » versus « mécanique » (Bonnot, 1990a), ou bien issue des théories « tout phonologique » versus « tout dynamique » (Perrier et al., 2004). Les effets anticipatoires de la coarticulation semblent être le résultat d'une planification active (Whalen, 1990), alors que les effets persévérants sont en général plutôt considérés comme un phénomène bas-niveau de la production de parole qui serait dû à l'inertie mécanique des articulateurs (Fowler, 1980 ; Recasens, 1984) : il est à noter cependant que la coarticulation persévérante dans certains cas est aussi considérée comme étant contrôlée (voir par exemple, Whalen, 1990 ; Rossato et al., 2003). Dans tous les cas, il semblerait que ces deux phénomènes soient basés sur un principe d'économie articulatoire (O'Shaughnessy, 2000 ; Perkell et al, 2002) ; en effet, le fait de positionner ses articulateurs de manière progressive sur plusieurs segments (par un mouvement d'anticipation ou bien un geste de retour à une position « neutre » des articulateurs) devrait nécessiter moins d'effort que de produire le geste de manière subite dans un court intervalle de temps. Entre un contrôle purement central et des contraintes purement biomécaniques des articulateurs, Perrier et al. (2004) proposent un compromis pour l'anticipation coarticulatoire : ce phénomène serait en majeure partie le résultat d'une stratégie optimale de contrôle, mettant en jeu des modèles internes des commandes motrices, les contraintes articulatoires n'ayant qu'un rôle secondaire. Le point crucial réside plutôt dans le critère optimal, celui qui tend à minimiser l'effort du locuteur tout en garantissant une certaine efficacité perceptive. Nous revenons donc au final à la proposition générale de Bonnot (1990a, p. 128) décrivant le système de production de la parole comme « une machinerie efficace, qui utilise l'ensemble de ses possibilités pour parvenir à un résultat optimal, en dépensant une énergie minimale ». L'auteur poursuit : « Seul un système très sophistiqué est capable d'atteindre ce but. C'est pourquoi il est possible de considérer que les mécanismes d'encodage "font feu de tout bois", faisant aussi bien appel à la préprogrammation qu'à une réévaluation locale du timing articulatoire en cours de production, et tirant partie des propriétés biomécaniques du tractus vocal ».

II.2. La production de la parole coarticulée

Comment la parole coarticulée est-elle produite et contrôlée ? Différents modèles ont été établis pour expliquer la coarticulation, ce phénomène intrinsèque à la parole qui se retrouve dans toutes les langues (bien que ses caractéristiques puissent varier d'une langue à l'autre). Nous ne ferons pas une revue complète des différents modèles et théories proposés (pour une analyse critique détaillée, voir Bonnot, 1990a ; voir également Marchal & Farnetani, 1993 ; Hardcastle & Hewlett, 1999 ; pour un aperçu historique, voir Bonnot & Keller, 2004) mais présenterons dans cette partie le modèle de coarticulation d'Öhman, mettant en évidence les rôles distincts des consonnes et des voyelles dans la

nature coarticulée de la parole, avant de détailler plus spécifiquement les différents contrôles du langage parlé. Nous nous baserons par la suite sur ces conceptions des contrôles de la parole pour tenter d'expliquer comment les mouvements manuels de la LPC se coordonnent avec cette parole coarticulée (voir section VI.3.2).

II.2.1. Le modèle de coarticulation d'Öhman

Dans les différents modèles, l'unité de production en parole est souvent la syllabe. Öhman (1966, 1967) propose un modèle de coarticulation dans lequel il met l'accent sur le rôle des voyelles dans la production de parole : la parole apparaît comme une suite d'articulations de voyelle à voyelle, sur lesquelles se superposent les articulations des consonnes (voir une schématisation de ce modèle pour le décours temporel de l'aire intérolabiale sur la Figure 8). Son modèle est issu d'analyses acoustiques et articulatoires de séquences voyelle-consonne-voyelle (V_1CV_2) isolées produites par un locuteur suédois (ces résultats ont également été comparés avec ceux d'un locuteur américain et d'un locuteur russe). L'auteur a étudié la coarticulation des consonnes voisées occlusives [b], [d] et [g] avec les voyelles [y], [ø], [a] et [u]. En observant le patron formantique des séquences VCV dans les différentes conditions, il constate que les transitions du deuxième formant sont assez variables selon le contexte et dépendent en fait de tous les segments, c'est-à-dire que le patron formantique de la portion V_1C dépend aussi de la dernière voyelle (V_2) qui est anticipée : « […] a motion toward the final vowel starts not much later than, or perhaps even simultaneously with, the onset of the stop consonant gesture » (1966, p. 165). Ainsi, les voyelles sont coarticulées à travers la consonne occlusive intermédiaire : les voyelles sont produites en continu, c'est-à-dire par un geste de voyelle à voyelle, sur lequel le geste de la consonne se superpose. Plus précisément, dans des séquences VCV, les deux voyelles sont produites par un geste relativement lent du corps de la langue (*tongue body*) depuis la position de la voyelle initiale à la position de la voyelle finale (*diphtongal movement*). Sur ce geste vocalique, le geste articulatoire de la consonne intermédiaire vient se superposer (« […] the stop-consonant gestures are actually superimposed on a context-dependent vowel substrate that is present during all of the consonantal gesture. », p. 165) et modifier ainsi le geste vocalique (« […] the tongue is able to make a distorted vowel gesture, while it is executing the stop consonant. », 1966, p. 166). D'après l'auteur, les articulations des voyelles et des consonnes sont indépendantes au niveau des instructions neurales et peuvent donc être activées simultanément : la langue correspondrait en fait à trois systèmes articulatoires séparés contrôlant trois sous-ensembles de muscles. La position du corps de la langue (*tongue body*) définit les voyelles, alors que les constrictions apicales et dorsales de la langue définissent les consonnes ([d] et [g] dans cette étude). La coarticulation résulterait donc de la co-production de la consonne et des voyelles sous forme d'une somme complexe des différentes

instructions : « […] the dynamic response of the tongue to a compound instruction is a complex summation (neural, muscular, and probably mechanical also) of the responses to each of the components of the instruction », (1966, p. 166).

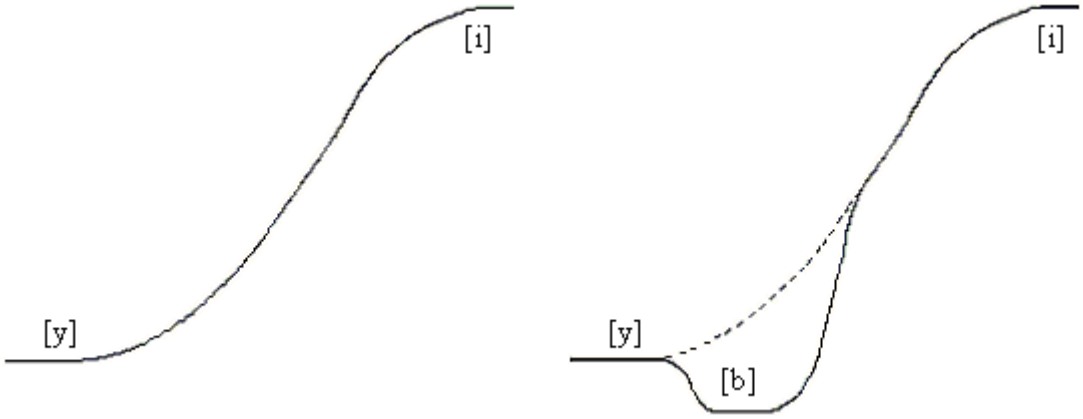


Figure 8. Schématisation du modèle d'Öhman montrant l'évolution dans le temps de l'aire aux lèvres pour un geste de voyelle à voyelle (transition de la voyelle [y] vers la voyelle [i] ; à gauche) et pour un geste vocalique avec consonne surimposée (transition [ybi] ; à droite) (figure tirée de Cathiard, 2003).

II.2.2. Syllabes et segments : les contrôles de la parole

En partant du modèle d'Öhman (1966) et en prenant également en compte un point de vue développemental (MacNeilage, 1998), Abry et al. (2002 ; voir aussi Abry et al., 2001 pour une version en français) proposent que la parole soit le résultat de deux types de contrôles (« […] the speech signal is biocybernetically a compound of (i) a carrier control, on proximal effectors, and (ii) a carried control, on distal end-effectors. », p. 228-229) :

– Le contrôle *proximal* de la porteuse de la parole (*carrier control*), la mandibule, qui produit le rythme syllabique. Au niveau du développement, c'est le premier contrôle à être acquis (durant le babillage canonique vers 7 mois) ; c'est ce que propose MacNeilage (Davis et MacNeilage, 1995 ; MacNeilage, 1998) dans sa théorie « Frame, then Content » pour expliquer l'évolution de la production de la parole humaine, au niveau de sa phylogenèse et de son ontogenèse. Ce rythme syllabique mandibulaire, qui maintient une certaine constance quelle que soit la vitesse d'élocution (6Hz±1, Sorokin et al., 1980), est à la base du contrôle de la parole humaine (« This initial rythmicity provides a basis for the control of speech throughout life. », p. 506).
– Le contrôle *distal* des articulateurs portés (*carried control*), la langue et la lèvre inférieure ; une fois le contrôle du cycle mandibulaire établi, ces articulateurs vont être contrôlés indépendamment de la mandibule et vont ainsi, en coopérant avec les autres articulateurs, créer des modulations sur ce cycle en produisant les *contenus* (MacNeilage, 1998) : les consonnes et les voyelles. Dans cette

théorie, le cadre et les contenus sont donc indépendants. Ceci permet d'expliquer l'acquisition non simultanée des différents types de contrôle dans la parole depuis le babillage canonique (aux alentours de 7 mois) jusqu'à la parole coarticulée de l'adulte (voir dans Vilain et al., 2000, une proposition en trois étapes pour le développement de la coarticulation de la parole, qui se ferait « […] from Frames, to Content, then to coarticulated Content », p. 84).

Le geste consonantique est le résultat d'un contrôle *local*, celui des constrictions, soient les contacts et les pressions des articulateurs sur différentes parties du conduit vocal (Vilain et al., 1999 ; Vilain, 2000). Par exemple, l'occlusion bilabiale caractéristique du [b] est le résultat d'une coordination entre la mâchoire et les deux lèvres : la lèvre inférieure est portée vers le haut par la mâchoire puis le contact et l'occlusion se produisent par le contrôle indépendant simultané de la lèvre inférieure qui exerce une pression sur la lèvre supérieure qui lui résiste (Tuller & Kelso, 1984). Le geste vocalique quant à lui est le résultat d'un contrôle postural *global* du conduit vocal qui permet de transiter d'une voyelle à l'autre (Öhman, 1967). Ce geste de voyelle à voyelle correspond à une mise en forme globale du conduit vocal, pouvant impliquer les différents articulateurs des lèvres au larynx. Passer d'une voyelle à l'autre consiste à modifier cette configuration globale. Par exemple, les contours sagittaux pour les séquences [ubu] et [aba] (Figure 9 et Figure 10 ; Vilain, 2000) montrent deux configurations globales différentes du conduit vocal : pour la séquence [ubu], les lèvres sont protruses, la mandibule est en position haute, le dos de la langue est monté vers la région vélaire et le larynx est abaissé ; pour la séquence [aba], la mandibule est en position basse, la langue est centrale, abaissée et plate et le larynx est élevé. Dans les deux cas, nous pouvons constater que le geste consonantique du [b], qui vient se superposer sur le geste vocalique (Öhman, 1967), correspond à un contrôle local du contact des deux lèvres. La forme interne globale du conduit vocal reste identique sur toute la séquence ; bien que la mandibule monte dans [aba] pour produire l'occlusion bilabiale du [b], la langue reste en position basse pour la voyelle. Le contrôle global de la voyelle est donc constant à travers la production de la consonne. En résumé, la production d'une voyelle implique le contrôle de tout le corps de la parole (lèvre, mandibule, langue, larynx) alors que la production d'une consonne est contrôlée localement (certains segments seulement sont impliqués).

En ce qui concerne le *phasage* de ces différents contrôles (Browman & Goldstein, 2000 ; Sato et al., 2002), les contrôles de la mandibule et de la mise en forme globale du conduit vocal pour la voyelle sont *en phase*. Dans le cas d'une consonne, le contrôle de la constriction est en phase avec la voyelle quand la consonne est en position d'attaque de syllabe (syllabe CV), les deux gestes pouvant être effectués en synchronie, mais il peut aussi être déphasé par rapport à la voyelle quand la consonne est en position de coda de la syllabe (syllabe VC) (le geste consonantique ne peut alors pas être anticipé

durant la voyelle). Pour les groupes consonantiques en position d'attaque ou de coda, le contrôle des constrictions peut être en phase (par exemple, [psa] ou [aps] : l'ouverture des lèvres durant l'explosion du [p] et le placement de la position antérieure de la langue vers les alvéoles pour la constriction du [s] sont partiellement synchrones) ou déphasé (par exemple, [spa] ou [asp]), l'ensemble pouvant être en phase avec la voyelle ([psa] ou [spa]) ou déphasé ([aps] ou [asp]) (Sato, 2004).

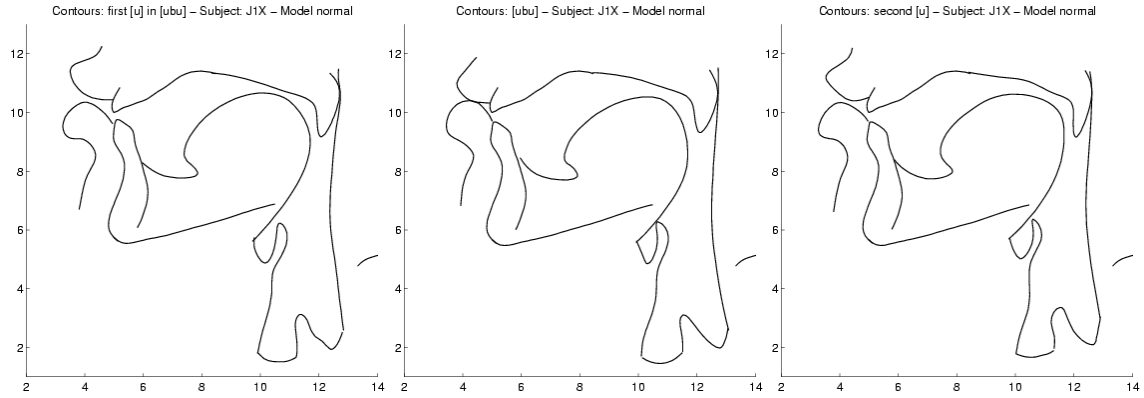


Figure 9. Contours sagittaux pour la production de [u], [b], [u] dans [ubu]. Figure tirée de Vilain, 2000.

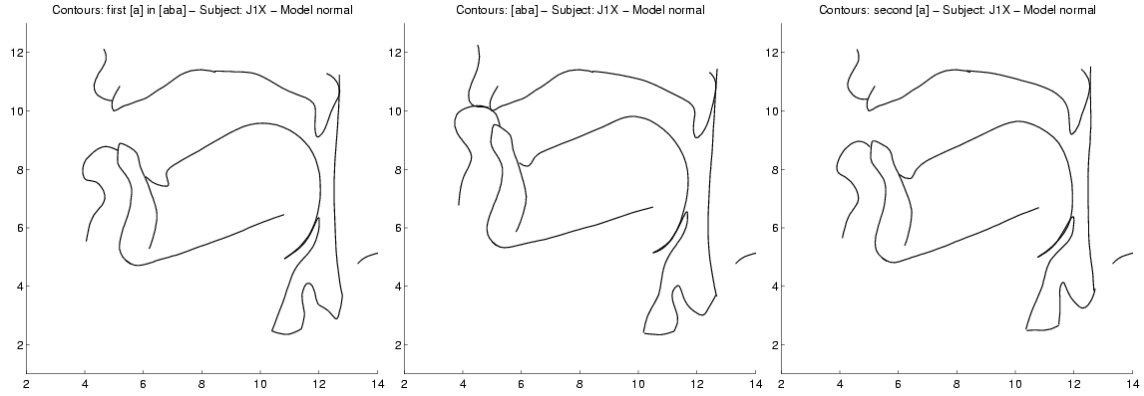


Figure 10. Contours sagittaux pour la production de [a], [b], [a] dans [aba]. Figure tirée de Vilain, 2000.

II.3. Les modèles d'anticipation labiale

Il est intéressant de décrire les différents modèles qui ont été établis pour expliquer le phénomène d'anticipation labiale, qui ont toute leur importance en raison de notre intérêt pour la lecture labiale. Ce phénomène peut en effet donner des indices sur la façon dont les unités linguistiques sont organisées au cours du temps dans la parole. Nous examinerons par la suite, dans certaines de nos expériences, la situation de parole codée dans ce cadre général (CHAPITRE VII). Les comportements labial et manuel observés pourront ainsi être comparés à ces modèles de production afin de comprendre le contrôle de la coproduction main-parole dans un tel contexte.

II.3.1. Trois modèles de production pour l'anticipation

L'anticipation vocalique d'arrondissement a été largement étudiée sur le plan articulatoire, à partir du geste de la lèvre supérieure, pour des séquences $V_1C_nV_2$, où V_1 est une voyelle étirée comme [i], V_2 une voyelle arrondie comme [u] et C_n, une ou plusieurs consonnes.

Le modèle dit « *Look-ahead* » de Henke (1967) prédit que le geste de protrusion/arrondissement du [u] dans des séquences [iC_nu] peut démarrer dès la fin acoustique du [i]. Ce modèle s'appuie sur la caractéristique d'arrondissement des phonèmes : les segments caractérisés par le trait arrondi sont notés par la valeur « + », ceux caractérisés par le trait non arrondi sont notés « - » et les segments non spécifiés (*unspecified*) au niveau de ce trait sont notés « 0 ». Dans la production d'une suite de phonèmes, la caractéristique + ou - arrondie pour un phonème va s'étendre à tous les segments précédents qui ne sont pas spécifiés, c'est-à-dire à tous ceux qui ont la valeur 0. Ainsi, dans une séquence [iC_nu], dans laquelle les consonnes sont « neutres » du point de vue de l'arrondissement, l'arrondissement du [u] va commencer dès la fin du [i]. Benguérel et Cowan (1974) ont par ailleurs démontré, pour des locuteurs français, que le geste de protrusion des lèvres, dû à la voyelle arrondie [y] ou [u], peut commencer à la première consonne d'une suite de six consonnes successives dans des séquences de type [i$C_{1...6}V_{arrondie}$]. En outre, dans des séquences de ce type n'incluant que quatre consonnes intervocaliques, le geste de protrusion peut même démarrer durant la production de la voyelle non arrondie [i]. Ainsi dans ce modèle, le geste de protrusion est déterminé par le contexte acoustico-phonétique (dans la mesure où la caractéristique d'arrondissement s'étend à tous les segments précédents jusqu'à rencontrer une articulation opposée) et augmente donc avec la longueur de la séquence consonantique.

Le modèle « *Time-locked* » (aussi appelé *frame model* ou *modèle de coproduction,* voir Boyce et al., 1990) proposé par Bell-Berti et Harris (1981, 1982) prédit au contraire, pour des séquences [iC_nu], que le geste de protrusion des lèvres pour former la voyelle [u] ne dépend pas de la fin acoustique de la voyelle non arrondie, mais commence à date fixe avant le début acoustique du [u], quel que soit le nombre de consonnes intervocaliques. Le début du geste de protrusion peut éventuellement dépendre du rythme de parole mais en aucun cas de la longueur de la séquence consonantique intervocalique. Dans ces séquences, c'est donc l'intervalle entre la fin acoustique du [i] et le début du geste de protrusion du [u] qui varie ; le geste de protrusion est quant à lui invariant. D'après ces auteurs, le geste de protrusion est relié essentiellement à l'articulation de la voyelle arrondie (à sa dynamique gestuelle) et ne dépend pas du contexte avoisinant, et donc ne s'étend pas à l'articulation des consonnes non arrondies comme dans le modèle Look-ahead, mais ce geste est en fait « co-produit » avec l'articulation des consonnes (Fowler, 1980) : « […] the specification of lip position for the

consonants is not altered by a migrating vowel feature. Instead, […], we see the vowel-rounding gesture beginning at a relatively fixed time before the acoustic onset of the vowel and simply co-occuring with some portion of the preceding lingual consonant articulations. » (Bell-Berti & Harris, 1982, p. 454). Dans ce cas, il semblerait que ce ne soit pas le phénomène de coarticulation qui soit planifié mais plutôt le début de la voyelle (Whalen, 1990). Il est à noter que les auteurs (Bell-Berti & Harris, 1982) ont également observé ce patron temporel fixe pour les effets de persévération.

Perkell et Chiang (1986), s'inspirant des résultats obtenus par Bladon et Al-Bamerni (1982) pour la coarticulation vélaire (ces auteurs ont observé deux patrons différents pour le mouvement d'anticipation vélaire des consonnes nasales : un geste régulier et un geste à deux étapes), proposent un « *modèle hybride* » à deux phases pour le geste de protrusion labiale. Ils observent sur un locuteur américain, prononçant des séquences du type [$V_{non\text{-}arrondie}C_n\#C_mu$], un geste de protrusion anticipatoire des lèvres qui se fait en deux étapes : une première étape lente et progressive et une seconde étape plus rapide et plus importante, les deux étapes étant délimitées par un *point d'inflexion* sur la trajectoire correspondant au pic d'accélération le plus important. La première phase du geste suivrait le modèle Look-ahead et serait donc initiée dès que les contraintes du contexte phonétique avoisinant le permettent. La deuxième phase du geste serait en revanche indépendante du contexte (et donc de la longueur de la séquence consonantique intervocalique) et serait initiée à date fixe avant le début acoustique de la voyelle arrondie, se conformant ainsi plutôt au modèle Time-locked. Perkell (1986) observe également ce patron « segmenté » du geste de protrusion labiale chez des locuteurs américains, français et espagnols. Il interprète ce phénomène comme étant le reflet d'une *co-production* (Fowler, 1980) entre les segments adjacents : le mouvement entier serait le résultat de la « somme » des trajectoires articulatoires invariantes qui se chevauchent en partie (« […] it is possible to conceive of the overall movement pattern as being the result of some kind of 'summation' of overlapping, relatively invariant trajectories for each of the segments in the string. », p. 65).

Pour des séquences $V_{non\text{-}arrondie}C_nV_{arrondie,}$ ces trois modèles prédisent donc une coarticulation labiale différente (voir une représentation des différentes prédictions sur la Figure 11). Perkell (1990) a testé ces trois modèles chez quatre locuteurs américains produisant des transitions [iC_nu] dans des phrases de type « It's a lee coot again » (n=1) ou du type « It's a leaked scoot again » (avec n=4 consonnes intervocaliques). L'auteur observe chez ces locuteurs une grande variabilité dans le déploiement spatio-temporel du geste de protrusion labiale, mais il remarque cependant que ce geste de protrusion est la plupart du temps divisé en deux phases, une phase initiale lente et une seconde phase plus rapide, comme le proposait le modèle « hybride » (voir aussi Perkell & Matthies, 1992). Il exprime les différentes prédictions des modèles sous la forme de relations entre des intervalles temporels : (i)

l'intervalle acoustique, temps entre la fin acoustique du [i] et le début acoustique du [u] (*acoustic interval*), (ii) l'intervalle de mouvement, temps entre le début du mouvement de protrusion des lèvres et le début acoustique du [u] (*movement interval*) et (iii) l'intervalle d'accélération, temps entre le pic d'accélération et le début acoustique du [u] (*acceleration interval*). Les prédictions des trois modèles peuvent alors être représentées sous forme de relations linéaires entre deux intervalles.

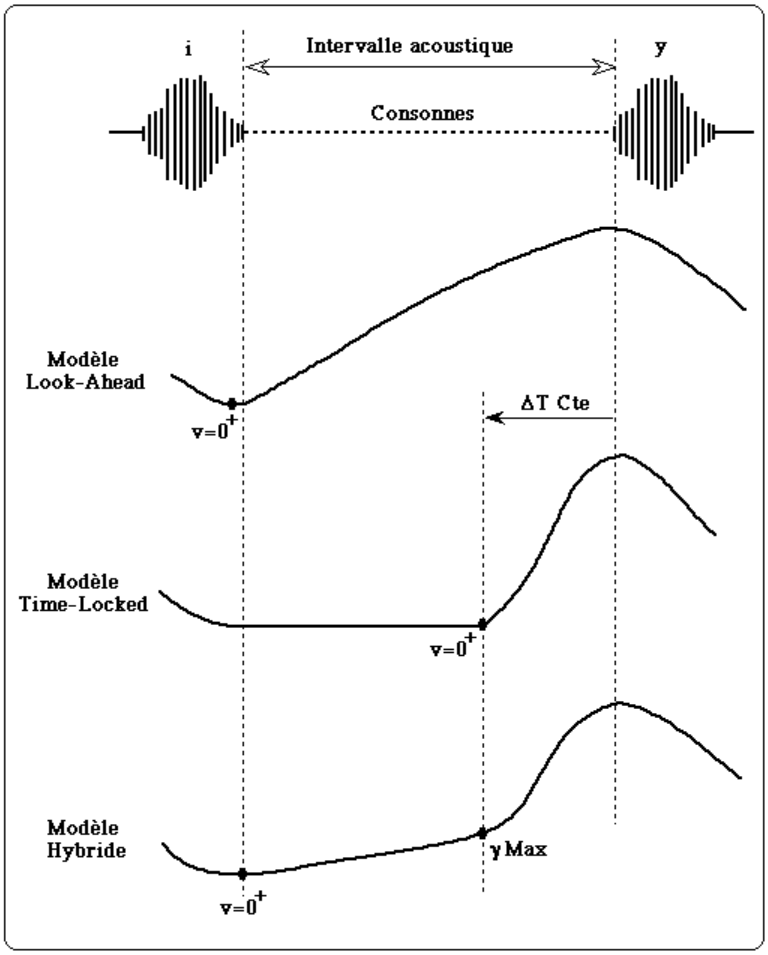


Figure 11. Représentation schématique des différentes prédictions des trois modèles d'anticipation du geste de protrusion labiale pour une séquence du type [iC_ny] (signal acoustique schématisé en haut). Chaque tracé représente la protrusion labiale de la lèvre supérieure en fonction du temps. Les points noirs indicés de $v=0^+$ correspondent au début du geste de protrusion, déterminé au moment où la vitesse de la courbe passe à 0. Le point noir indicé de γMax indique le moment du pic d'accélération, qui détermine le début de la 2ème phase du geste de protrusion dans le modèle hybride. (Figure tirée de Cathiard, 1994).

En ce qui concerne l'intervalle de mouvement, pour les modèles *Look-ahead* et *Hybride*, il devrait augmenter linéairement en fonction de l'intervalle acoustique (avec une pente de 1) et être indépendant (pente à 0) pour le modèle *Time-locked*. En revanche, l'intervalle d'accélération devrait être indépendant de l'intervalle acoustique (pente à 0) pour le modèle *Hybride*. L'auteur mesure donc ces différentes variables chez les locuteurs testés et s'aperçoit qu'en fait, aucun des modèles proposés ne peut expliquer entièrement la coarticulation labiale observée dans cette expérience : il rejette donc

ces trois modèles, du moins dans leur version forte. Perkell et Matthies (1992) proposent un compromis pour l'anticipation coarticulatoire du geste de protrusion labiale qui subirait une double influence du contexte acoustico-phonétique et des propriétés dynamiques des mouvements. Ils proposent finalement la compétition de trois contraintes pour ce geste de protrusion des lèvres : (1) le locuteur doit utiliser une durée préférentielle du geste (cette contrainte est en accord avec les modèles de coproduction comme le modèle Time-locked), (2) le geste de protrusion peut commencer dès que la contrainte pour la voyelle non arrondie est relâchée (cette contrainte est en accord avec le modèle Look-ahead) et (3) le geste de protrusion doit se terminer pendant la partie voisée de la voyelle arrondie (cette contrainte est en accord avec les deux modèles).

II.3.2. Le Modèle d'Expansion du Mouvement (M.E.M.)

L'analyse de l'évolution temporelle du geste de protrusion de la lèvre supérieure chez trois locuteurs français a amené Abry et Lallouache (1995a) à rejeter également ces trois modèles et à proposer une alternative pour le français, le « Modèle d'Expansion du Mouvement » (*Movement Expansion Model*) ou MEM. Formulé à la base pour le geste de protrusion labiale (MEM de protrusion), ce modèle peut également expliquer le timing de la constriction labiale, c'est-à-dire la diminution de l'aire intérolabiale (MEM de constriction ; Abry et Lallouache, 1995b).

II.3.2.1. Le MEM de protrusion

Le geste de protrusion de la lèvre supérieure a été analysé dans des séquences [iC_ny] (où C_n représente de 0 à 5 consonnes) insérées dans une phrase porteuse du type « Ces deux S**ixte scu**lptèrent ». Dans le cas de 0 consonne, la transition simple vocalique [iy] est étudiée (« ces deux scies utèrent ») puis [iky] (« ces deux scies cutèrent »), [ikky] (« ces deux Sikhs cutèrent »), [iksky] (« ces deux Sikhs sculptèrent »), [ikssky] (« ces deux Sixes sculptèrent ») jusqu'à « ces deux Sixtes sculptèrent » avec cinq consonnes intervocaliques. Les auteurs ont examiné le timing du geste de protrusion en relation avec l'*intervalle d'obstruence* (IO) déterminé par la fin acoustique du [i] et le début acoustique du [y]. Ils ont observé le timing relatif de différents événements temporels : le début du mouvement de protrusion, le maximum d'accélération, le maximum de vitesse et le maximum de protrusion. Chez les trois locuteurs, le début du mouvement peut se produire avant la fin acoustique du [i] et le maximum de protrusion (c'est-à-dire la fin du geste) se produit aux alentours du début acoustique de la voyelle arrondie [y]. En analysant en particulier la durée du mouvement en fonction de l'intervalle d'obstruence IO, les auteurs observent que la durée moyenne pour effectuer la transition simple [iy] sans consonne intermédiaire (dans ce cas IO=0 ms) est proche de celle nécessaire pour effectuer les séquences avec une consonne [iky] (dans ce cas IO=100 ms) et avoisine les 140-150 ms.

Cette valeur minimale est la durée incompressible du geste de protrusion, obtenue chez les trois locuteurs, et à partir de laquelle le mouvement, soit l'anticipation, va pouvoir s'étendre de manière linéaire en fonction de l'augmentation de l'intervalle consonantique (soit de la durée IO) (voir Figure 12). Cette augmentation se fait selon un « coefficient d'expansion du mouvement » qui est propre à chaque locuteur. Ainsi le mouvement serait « expansible sans être compressible très en deçà d'une constante [iy] » (Abry & Lallouache, 1995a, p. 97). En réponse aux différents modèles proposés, les auteurs concluent donc que « l'anticipation du mouvement de protrusion n'est pas déterminée par la fin de la voyelle non arrondie [i], pas plus qu'elle n'est déterminée de manière fixe par rapport au début de la voyelle arrondie [y]. » (p. 97). En revanche « […] *phénoménologiquement*, le mouvement de protrusion : (i) atteint son max. plus ou moins aux alentours du début de la voyelle arrondie [y] ; (ii) commence de plus en plus tôt, par rapport à [y], en fonction de l'augmentation du nombre de consonnes intervocaliques […] ; (iii) peut commencer après [i] […], ou dès le début de cette voyelle […]. » (p. 97).

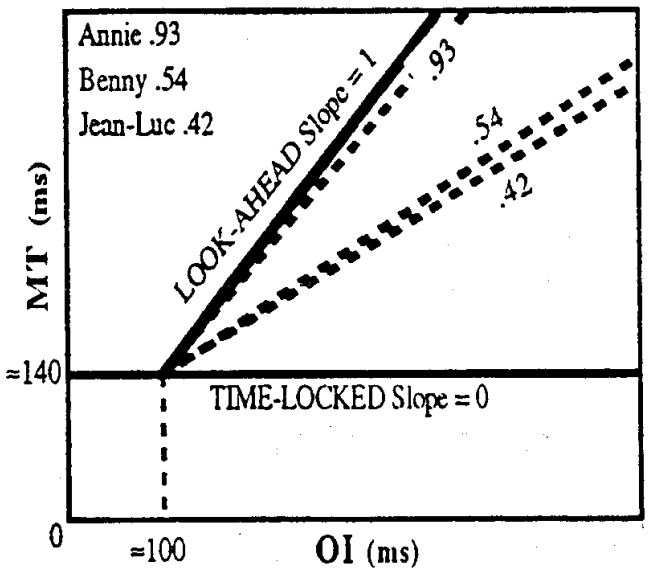


Figure 12. MEM de protrusion en relation avec les modèles Look-Ahead et Time-Locked : représentation de la durée du mouvement (MT) en fonction de l'intervalle d'obstruence IO. La durée de 140 ms désigne la durée minimale incompressible du geste de protrusion (pour une durée de IO de 0 et 100 ms). La pente de la droite représente le coefficient d'expansion du mouvement. Ce coefficient est propre au locuteur (ici, cas de trois locuteurs : Annie, Jean-Luc et Benny). Figure tirée de Abry et al., 1996a.

II.3.2.2. Le MEM de constriction

Les auteurs ont montré dans une autre étude utilisant le même corpus (mais avec un locuteur supplémentaire) (Abry & Lallouache, 1995b) que ce modèle peut aussi être appliqué à la constriction labiale pour la dimension d'arrondissement (c'est le seul modèle qui tient compte du timing de la constriction aux lèvres). Cette extension du MEM à la constriction se justifie par le fait que certains locuteurs n'ont pas de protrusion labiale (c'était le cas du 4ème locuteur étudié ici, Christophe). La diminution de l'aire aux lèvres est en revanche toujours présente pour la production de voyelles arrondies. Rappelons également que l'aire aux lèvres est un paramètre particulièrement pertinent, notamment sur le plan acoustique pour le maintien des effets de l'arrondissement (nous insisterons plus tard sur ce point pour justifier du choix de ce paramètre pour nos propres données). Sur les décours d'aire intérolabiale, les événements temporels suivants ont été repérés (voir Figure 13) : le maximum d'aire (pour le [i], événement 1 sur la figure) et le minimum d'aire (pour le [y], événement 4), ces deux événements donnant l'amplitude de la constriction ; les deux instants avant et après le minimum d'aire, où l'aire atteint 10% de l'amplitude (notés 10%aire.on et 10%aire.off, événements 3 et 5), ces deux événements délimitant une phase de tenue (*hold phase*) acoustiquement efficace du [y] ; et l'instant avant l'atteinte du maximum d'aire où l'aire atteint 90% de l'amplitude (90%aire.on, événement 2), cet instant représentant le début de la constriction du [y]. La phase délimitée par les instants 90%aire.on et 10%aire.on constitue le *Time falling* (TF), qui représente en fait le temps d'établissement du geste d'arrondissement des lèvres. La phase globale TF+H constituée par le time-falling et la tenue (*hold*) est étudiée en fonction de l'intervalle d'obstruence IO. Le timing de cette phase est très similaire aux résultats observés pour la durée du geste de protrusion : on retrouve une expansion linéaire de la durée de la constriction, à partir d'une constante minimale (environ 140 ms), et qui va croître différemment selon un coefficient propre à chaque locuteur (voir Figure 14). En observant le timing relatif du début de constriction (90%aire.on), les auteurs constatent la forte ressemblance comportementale avec le début du geste de protrusion : le début de la constriction pour le [y] peut se produire dans le [i] pour des durées petites de IO, mais se produit après le [i] pour des grandes durées de IO (typiquement supérieures à 300 ms pour cinq consonnes).

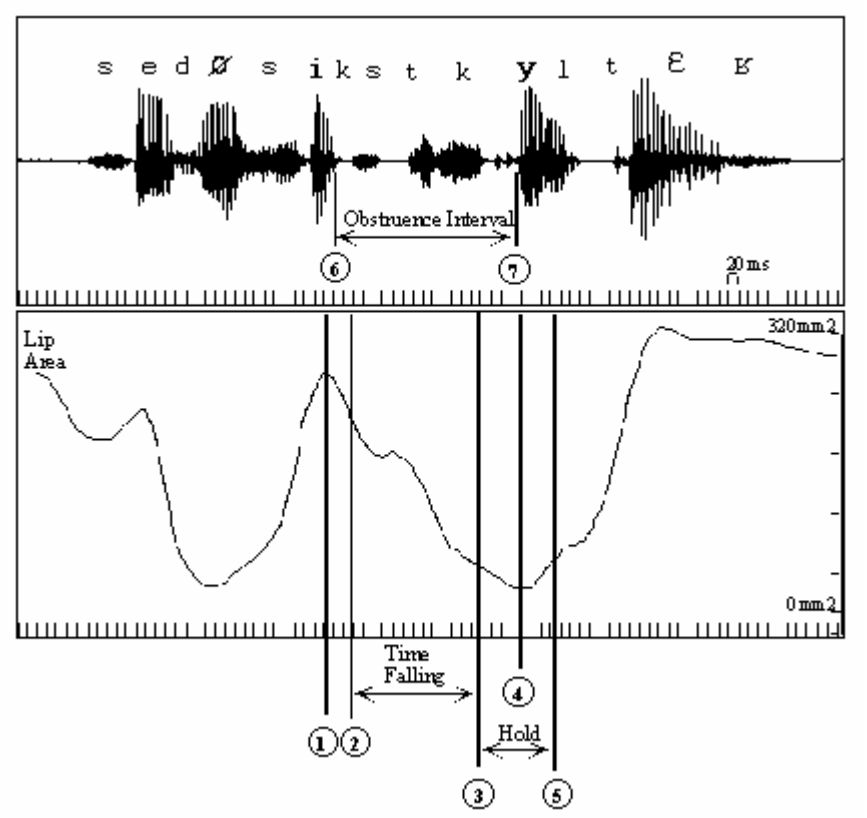


Figure 13. Signal acoustique et évolution temporelle de l'aire aux lèvres pour la séquence [sedøsikstkyltɛʁ]. Les événements temporels suivants sont repérés : (1) correspond au maximum de [i], (2) au 90%aire.on, (3) à 10%aire.on, (4) au minimum de [y], (5) à 10%aire.off, (6) à la fin acoustique du [i] et (7) au début acoustique du [y]. D'après C. Abry, publié dans Cathiard et al., 2003.

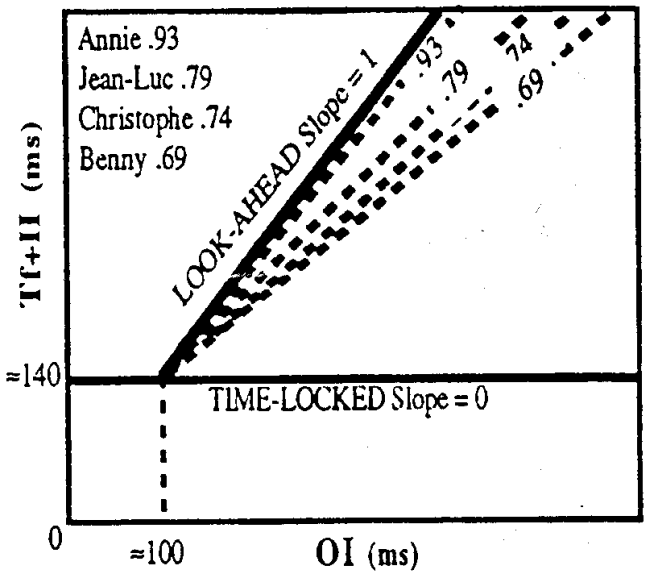


Figure 14. MEM de constriction : représentation de la phase de Time-Falling+Hold en fonction de la durée de l'intervalle d'obstruence pour quatre locuteurs. Figure tirée de Abry et al., 1996a.

La modélisation du geste d'arrondissement en protrusion et en constriction par le MEM nous donne donc une durée minimale du mouvement ainsi qu'une fonction d'expansion propre à chaque locuteur.

L'anticipation vocalique peut s'étendre au travers d'une suite de consonnes (sans forcément aller jusqu'au segment non arrondi), à un rythme propre à chaque locuteur.

Notons en outre que le MEM a également été testé pour le geste de base de voyelle à voyelle [i#y] sans consonne intermédiaire (Abry et al., 1996a). En effet, il apparaît que la fonction d'expansion calculée pour les réalisations ne contenant pas la consonne [s] (donc pour les transitions [iy], [iky] et [ikky] dans le corpus précédent ; le [s] est exclu car il semble bien influencer la phase de *Time-Falling* durant la constriction par un recrutement de la mâchoire propre au locuteur) peut également rendre compte du comportement de constriction labiale contrôlant l'aire à la sortie du conduit vocal pour des transitions [i#y] avec petite (100-150 ms), moyenne (150-300 ms) et longue pauses (450-650 ms ; notons cependant une plus grande variabilité pour ces longues pauses) (voir Fig. 2 et 3 dans Abry et al., 1996a). Le geste de constriction suit alors une fonction d'expansion en fonction de l'intervalle de pause selon un coefficient de 0,16 pour le locuteur testé (Jean-Luc).

Concernant l'anticipation de hauteur mise en évidence dans des transitions [i#a], le comportement est plus variable, lié à des stratégies articulatoires prosodiques différentes (Abry et al., 1996a). Un articulographe a été utilisé pour mesurer le mouvement propre du dos de la langue ainsi que le geste d'aperture dû à l'abaissement de la langue et de la mâchoire (dans cette expérience, l'évolution temporelle de l'aire intérolabiale n'a pas pu être mesurée). La phase de *Time-Falling* (suivant la même procédure que pour la constriction, de 90% à 10% de l'amplitude du mouvement) a été analysée en fonction de la durée de la pause intervocalique. Pour le mouvement d'aperture, on trouve une fonction d'expansion qui débute dans le passage entre courte et moyenne pause (soit aux environs de 300 ms), et qui reste valable pour une partie seulement des réalisations en longue pause. Notons que les longues pauses (supérieures à 500 ms) démontrent une forte variabilité : de fait, la majorité d'entre elles adoptent un coefficient d'expansion plus bas (voir Fig. 4 dans Abry et al., 1996a). Pour le mouvement propre de la langue, la tendance est encore différente : l'expansion du mouvement commence plutôt à partir de 400 ms (moyennes pauses). Ces différences viennent du comportement du locuteur qui adopte des stratégies bien différentes quand la pause s'allonge : le contrôle articulatoire de la prosodie durant ces longues pauses peut changer le profil du mouvement (ceci avait été également observé pour le contrôle de la jointure dans les suites de consonnes, Abry et Lallouache, 1991).

Ainsi le MEM, proposé comme une alternative aux autres modèles de l'anticipation, permet de rendre compte de la variabilité interlocuteur en affectant à chacun un coefficient d'expansion du mouvement spécifique à chaque comportement d'anticipation : la durée du mouvement est expansible, en fonction

de l'intervalle disponible entre les voyelles, selon un coefficient, dépendant du locuteur. En clair, l'initiation du geste vocalique ne dépend pas de la fin acoustique du segment non arrondi pas plus que du début du segment arrondi. Il n'en reste pas moins que ce contrôle est *orienté-vers-la-sortie,* comme le témoignent les études en perception qui ont validé le MEM à la fois au niveau acoustique et visuel (Cathiard, 1994 ; Abry et al., 1996b ; Ferbach-Hecker et al., 2001). Citons pour finir les mots de l'auteur du modèle C. Abry : « L'anticipation reste bien au contraire, selon le mot de Keele *et al.* (1990), une *pré-connaissance*, qui permet de réguler l'initiation des composantes de protrusion et constriction du geste vocalique, en fonction de la durée prosodique des éléments non vocaliques qui composent la séquence à exécuter – le coefficient d'expansion temporelle du mouvement de voyelle à voyelle étant supposé connu du locuteur. Il n'est donc pas nécessaire de connaître la fin d'un son (*look ahead*) ou d'un geste (*time-locked*) précédent pour commencer le suivant : il suffit de connaître l'empan temporel disponible pour l'extensibilité de son anticipation, un empan donné ici par le pas des consonnes entre les voyelles. » (Abry & Perrier, 1996).

CHAPITRE III.
Gestes et parole

« Gestures are an integral part of language as much as are words, phrases, and sentences – gesture and language are one system »

McNeill, 1992

Dans cette thèse, nous nous intéressons à la coordination entre les mouvements de la main et les gestes de la parole dans la production du code LPC. Avant d'aborder ce cadre particulier, il est intéressant de voir comment, dans la communication de tous les jours, des gestes naturels interviennent spontanément au cours de la parole, les gestes co-verbaux, et se coordonnent avec elle. Nous avons souligné précédemment le caractère multimodal de la parole ; la parole n'est pas seulement audible, mais elle est aussi visible. De plus, les gestes produits par le locuteur pendant qu'il parle ont également un rôle important. Nous allons voir dans ce chapitre que dans la communication parlée, les gestes et la parole sont liés par une forte interdépendance. Gestes et parole se coordonnent naturellement d'une manière particulière dans l'acte de communication spontanée. Nous verrons que la parole peut parfois s'ajuster à la durée du geste soit en se calant sur le geste et en l'attendant soit en se laissant entraîner par le rythme gestuel.

III.1. Une communication multimodale

« Gestures and speech are closely linked in meaning, function, and time: they share meanings, roles, and a comon fate »
McNeill, 1992

La parole permet pleinement de communiquer. En addition, les conduites non verbales (gestes, mimiques faciales, postures…) qui peuvent révéler certaines de nos émotions et pensées ont un rôle important dans l'interaction humaine. La gestualité et en particulier ses relations avec la parole ont fait l'objet de nombreuses études, la sémiologie du geste, sa fonction et sa synchronie avec la parole étant au centre des préoccupations. A l'exception des gestes autocentrés (gestes de grattage, gestes de confort, etc.) et des gestes ludiques tournés vers les objets (ex : jouer avec un stylo) qui se produisent durant la communication mais qui n'ont pas réellement une fonction communicative, les autres gestes sont en lien avec la parole par leurs significations, leurs fonctions et leurs relations temporelles et peuvent éclaircir la communication. Ils révèlent l'imagerie de la pensée du locuteur ; ainsi, en même temps, les gestes et les images sous-jacentes coexistent avec la parole (McNeill, 1992). On distingue généralement les gestes qui peuvent remplacer la parole et qui ont une signification hors contexte – ce sont les *emblèmes* qui peuvent être utilisés seuls – des gestes qui accompagnent la parole et qui sont produits en même temps que le locuteur parle, les *gestes co-verbaux* (pour plus de détails sur la classification des gestes par différents auteurs, voir McNeill et al., 1990).

III.1.1. Les emblèmes

Les emblèmes sont des gestes conventionnels (*codified* ou *conventionalized forms*) à forme standard (selon la culture) qui peuvent être utilisés indépendamment de la parole. Ils sont autonomes et porteurs de sens à eux seuls. Par exemple pour signaler à quelqu'un de s'arrêter, on peut lui dire « stop » tout comme on peut faire un geste de la main (main ouverte tendue face à l'interlocuteur) pour lui signifier de s'arrêter, mais on peut aussi faire les deux en même temps. Les emblèmes ont une fonction de communication : le locuteur exécute ces gestes de manière totalement consciente pour exprimer son intention. Il est à noter que ces gestes, bien que pouvant se substituer à la parole, ne forment cependant pas un système linguistique à part entière tel que la langue des signes pour les sourds par exemple ou les différents systèmes gestuels développés dans différentes communautés isolées (Kendon, 1997 ; pour plus de détails, voir Goldin-Meadow, 1999). Contrairement aux unités gestuelles de la langue des signes, les emblèmes ne se combinent pas entre eux en suivant des règles grammaticales pour former un énoncé.

III.1.2. Les gestes co-verbaux

Les gestes co-verbaux se distinguent des emblèmes par leur lien direct avec la parole. Ces gestes ne peuvent pas se substituer à la parole et ne peuvent pas être interprétés hors contexte. Ils se produisent de manière spontanée pendant que le locuteur parle. Ils entretiennent différents types de relation avec la parole qu'ils accompagnent au niveau phonologique, sémantique, syntaxique et pragmatique (voir McNeill et al., 1990 pour un détail sur les différents gestes co-verbaux et leurs relations avec la parole). Les gestes co-verbaux se rangent dans différentes catégories selon les entités, les actions, les espaces, les concepts auxquels ils font référence (classification de McNeill, 1992) :

– Les gestes iconiques (*iconic gestures*) : ces gestes ont un lien étroit avec le contenu sémantique de la parole ; ils peuvent avoir la même signification que le contenu verbal ou bien une signification complémentaire. Ils font référence à des entités concrètes (événements, objets ou actions) dont ils dépeignent certains aspects qui sont en même temps décrits dans la parole qu'ils accompagnent. Par exemple, le locuteur pourra faire un mouvement de la main vers le haut pendant qu'il décrira une personne très grande. La forme de ces gestes dépend strictement du contenu sémantique de la parole ; elle est donc différente selon l'entité décrite. La forme de ces gestes dépend également du point de vue du locuteur c'est-à-dire comment il se place par rapport à l'événement qu'il décrit ; il peut être observateur et décrire une scène d'un point de vue externe ou bien acteur et décrire la scène en jouant le rôle du personnage.

- Les gestes métaphoriques (*metaphoric gestures*) : ces gestes sont comme les gestes iconiques, mais diffèrent dans la nature de l'entité imagée qu'ils décrivent : ils font référence à une image mentale d'un concept abstrait, de connaissances (par exemple, le concept de « fou ») ;
- Les gestes bâtons ou battements (*beat gestures*) : ces gestes sont des rapides et petits mouvements bi-phasiques de la main ou des doigts qui marquent le rythme de la parole (souvent des mouvements de haut en bas ou d'avant-arrière). Souvent, ils ne représentent rien (leur fonction n'est pas signifiante en général) et ont donc le même aspect quel que soit le contenu de la parole. Ils ont en général une fonction pragmatique dans le discours (pour faire le lien entre différentes parties du discours, pour introduire un nouvel élément dans le discours ou pour mettre en évidence certains points importants) mais peuvent aussi se référer à la notion temporelle du passé (Boyer, 2001) ;
- Les gestes déictiques ou gestes de pointage (*deictic gestures*) : ils indiquent en pointant (avec la main ou souvent l'index) des unités concrètes ou abstraites (physiquement absentes) dans l'espace de communication. Par exemple, durant un discours, le locuteur peut faire référence à un objet présent dans la pièce et le montrer en pointant du doigt (« c'est ce téléphone qui n'arrête pas de sonner »). Il peut également pointer des régions dans l'espace sans viser particulièrement un objet présent dans la pièce ; il se réfère alors à des entités abstraites auxquelles il attribue une position dans son discours (c'est par exemple souvent le cas lorsqu'on raconte une histoire et qu'on définit dans l'espace des lieux différents). Ces gestes peuvent également avoir une fonction pragmatique (Kelly et al., 1999).

III.1.3. Les fonctions du geste

Tous ces gestes ne sont pas des mouvements aléatoires qui interviendraient au cours du discours : ils ont une signification liée à l'intention du locuteur et sont par ailleurs interprétés correctement et de manière fiable par l'interlocuteur (même dans le cas où chez les enfants le geste et la parole ne correspondent pas, cas de « mismatch », Goldin-Meadow & Momeni Sandhofer, 1999). Ces gestes ont différentes caractéristiques et fonctions selon leur relation avec la parole (Kendon, 1997 ; pour une revue, voir Goldin-Meadow, 1999). Ils ont clairement un rôle dans la communication aidant le locuteur et l'interlocuteur mais également un rôle dans la production des mots par le locuteur.

Quand les gestes co-verbaux ont un lien sémantique avec la parole, ils sont parfois porteurs de la même signification mais ils peuvent aussi transmettre des idées qui ne sont pas exprimées dans la parole du locuteur et ainsi enrichir la communication (voir Kelly et al., 1999, pour les gestes déictiques et iconiques). C'est par exemple le cas lorsque le locuteur indique un lieu en pointant du doigt, en

même temps qu'il énonce « le chat est là-bas ». Sans le geste de pointage, le sens de la phrase est incomplet car il n'indique pas à quel endroit exactement se trouve « le chat ». Le geste peut également améliorer la communication dans des situations où il est difficile de s'exprimer, à cause du bruit ambiant par exemple ou simplement parce que le locuteur et l'interlocuteur ne partagent pas la même langue (Faraco, 2000). Il nous est arrivé bien souvent de devoir renseigner un étranger sur la direction à prendre pour aller à un endroit précis. Dans ce cas-là, le peu de maîtrise de la langue étrangère est compensé par une gesticulation exagérée. Les enfants utilisent également certains gestes (gestes iconiques et déictiques) durant leurs premiers mots pour augmenter leur capacité à communiquer (Goldin-Meadow, 1999) : ils peuvent montrer une pomme et en même temps dire « donne ». L'information véhiculée par le geste précise ce que veut l'enfant et l'adulte est capable de comprendre ces gestes. En outre, cette façon de combiner le geste et la parole augmente sensiblement le répertoire de communication des enfants.

Le geste a également un rôle dans l'apprentissage de résolutions de problèmes (Goldin-Meadow et al., 1993). La non congruence entre le geste et la parole est caractéristique des enfants en train d'apprendre à résoudre une tâche (souvent le cas de problèmes en physique ou mathématiques) : le geste utilisé en même temps que la parole permet d'exprimer une autre idée. Il révèle l'état instable du sujet qui est en train d'apprendre. Ainsi cette phase durant laquelle le geste et la parole n'ont pas la même signification permettrait d'atteindre plus vite un raisonnement correct pour résoudre le problème.

Les gestes ont également une fonction de structuration pour le locuteur qui les exécute : « […] gestures do not just reflect thought but have an impact on thought. Gestures, together with language, help constitute thought » (McNeill, 1992, p. 245). Ils peuvent aider le locuteur à retrouver ses mots en mémoire (dans le cas de déficit lexical, d'hésitations) (pour une revue sur la fonction de recherche lexicale des gestes iconiques, voir Butterworth & Hadar, 1989 ; Morrel-Samuels & Krauss, 1992) et peuvent aussi influencer le locuteur dans le choix de ses mots (voir notamment Alibali et al., 2001b, sur l'influence de l'information motrice) et permettre d'accéder à de nouvelles idées. Les gestes ont en effet un format imagé de représentation, différent de celui de la parole : l'utilisation de ces gestes peut ainsi permettre au locuteur de représenter des idées qui ne sont pas encore assez développées pour être encodées en mots mais qui sont accessibles dans le format mimétique des gestes. Il est à noter que certains auteurs (pour une revue, voir Krauss & Hadar, 2001) considèrent que la fonction essentielle du geste n'est pas dans la communication mais plutôt dans la production de parole pour faciliter l'accès aux mots. D'autres auteurs (par exemple, Alibali et al., 2001a) soutiennent le fait que les gestes peuvent avoir à la fois un rôle dans la production de parole pour le locuteur et un rôle dans la communication pour l'interlocuteur.

III.2. Interdépendance du geste et de la parole

III.2.1. Les différentes phases du geste

Le geste est souvent décomposé en différentes phases de mouvement. Kendon (1980) a défini une unité de mouvement, la « *gesticular unit »,* qui commence quand la main (ou le bras) s'éloigne du corps et qui se termine quand la main revient à une position immobile. Entre le début et la fin, différentes étapes (les *gesticular phrase*) se succèdent : la préparation (*preparation*), le stroke (*stroke*), la tenue (*hold*) et le retour (*recovery*). Durant la phase de préparation, le membre se déplace jusqu'à la position initiale du stroke. Puis le stroke, c'est-à-dire la partie du geste qui est porteuse de sens et qui demande un effort plus important, est exécuté. La main est ensuite maintenue en position avant de revenir à une position immobile (de repos). Les différentes étapes du geste peuvent se synchroniser avec les différents niveaux prosodiques du discours (McNeill et al., 1990).

Kita et al. (1998 ; voir aussi de Ruiter, 2000) rajoute une phase optionnelle de tenue de la main avant le stroke (*pre-stroke hold*) et une phase de tenue après le stroke (*post-stroke hold*). Pendant le *pre-stroke hold*, la main se met en position initiale pour faire le stroke et elle attendrait que la parole soit produite (Kita et al., 1998). Cette phase se produit en général en même temps que les connecteurs de discours sont prononcés (tels que les pronoms, les pronoms relatifs par exemple) : elle permet d'établir une certaine cohésion entre le geste et la parole. Le stroke est ensuite exécuté en même temps que la syllabe ou le mot accentué : c'est la règle de synchronie phonologique (McNeill et al., 1990 ; McNeill, 1992). Si le geste finit trop rapidement, la main peut être maintenue en position durant la phase de *post-stroke hold* jusqu'à ce que la parole soit produite entièrement.

De manière générale, les phases sont déterminées par un changement de direction abrupt et une discontinuité dans le profil de vitesse. Pour l'identification du stroke, la force exercée étant plus importante, on peut se baser sur le profil d'accélération pour déterminer les frontières temporelles du stroke (Kita et al., 1998). Pour plus de détails sur les différentes phases possibles dans le geste et leur segmentation, on pourra se référer aux critères de McNeill (1992, pp. 374-377) et Kita et collaborateurs (1998, pp. 28-30).

III.2.2. Anticipation et synchronisation

III.2.2.1. Des règles de synchronisation

De manière générale, McNeill (McNeill et al., 1990 ; McNeill, 1992) énonce des règles de synchronisation[1] entre le geste et la parole :

- La règle de synchronie phonologique (*phonological synchrony*) postule 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. Dans le cas où le stroke est en avance, la main reste en position (phase de *post-stoke hold*) jusqu'au pic phonologique (*phonological peak*).
- La règle de synchronie sémantique postule que si le geste et la parole se produisent en même temps (*co-occur*), ils doivent être porteurs de la même signification (ou complémentaire) ;
- La règle de synchronie pragmatique postule que si le geste et la parole se produisent en même temps, ils doivent avoir une même fonction pragmatique.
- Dans le cas où des pauses sont insérées entre les mots d'une même expression et qui ont une même unité sémantique (ce cas se présente par exemple lors d'une hésitation, d'une recherche du mot approprié), le stroke continue pendant la pause afin de garder une cohérence sémantique.

La plupart de ces règles sont par ailleurs respectées dans différents pays qui ont chacun une structure linguistique différente (McNeill et al., 1990 ; McNeill, 1992 ; voir Özyürek, 2001 pour la règle de synchronie sémantique pour le turc). Pour les gestes qui n'ont pas de lien sémantique avec la parole (comme les gestes bâtons par exemple), le calage temporel se fait au niveau de la syllabe ou du mot accentué. Il semblerait que dans certains cas les indices suprasegmentaux de la parole entretiennent une relation étroite avec les gestes (voir McClave, 1998, pour une revue). En espagnol, il a été montré que dans le cas d'incises, les mouvements de main et la ligne de la fréquence fondamentale F0 étaient synchrones, la F0 et la main effectuant une montée et une tenue en même temps (Pietrosemoli et al., 2001). Boyer et al. (2001) ont également mis en évidence ce lien pour la focalisation, en considérant que « le point de rendez-vous d'indices intonatifs et gestuels à un endroit du continuum mettra en relief la partie concernée » (p. 460) : le point culminant du geste et la montée de F0 peuvent parfois être synchrones (la focalisation se fait sur la même syllabe), parfois non synchrones (dans ce cas, c'est

[1] Il est à noter que la synchronisation du geste et de la parole est une question problématique dans le sens où il est difficile de définir les frontières temporelles (début et fin) du geste. Le geste et la parole constituent, tous deux, deux intervalles temporels distincts produits par des articulateurs différents. En outre, il est parfois difficile pour certains types de gestes de définir avec précision à quel mot ou groupe de mots le geste se rapporte : il peut en effet se référer à un mot particulier de la phrase (ex : le chat est là, en pointant du doigt un endroit particulier) comme il peut également concerner plusieurs mots (ex : c'est un très grand cercle, en faisant un large geste circulaire du bras).

bien souvent le geste qui précède : la focalisation se fait alors par le geste et par la voix sur deux syllabes successives). Ces auteurs expliquent cette relation temporelle particulière entre les indices suprasegmentaux et les gestes comme faisant partie des différentes stratégies communicatives des locuteurs. De manière plus générale, McNeill et al. (2001) intègrent les mots, la prosodie et les gestes dans une même intention de communication : « […] the organization of discourse is inseparable from gesture and prosody. The three components are different sides of a single mental-communicative process. » (p. 480).

McNeill (1992) postule donc une relation temporelle stable entre le geste spontané et la parole (« a constant relationship in time »), le geste pouvant à la fois anticiper et être synchrone avec la parole, selon la phase du geste considérée (préparation ou stroke). Notons cependant que Butterworth et Hadar (1989), dans une note théorique en réponse à McNeill (1985), mettent en évidence le fait que ce qu'il considère comme une synchronisation parole-geste n'est en fait qu'un recouvrement temporel entre les deux intervalles (« […] there is some temporal overlap in the production of the speech and the gesture », p. 170) ; McNeill (1985) ne donne en effet aucun indice sur les paramètres temporels. Il ne considère en effet pas réellement le début du geste en rapport avec le début de la parole, qui sont d'après ces auteurs, deux événements temporels importants pour comprendre les mécanismes de la relation geste-parole : « The temporal relation between speech onsets and gesture onsets is, prima facie, a significant parameter because it may indicate the lower bounds on the time course of the putative underlying processes. […] Thus, despite the fact that gestures may continue until after speech onset, the beginnings of movements must be considered as events potentially separable from speech onsets […]. » (Butterworth et Hadar, 1989, p. 170).

III.2.2.2. Timing du début de geste

La phase de préparation peut être en avance sur la parole accompagnante (Morrel-Samuels et Krauss, 1992 ; Kendon, 1997). Butterworth & Hadar (1989), dans une revue de différentes études, mettent en évidence que le début du geste (il s'agit en fait surtout de gestes iconiques) peut se produire plus de 0,2 seconde avant le début de la parole. Cette anticipation révélerait le fait que le locuteur, alors qu'il articule un énoncé, serait déjà en train de formuler l'énoncé suivant (voir section IV.1.6).

Morrel-Samuels et Krauss (1992) ont observé que le début du geste co-verbal (principalement des gestes iconiques) se produisait en moyenne moins d'une seconde avant le début de la parole. Dans leur expérience, un locuteur était filmé en train de décrire des photos. Plus de 2000 gestes ont été identifiés à partir de ces vidéos et 129 sujets ont participé à l'analyse de ces gestes : ils devaient identifier le geste et la parole accompagnante ayant la même signification et devaient juger de la

familiarité du mot (*lexical affiliate*). Les résultats montrent que le début du geste peut être en avance sur le début de la parole ou synchrone avec lui mais ne se produit jamais après : « […] gestures are synchronized with speech and […] they are initiated before or simultaneously with (but not after) the onset of their lexical affiliates. » (p. 619). L'avance du geste sur la parole (*gesture-speech asynchrony*) peut varier de 0 à 3,8 secondes et est en moyenne de 0,99 sec.

Il est à noter par ailleurs que, dans des conditions moins naturelles que celles où les gestes co-verbaux sont spontanément produits avec la parole, soit, des conditions expérimentales où le geste et la parole sont exécutés ensemble comme une réponse à un même stimulus, le début du geste est également en avance sur le début de la parole (Holender, 1980 ; Levelt et al., 1985 ; voir Pashler, 1994 pour une revue théorique et expérimentale sur le paradigme de double-tâche ; Feyereisen, 1997). Dans une double tâche de dénomination et de « key-pressing » en réponse à un stimulus visuel (une lettre), Holender (1980) observe que le geste a une avance de 83 ms sur la parole. Pour des gestes de pointage exécutés en même temps qu'une expression déictique verbale, Levelt et al. (1985) observent de manière régulière une initiation du geste pouvant se produire de 300 à 600 ms avant le début de la parole selon le temps que le locuteur a pour répondre. Enfin, Feyereisen (1997), en répliquant les résultats de Levelt et al. (1985), observe une anticipation du geste de pointage sur la parole de 183 ms. Même avec une autre catégorie de gestes, des gestes iconiques, l'auteur observe encore une anticipation du geste sur la parole (de 283 ms à 336 ms selon la main utilisée pour faire le geste).

III.2.2.3. Timing du stroke

L'autre phase du geste, la phase de stroke qui est porteuse du sens, est d'après McNeill (1992), synchrone avec la parole. L'auteur observe ce patron temporel pour 90% des gestes analysés (il s'agit surtout de gestes iconiques et bâtons) produits par des locuteurs adultes américains décrivant six dessins animés différents (« 90% of all strokes occurred during the actual articulation of speech », p. 92). Les sujets dans cette expérience avaient pour consigne de regarder un dessin animé et ensuite de le décrire le mieux possible à un interlocuteur qui ne connaissait pas l'histoire et qui devait par la suite, à partir de la narration du sujet, décrire ce dessin animé à une autre personne.

Le point de vue de McNeill d'une stabilité de la relation entre geste et parole est renforcé par des expériences testant la production de gestes et de parole dans des conditions où le retour auditif du locuteur est retardé (*delayed auditory feedback, DAF ;* Lee, 1950). En effet, dans de telles conditions, le locuteur entend sa voix en écho et il a la sensation que sa propre parole est retardée par rapport à son articulation. Dans cette situation expérimentale particulière, la production de parole est dramatiquement perturbée : la parole du locuteur est nettement ralentie et est caractérisée par des

hésitations, des bafouillements et des bégaiements. McNeill (1992) demande aux sujets de raconter l'histoire d'un dessin animé qu'ils avaient visionné à un interlocuteur qui ne connaît pas l'histoire (les sujets n'ont pas de consigne particulière et peuvent donc s'aider de gestes). Les sujets portaient des casques par lesquels le retour auditif de leur propre parole était retransmis, soit de manière synchrone avec la production de parole du sujet (sans retard donc), soit retardé de 0,2 sec (retard qui correspond à la durée d'une syllabe et qui devrait avoir les effets les plus néfastes). Les résultats montrent, comme attendu, que le retard du feedback auditif perturbe la production de parole du sujet. En plus d'être ralentie et hésitante, la parole du locuteur est également simplifiée du point de vue de sa complexité linguistique habituelle (évaluée par le nombre de propositions emboîtées dans la même phrase). En revanche, la production des gestes dans cette situation n'est pas affectée ; au contraire, les sujets ont tendance à produire davantage de gestes et également des gestes plus complexes dans leur structure. Plus surprenant, la relation temporelle entre le geste et la parole reste inchangée par rapport à une condition normale ; le stroke du geste peut anticiper légèrement le mot mais cette différence ne suffit pas à amener l'auteur à dire que la synchronie geste-parole est rompue. Notons tout de même que les mesures de timing ne sont pas suffisamment précises pour réellement conclure sur la synchronie des deux éléments ; en effet, l'auteur définit les composantes du geste à partir de l'observation de vidéos des locuteurs et n'utilise donc pas de matériel lui permettant de mesurer et discuter des caractéristiques cinématiques des gestes (pour une discussion critique sur ce point, voir Butterworth & Hadar, 1989).

III.2.2.4. Variabilité

McNeill (1992) rapporte une seconde expérience utilisant cette méthodologie, menée par Mowrey et Pagliuca à l'université d'Ottawa (apparemment non publiée), et qui peut peut-être nuancer cette position. Ces auteurs ont testé sur eux-mêmes l'expérience de retard du retour auditif (DAF) durant la production de parole et de gestes appris, donc rappelés en mémoire. Ils ont appris une séquence de gestes iconiques et un texte correspondant à une séquence décrivant les étapes à effectuer depuis l'extraction d'un gâteau du four jusqu'à son service à table. Dans cette condition, les locuteurs n'ont pas à *créer* du sens mais seulement à répéter de manière continue une séquence apprise. Un des retards testé était de 0,2 sec comme dans l'expérience de McNeill (1992). Les auteurs observent comme précédemment une perturbation importante de la production de parole dans la condition DAF. De manière surprenante, ils observent également une rupture de synchronie entre le geste et la parole : le geste est largement en avance du mot auquel il se réfère (il se produit en fait durant le mot précédent). Il apparaît donc que pour une séquence apprise, c'est-à-dire une situation dans laquelle le

locuteur n'a pas à construire un énoncé, le geste n'est plus du tout synchrone avec la parole mais l'anticipe.

En fait, il apparaît que le timing geste-parole dépende en partie du geste considéré. En effet, Butterworth et Hadar (1989) ont mis en évidence la différence entre les gestes iconiques et les gestes bâtons dans les relations temporelles entre le geste et la parole. Ces deux types de gestes ont des fonctions différentes ; les gestes iconiques semblent avoir un rôle dans la recherche lexicale alors que les gestes de battements semblent plutôt dédiés à des fonctions d'insistance, de mise en évidence de certains points de la phrase. Ils ont également un timing différent par rapport à la parole : les gestes iconiques anticipent la parole surtout dans le cas de pauses reflétant une recherche de mots, alors que les gestes bâtons sont synchrones avec la parole. Les auteurs concluent donc sur une dissociation entre gestes iconiques et bâtons : « Speech dysfluencies and hesitations dissociate beats from iconic gestures. When lexical retrieval delays speech output, the onset of the iconic gesture is unaffected, causing the characteristic gesture precedence (Butterworth & Beattie, 1978 [cités par les auteurs]). However, a hesitation disrupts the flow of speech and delays the occurrence of the stressed item, causing beats, still synchronized with the stressed item, to occur after the dysfluency (Dittman, 1972 ; Hadar et al, 1984 [cités par les auteurs]) » (Butterworth & Hadar, 1989, p. 173). La production de ces deux types de gestes serait en fait dépendante de deux processus de traitement différents dans la production de parole (voir section IV.1) : la production des gestes iconiques se ferait après la construction précoce du message et la production des gestes bâtons serait dépendante d'une étape plus tardive où la spécification phonétique de la phrase serait complète avec tous les paramètres de timing spécifiés, y compris les marques prosodiques.

III.2.3. Coordination du geste et de la parole

On peut se demander comment les deux systèmes de production de gestes et de parole se régulent. Il apparaît que le geste a un plus grand effet sur la parole quand ils sont produits ensemble.

Dans une tâche, dans laquelle à la fois une réponse gestuelle et orale est demandée, Holender (1980) montre que le geste a un effet sur la parole. Dans cette double tâche, il présente visuellement aux sujets une lettre-stimulus et mesure les temps de réaction des sujets pour nommer cette lettre et appuyer sur un bouton correspondant à cette lettre (les sujets doivent répondre le plus vite possible). Les sujets ont le choix entre quatre lettres « L », « N », « R » et « S », commençant oralement toutes par le même phonème ([ɛ]). Les quatre boutons correspondants (dans l'ordre) sont placés les uns à côté des autres ; le sujet peut les activer avec un de ses doigts (dans l'ordre, majeur et index de la main gauche et majeur et index de la main droite). L'auteur observe que le geste est effectué 83 ms en

avance par rapport à la réponse verbale. Il apparaît donc que, même dans le cas d'un geste qui ne se produit pas de manière spontanée et naturelle avec la parole, le geste semble débuter avant l'émission de son. Les résultats de cette double tâche sont comparés aux temps de réaction mesurés dans chacune des tâches simples : nommer oralement la lettre ou appuyer sur le bouton seulement. Les résultats obtenus sont assez surprenants. Alors que le geste a exactement le même comportement (il se produit environ 475 ms après la présentation du stimulus) dans les deux types de tâche (simple et double), dans la tâche simple, la réponse orale est plus rapide que la réponse gestuelle (de l'ordre de 75 ms). Ainsi, utilisés ensemble au cours de la double tâche, la parole n'a pas d'effet sur le geste alors que le geste ralentit la production de la parole (la parole est ralentie de 150 ms environ entre les deux tâches). Dans une autre expérience du même type (expérience 2 dans Holender, 1980), l'auteur donne comme consigne aux sujets soit de synchroniser leur réponse gestuelle et orale, soit de donner d'abord la réponse orale. Les sujets n'arrivent pas alors à synchroniser les deux types de réponse. Même dans le cas où les sujets ont du temps pour donner leur réponse, ils donnent spontanément d'abord la réponse manuelle puis la réponse verbale (voir expérience 3 dans Holender, 1980). Dans le cas où geste et parole sont utilisés en même temps, il apparaît donc que leur relation temporelle est ajustée par la parole : c'est la parole qui s'aligne sur le geste et non le contraire ; la relation entre les deux modalités est donc unidirectionnelle.

Levelt et al. (1985) ont également observé cet effet du geste sur la parole dans leur tâche de pointage gestuel et oral. Les auteurs mènent une série d'expériences où des gestes de pointage sont co-produits avec des expressions déictiques afin de comprendre comment les deux systèmes moteurs opèrent pour créer une telle interdépendance geste-parole. Ils étudient expérimentalement la synchronisation du geste et de la parole en examinant les temps de réaction des deux systèmes dans une tâche de pointage de cible. Les cibles sont des diodes lumineuses placées en face du sujet et réparties de gauche à droite (dans le champ ipsilatéral et controlatéral par rapport à la main utilisée pour le pointage). Le sujet a pour tâche de pointer du doigt l'une d'entre elles, celle qui s'allume, et en même temps d'énoncer l'expression déictique « dat lampje » (cette lampe). L'allumage de l'une des quatre diodes est déclenché par le sujet qui appuie sur un bouton en position de repos ; l'enregistrement des données (c'est-à-dire le mouvement du doigt par ses coordonnées x et y transmises par une diode infrarouge posée sur le doigt du sujet et le signal acoustique) est alors enclenché et le sujet a pour consigne de fournir une réponse le plus rapidement possible. Trois variables sont mesurées : Ti, l'intervalle temporel entre l'allumage de la diode et le moment d'initiation du geste de pointage, Ta, l'intervalle temporel entre l'allumage de la diode et le moment où le geste de pointage atteint son extension maximale, son *apex* et Tv, l'intervalle temporel entre l'allumage de la

diode et le moment de début de la réponse verbale. Dans une de leur expérience (voir expérience 1 dans Levelt et al., 1985), les auteurs démontrent que le geste et la parole sont synchronisés : quand la distance de la cible à pointer est plus grande, le geste est aussi plus long et la parole se produit également plus tard. De manière générale, l'apex du geste et le début de la parole co-varient en fonction des différentes positions de la diode. Dans une autre de leurs expériences (voir expérience 2), ils comparent les résultats de la double tâche de pointage et de dénomination verbale de la cible aux deux tâches simples (soit pointer du doigt seulement, soit nommer oralement la cible). Les résultats montrent que le geste, utilisé en même temps que la parole, ralentit celle-ci de 99 ms (par rapport à la situation où la réponse est seulement orale) ; ce qui confirme les résultats obtenus par Holender (1980). En revanche, un léger effet de la parole sur l'initiation du geste est également observé. Par rapport à la tâche simple de pointage manuel, le geste débute 14 ms après, quand le sujet doit également donner une réponse verbale. Il est à noter cependant que la parole n'affecte pas la durée du geste mais seulement son initiation. La relation d'interdépendance du geste et de la parole est donc bien unidirectionnelle, mais ne concerne que la phase d'exécution du geste. Quand geste et parole sont co-produits, le début de la parole dépend du geste alors que l'exécution du geste est indépendante.

Ces résultats ont été, par ailleurs, répliqués par Feyereisen (1997) avec une procédure expérimentale différente : un programme informatique (en Basic) est utilisé pour enregistrer le début du geste et le début de la parole. Les sujets sont assis en face de l'écran d'un ordinateur (un Apple II). Ils doivent attendre qu'un signe « + » apparaisse à gauche ou à droite de l'écran pour répondre. Leur réponse peut être soit un geste (pointer la position du stimulus sur l'écran), soit un mot (« ti » ou « ta » selon que le signe est à gauche ou à droite de l'écran), soit les deux types de réponses en même temps. Les sujets ont pour consigne de maintenir enfoncée la touche « pomme » du clavier jusqu'à l'apparition du signe sur l'écran ; ils doivent ensuite répondre le plus vite possible en utilisant la main qui maintenait la touche enfoncée (la pression continue de la touche « pomme » laisse le programme en pause ; dès que la pression est relâchée, le programme est lancé et le moment d'initiation du geste est ainsi enregistré). Les 24 sujets sont répartis dans deux groupes, un groupe utilisant la main gauche pour faire le geste et l'autre groupe la main droite ; l'auteur veut ainsi tester un éventuel effet de la latéralité. Chaque sujet est testé sur 24 stimuli (en fait, il y a 20 stimuli de tests et quatre stimuli d'entraînement avant le test) dans deux tâches simples (geste ou parole) et une tâche double. Contrairement à ce qu'avaient observé Levelt et al. (1985), il n'y a pas ici d'effet de la latéralité. En revanche, il y a un retard dans l'initiation du geste et de la parole dans la tâche double par rapport à la tâche simple : le début du geste est retardé de 42 ms et le début de la parole de 91 ms. Tout comme Levelt et al.

(1985), l'effet d'interférence est beaucoup plus important pour la parole que pour le geste ; ce qui suggère une interdépendance geste-parole avec une influence plus importante du geste sur la parole.

Dans une autre expérience (voir expérience 2), Feyereisen (1997) teste si cette relation temporelle reste la même quand des gestes iconiques sont utilisés à la place des gestes de pointage. Quatre symboles différents sont utilisés (« \ », « o », « # » et « ◊ »). Les sujets ont appris quatre formes de main différentes et quatre mots correspondant à chaque symbole : pour le symbole « \ », la main oblique et tendue et le mot « barre », pour « o », une opposition entre le pouce et l'index et le mot « boule », pour « # », un geste illustrant une personne tenant un cube sur le côté et le mot « bloc » et pour « ◊ », le dos de la main et le mot « blanc ». Tous les sujets sont testés sur cinq séries de 40 stimuli (huit stimuli d'entraînement et 32 stimuli de tests) : une série pour la simple tâche de parole, deux séries pour le geste et deux séries pour la double tâche geste-parole (une avec la main gauche et une avec la main droite). Contrairement à l'expérience précédente, l'auteur n'observe aucun effet d'interférence de la parole sur le geste : le geste est initié au même moment qu'il y ait accompagnement ou non de parole. En revanche, pour la parole, il y a un effet dû à la tâche (le début de la parole se produit plus tôt dans la tâche simple que dans la tâche double) et un effet de la latéralité (l'interférence du geste sur la parole est plus grande quand c'est la main droite qui est utilisée pour faire le geste). Dans la double tâche, la réponse orale peut être retardée de 150 à 200 ms selon la main utilisée. L'interdépendance de la relation geste-parole est donc unidirectionnelle, comme ce qui était démontré par Holender (1980).

Cette interdépendance geste-parole peut être expliquée (Holender, 1980 ; Levelt et al., 1985 ; Feyereisen, 1997) par un processus de compétition entre les deux systèmes moteurs, manuel et oral, sur des ressources communes de traitement. Il semblerait que les sujets essaient de synchroniser le début de la parole (Holender, 1980) avec l'apex du geste (Levelt et al, 1985). Pour cette coordination, les sujets regroupent les deux types de réponse motrices : la réponse orale n'est donnée qu'une fois que le geste a débuté (Feyereisen, 1997). Les différents résultats obtenus peuvent permettre de comprendre comment se fait cette interdépendance ; plusieurs interprétations différentes ont été émises concernant le moment où la relation entre le geste et la parole est établie (pour une revue, voir Feyereisen, 1997 ; Furuyama et al., 2002). Le lien est-il déterminé dès la phase de planification avec aucune interaction des deux systèmes par la suite ou bien le lien se fait-il de manière interactive même durant la phase d'exécution motrice ? Ces deux points de vue sont respectivement celui d'une théorie « balistique » qui postule une certaine indépendance des deux systèmes moteurs et donc une architecture « modulaire » au sens de Fodor (1983) (la relation geste-parole étant déterminée dès le début ; Levelt et al., 1985 ; de Ruiter, 2000) et celui d'une théorie « interactive » dans laquelle les deux

systèmes moteurs interagissent constamment grâce à un feedback continu durant la phase de planification et durant l'exécution motrice de façon à accomplir un but commun (McNeill, 1992 ; Furuyama et al., 2002).

Les études de de Ruiter et Wilkins (1998) sur les gestes co-verbaux produits par des locuteurs néerlendais et arrernte (langage australien aborigène) supportent plutôt une théorie modulaire ; les locuteurs arrernte ont une phase de préparation plus longue du geste mais ne retardent pas pour autant le début de la parole (le timing de la parole ne s'adapte pas au timing du geste) durant l'exécution, ce qui démontre que geste et parole sont deux systèmes indépendants durant l'exécution. Au contraire, Furuyama et al. (2002 ; voir aussi Seyfeddinipur & Kita, 2001, pour une évidence à partir de l'étude des gestes dans le phénomène d'auto-correction en parole) proposent une interaction continue des deux systèmes, les relations temporelles entre geste et parole étant finement contrôlées ; ils ont analysé des gestes de pointage accompagnant la parole et impliquant différentes durées de phase de préparation (pour des distances de gestes plus ou moins grandes) et ont démontré le maintien d'une certaine synchronie entre le geste et la parole : « […] speakers can maintain tight synchrony between speech and gesture even as preparation phase length varies widely. Such behavior is possible only if speakers exert careful control over the temporal relationship between the various parts of their simultaneously unfolding speech and gestures », (Furuyama et al., p. 5). Levelt et al. (1985) proposent quant à eux un point de vue intermédiaire, une théorie mixte qui postulerait une interaction entre les deux systèmes moteurs durant la phase de planification et une indépendance des deux systèmes durant la phase d'exécution motrice. Les auteurs ont démontré en effet dans une de leurs expériences (voir expérience 4 dans Levelt et al., 1985), dans laquelle le geste était gêné par une charge appliquée à un moment inattendu durant le mouvement, que la parole pouvait s'ajuster sur le geste seulement si cette perturbation avait lieu au début de la phase d'exécution du geste. Plus tard dans l'exécution (au milieu du mouvement), le locuteur ne pouvait plus ajuster les deux systèmes, ce qui entraînait une avance du début de la parole sur l'apex du geste dans cette situation.

III.2.4. Quand gestes et parole sont étroitement liés

Nous présentons ici deux autres situations où geste et parole sont interdépendants même s'ils ne sont pas directement impliqués dans la communication : ils révèlent une coordination typique de la production des formulettes et enfantines[2].

III.2.4.1. Formulette d'incantation

Berthier et al. (1991) ont étudié la coordination rythmique du geste et de la parole dans le cadre de la fabrication traditionnelle de sifflets d'écorce de frêne en Rhône-Alpes. La fabrication de ces sifflets se fait manuellement en plusieurs étapes. L'une de ses étapes, qui a pour but de détacher l'écorce du tronçon de frêne, implique une coordination rythmique très précise entre la parole et le geste manuel de l'artisan. Durant cette étape, le sujet chantonne en dialecte une formulette d'incantation à la sève en même temps qu'il bat en rythme le tronçon de bois à l'aide de son couteau pour faire sortir la sève (plus précisément, le couteau est tenu par la lame et le battement se fait au niveau de la virole du couteau). La séquence de percussion produite par le couteau sur le bois forme un cycle de « geste de volée » qui se décompose en trois phases (voir Figure 15) : le *lancé* (depuis le début du geste jusqu'au moment où le couteau frappe le bois), le *percuté* (depuis le contact jusqu'au moment d'extension maximale entre la main et la lame du couteau) et le *relevé* (depuis ce moment d'extension maximale jusqu'au début du geste suivant). Un sujet assis pratiquant cette fabrication traditionnelle a été filmé : le sujet fait tourner le tronçon de bois avec la main gauche tandis qu'il déplace sa main droite en maniant le couteau. Le mouvement de la main en fonction du temps a été analysé et différents paramètres ont été mesurés pour étudier cette coordination (nous voyons sur la Figure 15 l'évolution au cours du temps de l'angle entre la phalange et la lame). En ce qui concerne le cycle de volée, une organisation temporelle typique de ce genre de geste a été observée : une durée moyenne du cycle de 260 ms, soit l'équivalent de quatre cycles environ par seconde, avec une répartition temporelle de 31% pour le lancé, 23% pour le percuté et 46% pour le relevé. En ce qui concerne la coordination avec la parole, les auteurs ont observé une contrainte de couplage entre le geste manuel et la parole : la percussion tombe pendant la consonne (on voit sur la Figure 15 que le coup se produit durant la consonne [s]) et dans tous les cas ne se produit jamais avant la fin de la voyelle précédente, et cela, quelle que soit la durée de la syllabe. Ceci amène les auteurs à conclure qu'il y a « un calage réciproque de la parole et

[2] Nous présentons ici seulement quelques études de sortes de chants vocaux-gestuels. Il est à noter cependant que d'autres études dans ce domaine ont également montré une étroite relation entre gestes et parole : c'est le cas par exemple d'une étude du chant haka effectué par les rugby-men néo-zélandais avant les matchs de rugby (Chafcouloff et al., 2001). L'analyse de ce chant issu de la tradition maorie révèle deux types de synchronisation voix-gestes : il peut y avoir simultanéité d'une frappe, résultant d'un contact entre deux parties du corps, et de la phase de détente des consonnes occlusives sourdes [p] et [k], ou bien il peut y avoir synchronisation du mouvement gestuel (qui est à ce moment-là plus lent) avec la durée du segment associé.

du geste » (p. 35). Au démarrage, le geste se cale sur la parole, puis la parole se règle sur le geste qui impose son rythme de battement et ralentit la parole (la parole est « *entraînée* par la cadence du bras »).

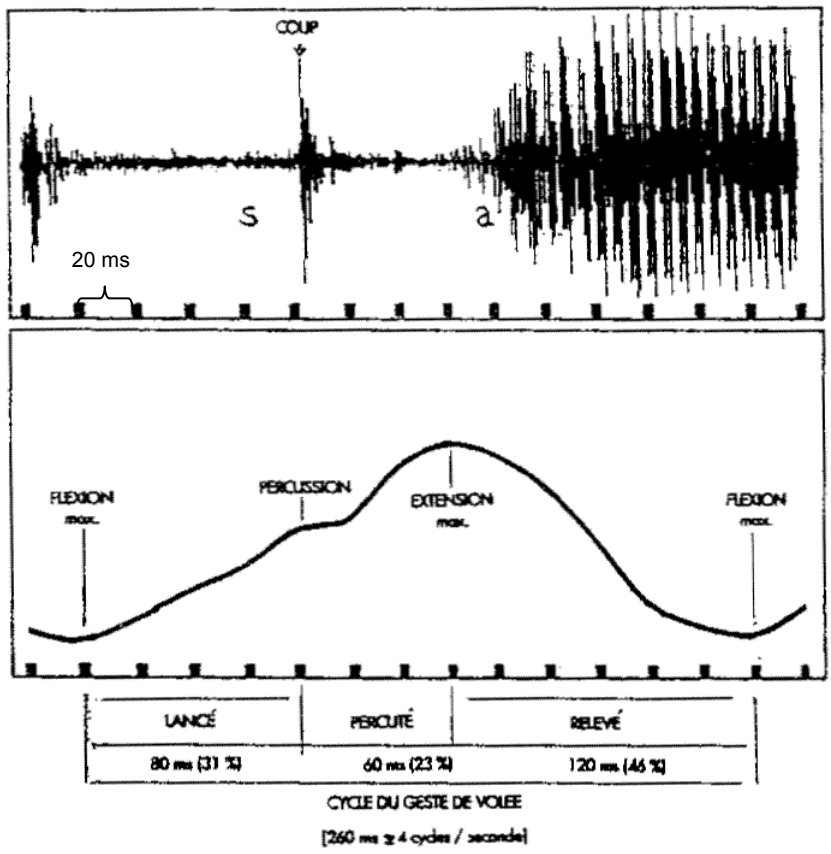


Figure 15. Décours temporel du cycle de volée en correspondance avec le signal acoustique de parole. En haut, portion de signal acoustique [sa] extraite de la formulette d'incantation. En bas, évolution au cours du temps de l'angle entre la phalange du sujet et la lame du couteau avec les trois phases du cycle du geste de volée. Sur le signal acoustique est repéré le moment où se produit la percussion (coup), soit durant la consonne. Figure tirée de Berthier et al., 1991.

III.2.4.2. La gestualité dans les enfantines

Chauvin-Payan (1999a) a étudié la gestualité chez les enfants dans un registre particulier, le registre poétique du folklore enfantin. Elle a étudié les différents jeux enfantins (comptines, formulettes, enfantines, etc.) qu'elle a pu observer dans les cours de récréation d'écoles maternelles et primaires de la région Rhône-Alpes. Le rythme (et sa régularité) étant « une constante de la tradition orale enfantine » (Chauvin-Payan, 1999a, p. 21), l'auteur a étudié les rapports existants entre la structure gestuelle et la structure rythmique et mélodique des enfantines. Pour rendre compte de ces relations, elle a mis au point un code iconique pour la description multimodale des jeux prenant en compte à la fois les paroles, la musique et les gestes (voir description détaillée dans Chauvin-Payan, 1999a). Elle fait donc une description de tous les jeux en mettant en relation la musique (notes de musique), les

paroles décrites en suite de syllabes et le geste accompagnant (par exemple, un frappement de main). Pour la plupart des « tape-mains » qui sont des jeux chorégraphiques où les enfants se placent l'un en face de l'autre et frappent dans leurs mains en chantant (par exemple : « trois p'tits chats »), un lien étroit entre la syntaxe des énoncés et le rythme gestuel et verbal est mis en évidence : les structures rythmiques peuvent respecter les frontières syntaxiques, le rythme, dans ces cas-là, permettant de *lier* la mélodie, les paroles et les gestes dans les enfantines (Blondel & Chauvin-Payan, 2002). La présence d'un schème gestuel ou « cycle de frappe » constitué de plusieurs unités qui peuvent se répéter plusieurs fois a été mis en évidence dans tous les jeux : par exemple, pour le tape-mains « trois p'tits chats », le cycle de frappe est constitué de trois unités. Le cycle de frappe peut coïncider avec les frontières de syntagmes (c'est le cas pour « Trois p'tits chats ») ou non (c'est le cas pour « Fanny » ; Chauvin-Payan, 1999b), mais dans tous les cas, chaque frappe se produit en même temps qu'une syllabe et qu'une note. Il apparaît que la répétition de ces cycles de frappe permettrait à la fois « la mémorisation et la coordination avec les paroles et avec la gestuelle des autres joueurs » (Blondel & Chauvin-Payan, 2002, p. 108).

Blondel et Chauvin-Payan (2002) ont comparé cette gestualité à celle des enfantines signées (exécutées par des enfants sourds ou par des adultes de manière pédagogique vers des enfants sourds signant). La gestualité dans les enfantines signées a un rôle double : elle peut être linguistique en faisant partie des signes de la Langue des signes mais elle peut aussi faire partie d'une autre gestuelle accompagnant les signes (des mimiques par exemple). Dans les enfantines signées, on ne trouve pas de gestes rythmiques (cycles de frappe) comme dans les enfantines orales : « quand tout le gestuel est ''parole'', il ne parait pas très étonnant que le gestuel purement rythmique soit absent » (p. 111). Cependant, il y a de nombreux points communs entre les deux types d'enfantines. Les schèmes rythmiques sont également présents dans les enfantines signées. Il peut y avoir une répétition des mots ou phrases comme dans les enfantines orales. Le schéma rythmique et mélodique des enfantines orales qui est fondé sur les valeurs de durées, d'accents et de hauteur est remplacé ici par un flux gestuel qui varie en intensité et en durée avec des gestes accentués par l'introduction de pauses, de tenues et de ralentis. De plus, de même que le cycle de frappe coïncidait avec les frontières syntaxiques dans les enfantines orales, les séquences signées coïncident avec les syntagmes.

Ainsi, il apparaît que les gestes co-verbaux ne sont pas seulement liés au contenu de la parole qu'ils accompagnent mais ils entretiennent aussi une relation temporelle avec la parole. Leur coordination peut être très variable dans le sens où nous trouvons certaines contradictions dans la littérature (c'est par exemple le cas pour de Ruiter & Wilkins, 1998, versus Furuyama et al., 2002). Cependant il semble que de nombreux facteurs soient à l'origine de ce constat, comme par exemple le type de geste, les outils d'analyse, la culture des sujets, etc. Néanmoins, ce qui ressort fortement est le fait que la plupart du temps, le début du geste anticipe le début de la parole alors que l'apex du geste est plus ou moins synchrone avec la parole.

CHAPITRE IV.

Modélisation de la production de parole et de gestes

Nous allons présenter maintenant une architecture cognitive générale de la production de mots et de gestes co-verbaux. Cette production multimodale sera expliquée dans le cadre du modèle « Speaking » de Levelt (1989, 1994) augmenté par un versant gestuel proposé par de Ruiter (2000). Le modèle de production de mots de Levelt explique comment à partir d'une intention communicative, le locuteur forme et articule un message approprié. De Ruiter reprend ce cadre général et y ajoute un modèle de production de gestes, le « Sketch Modèle », qui explique comment les différents gestes sont initiés et produits en relation avec la parole. Les deux modèles s'insèrent dans une approche de traitement de l'information (*information-processing approach*). L'architecture générale est constituée de modules de traitement (*boxologies*) qui opèrent sur des représentations c'est-à-dire des informations stockées auxquelles on peut accéder.

Cette partie permettra en discussion d'émettre des hypothèses sur la façon dont le code LPC est planifié en relation avec les segments de la parole. Nos études, comme nous le verrons, vont en effet mettre en évidence un lien très fort entre geste LPC et parole. Cette parole étant systématiquement resyllabifiée en suites CV, nous pourrons nous interroger sur la façon dont le code LPC impose sa structure à la parole. Le modèle de Levelt (1989) nous a semblé pertinent car – en plus du fait qu'il contient toutes les étapes de traitement qui se retrouvent généralement dans les modèles de production de mots (voir Butterworth & Hadar, 1989) – il met l'accent sur la syllabe qui est l'unité de production en LPC et décrit un stock de gestes syllabiques par le biais du syllabaire. Par ailleurs, c'est également ce modèle que de Ruiter a choisi d'augmenter par un versant gestuel pour la production des gestes co-verbaux. Ces deux modèles représentent donc une base sur laquelle nous pourrons tenter d'ajouter un versant expliquant la production des gestes de la LPC en relation avec la parole et les gestes co-verbaux.

IV.1. Un modèle psycholinguistique de production de mots : le modèle « Speaking »

Levelt (1989, 1994 ; voir aussi Segui & Ferrand, 2000) explique le processus de génération de mots dans la parole en plusieurs étapes cognitives et propose un modèle à modules encapsulés correspondant à différents niveaux de traitement (voir Figure 16) : la **conceptualisation** (*conceptual preparation*) de ce que le locuteur veut dire qui est gérée par le *conceptualizer*, la **formulation** de l'intention du locuteur avec les bons mots gérée par le *formulator* et l'**articulation**, étape durant laquelle le locuteur produit le mot et qui est gérée par l'*articulator*. L'étape de formulation est elle-même divisée en trois sous-processus : la sélection lexicale (*lexical selection*) durant laquelle le locuteur récupère les informations sémantiques et syntaxiques d'un mot, l'encodage phonologique

(*phonological encoding*) qui donne la forme phonologique du mot, l'encodage phonétique (*phonetic encoding*) qui associe à chaque syllabe du mot un geste articulatoire correspondant. Le locuteur peut de plus contrôler les sorties et se corriger si nécessaire (*self-monitoring*).

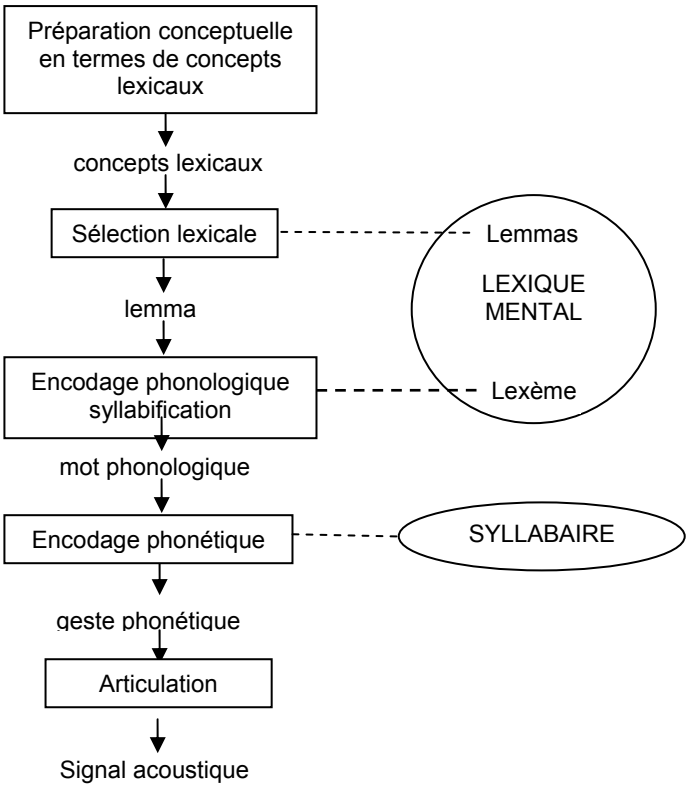


Figure 16. Les étapes de la production de mots dans le modèle « Speaking ». Figure adaptée de Levelt, 1994.

IV.1.1. La conceptualisation

Dans ce modèle, la première étape est la conceptualisation. « Speaking is a form of goal directed behavior » (Levelt, 1994). Le locuteur a une intention de communication et il doit la formuler de façon à ce que l'interlocuteur la reconnaisse : c'est son but. L'intention générale peut être divisée en une séquence de buts et de sous-buts (*subgoals*) : le locuteur doit alors pour chaque sous-but sélectionner le concept correspondant : c'est la phase de *macro-planning*. La sélection des concepts élémentaires porteurs de sens se fait par activation par le *conceptualizer* à partir d'une base de connaissances. Les concepts sont ordonnés et constituent une sorte de structure propositionnelle qui forme le concept lexical (*lexical concept*) : c'est la phase de *micro-planning*. Le concept lexical est un message *préverbal* (Segui & Ferrand, 2000), non linguistique, qui contient des concepts pour lesquels il existe un mot défini dans le lexique mental du locuteur. La façon d'ordonner les concepts dans un format propositionnel constitue le *perspective taking* : c'est le point de vue particulier que peut prendre le locuteur pour décrire une action ou un ordonnancement spatial. Par exemple pour décrire une scène mettant en jeu un chat et une table côte à côte, le locuteur peut dire que « le chat est à côté de la

table » ou que « la table est à côté du chat » (voire même encore d'autres descriptions). Cette multitude de choix vient du fait qu'il n'y a pas de relation « fixe » et unique entre un référent et le concept lexical qui va être sélectionné (Levelt, 1989).

A partir de la structure conceptuelle issue du *conceptualizer*, le *formulator* forme une structure linguistique correspondante, c'est-à-dire une structure syntaxique contenant les mots traduisant les concepts. Pour arriver à un tel résultat, trois étapes interviennent dans la formulation des « bons mots » : la sélection lexicale, l'encodage phonologique et l'encodage phonétique.

IV.1.2. La sélection lexicale

La sélection lexicale ou encodage sémantique et syntaxique (encore appelé *grammatical encoding*), permet de récupérer dans un lexique mental comportant un nombre important de mots ceux dont le sens traduit l'idée du locuteur, les *lemmas*. Cette récupération est rapide (environ 125 ms) et se fait par activation en parallèle des mots dont la sémantique est proche, un mécanisme de convergence permettant de faire la sélection finale du bon mot. Les *lemmas* correspondent plus précisément aux propriétés sémantiques et syntaxiques d'un mot. Cette notion permet d'expliquer le phénomène curieux du « mot sur le bout de la langue » qui nous donne la sensation désagréable de connaître toutes les propriétés d'un mot (son sens, le type de mot, si c'est un verbe, un nom, un adjectif, son genre…en un mot, le *lemma*) sans pouvoir pour autant retrouver la forme « phonologique » (sonore) de ce mot (le *lexème*). Un déficit de transmission d'informations entre les représentations sémantiques et syntaxiques et les représentations phonologiques des mots serait à l'origine de ce phénomène.

IV.1.3. L'encodage phonologique

Une fois que le lemma est sélectionné, il faut récupérer les informations phonologiques associées, le lexème. C'est l'étape d'encodage phonologique (voir Figure 17) qui attribue à chaque mot une représentation phonologique canonique, c'est-à-dire une décomposition syllabique de l'item lexical en phonèmes avant d'aboutir à sa forme « articulatoire » (durant l'encodage phonétique) qui correspond à la suite de gestes articulatoires à produire. Les lexèmes sont récupérés dans le lexique mental par assemblage d'unités plus petites et de morphèmes (affixes et racines de mots morphologiquement complexes). Le lexème est porteur d'une information segmentale sur le mot et d'une information métrique (*frame*). L'information segmentale est relative à la structure phonémique du mot c'est-à-dire sa décomposition en phonèmes consonantiques et vocaliques (par exemple pour le mot « avion », sa décomposition en segments phonologiques est V.C.C.V. /a/ /v/ /j/ /ɔ̃/). L'information métrique du lexème est la trame qui renseigne sur le nombre de syllabes du mot et sur son patron accentuel (par exemple pour le mot « avion » ce serait « σσ », si σ représente une syllabe). La syllabe est considérée

comme une unité de production de parole (Segui & Ferrand, 2000). Elle est représentée de manière mixte (Ferrand & Segui, 1998), à la fois sous forme d'unité ou *chunk* à contenu segmental fixe (ex : « ba », « bal ») et sous forme de patron ou *schéma* abstrait (ex : CV ou CVC). Les trames métriques des mots se combinent avec les trames des mots adjacents dans le flux de la parole (en suivant les règles spécifiques à la langue) par un processus incrémental de syllabation de gauche à droite pour donner des unités plus grandes (ou plus petites) selon le contexte, les formes phonologiques (*phonological word*). Il n'est alors pas tenu compte des frontières lexicales des mots (ex : « mon avion » « σ + σσ » donne « mon.na.vion » « σσσ »). Enfin, les séquences segmentales (suite de consonnes et voyelles phonologiques) sont rattachées aux formes phonologiques créant ainsi une forme syllabique phonologique du mot (/m/ /ɔ̃/ /a/ /v/ /j/ /ɔ̃/ + σσσ donne /mɔ̃.na.vjɔ̃/).

IV.1.4. L'encodage phonétique

A partir de ces syllabes phonologiques du mot, les gestes articulatoires (les syllabes phonétiques ou articulatoires) correspondants sont récupérés, durant l'encodage phonétique, dans un répertoire mental de gestes syllabiques unitaires, le *syllabaire* (*syllabary*) (Levelt & Wheeldon, 1994 ; voir aussi Cholin et al., 2004). Ce stock de programmes moteurs contient les gestes syllabiques les plus fréquents, spécifiés à la fois par une forme phonologique et par un geste articulatoire correspondant. Le geste articulatoire est une représentation des tâches à effectuer par les différents articulateurs entrant en action dans la production de parole ; il s'apparente au *gestural score* de Browman et Goldstein (2000 ; voir Levelt & Wheeldon, 1994). C'est une représentation abstraite de la tâche qui doit être exécutée ; il n'est pas dit comment les articulateurs de la parole doivent se coordonner pour exécuter cette tâche (par exemple, si la tâche est de fermer les lèvres, comme pour produire un [pa], le gestural score indique seulement que les lèvres doivent être fermées mais pas comment machoire, lèvres supérieure et inférieure se coordonnent pour y arriver). Les syllabes phonologiques du mot sont envoyées en entrée du syllabaire. Il y a alors une procédure de vérification sur la correspondance entre les syllabes phonologiques et les syllabes stockées dans le syllabaire. Chaque syllabe du syllabaire possède un seuil d'activation dépendant de la fréquence de la syllabe. Le choix du geste syllabique se fait finalement à la manière de la sélection du lemma par activation et convergence sur un geste unique. L'ensemble des gestes syllabiques correspondant aux syllabes phonologiques forme une séquence articulatoire syllabique (*un plan phonétique*) qui correspond également à la notion de parole interne (*internal speech*). Elle est ensuite envoyée au module d'articulation, l'*articulator*.

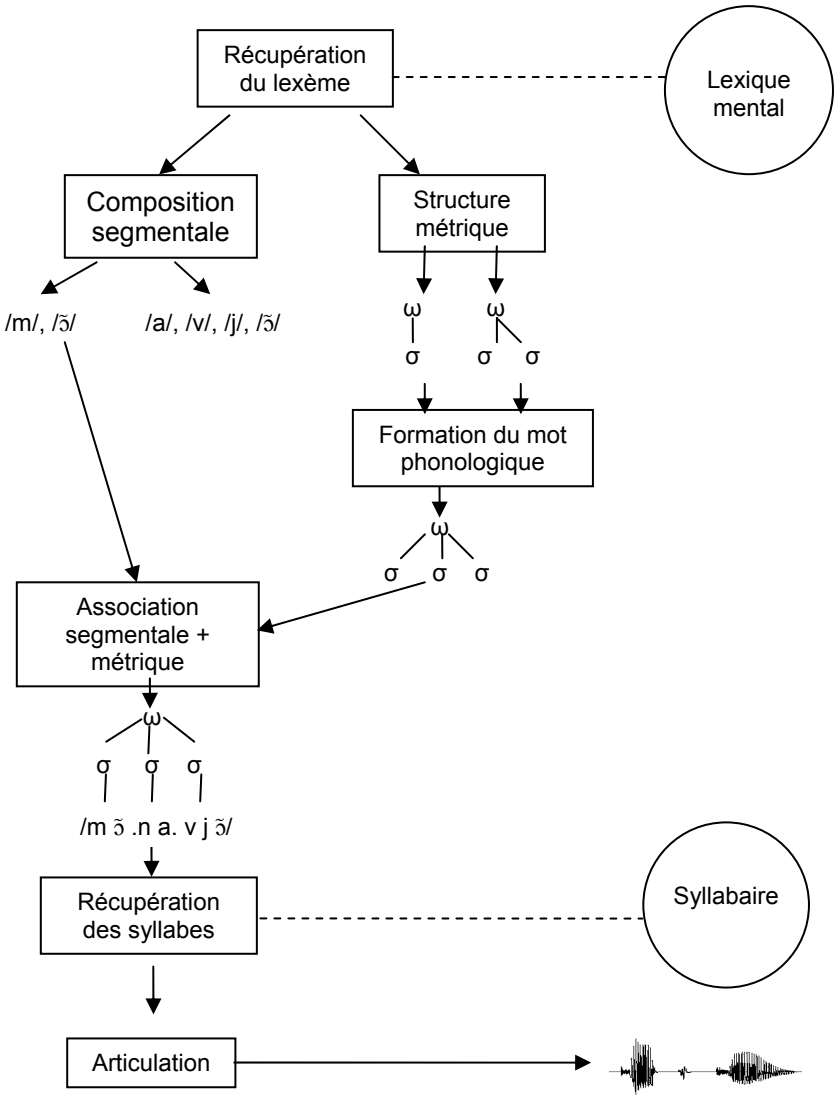


Figure 17. Détail de l'étape d'encodage phonologique dans le modèle « Speaking » + encodage phonétique et articulation. Adaptée de Levelt & Wheeldon, 1994 (voir également Segui & Ferrand, 2000).

IV.1.5. L'articulation

Le système de production (le système articulatoire) contrôle l'articulation c'est-à-dire la coordination d'un grand nombre de muscles créant la parole coarticulée. La séquence articulatoire syllabique n'est pas entièrement spécifiée car formée de *gestural scores* : elle n'indique donc pas avec précision comment articuler les différentes syllabes. Le modèle de Levelt n'a pas pour but d'expliquer la production de cette articulation mais nous pouvons nous référer à plusieurs modèles articulatoires qui expliquent comment, à partir d'une représentation minimale d'un geste articulatoire, l'articulation exacte émerge (voir Levelt, 1989 pour les différentes propositions de l'auteur sur ce point ; voir aussi Bonnot, 1990a, pour une revue des différents modèles de production de parole ; Sock, 1998). Remarquons que Bonnot (1990b) insiste sur la nécessité de prendre en compte la composante motrice dans un modèle psycholinguistique global. « Puisque tout locuteur est en principe en mesure de décoder les messages

qui lui sont soumis, c'est que des traces de ces mécanismes abstraits de la représentation de la parole *"interne"* subsistent aux niveaux les plus périphériques. Ces *traces cognitives* sont véhiculées par les systèmes chargés du contrôle final de la production de parole et se manifestent notamment sous la forme de variations du timing de l'activité neuromusculaire et des gestes articulatoires » (Bonnot, 1990b). Nous avons en effet déjà insisté sur le caractère flexible de la parole coarticulée, qui pouvait se manifester par une variabilité importante du timing des gestes articulatoires (CHAPITRE II, en particulier section II.3). Nous verrons en pratique, dans nos propositions sur la production de parole codée, qu'il est indispensable de spécifier et d'inclure au moins en partie les contrôles moteurs de la parole afin de rendre compte de cette activité *timée* coordonnée.

IV.1.6. La coopération temporelle de ces modules

Le traitement effectué par chacune de ces composantes est automatique. Il est sériel (suivant l'ordre des modules) mais également parallèle si l'on considère que les modules peuvent fonctionner en même temps. En effet, avant de commencer à articuler les mots, le locuteur n'a pas planifié la totalité du message mais cela vient au fur et à mesure. Dès qu'un premier concept ou notion est disponible, l'encodage grammatical débute. Il y a alors activation des mots appropriés et le début d'une construction de phrase. Le début de cette phrase est envoyé à l'encodage phonologique. Le locuteur peut alors commencer à articuler même avec peu de syllabes disponibles. Le *self-monitoring* peut entrer en action dès que des bribes de parole interne sont rendues disponibles. Par ailleurs, la modularité du système est une sorte de protection contre les erreurs de traitement : chaque composante fait son travail indépendamment de ce que font les autres composantes dans le même temps (pour une discussion sur la modularité de ce système, voir Segui & Ferrand, 2000 ; voir aussi de Ruiter, 2000).

Ce qu'il faut retenir de ce modèle, c'est bien que la syllabe apparaît comme une unité de production en parole : elle constitue une unité de planification de la parole, qui se produit de manière sérielle (les syllabes sont encodées phonologiquement de gauche à droite ; Meyer, 1990, 1991 ; Meyer & Schriefers, 1991) et représente une unité naturelle de production de parole pour différentes langues (pour le français, Ferrand et al., 1996, pour l'anglais, 1997 ; pour une revue, voir Segui & Ferrand, 2000). Elle est généralement considérée sous deux points de vue différents : elle peut être conçue comme une unité ou *chunk* représentant des séquences fixes de segments phonologiques (ex : « ba », « bal ») ou bien comme un patron ou *schéma* abstrait représentant une structure syllabique indépendante du contenu segmental (ex : CV ou CVC) (pour une revue, voir Sevald et al., 1995). Ferrand et Segui (1998) plaident en faveur d'un modèle mixte pour la syllabe, dépendant de la tâche à effectuer : les locuteurs ont accès à la fois aux représentations segmentales et abstraites liées aux

mots. Il semblerait donc qu'il y ait une certaine indépendance de la structure de la syllabe et de son contenu segmental.

IV.2. Un modèle pour les gestes et la parole : le Sketch Model

De Ruiter (2000) propose une extension à ce modèle (voir Figure 18). A la manière de Levelt, il propose une architecture générale modulaire de production de gestes spontanés en relation avec la parole. Son modèle permet d'expliquer les gestes déictiques (gestes de pointage) et les gestes iconiques dans lesquels il inclut les gestes métaphoriques (voir McNeill, 1992, voir aussi section III.1.2). Il prend en compte également les emblèmes et ajoute les pantomimes qui correspondent à des gestes qui imitent des activités motrices fonctionnelles et qui ne peuvent pas être issus directement de l'imagerie mentale (il est à noter que dans la typologie de McNeill, les pantomimes font partie des gestes iconiques). De Ruiter a une approche basée sur les représentations et les traitements (*Representations and Processes*) ; il s'intéresse donc aux représentations qui sont à la base du traitement des gestes et non pas à la sémiologie des gestes (cela explique sa motivation dans les quelques changements de typologie).

C'est le *conceptualizer* du modèle de Levelt (voir section IV.1.1) qui est à l'origine de l'initiation des gestes en synchronie avec les concepts lexicaux. Pour tous les gestes, le conceptualizer forme une représentation du geste à effectuer, le « sketch », et l'envoie au planificateur de gestes (*gesture planner*) qui sera en charge de construire un programme moteur à partir de ce sketch. L'exécution du geste se fera finalement par les unités du contrôle moteur.

IV.2.1. La conception du sketch

En se basant sur le fait que ces gestes spontanés sont produits en relation étroite avec la parole, de Ruiter fait l'hypothèse d'un module de traitement commun[3] pour l'initiation du geste et de la parole, le *conceptualizer.* Ce module a en effet accès à une base de connaissances, la mémoire de travail, qui contient entre autres des informations conceptuelles pour les mots et des informations imagées (spatio-temporelles) pour les gestes. De Ruiter ajoute au modèle un stock de gestes conventionnels, le gestuaire (*gestuary*), auquel le conceptualizer peut également accéder. Le conceptualizer forme le sketch qui peut contenir différentes informations selon le type de gestes à encoder : des informations

[3] Il est à noter que de nombreux auteurs ne partagent pas ce point de vue d'un module de traitement commun pour l'initiation du geste et de la parole : ils proposent alors une interaction entre gestes et parole à plusieurs étapes du traitement (voir une autre proposition de module commun p. 22 ; voir aussi de manière plus générale la discussion p. 22). Le sujet principal de cette thèse ne concernant pas directement les gestes co-verbaux, nous ne discuterons pas de la position de de Ruiter pour son modèle.

spatio-temporelles sur les images mentales pour les gestes iconiques, des informations sur la direction d'un référent (pour les gestes déictiques), des informations sur la forme du geste récupérées dans le gestuaire (pour les gestes déictiques et les emblèmes) et des informations relatives aux connaissances motrices (*action schema*) pour les pantomimes.

Selon l'intention communicative du locuteur, le conceptualizer peut si nécessaire activer, en plus des concepts appropriés, des images (à partir de la mémoire de travail) relatives à ces concepts qui seront à la base du geste. C'est le cas pour les gestes iconiques. L'intention communicative a alors une composante conceptuelle (propositionnelle) qui permet de former le message préverbal et une composante « imagée » (avec une organisation spatio-temporelle) qui est à l'origine de la formation du geste. Un seul geste peut remplacer une quantité importante de mots, ainsi les concepts difficiles à exprimer avec des mots sont plus facilement exprimés à l'aide de représentations imagées ; ceci explique la grande utilisation des gestes iconiques pendant les narrations. La forme du geste iconique est déterminée par l'information imagée. Le conceptualizer extrait de l'image mentale les caractéristiques pertinentes et crée une représentation qui est stockée dans le sketch. Cette représentation, appelée « trajectoire » (*trajectorie*), peut contenir une ou plusieurs caractéristiques spatiales et temporelles (mouvements) de la représentation imagée. Pour les gestes iconiques, le sketch contient également la position relative du locuteur par rapport aux trajectoires ; ceci permettra au locuteur d'adopter une certaine position spatiale par rapport au contenu sémantique du geste (voir section IV.1.1, la *perspective taking*). Pour les emblèmes, le conceptualizer accède au gestuaire qui contient un ensemble de programmes moteurs abstraits (c'est-à-dire pas entièrement spécifiés) des gestes conventionnels, les *templates*, indexés par le concept qu'ils représentent. Si le conceptualizer trouve dans le gestuaire un emblème qui peut exprimer le concept lexical de l'intention communicative du locuteur, alors il stocke dans le sketch un pointeur (*pointer*) qui fait référence à cet emblème. Pour les gestes déictiques, le conceptualizer stocke dans le sketch un vecteur qui indique la direction de la position du référent que l'on doit pointer. Comme la forme des gestes de pointage est souvent prédéterminée (conventionnelle selon la culture : c'est souvent l'index qui est utilisé mais ce peut être aussi la main entière, le pouce, etc.), le conceptualizer stocke également dans le sketch un pointeur qui fait référence à un template de pointage contenu dans le gestuaire. Finalement, une fois que le sketch est généré, il est envoyé au planificateur de gestes.

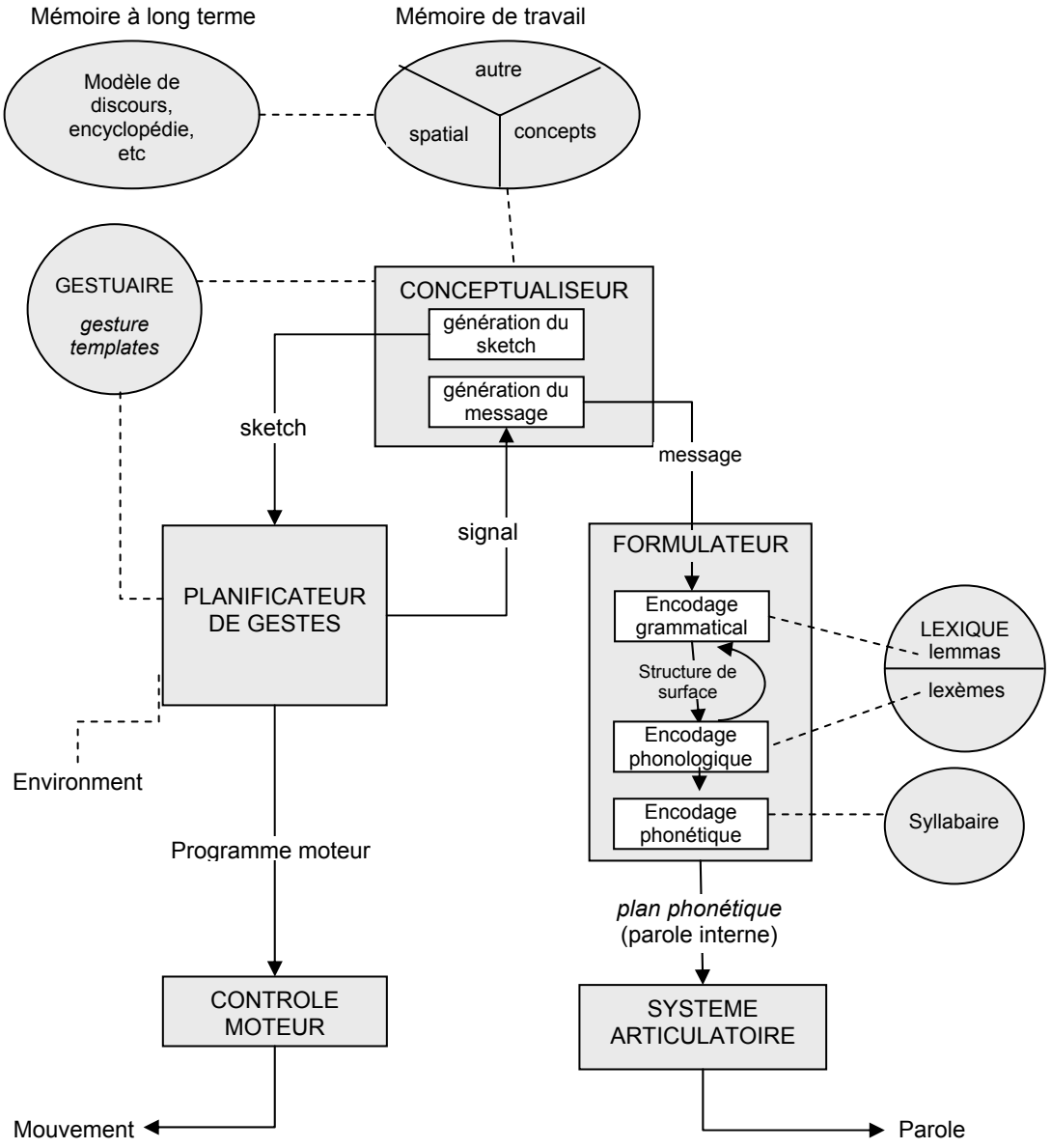


Figure 18. Le Sketch Model. Figure adaptée de de Ruiter, 2000.

IV.2.2. Le planificateur de gestes

Le planificateur de gestes est chargé de construire un programme moteur à partir du sketch et des informations qu'il contient. Pour ce faire, il a accès au gestuaire (pour retrouver les pointeurs), aux procédures motrices et à certaines informations extérieures sur les contraintes de l'environnement. Il doit construire le programme moteur en tenant compte de ces contraintes environnementales (de façon à ne pas heurter des objets ou personnes durant l'exécution des gestes). Il complète ainsi les programmes moteurs (templates) du gestuaire (pour les gestes déictiques et emblèmes) en spécifiant les caractéristiques non encodées dans le gestuaire et qui dépendent du contexte et de la situation telles que la partie du corps qui va exécuter le geste, la durée du geste, la position du membre dans l'espace, etc. C'est en effet lui qui gère les différentes parties du corps (*body-part allocation*) en

choisissant le membre qui va effectuer le geste (ce peut être le doigt, la main, l'autre main, mais aussi la tête, le regard, la jambe, etc.). Pour les gestes iconiques, le planificateur de geste peut programmer la main pour qu'elle « retrace » la trajectoire spatio-temporelle. Il peut également se produire des « fusions » de gestes de types différents (voir un exemple détaillé de fusion entre un geste iconique et un geste déictique dans de Ruiter, 2000, pp. 303-304). De Ruiter fait alors l'hypothèse que si l'un des gestes implique un pointeur sur un template dans le gestuaire, le planificateur de gestes va utiliser les caractéristiques non spécifiées du programme moteur pour effectuer les autres gestes fusionnés.

Finalement, une fois que le programme moteur est entièrement spécifié, il est envoyé aux modules de contrôle moteur bas niveau qui vont exécuter le mouvement.

IV.3. Coordination temporelle des deux modèles

La coordination entre le geste et la parole est assurée par le conceptualizer qui a la possibilité d'envoyer des signaux de synchronisation. Comme il a été dit précédemment, le début du geste précède souvent le début de la parole. Pour rendre compte de ce phénomène, de Ruiter (2000) postule que le message préverbal reste en attente dans le conceptualizer tant que le programme moteur généré à partir du sketch n'est pas entièrement formé. Le sketch et le message préverbal sont formés à peu près en même temps dans le conceptualizer, mais le conceptualizer peut envoyer le sketch au planificateur de gestes même si le message préverbal n'est pas encore entièrement formé. Le conceptualizer attend ensuite que le planificateur de gestes lui envoie un signal (dès que le programme moteur est prêt) pour transmettre le message préverbal au formulator. Une fois que le message préverbal est envoyé, le conceptualizer envoie un signal résumé (*resume*) au planificateur de gestes pour lui signifier de continuer à envoyer le reste du programme moteur (c'est ce qui explique le phénomène de pré-stroke dans les gestes de pointage, voir section III.2.1). Enfin, le conceptualizer attend que le message verbal soit entièrement produit pour signaler au planificateur de gestes la fin du mouvement (soit la fin de la phase de post-stroke durant laquelle la main reste immobile le temps que la parole soit entièrement produite, soit la fin du geste répétitif qui est exécuté en boucle tant que la parole n'est pas terminée).

Au terme de cette première partie, nous pouvons dresser le portait suivant. La parole est un produit complexe co-articulé clairement orienté-vers-la-sortie. Elle est le résultat d'une coordination spatio-temporelle, finement contrôlée, des différents gestes articulatoires du conduit vocal, qui vise à délivrer une information linguistique qui peut être récupérée auditivement mais pas uniquement. Sa récupération peut se faire en effet par d'autres modalités ; en particulier, la vision a un rôle de premier ordre. Informations visuelles et auditives sont ensuite intégrées et fusionnées pour aboutir à l'identification d'un unique percept multimodal linguistique. La communication parlée ne se limite pas à la parole audio-visuelle mais elle inclut tout un ensemble de gestes, dits *co-verbaux*, qui sont produits naturellement par le locuteur et qui accompagnent cette parole. L'étude de ces gestes nous enseigne qu'ils sont coordonnés avec le message linguistique qui est délivré et que leur organisation temporelle peut être différente selon la catégorie de gestes.

La Langue française Parlée Complétée (LPC) est un système artificiel, qui délivre des clés manuelles avec la parole pour les sourds. Ce code manuel est complètement dépendant de la parole : ses composantes – la forme de main et sa position près du visage – sont déterminées par l'information phonologique et syllabique et ce système, à lui seul, ne permet pas de percevoir ce qui est dit. C'est bien l'association LPC-parole qui permet au sourd de récupérer un percept unique. Sa récupération met en jeu surtout la modalité visuelle mais elle implique également (et de plus en plus) la modalité auditive avec le développement croissant des implants cochléaires (les restes auditifs amplifiés ou non par les appareils ne sont pas non plus à négliger). A ce jour, aucune étude n'a été menée en production permettant de comprendre la coordination de la main, des lèvres et du son. Ce sera l'objet de notre recherche expérimentale. Par rapport aux gestes co-verbaux, il apparaît clairement que le code LPC doit être davantage *ancré* sur la parole : nous verrons en particulier que le geste de la main lors du codage LPC suit une organisation temporelle spécifique en coordination étroite avec la parole audio-visuelle.

Partie II
Partie expérimentale

« In running speech, the consonant-vowel hand cues are coarticulated in a one-to-one relation ship with the syllables of the language »
Nicholls & Ling, 1982

Ainsi que nous l'avons précédemment exposé, la Langue française Parlée Complétée délivre une information labio-manuelle suffisamment complète pour permettre au sourd de percevoir correctement la parole. Mais quel est le secret de ce système ? La coordination entre les mouvements orofaciaux de la parole et les mouvements manuels du code LPC semble jouer un rôle crucial. C'est en tous cas ce que l'on peut retenir des diverses tentatives de mise au point de système de synthèse LPC ; la précision de cette coordination temporelle pourrait être un facteur majeur. Les études de simulation de Bratakos et al. (1998) indiquent notamment que la synchronisation temporelle entre la main et le son semble critique : le sujet décodeur accepte mal que la clé manuelle soit trop en retard sur le son. Une avance de 100 ms comme celle proposée empiriquement par l'équipe de Duchnowski et al. (2000) améliore en effet significativement la réception des séquences codées. Le fondateur de la méthode, Orin Cornett, insistait quant à lui sur une certaine « synchronisation » entre le geste manuel du Cued Speech et le son. La lecture de la littérature ne nous a cependant pas permis d'établir un patron de coordination clair pour le Cued Speech car il n'existe en réalité aucune étude de la production du Cued Speech, ni de son équivalent français.

Notre étude s'attachera à décrire la coordination des gestes manuels et orofaciaux au cours de la production de séquences de syllabes CV. Ce type de séquences constitue une base pour étudier l'organisation de ce système qui a pour unité la syllabe consonne-voyelle. Nous examinerons avec précision les relations temporelles entre la main, les lèvres et le son produit, en nous appuyant sur des événements temporels pertinents repérables sur les signaux manuels, articulatoires et acoustiques. La main effectuant des transitions d'une position à l'autre du visage, nous pouvons de manière fiable repérer les débuts et fins de transition. Les lèvres constituent la partie du conduit vocal la plus facile d'accès. L'aire intérolabiale du contour interne des lèvres nous donnera un bon aperçu des voyelles et notamment de leur cible. Enfin, nous pourrons relier ces événements au signal acoustique de parole et ainsi établir un patron général de coordination du geste de la main en rapport avec la parole. Derrière

les patrons temporels dégagés se profile la question du contrôle. C'est une question qui aboutira à une hypothèse de contrôle s'appuyant sur l'organisation motrice de la parole et du code LPC.

C'est d'abord par l'étude du comportement d'un seul locuteur-codeur que nous établirons ce patron temporel (CHAPITRE V). L'étude de trois autres sujets permettra de confirmer nos résultats et plus généralement d'étudier la variabilité dans la production de la LPC (CHAPITRE VI). Nous examinerons également la situation particulière de l'anticipation labiale dans la coarticulation et comment la main se comporte dans un tel contexte (CHAPITRE VII). Finalement nous verrons par une étude perceptive comment cette coordination particulière manuelle et orofaciale est perçue par le sourd décodeur (CHAPITRE VIII).

CHAPITRE V.

Etude pilote de la production de syllabes codées

Ce chapitre a fait l'objet d'une publication dans la revue *Speech Communication*

Attina V., Beautemps D., Cathiard M.-A. & Odisio M. (2004).
A pilot study of temporal organization in Cued Speech production of French syllables:
Rules for a Cued Speech synthesizer.
Speech Communication, 44, pp. 197-214.

Ce chapitre concerne l'étude complète des coordinations manuelles et oro-faciales dans la production de code LPC par une locutrice-codeuse (GB).

Nous exposerons tout d'abord, dans l'expérience 1, les relations temporelles entre la main, les lèvres et le son pour des séquences consonne-voyelle CV (sans sens). Nous rappelons qu'une syllabe de type CV se code à la position cible correspondant à la voyelle V avec la configuration de la main correspondant à la consonne C (voir section I.4 et Figure 5 p. 22). Cette étude permettra de comprendre comment la main est coordonnée à la parole dans la production de séquences syllabiques « simples », c'est-à-dire n'impliquant aucun mouvement de doigts (aucun changement de configuration consonantique). Cette étude se focalise donc uniquement sur le mouvement de la main de position à position. Ces transitions manuelles d'une position à l'autre du visage définissent donc le geste de base.

Nous verrons ensuite, dans l'expérience 2, comment les articulateurs de la LPC se coordonnent quand il est nécessaire de changer de configuration consonantique au cours de séquences syllabiques codées. Dans cette seconde étude, nous analyserons, en plus des mouvements de la main, ceux des doigts pour la formation des différentes configurations consonantiques, en relation avec les mouvements labiaux et le son. Ces deux études nous permettront d'établir un patron temporel général de la production de syllabes CV codées par un même sujet.

V.1. La locutrice-codeuse : GB

A l'époque de l'enregistrement, la locutrice GB est une femme française âgée de 36 ans qui pratique la LPC quotidiennement depuis décembre 1992, en raison de la surdité de son enfant. Elle n'a aucun handicap et une élocution normale. Elle a été sélectionnée pour cette étude par l'intermédiaire de Martine Marthouret, orthophoniste au CHU de Grenoble, pour ses qualités de codage (fluidité du code, visibilité de la main et des formes labiales). GB codait à la maison environ trois heures par jour pour son enfant. Elle a obtenu son diplôme de codeuse professionnelle[4] en mai 1996 et depuis code en classe tous les jours selon les besoins de codage des élèves (la durée de codage peut en effet varier

[4] Ce diplôme national est attribué par un jury (constitué de professionnels de la surdité, de la LPC et de sourds décodeurs) qui évalue la précision des codes manuels et la fluidité et la rapidité des mouvements. Il évalue également la capacité de la personne à s'adapter aux sourds décodeurs (grande différence entre un enfant en maternelle et un adulte) ainsi que sa capacité à résumer et synthétiser un discours (ceci est nécessaire parfois quand le rythme et le niveau de parole du locuteur sont trop soutenus). Enfin des questions autour de la surdité et des problèmes engendrés sont posées. Le nombre d'enfants utilisant la LPC est en augmentation chaque année, ainsi il y aurait un besoin estimé de 1000 codeurs en France alors qu'à l'heure actuelle, seulement 220 codeurs sont diplômés (Maunoury, 2004). Il est à noter qu'une Licence Professionnelle qualifiante sur un an a été ouverte en septembre 2005 à Paris (Université Pierre et Marie Curie, Paris 6) pour la formation des futurs codeurs : cette formation est à la fois théorique (sur le développement de l'enfant, sur la communication, l'éducation et la pédagogie), technique et pratique (LPC).

d'un enfant à l'autre et ceci dépend également de la matière enseignée). A titre d'indication, pour un enfant en CE2, elle assurait environ six à huit heures de codage par semaine.

V.2. Expérience 1 : étude des transitions manuelles avec configuration consonantique fixe

Cette étude vise à définir les relations temporelles entre main, lèvres et son pour des syllabes CV. Nous pourrons dans cette partie juger du décalage temporel entre le son et la position manuelle correspondante, et répondre à la question générale de la coordination main-lèvres-son dans l'organisation de la parole coarticulée.

V.2.1. Corpus d'étude

Nous avons constitué un corpus permettant d'étudier des transitions manuelles d'une position à l'autre du visage avec maintien d'une même clé consonantique tout au long de la séquence. Cette étude explore les cinq positions cibles de la LPC et utilise les configurations consonantiques n° 1 (index uniquement visible) et n°5 (main entièrement visible). Les séquences sont de la forme [Ca.CV_1.CV_2.CV_1] avec les consonnes [m], [p] et [t] pour C et les voyelles [a], [i], [u], [ø] et [e] pour V_1 et V_2. Les consonnes retenues ont l'avantage d'être bien identifiables sur les signaux acoustiques et articulatoires : la consonne occlusive [t] est caractérisée par une période de silence dans le signal acoustique et la consonne [m] est caractérisée par une occlusion bilabiale nettement visible sur le signal articulatoire (l'aire aux lèvres est nulle durant la production du [m]). La consonne occlusive [p] peut être repérée facilement à la fois au niveau acoustique et articulatoire. Finalement les cinq voyelles sélectionnées correspondent chacune à une position cible LPC (la voyelle [a] pour la position « côté », [i] pour la position « bouche », [u] pour la position « menton », [ø] pour la position « pommette » et [e] pour la position « cou », voir Figure 5) et ont été choisies en raison de leur forme labiale claire. La configuration consonantique étant fixée durant toute la production de chaque séquence, en variant les voyelles, nous obtenons 20 séquences de quatre syllabes de type [Ca.CV_1.CV_2.CV_1] pour chaque consonne (par exemple, [mamamima] pour la consonne [m] et les voyelles [a] et [i]), soit un total de 60 séquences. En variant les voyelles de la même façon, nous avons en plus enregistré 20 séquences de type [ma.V_1.V_2.mV_1] (par exemple, [maaima]) qui nous permettront d'étudier la coordination du geste labial de voyelle à voyelle sans perturbation par la consonne avec le geste manuel « de base » de position à position. Pour toutes les séquences (80 au total), la première syllabe sert à mettre la main en position initiale de codage et n'est donc pas analysée. Nous traitons en fait la syllabe englobée S_2 (c'est-à-dire les transitions manuelles de S_1 à S_2 et de S_2 à S_3) de chaque séquence « $S_0S_1S_2S_3$ » afin d'éviter les éventuels effets prosodiques liés aux début et fin de séquence.

V.2.2. Acquisition des données

V.2.2.1. Description du matériel

L'enregistrement audiovisuel a été effectué dans la chambre sourde de l'I.C.P. au moyen du poste Visage-Parole (Lallouache, 1991). La locutrice est assise sur une chaise, sa tête étant maintenue immobile par un casque solidaire du siège afin d'éviter tout mouvement hors du champ des caméras. Elle porte des lunettes opaques (qui protègent sa vue du fort éclairage nécessaire au bon contraste vidéo) marquées d'un point coloré, celui sur le verre gauche servant par la suite de point de référence aux différentes mesures (voir Figure 19). Afin de suivre les mouvements manuels de la locutrice dans le plan, nous avons collé une pastille de couleur bleue sur le dos de la main utilisée pour coder (il s'agit de la main droite pour cette codeuse). Ses lèvres sont maquillées en bleu (très saturé) afin de nous permettre par la suite de détecter de façon automatique les contours labiaux sur chaque image.

Deux caméras vidéos placées en face du sujet ont été utilisées : l'une en champ large permettant de filmer le visage de la codeuse et tous les mouvements de main et l'autre en mode zoom sur les lèvres afin de permettre une détection précise du contour intérolabial. Les deux caméras sont synchrones et filment à une fréquence de 50 Hz, soit une trame vidéo toutes les 20 ms. L'enregistrement de chaque caméra est stocké sur une cassette Bétacam SP par un magnétoscope. Le son est enregistré sur la bande audio de la vidéo et de fait synchrone avec l'image.

Au début de l'enregistrement, un « bip » est émis par pression sur un bouton-poussoir : simultanément, un pavé de huit diodes (LED) est allumé dans le champ des deux caméras et enregistré sur la trame impaire de chacune des deux images vidéo. Ce bip sert en fait de point de référence temporelle pour chacun des deux enregistrements, les deux bandes vidéo ayant chacune leur propre time-code. Ainsi, le repérage du bip sur les deux bandes vidéos permettra, une fois l'enregistrement terminé, de mettre en correspondance les deux enregistrements. Enfin, un tableau quadrillé (carreaux de 1 cm x 1 cm) est placé dans le champ des deux caméras afin de permettre la conversion des tailles de pixel en centimètre.

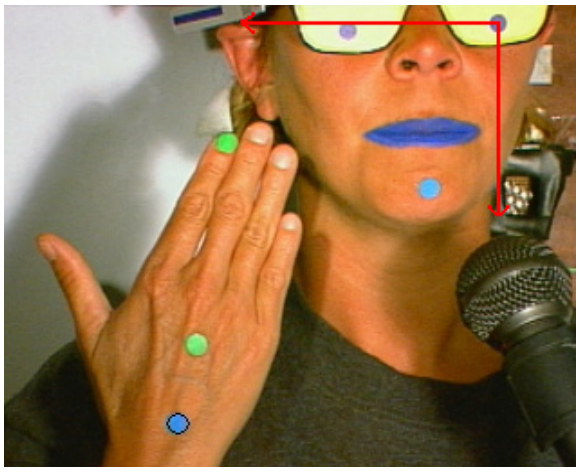


Figure 19. Photo de la locutrice-codeuse lors de l'enregistrement avec les positions des pastilles colorées sur la main et sur le visage.

V.2.2.2. Traitement des données

Avant de pouvoir analyser les données, il est nécessaire de faire des traitements préliminaires sur les enregistrements. Chaque séquence codée est repérée sur la vidéo par ses time-codes de début et de fin (un time-code consiste en une chaîne de caractères identifiant de manière unique l'image sous la forme « heures : minutes : secondes : images »). Nous utilisons ensuite un système de capture d'images et de son (système CAPTURE interne au laboratoire ; Audouy, 2000) pour numériser[5], à partir des time-codes, les trames des séquences vidéo (sauvées sous forme d'images en format bitmap 24 bits et codées en composantes rouge, vert, bleu, RVB) à une fréquence de 25 Hz et le son de manière synchrone à la vidéo à une fréquence de 22050 Hz.

Sur les images extraites des séquences en mode zoom pour les lèvres, nous effectuons un suivi automatique des contours labiaux grâce au logiciel TACLE (Traitement Automatique du Contour des LEvres, logiciel interne au laboratoire ; Audouy, 2000). Cette application permet d'extraire, à partir d'une séquence d'images numérisées, des paramètres descripteurs des lèvres vues de face, tels que l'aperture des lèvres (séparation verticale), l'étirement (séparation horizontale) et l'aire intérolabiale par exemple. Ce logiciel extrait tout d'abord les trames paires et impaires de chaque image pour reconstituer deux images séparées de 20 millisecondes, les lignes manquantes étant reconstituées par interpolation. Le traitement nécessite ensuite plusieurs étapes :

1. Un seuillage numérique (ou chroma-key) des images est d'abord appliqué sur les images brutes afin de noircir les zones maquillées en bleu. Nous avons au préalable, pour chaque séquence,

[5] L'acquisition se fait à partir de la bande vidéo Betacam par le biais d'une carte Matrox Meteor I. Une image est constituée de deux trames successives et entrelacées, c'est-à-dire les lignes impaires correspondent à la première trame et les lignes paires à la seconde. Le résultat est une image entrelacée à 25 Hz.

déterminé manuellement deux fenêtres de traitement pour détecter la pastille de référence sur la lunette et le contour des lèvres. Pour chacune de ces fenêtres, nous indiquons manuellement un point de départ par un clic de la souris (nécessaire au logiciel pour faire une recherche automatique du contour) et choisissons des paramètres de réglage concernant la teinte, la saturation et la luminance (codage TSL[6] ou HSL en anglais) pour le seuillage automatique de la couleur bleue des lèvres et de la pastille de référence. Ces paramètres de seuillage correspondent aux valeurs de teintes minimales et maximales ainsi qu'à la saturation maximale. Chaque pixel de la fenêtre de traitement satisfaisant les conditions précisées est ensuite noirci.

2. Un filtrage médian est ensuite appliqué afin de régulariser les zones noires (élimination des pixels noirs isolés qui empêcheraient une bonne détection des contours par la suite).

3. Les contours des masses noires des lèvres et de la pastille de référence sont ensuite déterminés automatiquement suivant un algorithme de détection (pour plus de détails voir Audouy, 2000, p. 29-31).

4. Enfin la dernière étape consiste à calculer les paramètres labiaux à partir des points constituant les contours interne et externe ; l'aire intérolabiale S est donnée par la formule suivante du calcul d'aire délimitée par un contour discret (dans ce cas, il s'agit du contour interne), en considérant que le contour est un ensemble de points (X_k, Y_k) (Lallouache, 1991) :

$$S = \frac{1}{2}\sum_k X_k Y_{k+1} - Y_k X_{k+1}.$$

Parmi les mesures labiales de face (aperture, étirement, aire intérolabiale), c'est l'aire aux lèvres que nous avons sélectionnée et que nous utiliserons dans la suite pour toutes nos études. Au niveau de la détection et de la précision, ce paramètre articulatoire est sans doute le plus robuste des trois. Ce paramètre, fortement lié aux autres dimensions labiales, représente de fait un bon paramètre articulatoire descripteur des lèvres (Abry et al., 1980 ; Abry & Boë, 1986). Résultat d'une coopération entre les lèvres et la machoire par un *contrôle orienté vers la sortie* (Abry et al., 1980), il nous donne une image précise des segments articulés au niveau des lèvres au cours du temps ; l'aire aux lèvres permet notamment de distinguer totalement les voyelles arrondies des voyelles non arrondies (Abry et al., 1980) et les voyelles ouvertes des voyelles fermées (Robert-Ribes et al., 1998). Ce paramètre a de plus une grande importance au niveau acoustique (Stevens & House, 1955) et c'est le paramètre visuel

[6] Les images sont codées dans l'espace TSL et non pas dans l'espace RVB, essentiellement pour des raisons d'ergonomie au niveau de l'interface utilisateur. Le modèle TSL est en effet plus proche de la perception « naturelle » des couleurs par l'humain.

le mieux corrélé à l'acoustique du conduit vocal (voir notamment, Badin et al., 1994). Grâce au logiciel TACLE, qui effectue la conversion pixel-cm, nous obtenons donc pour chaque séquence une valeur en cm^2 de ce paramètre articulatoire labial toutes les 20 ms ; le tracé de ces valeurs successives nous permet de visualiser le décours temporel de l'aire intérolabiale.

Pour le mouvement de la main dans le plan, nous avons développé un programme (Matlab) qui permet de suivre les pastilles colorées posées sur le dos de la main de la locutrice à partir des séquences d'images en plan large (main + visage). Avant de commencer le suivi, ce programme détrame (et reconstitue par interpolation) les images de façon à avoir une information toutes les 20 ms comme pour les lèvres. Le suivi de pastilles est effectué sur les valeurs RVB des pixels composant les pastilles : un seuil pour chacune des composantes rouge, vert et bleu des pixels est fixé (ce seuil est suffisant étant donné que les couleurs des pastilles ont été choisies de manière à être très différentes de la couleur naturelle de la main). Ce seuil correspond en fait aux valeurs maximales et minimales (calculées sur un échantillon d'images) que les pixels de la pastille peuvent avoir selon l'orientation de la main et la luminosité. A partir d'un point de départ sur la pastille, le programme délimite une fenêtre de traitement qui contient la pastille à détecter, puis parcourt cette fenêtre et stocke tous les pixels de l'image dont les niveaux RVB sont compris dans les intervalles RVB de recherche. Les pixels composant la pastille étant repérés, le programme détermine son contour et calcule ensuite le centre de gravité de la pastille. Ce centre est ensuite stocké dans un vecteur résultat et le suivi se poursuit sur l'image suivante (en prenant automatiquement comme point de départ les coordonnées du centre de gravité de la pastille de l'image précédente). Les coordonnées du centre de gravité de la pastille de référence sur les lunettes sont calculées de la même façon. Ainsi le suivi de pastille peut s'effectuer automatiquement pour toute une séquence d'images. Pour chaque séquence de l'enregistrement, il est nécessaire de déterminer manuellement le point de départ du suivi sur la première image de la séquence ; ce repérage est fait pour toutes les séquences (par le biais d'un programme qui affiche chaque première image de séquence et qui enregistre tous les points de départs cliqués à l'écran avec la souris) et permet par la suite d'effectuer un suivi en boucle pour la totalité de l'enregistrement. Le résultat du suivi donne pour toutes les séquences les coordonnées x et y dans l'image (pixels) de la pastille colorée calculées par rapport à la pastille de référence sur la lunette gauche (cette opération est nécessaire afin d'éliminer les éventuels mouvements de tête du sujet), et qui sont au final converties en centimètres. Nous obtenons ainsi, comme pour les paramètres labiaux et de manière synchrone, une position (en cm) toutes les 20 ms : le tracé reflète donc la trajectoire de la main dans le plan 2-D.

Nous obtenons donc, à la fin des différents traitements, quatre signaux synchrones au cours du temps, illustrés sur la Figure 20 pour la séquence [tatatuta] : (1) le décours temporel de l'aire aux lèvres (cm^2),

(2) la trajectoire en x de la pastille colorée de la main (cm), (3) la trajectoire en y de la pastille colorée de la main (cm) et (4) le signal acoustique correspondant. En ce qui concerne le paramètre labial, nous retrouvons un décours caractéristique de la séquence prononcée [tatatuta] : l'aire intérolabiale est élevée pour la production de la voyelle [a] (ici aux environs de 3 cm^2) caractérisée par une grande aperture aux lèvres et beaucoup plus petite (proche de 0,5 cm^2, voire même inférieure) pour la production de la voyelle arrondie protruse [u] (nous pouvons constater sur le signal une nette diminution de l'aire pour la syllabe du milieu). En ce qui concerne la trajectoire manuelle, nous pouvons remarquer sur la figure que les valeurs de position en x et en y de la pastille au cours du temps forment des trajectoires caractérisées par des transitions (mouvement de la main vers une position cible) et des plateaux (atteinte et tenue de la cible) (voir section III.2.1 pour une comparaison avec les phases du geste co-verbal). Les valeurs de position sont calculées en fonction du point de référence sur la lunette du sujet dans le repère superposé sur la Figure 19. Ainsi, sur la Figure 20, une diminution en x correspond à un rapprochement horizontal de la main vers le point de référence et une augmentation en y correspond à un éloignement vertical de la main vers le bas : c'est ce qui se produit lorsqu'on passe de la position « côté » codant la voyelle [a] à la position « menton » codant la voyelle [u].

Sur ces différents signaux, nous repérons certains événements temporels concernant la syllabe S_2 (en particulier) de chaque séquence « $S_0S_1S_2S_3$ » ; ces événements qui marquent les débuts et fins de geste sont un reflet des stratégies de contrôle moteur (Perkell, 1990 ; Perkell & Matthies, 1992). Sur le signal acoustique (en nous appuyant quand nécessaire sur la structure formantique et sur le spectrogramme), nous repérons tous les débuts et fins de consonnes et voyelles de chaque séquence (tous les segments sont étiquetés afin de calculer par la suite les durées de consonne et de syllabe). En particulier, pour l'analyse temporelle, l'étiquette A1 représente le début (instant du début de l'occlusion) de la consonne de la syllabe S_2. Sur le décours de l'aire aux lèvres, nous repérons la cible vocalique de la syllabe S_2, étiquetée L2, grâce au pic d'accélération (voir ci-dessous et voir Figure 20). Sur les trajectoires en x et en y de la main, nous définissons les débuts et fins de transitions manuelles à partir des pics d'accélération et de décélération (ces pics permettant de déterminer les phases de transition et de tenue d'un geste, voir Schmidt, 1988 ; Perkell & Matthies, 1992) : l'étiquette M1 correspond au début de la transition manuelle vers la position cible codant S_2, M2 correspond à l'atteinte de cette cible (et donc à la fin de la transition manuelle), M3 correspond au début du mouvement vers la cible suivante codant la syllabe S_3 et M4 à la fin de cette deuxième transition. Lorsque les valeurs de position en x et en y ont des débuts et fins de mouvements qui ne correspondent pas, nous tenons compte du premier instant dans le temps sur l'une des deux

trajectoires pour le début de mouvement et du dernier instant pour la fin de mouvement de la phase considérée (voir Figure 20).

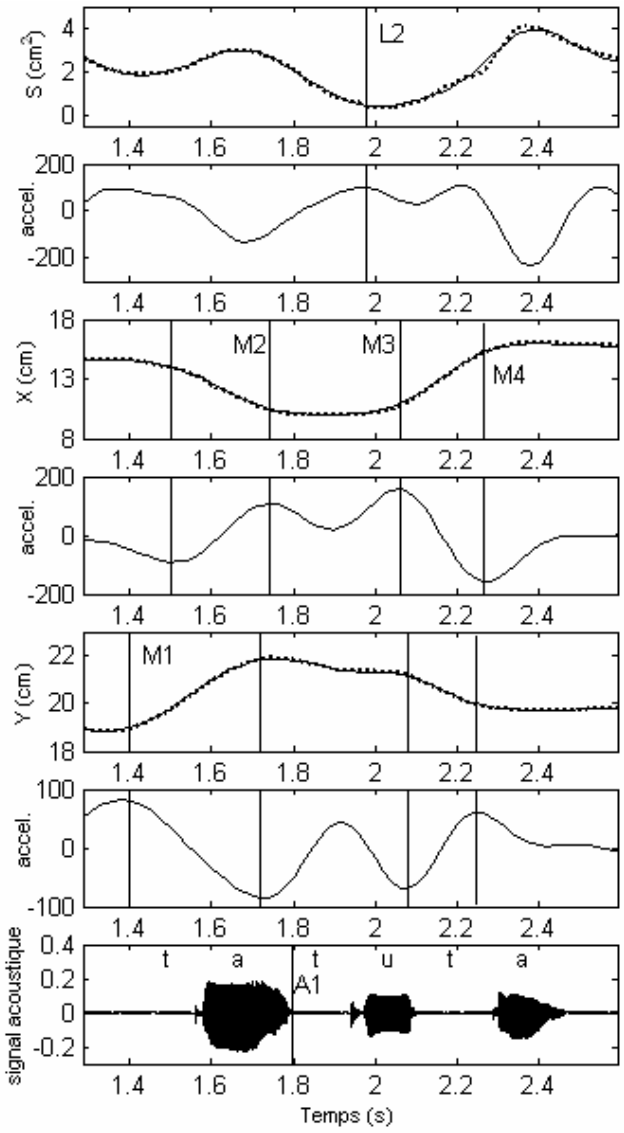


Figure 20. Tracé des différents signaux au cours du temps pour la portion [tatuta] extraite de la séquence [tatatuta]. De haut en bas : (1) décours temporel de l'aire aux lèvres S (cm^2), avec en dessous (2) le profil d'accélération correspondant ; (3) tracé de la position de la coordonnée x de la pastille sur le dos de la main (cm) avec (4) son profil d'accélération ; (5) tracé de la position de la coordonnée y de la pastille sur le dos de la main (cm) avec (6) son profil d'accélération ; (7) signal acoustique correspondant. Les signaux en traits pointillés correspondent aux signaux bruts et les signaux en traits pleins correspondent aux signaux filtrés que nous utilisons pour calculer le profil d'accélération nécessaire à l'étiquetage (voir texte).

Pour chacun des signaux de position, nous avons tracé sur la Figure 20 à la fois les signaux bruts, indiqués par des tracés en pointillés et les signaux filtrés, indiqués en traits pleins (on peut remarquer cependant que les signaux bruts sont très peu bruités). Le filtrage des données est nécessaire pour calculer la courbe d'accélération correspondante sur laquelle nous nous appuyons pour étiqueter nos

signaux : nous appliquons sur nos données (aire aux lèvres S et coordonnées en x et y de la pastille de la main) un filtre passe-bas (type Chebyshev) avec une fréquence de coupure de 4 Hz[7].

Le profil d'accélération est ensuite obtenu par la formule suivante, utilisée généralement dans l'étude du contrôle moteur (voir par exemple, Winter, 1990, p. 48) et qui tient compte du point courant et des points précédent et suivant pour le calcul de l'accélération en un point donné, ce qui a pour avantage de ne pas induire de déphasage :

$$\left.\frac{d^2x(t)}{dt^2}\right|_{t=t_0} \approx \frac{x(t_0+\Delta)-2\cdot x(t_0)+x(t_0-\Delta)}{\Delta^2}$$

Notons que cette formule s'obtient à partir de la somme ([1]+[2]) des développements limités de 2nd ordre à chaque instant t_0 des valeurs de position (filtrées) de $x(t_0+\Delta)$ et $x(t_0-\Delta)$ ci-dessous et en négligeant les termes d'ordre supérieur à 2, $o_1(\Delta^2)$ et $o_2(\Delta^2)$:

$$x(t_0+\Delta) = x(t_0) + \Delta\cdot\left.\frac{dx(t)}{dt}\right|_{t=t_0} + \frac{\Delta^2}{2}\cdot\left.\frac{d^2x(t)}{dt^2}\right|_{t=t_0} + o_1(\Delta^2) \qquad [1]$$

$$x(t_0-\Delta) = x(t_0) - \Delta\cdot\left.\frac{dx(t)}{dt}\right|_{t=t_0} + \frac{\Delta^2}{2}\cdot\left.\frac{d^2x(t)}{dt^2}\right|_{t=t_0} + o_2(\Delta^2) \qquad [2]$$

V.2.2.3. Caractéristiques de production

Comme nous l'avons précisé précédemment, c'est la syllabe S_2 de chaque séquence « $S_0S_1S_2S_3$ » que nous analysons. A partir de la segmentation du signal acoustique, nous avons calculé la durée moyenne de la syllabe CV S_2 (définie par l'instant de début de l'occlusion de la consonne de S_2 jusqu'au début de l'occlusion de la consonne de S_3), ainsi que la durée moyenne de la consonne de S_2 (délimitée par le début de l'occlusion de la consonne de S_2 jusqu'au début de la structure formantique de la voyelle de S_2), pour l'ensemble des séquences.

En ce qui concerne la coordination temporelle des différents articulateurs de la LPC, nous avons calculé différentes durées d'intervalles temporels à partir des étiquettes décrites précédemment pour chaque syllabe S_2 :

[7] Nous avons par ailleurs vérifié la pertinence du choix de cette fréquence de coupure sur un échantillon de signaux en analysant le résidu de la différence entre les signaux bruts et les signaux filtrés comme une fonction de la fréquence de coupure du filtrage. Nous pouvons ainsi déterminer la fréquence de coupure la mieux appropriée, c'est-à-dire un compromis entre la quantité de distorsion du signal et la quantité de bruit (voir Winter, 1990, pp. 41-42).

- l'intervalle M1M2 désigne la durée de la transition manuelle pour coder S_2, c'est-à-dire le temps entre le début du mouvement et l'atteinte de la position cible codant la voyelle de S_2 ;
- l'intervalle M2M3 désigne la durée du maintien de la main en position cible codant S_2 ;
- M1A1 désigne la durée entre le début du mouvement manuel et le début acoustique de la consonne de la syllabe S_2 ;
- A1M2 désigne la durée entre le début acoustique de la consonne et l'atteinte de la position manuelle cible (la fin du mouvement) ;
- M2L2 désigne la durée entre l'atteinte de la cible manuelle et l'atteinte de la cible vocalique labiale ;
- M3L2 désigne la durée entre l'atteinte de la cible vocalique aux lèvres et le départ de la main vers la position suivante (qui code la syllabe S_3) ;
- M3M4 désigne la durée de la transition manuelle pour coder la syllabe suivante S_3.

Chaque intervalle est obtenu par soustraction des valeurs temporelles d'une paire d'étiquettes (par exemple, M1A1= A1 – M1 (ms)). Avec cette définition, si on considère un intervalle temporel XY, une valeur positive pour cet intervalle indique donc que l'événement X se produit avant l'événement Y. A l'opposé, une valeur négative pour cet intervalle indique que l'événement X se produit après l'événement Y.

Notons que le lecteur pourra trouver dans le texte des intervalles supplémentaires : ceux-ci sont déduits à partir des intervalles indiqués ci-dessus (par exemple, l'intervalle A1L2, entre le début acoustique de la consonne et la cible labiale de la voyelle, provient de l'addition des intervalles A1M2 et M2L2).

Pour la comparaison statistique des résultats, nous utilisons le test t de Student (test bilatéral) pour la comparaison de deux moyennes ainsi que l'ANOVA, si besoin. Les conditions d'application de ces tests sont vérifiées au préalable par le test de Lilliefors pour la normalité des distributions et par le test de Levene pour l'homogénéité des variances (l'ensemble du traitement statistique des données a été fait grâce à des fonctions Matlab pré-existantes, version 6.5, ou que nous avons développées si besoin). De manière générale, il est habituel de faire confiance à la robustesse de ces tests même si les conditions ne sont pas complètement respectées (Scheffé, 1959 ; Hays, 1988 pour les conditions d'homoscédascité ; Serniclaes, 2005, communication personnelle). Cependant, dans le cas où nos données ne respectent pas strictement ces conditions, nous appliquons également un test non paramétrique équivalent afin de vérifier la significativité du test (ces résultats seront indiqués en note dans le texte).

V.2.3. Résultats

V.2.3.1. Séquences VCV

Sur l'ensemble des réalisations, la syllabe CV mesurée à partir du signal acoustique dure en moyenne 399,5 ms (s= 95,6 ms[8]). Cette durée nous permet de calculer un rythme moyen de parole codée de 2,5 Hz. On remarque d'emblée que le rythme de cette locutrice-codeuse est assez lent par rapport à un rythme normal de parole qui engendre environ 6±1 syllabes par seconde (Sorokin et al., 1980). Ce net ralentissement peut s'expliquer en partie par l'ajout du code manuel à la parole (nous discuterons de ce point plus loin).

En ce qui concerne le code manuel, nous obtenons une moyenne de 276 ms (s= 54 ms) pour l'intervalle M1M2, soit la durée de la transition manuelle vers la position cible et une moyenne de 205 ms (s= 109 ms) pour l'intervalle M2M3 qui représente la tenue de la main en position cible.

Comment ce code manuel se place-t-il par rapport aux événements acoustiques et articulatoires dans le flux de la parole ? Nous avons obtenu une durée moyenne de 239 ms (s= 87 ms) pour l'intervalle M1A1 ; ce qui signifie clairement que la main débute sa transition bien avant le début de la production acoustique de la syllabe. La durée moyenne de 37 ms (s= 76 ms) pour l'intervalle A1M2 indique que la main arrive en position cible juste après le début acoustique de la syllabe CV (quasiment en synchronie), en fait au tout début de la consonne. Plus précisément, la durée de l'intervalle A1M2 représente 16% de la durée totale de la consonne (la consonne durant en moyenne 234 ms, s= 68 ms). Par rapport aux lèvres, la position manuelle est atteinte 256 ms (s= 101 ms) avant la cible labiale vocalique (M2L2). Cette avance significative de la main sur les lèvres indique que l'information sur la voyelle transmise par le code manuel (indiquée par la position) est donnée et donc disponible avant l'information vocalique aux lèvres.

La durée moyenne de 51 ms (s= 60 ms) obtenue pour l'intervalle M3L2 indique que la main repart vers la position cible suivante (qui code la syllabe S_3) avant même que la cible vocalique ne soit réalisée aux lèvres. Enfin, en ce qui concerne la transition manuelle, nous avons obtenu une durée moyenne de 282 ms (s= 51 ms) pour l'intervalle M3M4 ; cette durée, très proche de la durée de la transition manuelle précédente (intervalle M1M2, m= 276 ms), suggère une certaine stabilité dans le rythme

[8] Les écarts-types donnés dans cette thèse correspondent aux estimations non biaisées des écarts-types au niveau de la population mère calculées à partir d'un échantillon de n individus. Ils sont définis par : $s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x_n})^2}{n-1}}$, où $\bar{x_n} = \frac{1}{n}\sum_{i=1}^{n} x_i$, x_i étant la valeur du paramètre étudié pour le $i^{ème}$ individu de l'échantillon considéré.

manuel général (la comparaison des deux moyennes par un test t indique une différence non significative au risque α= 5% ; le t calculé $|t_c|$= 0,6 ddl= 118).

V.2.3.2. Séquences V-V

En ce qui concerne les 20 séquences de type « $maV_1V_2mV_1$ », nous avons calculé les mêmes intervalles temporels pour la syllabe S_2. Au dépouillement de nos signaux acoustiques, nous avons constaté que cette syllabe était constituée de la voyelle V_2 précédée d'une attaque glottale (« coup de glotte ») que la locutrice a produite systématiquement dans toutes les réalisations (au début de chaque syllabe S_2 entre les deux voyelles consécutives V_1 et V_2). Ces attaques glottales révèlent une stratégie de parole claire (« clear speech » ; Picheny et al., 1986) avec séparation des frontières vocaliques, ce qui permet en codage LPC de clairement marquer les positions vocaliques. Nous avons donc repéré le début de la portion de silence de cette attaque (qui prend virtuellement la place de la consonne) par l'étiquette A1. Le repérage de l'attaque glottale et de la voyelle délimite une syllabe d'une durée moyenne de 383,6 ms (s= 61,4 ms). Nous obtenons, en ce qui concerne le code manuel, une moyenne de 267 ms (s= 55 ms) pour la durée de l'intervalle M1M2 et une moyenne de 157 ms (s= 97 ms) pour la durée de M2M3.

En ce qui concerne les relations entre la main et le son, nous avons obtenu une durée moyenne de 183 ms (s= 79 ms) pour l'intervalle M1A1 et de 84 ms (s= 64 ms) pour l'intervalle A1M2, soit une durée M1M2 de 267 ms. La main débute donc sa transition avant le début du silence de l'attaque glottale insérée devant la voyelle et atteint la position cible durant ce silence (le silence entre les deux voyelles dure en moyenne 188,5 ms, s= 39,5 ms ; la position est donc atteinte dans les 45% de ce silence). La valeur moyenne de 73 ms (s= 66 ms) pour l'intervalle M2L2 indique que la position cible manuelle est atteinte avant la réalisation de la cible vocalique aux lèvres. Enfin, la main repart vers la position suivante 84 ms après que la voyelle soit visible aux lèvres (M3L2, m= - 84 ms, s= 68 ms). La transition manuelle jusqu'à la cible suivante dure en moyenne 276 ms (M3M4, s= 56 ms) ; de nouveau nous pouvons observer une similitude dans la durée des transitions manuelles d'une position à l'autre (il n'y a en effet pas de différence significative entre les deux moyennes M1M2 et M3M4 au risque α= 5% : $|t_c|$= 0,5 ddl= 38).

V.2.4. Vers un schéma de coordination « position-lèvres-son »

Les résultats obtenus dans cette étude révèlent une coordination main-lèvres-son particulière dans le décours de la parole coarticulée codée. Au niveau du geste manuel, nous obtenons des valeurs de transition (M1M2 et M3M4) de 276 ms et 282 ms pour les séquences VCV et de 267 ms et 276 ms pour les séquences V-V. Il apparaît donc que ces durées de transitions manuelles sont très proches

dans les deux conditions et cela, que ce soit pour aller vers la position cible de la syllabe S_2 ou bien pour repartir vers la position suivante. Les valeurs de tenues, quant à elles, s'élèvent à 205 ms en moyenne pour les séquences VCV et à 157 ms pour les séquences V-V : la durée de la tenue de la main en position cible montre donc un peu plus de variabilité.

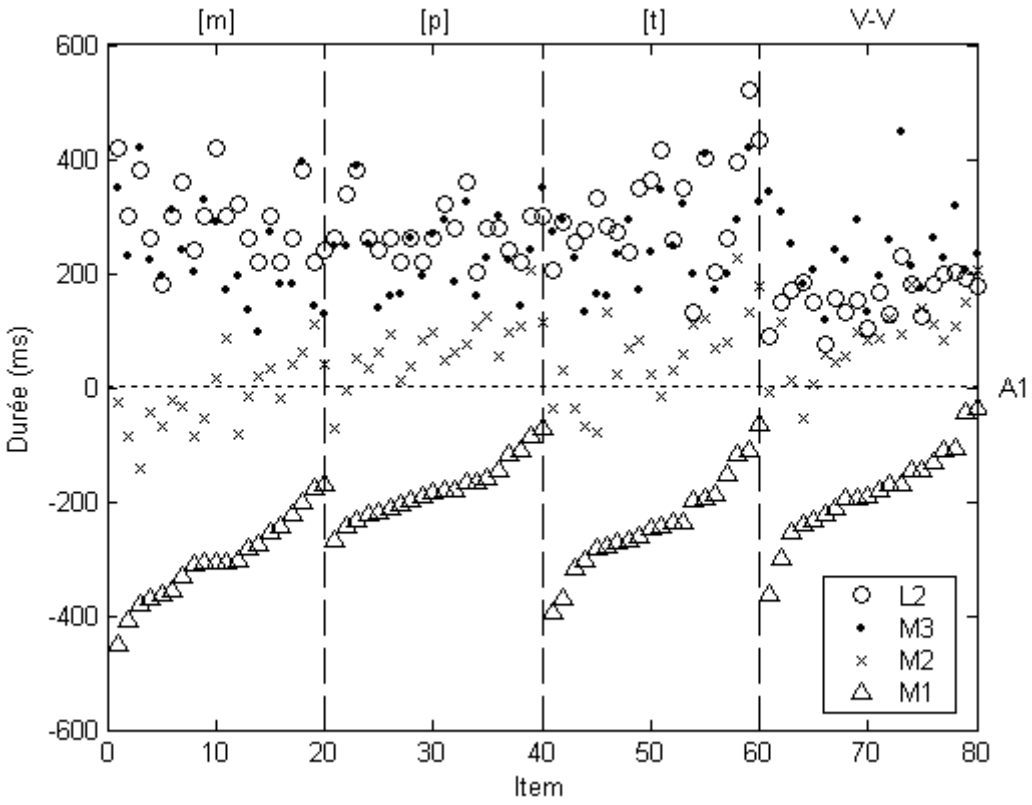


Figure 21. Positionement temporel des événements M1 (début du geste manuel), M2 (fin du geste manuel), L2 (cible labiale vocalique) et M3 (début du geste manuel vers la position suivante) par rapport à A1 (repéré par 0), début acoustique de la syllabe, pour les différents types de séquences (avec et sans consonne). Les items sont ordonnés dans l'ordre croissant de M1 (ce qui correspond à une courbe de fréquence cumulée avec ici la variable en y et le rang en x).

Au niveau de la coordination de la main avec la parole, un patron temporel relativement stable se dégage des résultats. Notre résultat majeur est le fait que **le geste manuel anticipe sur l'acoustique et sur le geste labial**. L'initiation du geste est en avance sur le début de la syllabe de 183 ms à 239 ms en moyenne (M1A1) dans cette étude. La Figure 21 montre plus précisément, pour chacune des séquences étudiées, le positionnement temporel, par rapport au début acoustique de la syllabe (A1, soit le début de la consonne ou de l'attaque glottale), du début du geste manuel (M1), de la fin du geste manuel (M2), de la cible labiale vocalique (L2) et du début du geste manuel vers la position suivante (M3). Nous pouvons voir que pour toutes les séquences produites avec ou sans consonne, l'événement M1 se produit toujours en avance par rapport à l'événement A1 (sur la figure, tous les triangles qui représentent M1 sont effectivement en-dessous de la ligne en pointillés qui représente

A1). La valeur d'anticipation observée selon la séquence testée est très variable (de 38 ms à 453 ms ; ces valeurs correspondent aux valeurs minimum et maximum de M1A1) mais dans tous les cas reste positive.

Nous avons vu que la main termine son mouvement de 37 ms à 84 ms en moyenne après le début acoustique de la syllabe (A1M2). La Figure 21 montre la répartition de la cible LPC (M2, représentée par des croix) par rapport au début acoustique de la syllabe. Globalement cet événement se produit aux alentours de A1. La position cible LPC peut même parfois être atteinte avant le début acoustique de la syllabe : c'est le cas surtout pour les séquences incluant la consonne [m] (beaucoup de croix sur la figure sont placées en-dessous de l'événement A1). Dans l'ensemble, nous pouvons remarquer que la majorité de la distribution se retrouve dans les 100 premières millisecondes de la syllabe, donc au début de la consonne.

En moyenne la position manuelle qui code la voyelle est atteinte de 73 ms à 256 ms avant la cible labiale (M2L2). La Figure 21 montre que globalement sur l'ensemble des séquences la position (M2, représentée par des croix) est atteinte avant la cible labiale (L2, représentée par des cercles). C'est toujours le cas pour les séquences VCV avec consonne. Nous pouvons remarquer un comportement un peu plus variable pour les séquences V-V sans consonne : la réalisation de la cible labiale L2 semble beaucoup plus proche de l'atteinte de la cible LPC M2. Nous pouvons même noter que pour trois séquences, la position LPC est atteinte juste après la réalisation de la cible vocalique labiale. Cette variabilité est en fait principalement due à la cible vocalique aux lèvres : par rapport au son, cette cible est clairement anticipée dans les séquences V-V (sur la Figure 21, la position des cercles représentant la cible labiale L2 est beaucoup plus proche du début acoustique de la syllabe A1 pour les séquences V-V avec attaque glottale que pour celles contenant une consonne). Au niveau statistique, une ANOVA[9] à un facteur appliquée aux valeurs de durées entre le début acoustique de la syllabe A1 et la cible labiale vocalique L2 selon le type de séquences ([m], [p], [t] et V-V) montre une différence significative entre les conditions ($F(3, 79)= 22,44$ $p< .01$). Des comparaisons multiples a posteriori par le test de Scheffé montrent que ce sont bien les séquences V-V qui diffèrent des autres ($p< .01$). La cible labiale de la voyelle est donc anticipée pour ces séquences sans consonne. C'est le phénomène d'anticipation coarticulatoire dont nous avons parlé précédemment (voir notamment section II.3) ; dans cette condition, les lèvres profitent du silence acoustique précédant la voyelle pour se positionner à l'avance pour l'articulation de cette voyelle. L'attaque glottale est un geste qui, comme son nom

[9] Les variances n'étant pas toutes homogènes, nous avons recomparé les moyennes par le test non paramétrique de Kruskal-Wallis : nous trouvons également une différence significative entre les moyennes ($p< .01$).

l'indique, se produit au niveau du larynx ; il s'agit d'une fermeture des cordes vocales avec forte tension. Dans le même temps, il est possible de produire au niveau supraglottique la forme labiale de la voyelle. Au contraire, nous observons dans les transitions VCV, une mise en place de la cible labiale qui se produit plus tard. Rappelons que parmi les trois consonnes que nous avons utilisées [m, p, t], les consonnes [m] et [p] sont des bilabiales, caractérisées par une occlusion des lèvres (aire aux lèvres nulle). Le geste labial consonantique est donc fortement masquant pour la voyelle qui suit, puisqu'il faut attendre nécessairement la fin de l'occlusion consonantique pour pouvoir réaliser l'ouverture caractéristique de la voyelle. Pour les séquences incluant la consonne [t], nous trouvons un peu plus de variabilité. La consonne occlusive coronale [t] est caractérisée par un contact de la pointe de la langue au niveau des alvéoles dentales ; contrairement aux autres consonnes, elle masque moins la voyelle au niveau labial. En tous les cas, l'ouverture labiale peut s'initier au cours de la réalisation de la consonne. Cependant, la production du [t] impliquant le contrôle de la pointe de la langue peut modifier ou retarder la mise en forme de la voyelle qui suit ; ce qui explique pourquoi nous ne trouvons pas le même patron de résultats que pour les séquences V-V.

Enfin, nous avons observé une grande variabilité concernant les relations entre la cible labiale vocalique et le début du geste manuel vers la position suivante selon le type de séquence : la main démarre son geste en moyenne 51 ms avant la cible labiale pour les séquences VCV et en moyenne 84 ms après pour les séquences V-V. La main repart donc vers une autre position aux environs de la réalisation de la cible labiale, soit avant ou après cette cible. Nous pouvons observer sur la Figure 21 que cette tendance moyenne est globalement conservée séquence par séquence : pour les séquences VCV, la main semble initier son mouvement avant la réalisation de la cible labiale (sur la Figure 21, les points représentant le début du geste manuel M3 sont répartis en-dessous des cercles représentant la cible labiale L2 pour les séquences avec consonne [m], [p] et [t]) alors que pour les séquences V-V, le patron est inversé (les points sont plutôt positionnés au-dessus des cercles). Cette différence s'explique complètement par la variabilité liée à la cible labiale (voir ci-dessus).

V.2.5. En résumé

Nous pouvons donc proposer le patron général suivant (schématisé sur la Figure 22) : pour une syllabe de type CV, pour laquelle aucun changement de configuration consonantique n'est nécessaire, la main débute sa transition bien avant le début acoustique de la syllabe avec une avance de l'ordre de 200 ms et atteint la position cible en début de consonne, soit bien avant la cible labiale vocalique. Après une phase de tenue en position cible de 150 à 200 ms, la main repart ensuite vers la position suivante durant la production de la voyelle de la syllabe CV. Il y a donc bien dans la production de la LPC française une avance de la main sur les gestes orofaciaux.

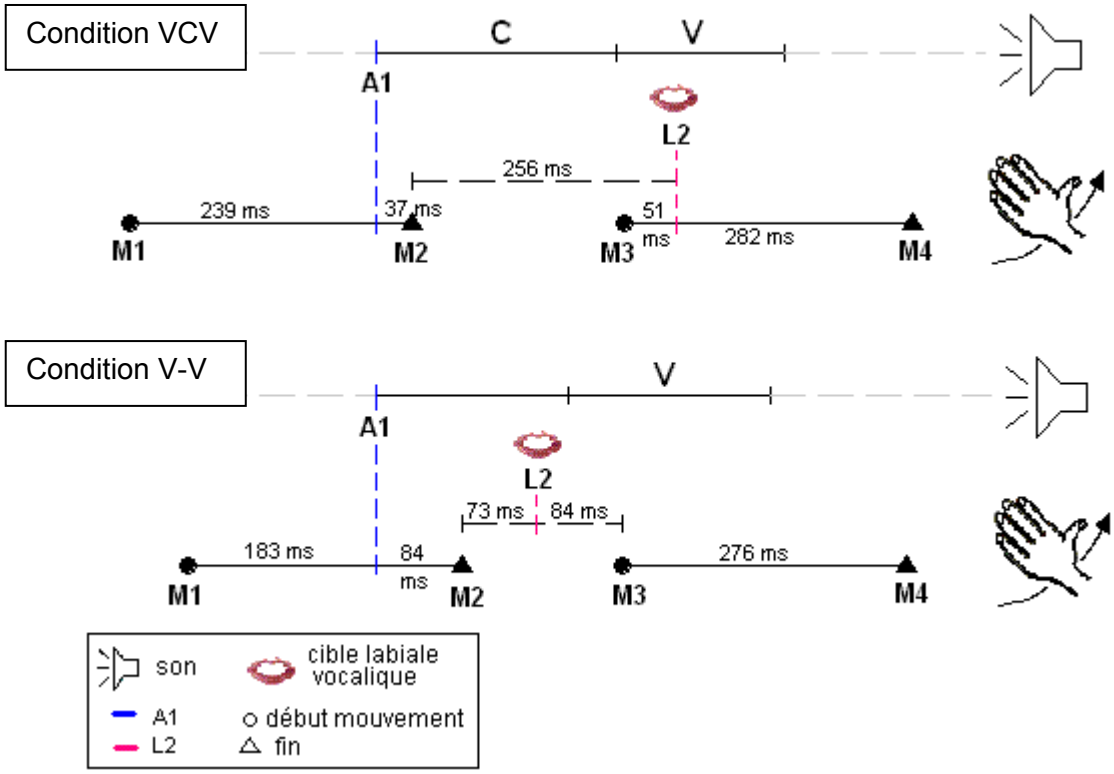


Figure 22. Schéma de la coordination temporelle main-lèvres-son obtenue dans l'expérience 1 pour les deux conditions pour la codeuse GB. En haut, schéma temporel des résultats de la condition VCV pour la production de syllabe CV. En bas, schéma temporel des résultats de la condition V-V.

V.3. Expérience 2 : formation des clés manuelles

Est-ce que le patron observé précédemment pour des séquences codées simplement par des transitions manuelles d'une position à une autre est maintenu dans le cas de séquences codées impliquant à la fois un changement de position et un changement de configuration de la main (clé consonantique) ? C'est à cette question que tente de répondre l'expérience 2. Plus précisément, la main est-elle toujours en avance sur le son et les lèvres lorsque des changements de clés consonantiques (impliquant des mouvements de doigts) viennent s'ajouter aux transitions manuelles ? Cette expérience nous renseigne également sur l'organisation temporelle de la transmission de l'information consonantique et vocalique dans le code LPC. Les deux informations suivent-elles une organisation séquentielle où l'on trouverait pour une syllabe CV, d'abord la consonne puis la voyelle ou bien sont-elles organisées différemment ? Nous avons vu dans le CHAPITRE II, que la parole n'est pas, comme on pourrait intuitivement le penser, organisée de cette manière. Qu'en est-il des gestes manuels de la LPC ? Le changement de configuration consonantique se fait-il bien en milieu de transition manuelle comme l'ont défini Duchnowski et al. (2000) dans leur affichage « synchronous » (voir section I.8.4.2.2) ?

V.3.1. Corpus d'étude

Nous avons constitué dans cette étude deux corpora différents nous permettant d'une part, d'étudier la formation des configurations consonantiques effectuées à une même position du visage, et d'autre part, d'étudier la formation de ces clés consonantiques quand elles s'ajoutent aux transitions manuelles d'une position à l'autre du visage. Le premier corpus nous permettra de voir à quel moment la clé consonantique apparaît par rapport au début de la réalisation articulatoire de la consonne quand il n'y a pas de déplacement de la main. Et le deuxième corpus nous permettra de voir à quel moment la clé consonantique apparaît dans une transition de voyelle à voyelle, quand il y a un déplacement de la main d'une position à une autre.

Dans cette étude, nous utilisons un gant de données pour capturer les mouvements des doigts. Afin de faciliter le dépouillement et la lecture des données, nous nous imposons la contrainte de n'avoir qu'un seul doigt qui bouge durant le passage d'une configuration à l'autre dans la construction du corpus.

V.3.1.1. Changements de configurations consonantiques

Ce corpus a été constitué de manière à avoir un certain nombre de changements de clés consonantiques à une même position cible manuelle. Les séquences sont de la forme [mV.C_1V.C_2V] avec [a] ou [ɛ] pour la voyelle V (la voyelle, qui détermine la position, étant fixée durant la séquence) et les paires {[p], [k]}, {[s], [b]} et {[b], [m]} pour les consonnes C_1 et C_2. Il y a donc deux choix possibles de positions : le « côté » pour les séquences contenant la voyelle [a] et le « menton » pour celles contenant la voyelle [ɛ]. De plus, les paires de consonnes sont choisies de façon à impliquer un seul mouvement de doigt ; c'est le majeur pour le passage de [p] à [k], l'index pour le passage de [s] à [b] et le pouce pour le passage de [b] à [m]. Par exemple, une séquence [mabama] est codée sur la position « côté » et implique un geste d'effacement puis de réapparition du pouce (voir Figure 23). Nous avons ainsi six séquences différentes qui sont répétées dix fois ; nous obtenons donc un total de 60 séquences. Comme précédemment, nous analysons la syllabe du milieu, soit S_2 dans chaque séquence « $S_1S_2S_3$ ».

Figure 23. Code de la séquence [mabama]. Toute la séquence est codée en position de main sur le « côté ». Durant la séquence, la configuration consonantique est changée pour le codage de la consonne [b] (passage de la configuration n°5 à la n°4, puis retour à la n°5) : ceci est caractérisé par un geste d'effacement puis de réapparition du pouce.

V.3.1.2. Changements de configuration consonantique avec transitions manuelles

Ce corpus a été constitué de manière à avoir à la fois des changements de clé consonantique et des déplacements de main d'une position cible à une autre. Les séquences sont de la forme [$mV_1.C_1V_2.C_2V_1$] avec les paires {[a], [u]}, {[a], [e]} et {[u], [e]} pour les voyelles V_1 et V_2, et les paires {[p], [k]}, {[ʃ], [g]}, {[s], [b]} et {[b], [m]} pour les consonnes C_1 et C_2. Trois positions sont donc ici concernées : le « côté » pour [a], le « menton » pour [u] et le « cou » pour [e]. Pour les configurations consonantiques, le doigt étudié est le majeur pour le passage de [p] à [k] et de [ʃ] à [g], l'index pour le passage de [s] à [b] et le pouce pour le passage de [b] à [m]. Comme illustré sur la Figure 24 pour la séquence [mabuma], le codage de ces séquences implique un déplacement de la main d'une position à une autre du visage (dans l'exemple, la main se déplace de la position « côté » qui code la voyelle [a] à la position « menton » pour coder la voyelle [u] puis retourne sur le « côté ») et en même temps, un changement de configuration de main (dans l'exemple, la main passe de la configuration n°5 à la n°4 par la disparition du pouce, puis à la n°5 par la réapparition du pouce). Nous avons donc douze séquences différentes, répétées cinq fois ; ce qui nous donne au total 60 séquences. Durant cet enregistrement, une erreur de codage s'étant produite pour l'une des séquences, nous n'avons donc analysé que 59 séquences au total. Nous étudions la syllabe du milieu S_2 dans chaque séquence « $S_1S_2S_3$ ».

Figure 24. Code de la séquence [mabuma]. La séquence implique à la fois un changement de clé consonantique pour la consonne [b] (disparition du pouce) et un déplacement de la main (de la position « côté » vers la position « menton »).

V.3.2. Acquisition des données

V.3.2.1. Description du matériel

Le même dispositif expérimental que dans l'expérience 1 a été utilisé pour cette étude. La codeuse enregistrée est aussi GB. Le seul changement réside dans l'utilisation d'un gant de données (*Cyberglove*) pour acquérir les mouvements des doigts produits durant les séquences codées. Ce gant de données est muni de 18 capteurs angulaires placés au niveau des articulations (et aussi entre les doigts). Il nous fournit un fichier de données brutes linéairement reliées à la déviation de l'angle formé par les phalanges au niveau de l'articulation. Sa fréquence d'échantillonnage est de 64 Hz. Pour synchroniser les données du gant avec le reste du dispositif, nous avons mis au point un système

particulier de synchronisation. A chaque début de séquence, le sujet doit presser sur un « bip » à l'aide de son pouce et de son index. A l'instant du contact des deux doigts, un signal audio (bip de synchronisation) est enregistré sur la bande audio de la vidéo. Ce contact est caractérisé par un plateau dans les données brutes du gant (pour les capteurs sensibles au mouvement de ces deux doigts) ; le début du plateau et du bip sonore repèrent l'instant du contact, ce qui permet ainsi de synchroniser l'acquisition issue du gant de données avec l'enregistrement vidéo. Pour suivre les mouvements de la main dans le plan 2-D, nous avons placé des pastilles colorées sur le dos du gant de données porté par le sujet (voir Figure 25).

Figure 25. Photo de la locutrice-codeuse portant le gant de données et positions des pastilles de couleur utilisées pour le suivi des mouvements. Le repère référentiel utilisé est tracé en superposition sur la photo.

V.3.2.2. Traitement des données

Le traitement des données de cette expérience s'est déroulé de la même façon que dans l'expérience 1 pour les mouvements labiaux, le signal acoustique et le mouvement de la main dans le plan (pour le corpus 2). En ce qui concerne les mouvements des doigts, pour chaque séquence nous prenons en compte les données issues du capteur placé sur le doigt en mouvement dans la séquence.

Ainsi, nous obtenons cinq signaux[10] synchrones au cours du temps, illustrés sur la Figure 26 pour la séquence [mabema] : (1) le décours temporel de l'aire aux lèvres (à une fréquence de 50 Hz), (2) la trajectoire en x de la pastille colorée placée sur le dos du gant (à une fréquence de 50 Hz), (3) la trajectoire en y de la pastille colorée (à une fréquence de 50 Hz), (4) la trajectoire du doigt en mouvement dans la séquence (à une fréquence de 64 Hz) et (5) le signal acoustique correspondant (à une fréquence de 22050 Hz). Comme nous pouvons le remarquer sur la figure, pour la séquence [mabema], une valeur d'aire nulle caractérise l'occlusion labiale pour la production des consonnes [m]

[10] C'est le cas pour le corpus 2 où la main se déplace entre deux positions. Pour le corpus 1, nous avons seulement trois signaux.

et [b] alors que pour les voyelles [a] et [e] cette aire est grande (autour de 5 cm^2 pour la voyelle [a]). En ce qui concerne le mouvement manuel, nous pouvons remarquer, comme dans l'expérience 1, des trajectoires caractérisées par des transitions et des plateaux en position cible. Les valeurs de position sont calculées en fonction du repère superposé sur la Figure 25 ; dans ce repère, un rapprochement horizontal de la main vers le point de référence est traduit par une diminution de la valeur de la position en x et un éloignement vertical de la main du point de référence est traduit par une diminution de la valeur de la position en y ; c'est ce que nous pouvons observer sur la figure pour le codage de la séquence [mabema] durant laquelle la main passe de la position « côté » pour la voyelle [a] à la position « cou » pour la voyelle [e]. En ce qui concerne le mouvement digital, nous avons sur la figure une représentation des données brutes issues du capteur du pouce ; c'est en effet ce doigt qui est actionné pour le changement de clé des consonnes [m] et [b] (clés n°5 et n°4). Nous pouvons remarquer la présence de transitions et de plateaux qui caractérisent le mouvement du doigt pour changer de configuration de main et le maintien de la configuration consonantique. Avant étiquetage, ces signaux sont filtrés (filtre passe-bas avec une fréquence de coupure de 4 Hz) pour le calcul de l'accélération.

Nous repérons ensuite les événements temporels par rapport à la syllabe S_2 de chaque séquence « $S_1S_2S_3$ ». Sur le signal acoustique, nous repérons tous les débuts et fins de consonnes et voyelles de chaque séquence. En particulier, pour l'analyse temporelle, l'étiquette A1 représente le début de la consonne de la syllabe S_2. Sur le décours de l'aire aux lèvres, nous repérons la cible vocalique de la syllabe S_2, étiquetée L2, grâce au pic d'accélération. Sur les trajectoires en x et en y de la main, nous repérons les débuts et fins de transitions manuelles à partir des pics d'accélération et de décélération : l'étiquette M1 correspond au début de la transition manuelle vers la position cible codant S_2, M2 correspond à l'atteinte de cette cible, M3 correspond au début du mouvement vers la cible suivante codant la syllabe S_3 et M4 à la fin de cette deuxième transition. Pour les mouvements de doigts, nous repérons, toujours à l'aide du profil d'accélération (voir Figure 26), les débuts et fins de gestes de formation de la configuration manuelle : D1 marque le début de cette formation pour la syllabe S_2, D2 la fin de la mise en place de la configuration consonantique, D3 le début de la formation de la configuration suivante qui code la consonne de S_3 et D4 la fin de cette configuration.

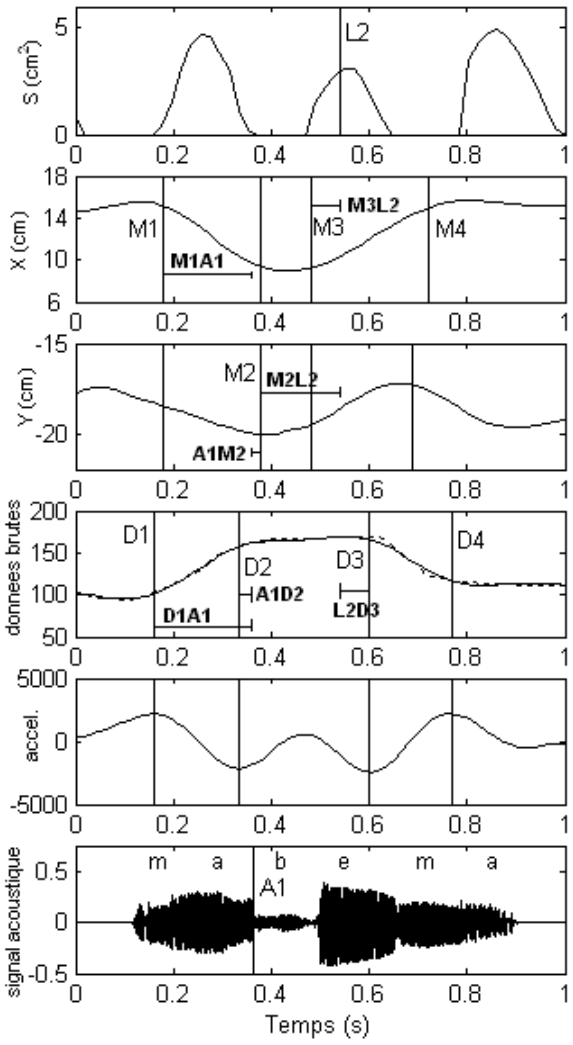


Figure 26. Tracé des différents signaux pour la séquence [mabema]. De haut en bas : (1) décours temporel de l'aire aux lèvres (cm²) ; (2) position de la coordonnée x de la pastille sur le dos du gant au cours du temps (cm) ; (3) position de la coordonnée y de la pastille sur le dos du gant au cours du temps (cm) ; (4) données brutes issues du capteur du gant (capteur sur le pouce) (en pointillés, données non filtrées ; en trait plein, données filtrées pour le calcul de l'accélération), (5) avec en-dessous le profil d'accélération correspondant ; (6) signal acoustique. Sur chacun des signaux sont superposés les étiquettes et les intervalles étudiés (voir texte).

V.3.2.3. Caractéristiques de production

Pour l'ensemble des séquences, nous avons calculé le rythme moyen syllabique de parole à partir de la segmentation du signal acoustique de la syllabe S_2 de chaque séquence « $S_1S_2S_3$ ». En ce qui concerne la coordination entre les gestes des doigts et la parole pour la syllabe S_2, nous avons calculé différentes durées d'intervalles temporels (différence arithmétique en millisecondes entre deux événements temporels) pour les corpus 1 et 2 :

– L'intervalle D1D2 désigne la durée pour former la configuration digitale codant la consonne de la syllabe S_2 ;

- D1A1 désigne la durée entre le début de la formation de la configuration digitale et le début acoustique de la consonne correspondante ;
- A1D2 désigne la durée entre le début acoustique de la consonne et la fin de la formation de la clé digitale ;
- D2L2 désigne la durée entre la fin de la formation de la configuration digitale et la cible vocalique labiale ;
- L2D3 désigne la durée entre la cible vocalique labiale et le début de la formation de la clé de la syllabe suivante S_3 ;
- D3D4 désigne la durée pour former la configuration digitale codant la consonne de la syllabe suivante S_3.

Pour le corpus 2, impliquant également des transitions manuelles près du visage, nous calculons en plus les intervalles suivants :

- M1M2 désigne la durée de la transition manuelle pour coder S_2, c'est-à-dire le temps entre le début du mouvement et l'atteinte de la position cible codant la voyelle de S_2 ;
- M2M3 désigne la durée du maintien de la main en position cible codant S_2 ;
- M1A1 désigne la durée entre le début du mouvement manuel et le début acoustique de la consonne de la syllabe S_2 ;
- A1M2 désigne la durée entre le début acoustique de la consonne et l'atteinte de la position manuelle cible (la fin du mouvement) ;
- M2L2 désigne la durée entre l'atteinte de la cible manuelle et l'atteinte de la cible vocalique labiale ;
- M3L2 désigne la durée entre le départ de la main vers la position suivante (qui code la syllabe S_3) et l'atteinte de la cible vocalique aux lèvres ;
- M3M4 désigne la durée de la transition manuelle pour coder la syllabe suivante S_3.

V.3.3. Résultats

V.3.3.1. Coordination manuelle et orofaciale à positions vocaliques manuelles fixes

Pour les séquences du corpus 1 avec voyelle inchangée, nous avons obtenu une moyenne de 275 ms (s= 33 ms) pour la durée de la syllabe CV (syllabe S_2 de chaque séquence « $S_1S_2S_3$ » mesurée sur le signal acoustique), ce qui nous donne un rythme syllabique moyen de 3,6 Hz. En ce qui concerne la durée de formation de la configuration digitale pour S_2, nous avons obtenu une moyenne de 170 ms

(s= 25 ms) pour D1D2. Par rapport au début acoustique de la syllabe, le geste des doigts pour former cette configuration est initié 124 ms avant (D1A1, s= 34 ms) et se termine en moyenne 46 ms après (A1D2, s= 35 ms), soit durant la première partie de la consonne ; la durée moyenne de la consonne étant de 152 ms, (s= 28 ms), la configuration digitale est en fait formée durant le premier tiers de la consonne. Par rapport aux indices labiaux, cette configuration digitale est entièrement formée 149 ms (D2L2, s= 50 ms) en moyenne avant la cible de la voyelle. En ce qui concerne la formation de la configuration consonantique suivante codant la syllabe S_3, nous avons obtenu une moyenne de 34 ms (s= 41 ms) pour l'intervalle L2D3, ce qui indique que la formation de cette clé débute une fois que la cible de la voyelle est réalisée aux lèvres. La durée moyenne de 162 ms (s= 15 ms) obtenue pour l'intervalle D3D4, très proche de la valeur obtenue pour D1D2, révèle finalement peu de variabilité dans la durée du geste digital pour former les différentes configurations consonantiques.

V.3.3.2. Coordination manuelle et orofaciale avec changement de positions vocaliques

Pour les séquences du corpus 2 avec transitions manuelles, nous avons obtenu une moyenne de 316,3 ms (s= 43,6 ms) pour la durée de la syllabe CV, équivalent à un rythme syllabique moyen de 3,2 Hz. En ce qui concerne la mise en place de la configuration manuelle, nous avons obtenu une moyenne de 168 ms (s= 22 ms) pour l'intervalle D1D2. Par rapport au son, la formation de la configuration débute en moyenne 171 ms (D1A1, s= 48 ms) avant le début acoustique de la syllabe et se termine également 3 ms avant (A1D2, m= - 3 ms, s= 45 ms), soit en quasi-synchronie avec le début de la consonne. Cette configuration de main est formée 208 ms (D2L2, s= 64 ms) en avance par rapport à la cible de la voyelle aux lèvres. En ce qui concerne la configuration digitale suivante qui code la consonne de la syllabe S_3, sa formation dure en moyenne 160 ms (D3D4, s= 9 ms) et débute 53 ms (L2D3, s= 54 ms) après la cible vocalique labiale.

Comment s'organisent temporellement les transitions manuelles par rapport à ce schéma ? Nous avons obtenu une moyenne de 238 ms (M1M2, s= 62 ms) pour la durée de la transition et une moyenne de 129 ms (M2M3, s= 70 ms) pour la durée de la tenue de la clé manuelle en position cible. Par rapport au début acoustique de la syllabe, la main débute sa transition en moyenne 205 ms (M1A1, s= 54,5 ms) en avance et atteint sa position 33 ms (A1M2, s= 50 ms) après le début acoustique de la syllabe : l'arrivée de la main en position se produit donc dans la première partie de la consonne acoustique (17%, la durée moyenne de la consonne s'élevant à 194 ms, s= 45 ms). Cette position est donc atteinte bien avant la cible vocalique labiale, en moyenne 172 ms (M2L2, s= 67 ms) en avance. La transition manuelle vers la position suivante (qui code la voyelle de la syllabe S_3) débute 43 ms

(M3L2, s= 76 ms) avant que cette cible vocalique ne soit réalisée aux lèvres et dure en moyenne 257 ms (M3M4, s= 38 ms).

V.3.4. Vers un schéma de coordination « position-clé-lèvres-son »

Nous avons obtenu un rythme syllabique moyen de 3,6 Hz (pour les séquences à positions fixes) à 3,2 Hz (pour les séquences à positions variables), puisqu'il faut effectuer dans cette dernière condition, en plus des changements de configurations manuelles, un déplacement de la main d'une position à une autre du visage. Il semble que le code manuel, nécessitant plus de temps dans la condition 2, ralentisse légèrement le rythme de parole par rapport à la condition 1.

En ce qui concerne les formations de configurations digitales (D1D2 et D3D4), nous avons obtenu des durées moyennes de 170 ms et 162 ms pour les séquences à positions fixes et de 168 ms et 160 ms pour les séquences avec transitions. Il apparaît que ces durées de formations de clés sont très proches les unes des autres. Elles ne sont en effet pas significativement différentes par le test de Kruskal-Wallis (test non paramétrique, équivalent de l'ANOVA, utilisé ici car les conditions de normalité et d'homoscédasticité n'étaient pas vérifiées : H_c= 8,5 < H_t= 11,345 pour α= 1% et ddl= 3). Ces résultats nous laissent donc supposer que le geste digital pour former les clés est relativement indépendant du fait qu'il y ait un déplacement ou non de la main d'une position à une autre du visage.

En ce qui concerne la coordination de ces clés avec le son, il apparaît, dans les deux conditions, que **la clé LPC est synchrone avec le début acoustique de la syllabe CV**, c'est-à-dire avec le début de la consonne. Plus précisément la configuration digitale commence à être formée 124 ms en moyenne avant le début acoustique de la syllabe (D1A1) et sera complètement mise en forme dès le début acoustique de la consonne (A1D2 s'élevant à 46 ms). Dans le cas où la main effectue un déplacement d'une position à l'autre (condition 2), ce schéma est maintenu : la main débute d'abord sa transition vers la position cible (en moyenne 205 ms avant le début acoustique de la consonne, M1A1) puis commence à former la configuration de la clé (en moyenne 171 ms avant le début de la consonne, D1A1). La clé est entièrement formée en synchronie avec le début acoustique de la consonne (A1D2, -3 ms) et la main atteint la position cible en début de consonne (A1M2, 33 ms). Ce patron de coordination est globalement conservé quand on examine plus précisément le comportement séquence par séquence. Le positionnement temporel des différents événements est illustré sur la Figure 27. Nous pouvons remarquer pour la condition 1, que le geste des doigts est toujours initié avant le début acoustique de la syllabe (les triangles représentant D1 sont tous en-dessous des croix représentant A1). La configuration de main est entièrement formée juste après le début acoustique de la consonne, dans les 100 premières millisecondes (à quelques exceptions près, on retrouve la

majorité des croix au-dessus de A1). Pour la condition 2, il apparaît également que le geste digital est toujours initié en avance par rapport au début acoustique de la consonne. En ce qui concerne la fin de la formation de la clé, celle-ci se termine plutôt aux alentours du début acoustique de la consonne : nous pouvons remarquer en effet dans cette condition, une distribution répartie de manière presque identique de part et d'autre de A1. La clé est donc formée en synchronie avec le début du son. En ce qui concerne les transitions de main, nous pouvons de manière générale constater que la main semble débuter sa transition avant le début de la formation de la clé digitale (les étoiles représentant M1 ont plutôt tendance à être en-dessous des triangles représentant D1). De la même manière, il apparaît que la position cible LPC est atteinte après la fin de la formation de la clé, durant le début de la consonne (de manière générale, les points représentant M2 se retrouvent au-dessus des croix représentant D2). Ainsi la formation de la clé est incluse dans la transition de main.

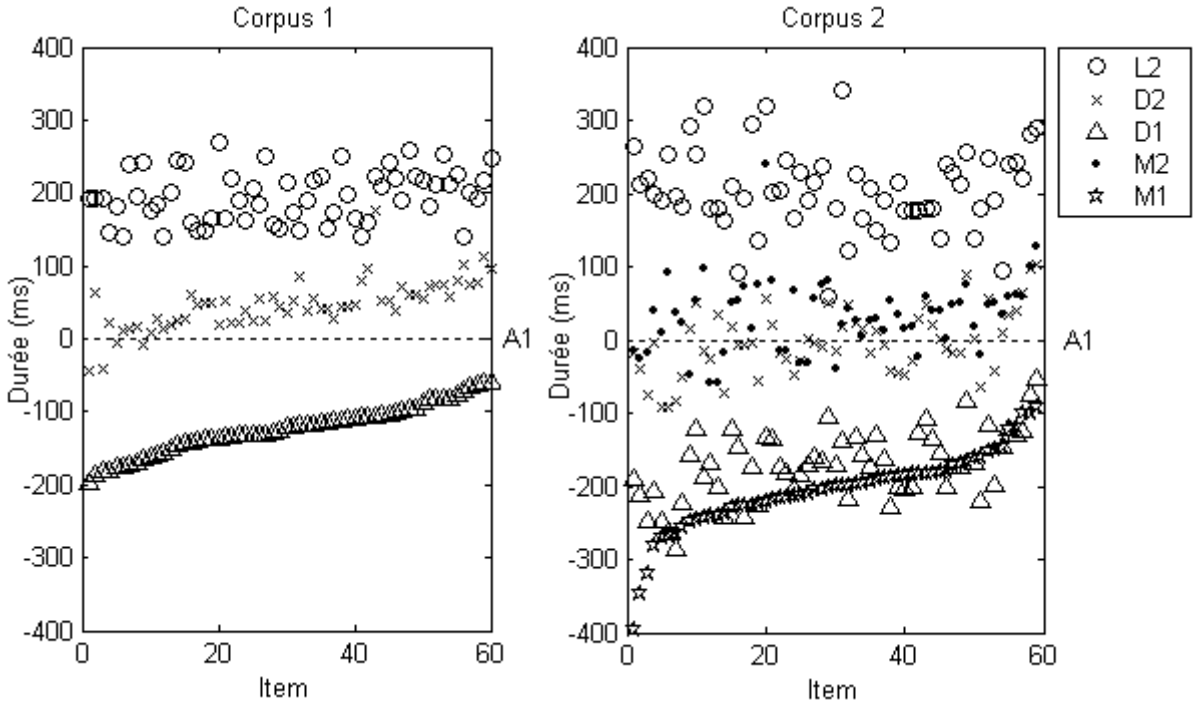


Figure 27. Positionnement temporel des événements suivants par rapport à A1 (repéré par 0), début acoustique de la consonne, pour les deux conditions avec transitions de main (corpus 2, à droite) et sans transition de main (corpus 1, à gauche) : D1 (début de la configuration des doigts), D2 (fin de la formation de la configuration), M1 (début de la transition de main), M2 (atteinte de la position cible) et L2 (atteinte de cible labiale). Les items sont ordonnés dans l'ordre croissant de D1 pour le corpus 1 et de M1 pour le corpus 2.

Nous avions noté que les durées de formation de clés LPC étaient similaires dans les deux conditions mais nous pouvons remarquer cependant une légère différence d'une condition à l'autre en ce qui concerne la relation temporelle entre la formation de la clé et le son. Dans le cas où la main doit se déplacer vers une autre position du visage (corpus 2), la formation de la clé manuelle débute un peu plus tôt (D1A1, 171 ms contre 124 ms avant le début acoustique de la consonne) et se termine

également plus tôt (A1D2, - 3 ms contre 46 ms) que dans le cas où la main reste à une même position cible (corpus 1). Dans les deux cas, la clé consonantique est entièrement formée à un moment très proche du début acoustique de la consonne. La différence entre les deux peut néanmoins s'expliquer par la nécessité d'effectuer en plus la transition manuelle ; tout se passe comme si la main devait arriver en position cible en début de consonne acoustique. C'est ce que nous avions déjà remarqué dans l'expérience 1. Ce but semble confirmé ici, puisque nous observons une atteinte de la position cible manuelle qui se produit au tout début de la consonne (33 ms). La formation de la clé se trouve donc légèrement décalée par rapport au son. L'information sur la consonne donnée par la forme de main est donc disponible en quasi-synchronie avec le début acoustique de cette consonne.

En ce qui concerne l'information sur la voyelle, pour le corpus 1, elle est donnée en même temps que l'information sur la consonne puisque la main reste à la même position cible, soit bien avant la cible de la voyelle aux lèvres (D2L2, 149 ms) (pour un rappel des règles de codage, voir section I.4). Nous retrouvons cette anticipation séquence par séquence : nous pouvons voir sur la Figure 27 que les croix représentant l'événement D2 sont toutes au-dessous des cercles qui représentent l'événement L2. Pour le corpus 2, cette information vocalique est donnée par la position manuelle, qui est atteinte en début de consonne acoustique, soit également bien avant la cible vocalique labiale (M2L2, 172 ms). De la même manière, nous retrouvons une anticipation de la position LPC sur la cible labiale en examinant le positionnement dans le temps de ces événements séquence par séquence : tous les points représentant M2 sont placés au-dessous des cercles représentant L2. Dans les deux cas, **l'information vocalique donnée par la main anticipe sur la cible labiale vocalique**. Cette anticipation manuelle est le résultat majeur que nous avions mis en évidence dans l'expérience précédente pour des séquences avec clés manuelles fixes et que nous retrouvons ici pour des séquences impliquant des changements de configurations de main.

Enfin, en ce qui concerne la syllabe suivante S_3, dans les deux conditions, la formation de configuration correspondante débute de 34 ms à 53 ms en moyenne après la réalisation de la cible vocalique labiale de la syllabe S_2 (L2D3). Dans la condition 2, la transition manuelle codant la syllabe S_3, débute en moyenne 43 ms avant cette cible articulatoire (M3L2). Cette coordination manuelle et labiale correspond à celle que nous avions mise en évidence dans l'expérience précédente pour des transitions de main avec configurations fixes : le geste de la main vers la position suivante est initié autour de la cible labiale, durant la voyelle acoustique. Ceci reste donc valide pour des transitions avec changement de configurations de main.

V.3.5. En résumé

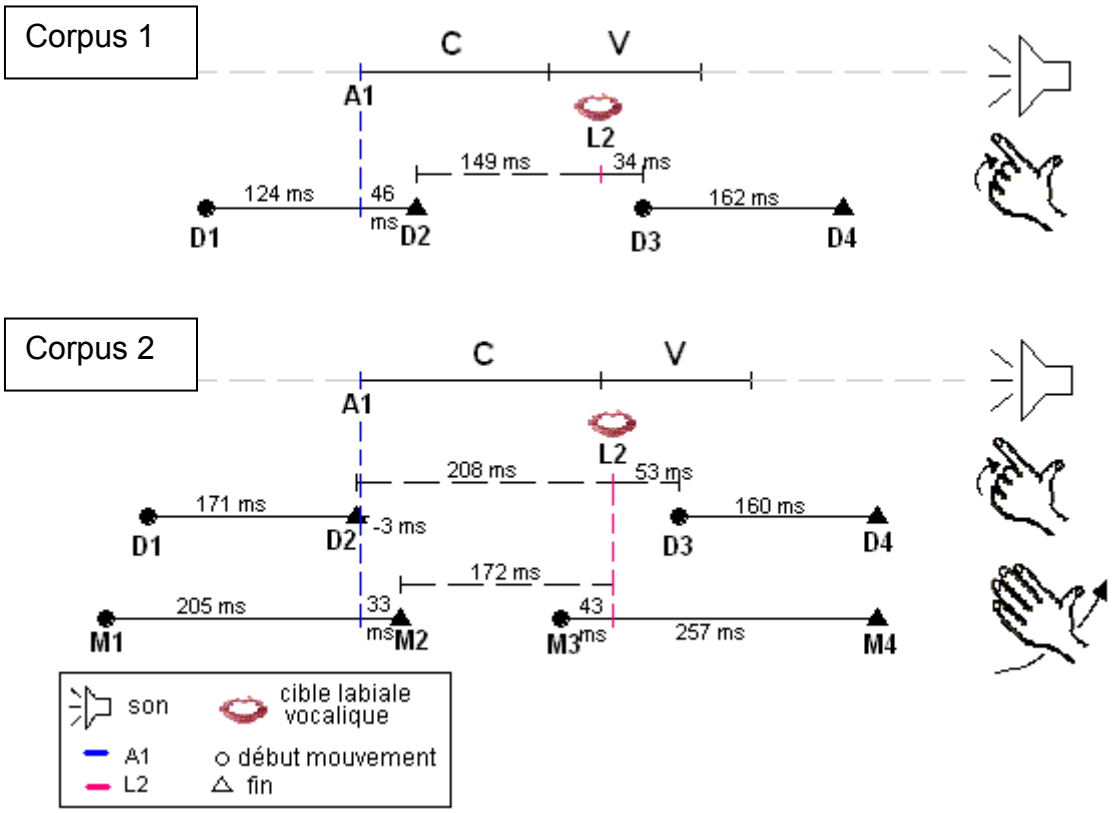


Figure 28. Schéma temporel de coordination entre les doigts, les lèvres, le son et éventuellement la main résumant les résultats de l'expérience 2 dans les deux conditions pour la codeuse GB. En haut, schéma résumé de l'étude 1 des changements de configuration consonantique sans transition manuelle. En bas, schéma résumé de l'étude 2 des changements de configurations avec transitions manuelles.

Nous pouvons donc proposer le patron général suivant à partir du corpus 2 (schématisé sur la Figure 28) : pour une syllabe CV, nécessitant à la fois un déplacement de main et un changement de configuration, la main débute sa transition en avance de 205 ms en moyenne par rapport au début acoustique de la consonne et atteint sa position en début de consonne, soit bien avant la cible labiale vocalique. La main reste en position 130 ms et repart ensuite vers la position suivante un peu avant la réalisation de la cible de la voyelle aux lèvres. La configuration de main, quant à elle, est formée durant la transition d'une position à l'autre du visage. Elle est temporellement incluse dans la transition manuelle, de façon à ce que la configuration soit complètement formée au début du son. En ce qui concerne la transmission des deux informations, pour une syllabe CV, il apparaît clairement que l'information sur la consonne et celle sur la voyelle sont données quasiment en même temps. L'information consonantique est superposée à l'information vocalique, et recouvre de plus une grande partie de la transition (71% environ, cette indication est donnée par le calcul du rapport D1D2/M1M2). Il apparaît donc, comme pour la parole, que les deux informations ne sont pas séquentielles mais

coarticulées. Contrairement à ce qu'avaient proposé Duchnowski et al. (2000) pour leur système de synthèse, nous n'avons pas spécifiquement trouvé un changement de configuration qui se produirait en milieu de transition.

V.4. Conclusion sur les deux expériences

V.4.1. Comparaison normalisée

Que pouvons-nous conclure sur l'ensemble des expériences 1 et 2 ? A première vue, ce qui apparaît est bien le fait que nous retrouvions un patron de coordination temporelle très similaire d'une expérience à l'autre. De manière plus précise, afin de voir si les changements de formes de main au cours du code modifiaient le schéma de coordination, nous avons comparé la condition 1 de l'expérience 1 (codage de syllabes CV avec configuration fixe et transitions manuelles) et la condition 2 de l'expérience 2 (codage de syllabes CV avec changement de configurations et transitions manuelles). Pour procéder à une comparaison objective, nous avons normalisé les différentes durées : nous avons exprimé les différents intervalles temporels dans le domaine de la syllabe acoustique CV, afin d'éliminer l'effet du rythme syllabique qui est différent dans les deux études (2,5 Hz et 3,2 Hz). Pour chaque séquence, les différents intervalles temporels sont exprimés en proportions de temps relatives à la durée de la syllabe CV de la séquence (sans unité, noté $\%_{rel}$). Nous aurons ainsi des valeurs comparables en fonction du cycle syllabique.

Les valeurs moyennes et les écarts-types obtenus sont donnés dans le Tableau 2. Nous avons également représenté graphiquement chaque distribution par des boîtes à moustaches (*Box & Whiskers Plot* ou diagrammes de Tuckey) sur la Figure 29. Ces diagrammes permettent de visualiser certains paramètres descriptifs tels que les quartiles, la médiane et la moyenne de chaque distribution ainsi que leurs positions respectives. Ils décrivent ainsi de façon synthétique les mesures de tendance centrale, l'étendue, la dispersion et la symétrie des données. Ils permettent également d'identifier les valeurs extrêmes (hors normes) de chaque distribution. En outre, nous avons tracé côte à côte les boîtes à moustaches des intervalles temporels des deux expériences (en utilisant la même échelle) : pour chaque intervalle, la comparaison des deux distributions est de cette manière très aisée. Sur chaque graphique de la Figure 29, nous avons ainsi les deux boîtes à moustaches définies pour chaque intervalle de durée des deux expériences, avec en ordonnée, la durée correspondante des intervalles (en millisecondes ou en pourcentages relatifs). La boîte centrale est construite à partir des 1er (25ème percentile) et 3ème quartiles (75ème percentile) (Q1 et Q3) qui définissent ses bords (la longueur de la boîte correspond à l'écart interquartile Q3-Q1). Le trait horizontal à l'intérieur de la boîte représente la médiane et la croix 'x' représente la moyenne. Ainsi la partie centrale représentée par la

boîte contient 50% des données. Les moustaches sont les lignes verticales qui s'étendent au-dessous et au-dessus de la boîte dont les traits horizontaux sont les extrémités. Elles correspondent aux valeurs minimum et maximum des données si la distribution ne contient aucune valeur hors norme. Les valeurs hors normes sont des valeurs supérieures à 1,5 fois l'écart interquartile à partir de Q3 et inférieures à 1,5 fois l'écart interquartile à partir de Q1 (le coefficient 1,5 est une valeur pragmatique) ; elles sont représentées par le signe '+'.

	Expérience 1		Expérience 2	
	Moyenne	Ecart-type	Moyenne	Ecart-type
M1M2	72	19	76	17
M2M3	51	24	40	21
M1A1	62	25	65	17
A1M2	10	19	10	15
M2L2	64	20	55	21
M3L2	13	16	15	25
M3M4	74	19	92	17
D1D2			54	12
D1A1			55	17
A1D2			-1	15
D2L2			66	19
L2D3			16	16
D3D4			57	8

Tableau 2. Moyennes et écart-types ($\%_{rel}$) des intervalles temporels exprimés en fonction du cycle syllabique CV pour les expériences 1 et 2.

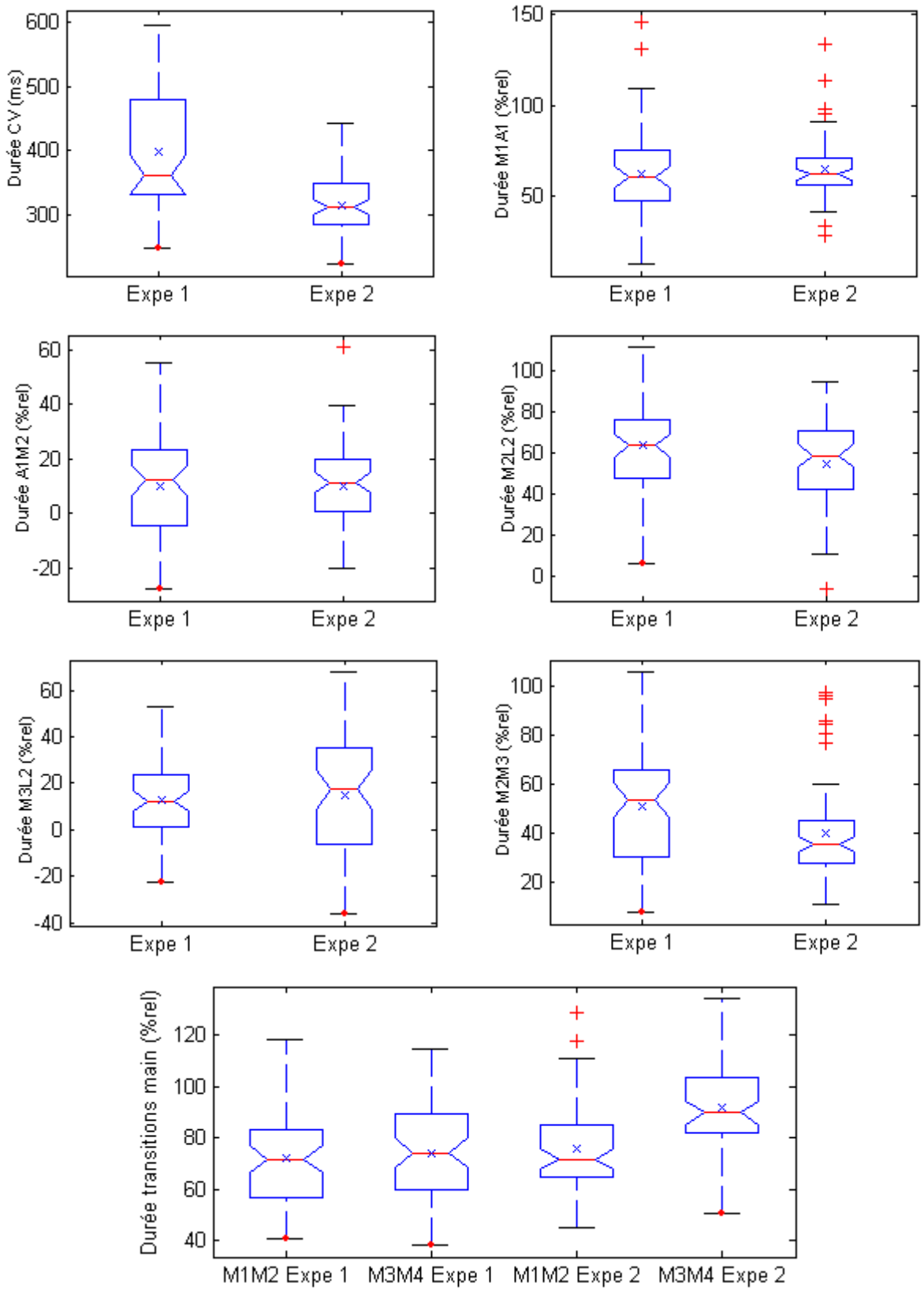


Figure 29. Boîtes à moustaches représentant la distribution de chaque intervalle temporel dans le cycle syllabique pour la codeuse GB dans les expériences 1 et 2. Chaque boîte indique la médiane, les 1er et 3ème quartiles, les valeurs frontières, les valeur hors normes ('+') et la moyenne ('x').

Comme nous l'avons souligné plus tôt, la codeuse GB a produit des syllabes de durées différentes dans les deux expériences. Cette différence est facilement visualisable sur la Figure 29 (graphique en haut à gauche) : la comparaison des deux boîtes à moustaches montre que la locutrice a produit des syllabes de durées très dispersées dans l'expérience 1, avec une tendance à être plus longues que celles de l'expérience 2 (cette différence est confirmée statistiquement[11] par un test t ; $|t_c|$= 6,09 ddl= 117, p< .01).

Les données obtenues dans le domaine syllabique confirment la similarité des résultats des deux expériences. Nous pouvons constater en effet certaines caractéristiques de codage relativement constantes du sujet GB. Tout d'abord, cette codeuse maintient un rythme de transitions manuelles relativement stable par rapport au cycle syllabique (dans 3 cas sur 4 autour de 72-76%$_{rel}$). Comme nous pouvons le voir en bas de la Figure 29, les boîtes à moustaches représentant les distributions des transitions manuelles M1M2 et M3M4 sont globalement similaires d'une expérience à l'autre. Seule la durée moyenne des transitions M3M4 de l'expérience 2 diffère statistiquement des autres : cette différence est clairement visible sur la figure, de plus, cette observation est confirmée par une ANOVA à un facteur sur les différentes transitions (groupes significativement différents, F(3, 237)= 15,2 p< .01 ; les tests post-hoc confirment que c'est bien le dernier type de transitions qui diffère des autres, p< .01). Malgré cette différence, nous pouvons remarquer cependant un écart interquartile (longueur des boîtes) très proche d'une distribution à l'autre. Ces durées d'intervalles sont distribuées de manière normale[12] pour l'expérience 1 et pour les transitions M3M4 de l'expérience 2 (nous pouvons noter une certaine symétrie des données autour de la médiane, ainsi qu'une correspondance médiane-moyenne), alors qu'elles sont légèrement asymétriques pour les transitions M1M2 de l'expérience 2 (il semble y avoir une plus grande dispersion des données pour les valeurs supérieures à la médiane). Dans tous les cas, la répartition des durées de transitions par rapport au cycle syllabique semble avoir une même étendue, avec une concentration de la moitié des données grossièrement entre 60 et 80%$_{rel}$ (en excluant le dernier groupe).

En ce qui concerne la coordination temporelle des différents événements, nous obtenons un patron très semblable d'une expérience à l'autre. En effet dans les deux expériences, la main débute sa transition vers la position correspondant à la voyelle bien avant le début acoustique de la syllabe CV. En outre, cette avance manuelle est en moyenne très proche pour les deux expériences (M1A1 de 62 à 65%$_{rel}$ en moyenne) : les moyennes ne sont en effet pas significativement différentes (test t, $|t_c|$= 0,79

[11] Quand les conditions d'application du test t de Student ne sont pas entièrement respectées, nous utilisons également le test non paramétrique de la somme des rangs de Wilcoxon (*Wilcoxon rank sum test) :* nous avons également obtenu une différence significative (p< .01).
[12] La normalité des distributions est vérifiée par le test de Lilliefors.

ddl= 117 au risque α= 5%). En regardant l'ensemble de la distribution sur la Figure 29, nous pouvons remarquer d'une part que toutes les valeurs d'intervalles sont positives ; ce qui signifie clairement que la main débute sa transition toujours avant le son, même dans le cas où des changements de clés se superposent aux transitions de main. D'autre part, il apparaît que, bien que les données soient beaucoup plus dispersées dans l'expérience 1, il semble y avoir dans les deux cas, une concentration de la majorité des données centrales entre 50%$_{rel}$ et 70%$_{rel}$. Ce résultat indique que **la main anticipe sur le début du son, avec une avance équivalent à plus d'une demi-syllabe**.

La position vocalique manuelle est ensuite atteinte en début de consonne acoustique dans les deux expériences. Les deux boîtes à moustaches représentant les distributions de A1M2 sur la Figure 29 sont très similaires : les deux jeux de données sont distribués normalement avec une variabilité plus importante pour l'expérience 1. Nous pouvons remarquer pour les deux distributions que le 1er quartile est situé autour de 0 (à - 4,6%$_{rel}$ pour l'expérience 1 et à 0,3%$_{rel}$ pour l'expérience 2) : la majorité des données se retrouve donc dans les valeurs positives, ce qui signifie clairement que la position cible LPC est en majorité atteinte après le début acoustique de la consonne. Les données positionnées dans la partie centrale (soit 50% des données totales) se retrouvent concentrées globalement entre 0 et 20%$_{rel}$ de la durée de la syllabe CV, soit clairement au début de la consonne acoustique. En moyenne, l'intervalle A1M2 représente 10%$_{rel}$ de la durée de la syllabe CV dans les deux expériences : les deux moyennes ne sont significativement pas différentes (test t, $|t_c|$= 0,05 ddl= 117 au risque α= 5%). Dans le domaine de la syllabe, le calcul des durées relatives des consonnes dans les deux expériences donne des valeurs moyennes de 59%$_{rel}$ (s= 11%$_{rel}$) pour l'expérience 1 et de 61%$_{rel}$ (s= 9%$_{rel}$) pour l'expérience 2. Ainsi il apparaît clairement que la position cible est atteinte dans la toute première partie de la consonne (de 16 à 17% [= A1M2/duréeConsonne] en moyenne sur les deux expériences). Il semblerait donc que ce comportement anticipatoire de la main soit motivé par un but, un *rendez-vous temporel* entre la position de main et le début acoustique de la consonne.

La position de main codant la voyelle est donc atteinte en quasi-synchronie avec le début acoustique de la consonne ; elle est de ce fait largement en avance sur la cible articulatoire de la voyelle, en moyenne de 55%$_{rel}$ à 64%$_{rel}$ (M2L2), moyennes significativement non différentes (test t, $|t_c|$= 2,3 ddl= 117 au risque α= 1%). L'examen des boîtes à moustaches sur la Figure 29 nous indique que cette anticipation de la cible manuelle sur la cible labiale est toujours positive, à l'exception d'une valeur hors norme de - 6%$_{rel}$ dans l'expérience 2 (identifiée sur la figure par un signe '+' en-dessous de la moustache inférieure de la boîte). Les deux distributions sont normales et concentrent la moitié des données en position centrale largement entre 40%$_{rel}$ et 80%$_{rel}$. Nous retrouvons donc au niveau de la cible l'anticipation d'une bonne demi-syllabe notée précédemment pour l'initiation du geste. Les durées

M2L2 sont toutefois plus dispersées que les durées M1A1 et démontrent de ce fait une plus grande variabilité, sans doute liée à la cible labiale.

La main est maintenue en position cible de 40%$_{rel}$ à 51%$_{rel}$ en moyenne (M2M3). Les distributions des durées de tenue de cible paraissent assez dissemblables sur la Figure 29 : nous pouvons remarquer en effet des durées distribuées normalement pour l'expérience 1, avec une très grande variabilité (les moustaches de la boîte s'étendent en effet de 8%$_{rel}$ à 105%$_{rel}$), alors que pour l'expérience 2, les données semblent moins dispersées (les moustaches de la boîte s'étendent de 11%$_{rel}$ à 60%$_{rel}$) mais présentent plusieurs valeurs hors normes. Ces différences sont confirmées statistiquement (test t, $|t_c|$= 2,55 ddl= 117 p< .05).

La main repart ensuite vers la position suivante un peu avant la réalisation de la voyelle aux lèvres en moyenne de 13%$_{rel}$ à 15%$_{rel}$ (M3L2) ; ces moyennes ne sont significativement pas différentes ($|t_c|$= 0,47 ddl= 117 au risque α= 5%). Nous pouvons voir d'après les boîtes à moustaches sur la Figure 29 que les deux distributions sont normales, les durées de l'expérience 2 étant plus dispersées. Le 1er quartile des deux distributions est positionné autour de 0 (1,3%$_{rel}$ pour l'expérience 1 et -6%$_{rel}$ pour l'expérience 2), ce qui indique que la majorité des deux distributions est répartie dans les valeurs positives : ainsi la main repart vers la position suivante avant la réalisation de la cible de la voyelle aux lèvres dans la majorité des cas.

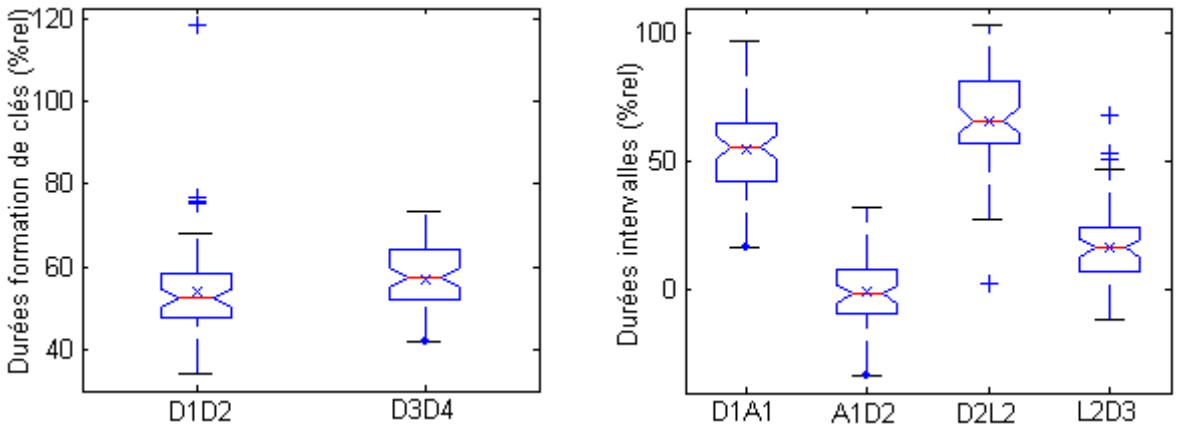


Figure 30. Boîtes à moustaches indiquant la distribution de chaque intervalle temporel pour l'étude de la configuration digitale relativement au cycle syllabique pour la codeuse GB dans l'expérience 2 (corpus 2).

En ce qui concerne la configuration consonantique pour l'expérience 2, nous pouvons noter une très grande proximité des durées de formation de clé (D1D2 et D3D4) : la visualisation des boîtes à moustaches correspondantes sur la Figure 30 nous permet de juger graphiquement de cette similarité. Les données sont globalement assez peu dispersées : l'écart interquartile (longueur de la boîte comprenant 50% des données) est de 11%$_{rel}$ pour D1D2 et de 12%$_{rel}$ pour D3D4. Les durées

moyennes de 54%$_{rel}$ et 57%$_{rel}$ obtenues pour les deux jeux de données ne sont pas significativement différentes (test t, $|t_c|$= 1,6 ddl= 116 au risque α= 5%). Nous pouvons remarquer que ces formations de configuration de main (D1D2, m= 54%$_{rel}$) durent en moyenne moins longtemps que les transitions manuelles (M1M2= 76%$_{rel}$) ; ceci reste valable séquence par séquence comme nous pouvons le constater sur le graphique en bâtons superposés représentant les durées de transitions de main et les durées de formations de clés sur la Figure 31 (les durées de transitions manuelles sont plus longues que les durées de formations de clés, à l'exception de cinq séquences).

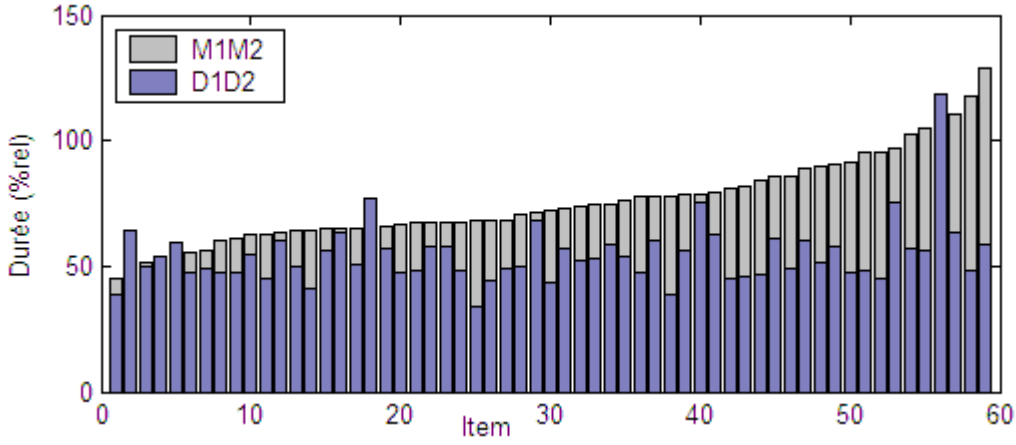


Figure 31. Graphique en bâtons superposés représentant les durées (en %$_{rel}$) des formations de configurations digitales (foncé) superposées sur les durées totales de transitions de main (clair) pour toutes les séquences. Les données sont triées dans l'ordre croissant de la durée M1M2.

Comme nous l'avons remarqué plus tôt, les formations de clés sont en fait superposées sur les transitions de main. De cette manière, il apparaît que le début de la formation de la clé se produit toujours avant le début acoustique de la consonne : la boîte à moustaches sur la Figure 30 montre que les valeurs de D1A1 sont toutes positives. La moitié des données, concentrées dans la partie centrale de la distribution, se retrouve entre 41,5%$_{rel}$ et 65%$_{rel}$, avec une moyenne s'élevant à 55%$_{rel}$: ce résultat indique que le début du geste de formation de la configuration de main est en avance d'une demi-syllabe par rapport au son. En ce qui concerne la fin de la formation de la clé, il apparaît qu'elle est synchrone avec l'acoustique (A1D2 s'élevant à -1%$_{rel}$) : cette moyenne n'est en effet pas statistiquement différente de 0 (test t, $|t_c|$= 0,05 ddl= 58 au risque α= 5%). Sa distribution est normale et répartit les données autour de 0. Ainsi cette codeuse forme les clés consonantiques alors qu'elle est en train de déplacer sa main vers une position cible avant la production de son ; les deux informations transmises par la main se chevauchent (plus précisément c'est l'information consonantique qui se superpose à l'information vocalique) et sont complètement disponibles en synchronie avec le début acoustique de la consonne. La clé est donc formée entièrement bien avant la réalisation de la voyelle aux lèvres, avec une avance moyenne de 66%$_{rel}$, soit bien plus qu'une demi-syllabe. Enfin, les résultats

montrent que la main débute la formation de clé suivante en moyenne 16%$_{rel}$ après la réalisation de la cible de la voyelle aux lèvres. La Figure 30 indique que le 1er quartile se trouve au-dessus de 0 (6,7%$_{rel}$), ce qui signifie que plus de 75% des données sont positives : la clé est donc formée dans la majorité des cas après la cible labiale.

V.4.2. Un schéma de coordination temporelle main-lèvres-son

Ainsi dans la production du code LPC, qu'il y ait superposition ou non de clé consonantique, la codeuse GB présente un patron de coordination manuelle et orofaciale très stable d'un enregistrement à l'autre. Ce patron peut se résumer de la façon suivante (voir également la Figure 32) : dans la production d'une syllabe CV codée, la main débute sa transition avec une avance moyenne d'une bonne demi-syllabe sur le début acoustique de la consonne et atteint sa position cible sur le visage en début de consonne, donc bien avant la cible vocalique aux lèvres. La main reste en position durant toute la consonne, puis redémarre son geste vers la position suivante, dans la voyelle. En ce qui concerne la clé consonantique, elle est formée durant la transition de la main d'une position à une autre du visage (l'information consonantique est donc superposée à l'information vocalique) et est entièrement réalisée en début de consonne. Ce qui ressort fortement de cette étude est la désynchronisation des deux composantes vocaliques : la position de main en LPC anticipe clairement la cible labiale en parole. Les deux composantes consonantiques, la consonne acoustique et la configuration digitale, sont quant à elles synchrones. Le phasage LPC-parole semble bien se faire entre la position de main et la consonne acoustique.

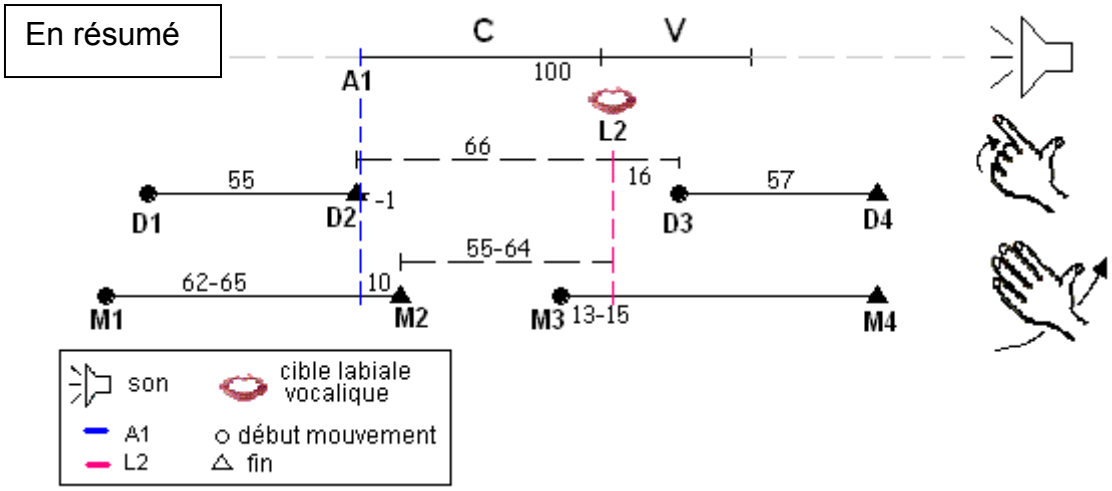


Figure 32. Schéma de coordination temporelle de la main, des doigts et des lèvres en relation avec le son durant la production d'une syllabe CV codée en LPC chez le sujet GB dans le domaine syllabique. Les intervalles indiqués sur la figure sont extraits des expériences 1 et 2 (voir également le Tableau 2) : les durées sont exprimées en pourcentages relatifs à la durée de la syllabe acoustique CV (unité : %$_{rel}$).

CHAPITRE VI.
Expérience 3 : généralisation à d'autres locuteurs-codeurs sur des séquences syllabiques CV

Attina V., Cathiard M.-A. & Beautemps D. (accepté).
Temporal measures of hand and speech coordination
during French Cued Speech production.
Lecture Notes in Artificial Intelligence, LNAI/LNCS, Springer Verlag.

Ce chapitre concerne la validation du patron temporel de coordination, obtenu chez la locutrice-codeuse GB, chez d'autres locuteurs-codeurs. Nous avons enregistré pour cela trois autres sujets en utilisant un corpus très large impliquant une grande variété de clés digitales et de transitions manuelles. Cette étude pourra nous apporter des informations supplémentaires sur le patron temporel général de coordination entre la main, les lèvres et le son durant la production de séquences syllabiques CV codées par des codeuses professionnelles de Langue française Parlée Complétée. Plus précisément, cette étude va nous permettre de répondre aux questions suivantes. Le schéma temporel de coordination décrit précédemment est-il spécifique à la codeuse GB ou bien est-ce une caractéristique plus générale de production de code LPC ? En clair, va-t-on retrouver le même patron temporel de coordination chez les trois autres locutrices-codeuses ? Quelle est l'importance de la variabilité intra- et inter-codeur ? Cette étude va donc nous permettre d'étudier la coordination temporelle en LPC chez d'autres codeuses et de comparer les différents patrons de coordination obtenus afin d'avoir des règles générales de production pour la LPC.

VI.1. Méthode

VI.1.1. Sujets

Les trois sujets enregistrés dans cette étude sont trois locutrices françaises titulaires du diplôme professionnel de codeur en LPC. AM est une femme âgée de 43 ans au moment de l'enregistrement. Elle a obtenu son diplôme de codeuse en 1991 et utilise régulièrement le code LPC depuis (de manière quotidienne avec son enfant sourd). Elle code au lycée pour des jeunes adolescents sourds en moyenne 17 heures par semaine. SC est une femme âgée de 30 ans, qui a obtenu son diplôme en 2001. Elle pratique le code de manière professionnelle uniquement, depuis quatre ans et code en moyenne 24 heures par semaine au lycée. Enfin, RV est une femme âgée de 45 ans, qui a eu son diplôme en 1997. Elle pratique le code depuis 14 ans (en raison de la surdité de son enfant) et a une pratique professionnelle de 17 heures en moyenne par semaine au collège.

VI.1.2. Corpus d'étude

Nous avons constitué le corpus de façon à explorer un grand nombre de transitions manuelles (les cinq positions sont utilisées) et de configurations consonantiques (les huit clés manuelles sont utilisées). Les séquences syllabiques utilisées sont de la forme [$C_1V_1.C_1V_1.C_2V_2.C_3V_1$] avec les consonnes [m] et [b] pour C_1 et les paires {[p], [j]}, {[s], [l]}, {[v], [g]} et {[b], [m]} pour C_2 et C_3 (avec V_1 différent de V_2 et C_2 différent de C_3) et les voyelles [a, i, u, ø, e] pour V_1 et V_2 (voir la liste complète des séquences en Annexe 1). En combinant les voyelles et les consonnes, nous obtenons au total 160 séquences de 4 syllabes (par exemple, [mamasila] avec C_1= [m], C_2= [s], C_3= [l] et V_1= [a], V_2= [i]). La syllabe étudiée

est la syllabe S_2 de chaque séquence $S_0S_1S_2S_3$. Au terme des enregistrements, nous avons sélectionné exactement 161 séquences pour la codeuse AM (qui a produit une séquence supplémentaire que nous avons gardée puisqu'elle était correcte), 159 séquences pour la codeuse SC (une erreur de codage s'est produite) et 155 séquences pour la codeuse RV (qui a produit cinq erreurs).

VI.1.3. Acquisition des données et traitement

Afin de simplifier les conditions expérimentales, le gant de données n'a pas été utilisé pour capturer les mouvements digitaux. Ainsi, cette étude se focalise sur les transitions manuelles de position à position (comme dans l'expérience 1 pour la codeuse GB), en supposant que pour ces trois codeuses, les changements de clés digitales superposées aux transitions de main ne modifient pas le patron de coordination, comme nous l'avons montré précédemment pour la codeuse GB. Le dispositif expérimental est le même pour les trois sujets et est également similaire à celui de l'expérience 1 pour la codeuse GB (voir section V.2.2.1). Chaque locutrice-codeuse a employé la main utilisée pour coder habituellement, c'est-à-dire la main gauche pour AM et SC et la main droite pour RV.

Nous avons utilisé la même méthode de traitement des données (numérisation des images avec le logiciel Capture, extraction des mesures labiométriques avec le logiciel Tacle et utilisation des différents programmes Matlab que nous avons développés pour le suivi de la pastille colorée du dos de la main ; pour un rappel, voir section V.2.2.2). Au terme de chaque enregistrement et traitements, nous obtenons ainsi pour chaque locutrice un ensemble de quatre signaux synchrones au cours du temps : (1) le décours temporel de l'aire intérolabiale (en cm^2 avec une information toutes les 20 ms), (2) la position en x de la pastille sur le dos de la main (en cm avec une information toutes les 20 ms), (3) la position en y de la pastille sur le dos de la main (en cm avec une information toutes les 20 ms) et le signal acoustique (échantillonné à une fréquence de 44100 Hz).

Pour chaque codeuse, les coordonnées de la pastille sur le dos de la main sont calculées par rapport à un point de référence sur les lunettes que portent les sujets (voir Figure 33 pour SC et voir Annexe 2). Par exemple, pour la locutrice SC, codant la séquence [ma.ma.be.ma] illustrée en Figure 34, une diminution en x correspond à un rapprochement horizontal de la main vers le point de référence et une augmentation en y correspond à un éloignement vertical de la main par rapport au point de référence sur la lunette : c'est ce qui se produit lorsque la main passe de la position « côté » qui code la voyelle [a] à la position « cou » pour la voyelle [e].

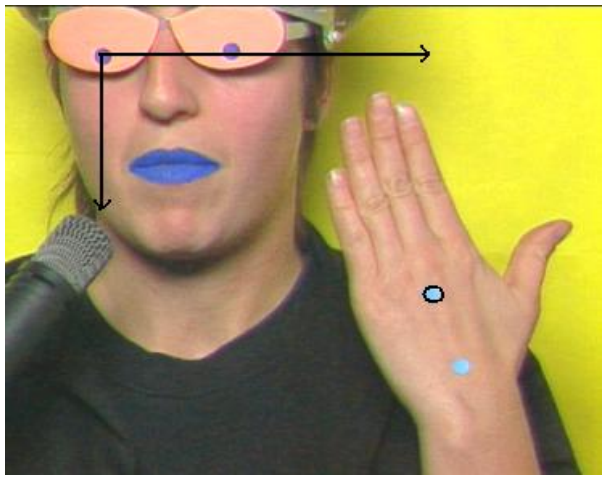


Figure 33. Photo de la locutrice-codeuse SC avec superposition du repère utilisé pour le calcul de la position de la pastille colorée sur le dos de la main.

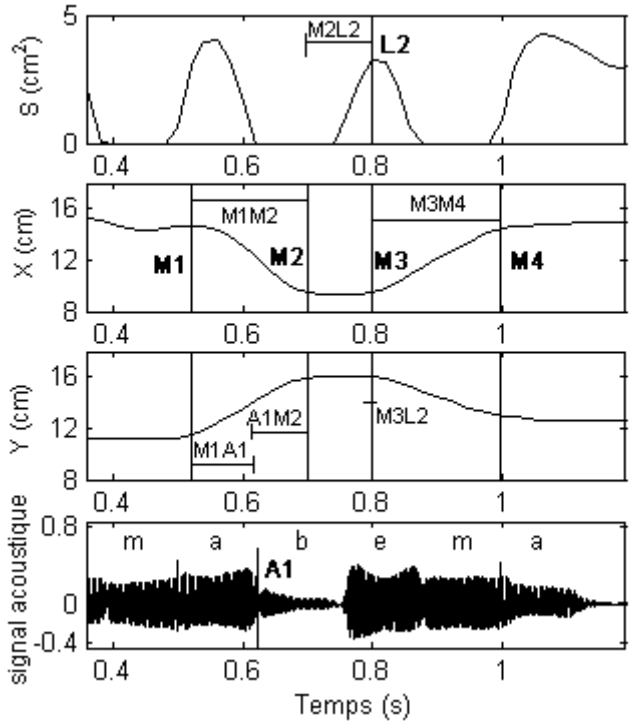


Figure 34. Tracé des différents signaux pour la portion [ma.be.ma] extraite de la séquence [ma.ma.be.ma] de la locutrice-codeuse SC. De haut en bas : (1) décours temporel de l'aire aux lèvres S (cm^2) ; (2) position de la coordonnée x de la pastille sur le dos de la main au cours du temps (cm) ; (3) position de la coordonnée y de la pastille au cours du temps (cm) ; (4) signal acoustique correspondant. Sur chacun des signaux sont superposés les événements temporels (en gras) et les intervalles étudiés.

Les événements[13] temporels suivants ont été repérés sur les différents signaux : pour chaque séquence « $S_0S_1S_2S_3$ », A1 est le début acoustique de la consonne de la syllabe S_2, L2 est la cible de la voyelle aux lèvres de la syllabe S_2, M1 est le début du mouvement manuel pour coder S_2 et M2 l'atteinte de la position cible codant la voyelle de la syllabe S_2, M3 est le début de la transition manuelle suivante codant S_3 et M4 la fin de cette transition.

Les caractéristiques de production étudiées sont (voir les intervalles sur la Figure 34) : (1) la durée de la syllabe S_2 indiquant un rythme moyen de syllabes codées, (2) les intervalles M1M2 et M3M4 désignant les durées de transition manuelle, (3) M2M3 la durée du maintien de la main en position cible codant S_2, (4) M1A1 la durée entre le début du geste manuel et le début acoustique de la consonne de S_2, (5) A1M2 la durée entre le début acoustique de la consonne et l'atteinte de la position cible manuelle, (6) M2L2 la durée entre l'atteinte de la position cible manuelle et la réalisation de la cible vocalique aux lèvres pour S_2 et (7) M3L2 désignant la durée entre la cible vocalique labiale et le début de la transition manuelle vers la position suivante codant S_3.

VI.2. Résultats

Nous avons effectué différentes analyses sur les résultats de cette étude. Nous donnerons tout d'abord les résultats en millisecondes afin d'avoir un schéma temporel complet pour les trois sujets, puis nous procèderons à une normalisation en pourcentages relatifs de la durée de la syllabe S_2 (pour éliminer la variabilité de durée syllabique intra et intersujet) afin de comparer les trois patrons temporels de coordination obtenus et de faire ressortir ainsi les régularités caractéristiques du codage LPC. Enfin l'étude de l'évolution de ces patrons de coordination (c'est-à-dire le positionnement relatif des différents événements temporels) en fonction de la durée de la syllabe CV nous permettra de mieux comprendre les stratégies de codage (les différents phasages LPC-parole) utilisées dans le domaine syllabique.

VI.2.1. Schéma temporel de coordination pour les trois codeuses

VI.2.1.1. Organisation temporelle main-lèvres-son

Les résultats moyens (et écarts-types) en millisecondes pour les trois sujets sont indiqués dans le Tableau 3 et schématisés sur la Figure 35 afin de donner un aperçu général des patrons temporels de coordination obtenus pour chaque sujet.

[13] Les événements cinématiques sont repérés manuellement à l'aide du profil d'accélération des signaux filtrés passe-bas à une fréquence de coupure de 4 Hz. Dans la plupart des séquences, l'événement temporel est étiqueté sur un pic d'accélération (ou de décélération). Quand le pic n'est pas bien marqué, l'événement étiqueté est approximé au point le plus proche temporellement sur le signal de position.

Durées moyennes en ms (écarts-types)	AM	SC	RV
Syllabe CV	252 (41)	253 (45)	258 (56)
Consonne	119 (37)	141 (41)	147 (51)
M1M2	170 (29)	174 (37)	192 (33)
M2M3	146 (75)	156 (62)	164 (60)
M1A1	153 (56)	145 (56)	143 (50)
A1M2	17 (51)	29 (55)	49 (49)
M2L2	155 (54)	143 (50)	123 (66)
M3L2	9 (73)	-13 (57)	-41 (64)
M3M4	183 (34)	175 (33)	197 (37)

Tableau 3. Durées moyennes et écarts-types (en ms) des différents intervalles temporels obtenus pour les trois locutrices-codeuses AM, SC et RV.

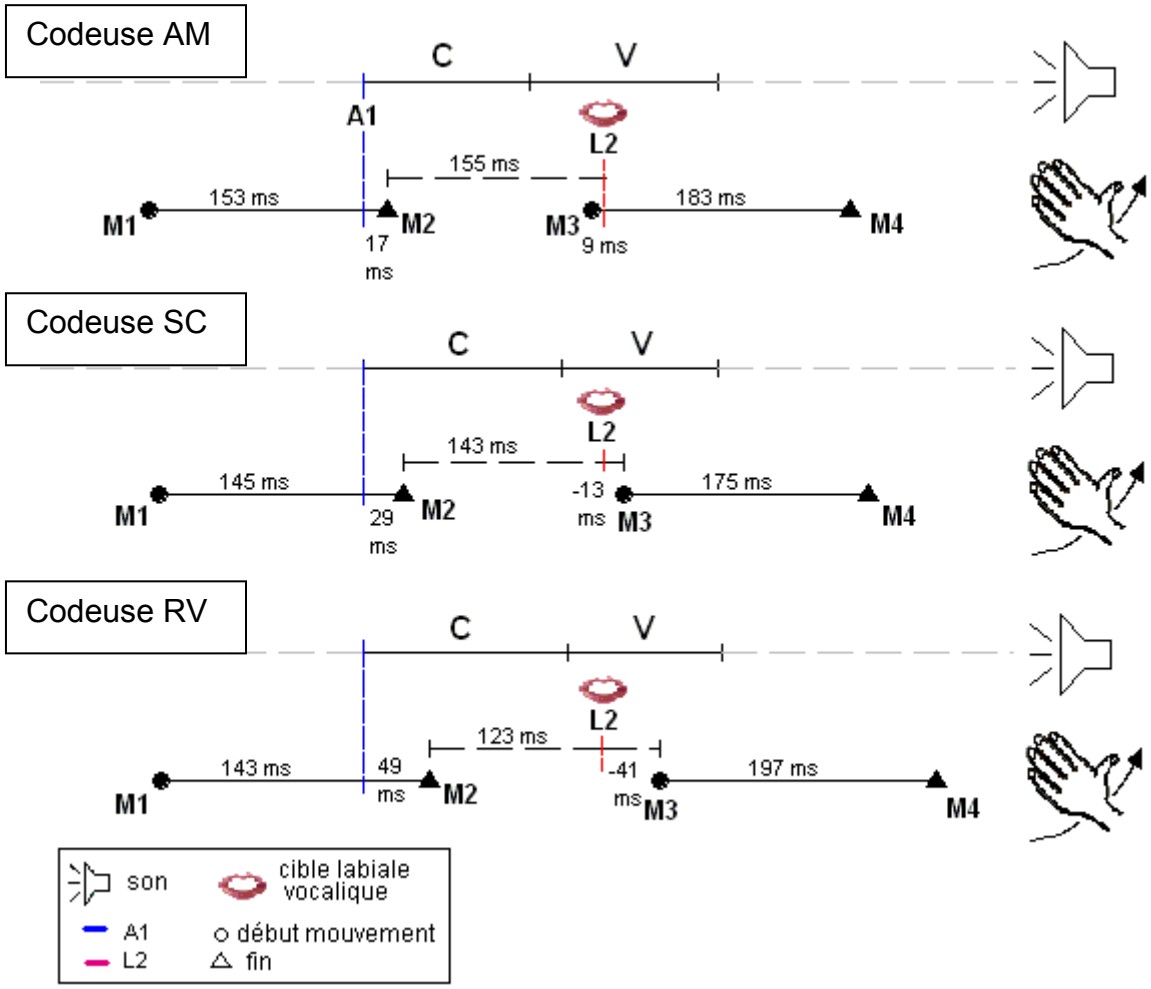


Figure 35. Schémas des patrons temporels de coordination main-lèvres-son pour chacune des codeuses AM, SC et RV.

Nous pouvons remarquer une grande similarité des caractéristiques de production ainsi que des patrons de coordination obtenus entre les trois codeuses. En particulier certaines durées semblent très proches (c'est le cas par exemple des durées de syllabes et des transitions de main, mais également de la relation temporelle entre le début du geste manuel et le début acoustique de la syllabe M1A1) tandis que d'autres, comme l'intervalle M3L2 par exemple, démontrent plus de variabilité (ces intervalles seront comparés statistiquement dans la section VI.2.1.2). Néanmoins, l'organisation temporelle générale entre la main, les lèvres et le son qui se profile pour les trois locutrices-codeuses semble être la suivante (voir sur la Figure 35 les schémas moyens et sur la Figure 36 le détail séquence par séquence).

La main débute son mouvement vers la position cible LPC avant le début acoustique de la consonne (M1A1). C'est toujours le cas pour la codeuse SC (sur la Figure 36, nous pouvons en effet remarquer que tous les triangles qui représentent le début du geste de la main M1 sont placés en-dessous de la ligne en pointillés représentant A1). La codeuse AM a quant à elle produit une seule séquence pour laquelle la main débute la transition après le début acoustique de la consonne et la codeuse RV a produit deux séquences où la main débute son mouvement en synchronie avec le début de la consonne. Ces trois exceptions mises à part, nous pouvons considérer que de manière générale, la main débute son geste de transition avant le début acoustique de la syllabe (les trois distributions peuvent en effet être considérées comme des densités de probabilité avec aux extrêmités les événements les moins probables) ; cette durée d'anticipation est assez variable (elle peut aller de 19 ms à 312 ms) et s'élève en moyenne à 143-153 ms pour les trois sujets.

La main arrive en position cible en moyenne en début de consonne acoustique (A1M2). Pour les trois sujets, cette atteinte de position de main par rapport au début acoustique de la consonne peut varier selon la séquence : la main peut parfois arriver en position cible LPC avant le début de la consonne (ceci est représenté par le fait que des croix représentant M2 sont positionnées en-dessous de A1 sur la Figure 36). Pour la plus grande partie des séquences, la main arrive en position cible LPC après le début acoustique de la consonne. Elle atteint la position de 17 ms à 49 ms en moyenne après le début de la syllabe. Plus précisément, la main atteint la position dans la première partie de la consonne ; en effet, l'intervalle A1M2 représente en moyenne 14% de la durée totale de la consonne acoustique pour le sujet AM, 21% pour le sujet SC et 33% pour le sujet RV (valeur donnée par le calcul [moyenne(A1M2) / moyenne(Consonne) x 100]). Nous retrouvons donc chez ces trois codeuses une certaine synchronie de la position manuelle avec le début acoustique de la consonne.

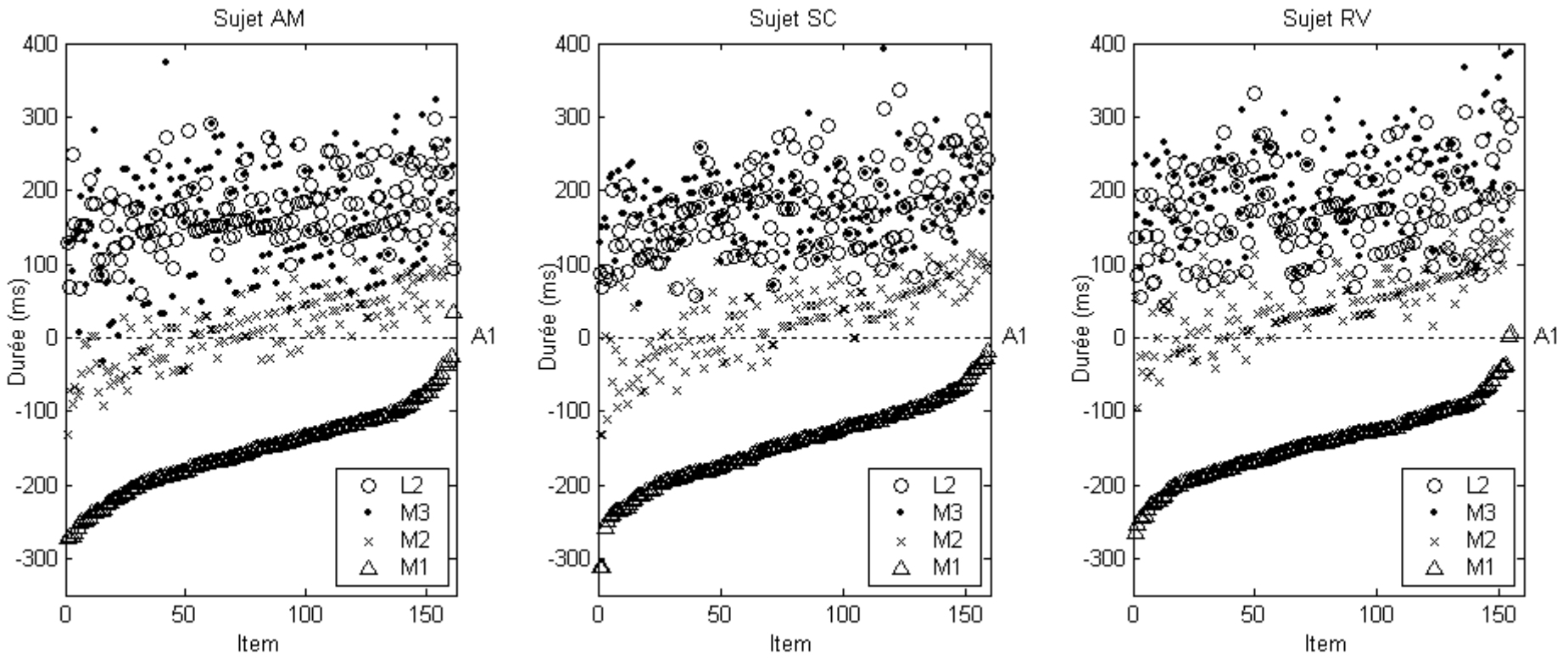


Figure 36. Positionnement temporel (en ms) des événements M1 (début du geste manuel), M2 (fin du geste manuel), L2 (cible labiale vocalique) et M3 (début du geste manuel vers la position suivante) par rapport au début acoustique de la consonne A1, pour les trois codeuses AM, SC et RV. Pour faciliter la lecture des graphiques, les items sont triés dans l'ordre croissant en fonction de M1.

La transition manuelle (M1M2) peut durer en moyenne de 170 ms à 192 ms selon le sujet et **semble relativement peu variable** d'une séquence à l'autre : nous pouvons remarquer en effet sur la Figure 36 que le positionnement des croix qui représentent la fin de la transition manuelle M2 semble *suivre* le positionnement des triangles, représentant M1 (l'espace vertical entre les triangles et les croix semble relativement constant au cours des séquences et définit un certain parrallélisme entre les deux distributions, ce qui signifie que M1 et M2 sont appariés).

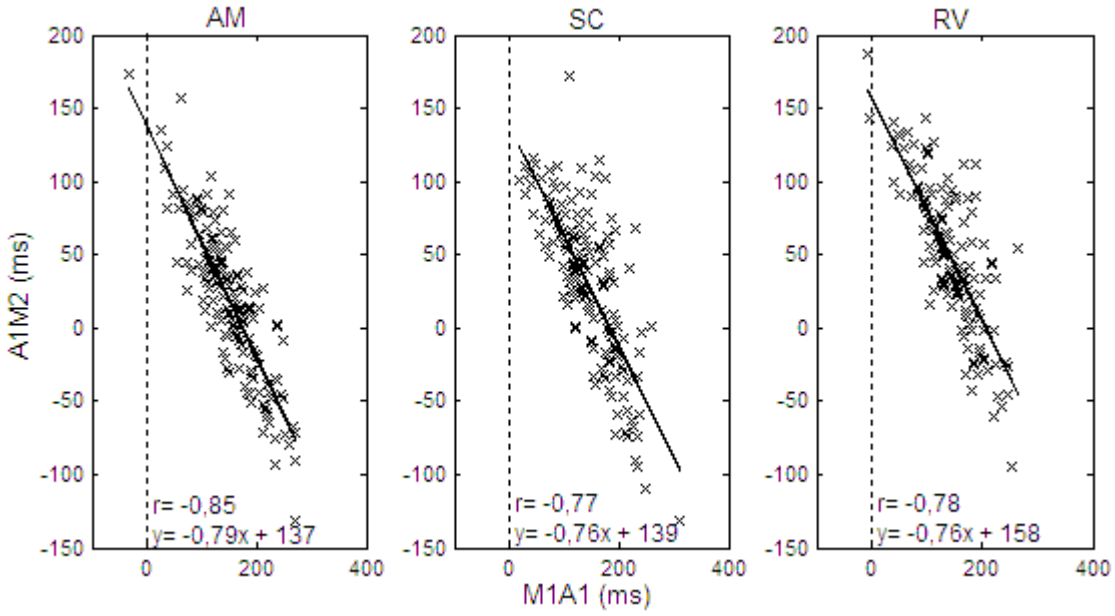


Figure 37. Evolution de l'intervalle A1M2 en fonction de l'intervalle M1A1 (ms) pour les trois codeuses AM, SC et RV. Sur chacun des graphes sont indiqués les coefficients de corrélation (tous significatifs à p< .01) et les équations de droites ajustées linéairement sur les données.

La Figure 37 montre plus précisément l'évolution de l'intervalle A1M2 (l'atteinte de la position de main par rapport au début acoustique de la consonne) en fonction de M1A1 (le début du geste de main par rapport au début de la consonne), soit le phasage de la transition de main M1M2 avec le début acoustique A1 de la consonne. Nous pouvons clairement voir que plus M1A1 augmente, plus A1M2 diminue (cette relation est très bonne pour les trois sujets car nous obtenons de fortes corrélations négatives ; voir sur les graphes de cette figure). La relation statistique mise en évidence signifie que la durée de la transition manuelle (c'est-à-dire la somme de M1A1 et A1M2) est statistiquement stable (invariance absolue, cf. infra VI.2.3) : la transition de main va en fait se déplacer autour du début acoustique de la consonne (A1). Notons que les coefficients sont très proches d'une codeuse à l'autre, de même que les valeurs d'intercepts (137 ms, 139 ms et 158 ms) qui indiquent le temps minimal de transition lorsque le geste démarre au début acoustique de la consonne. Ainsi, si la durée de la transition est stable, le calage de cette phase par rapport au début acoustique de la consonne est variable. Plus la main anticipe son début de mouvement de transition sur le début acoustique de la

consonne, plus la position cible LPC est atteinte précocement. Elle peut être atteinte avant même le début du son (c'est le cas pour les séquences pour lesquelles A1M2 est négatif) et dans ce cas les valeurs de M1A1, soit l'anticipation du début du geste manuel, sont grandes. Notons de plus que pour chacun des sujets, les durées de transitions manuelles au cours de la séquence sont en moyenne assez constantes (M1M2 est en effet proche de M3M4).

La main a donc atteint la position vocalique LPC avant la réalisation de la voyelle aux lèvres (M2L2). C'est toujours le cas pour la codeuse SC (pour qui les croix représentant M2 sont toutes en-dessous des cercles représentant L2 sur la Figure 36). Pour la codeuse AM, c'est toujours le cas à l'exception d'une séquence et pour la codeuse RV, c'est également le cas à l'exception de deux séquences. Dans l'ensemble, nous pouvons considérer que la position manuelle LPC est atteinte en avance par rapport à la réalisation de la voyelles aux lèvres : en moyenne cette avance s'élève de 123 ms à 155 ms pour les trois sujets. Ainsi pour la voyelle, le comportement anticipatoire de la position manuelle sur la cible labiale qui caractérisait le sujet GB, se retrouve également chez ces trois autres sujets.

La main reste en position durant toute la consonne (l'intervalle M2M3 – tenue de cible manuelle – étant en moyenne plus long que la durée moyenne de la consonne, voir Tableau 3) **et repart finalement vers la position suivante** aux alentours de la réalisation de la cible labiale vocalique (M3L2), soit **durant la voyelle acoustique**. L'examen de l'ensemble des séquences pour les trois sujets sur la Figure 36, nous permet de remarquer une assez grande dispersion de la cible labiale L2 (représentée par les cercles) et du début du geste vers la position suivante M3 (représentée par les points) : contrairement à M2, ces deux événements ne semblent pas être liés au début de la transition manuelle M1. Les deux événements sont dans l'ensemble confondus, ce qui montre que le début de la transition suivante se produit autour de la cible labiale. En moyenne, la main débute la transition suivante 9 ms avant la cible articulatoire pour la codeuse AM et de 13 ms à 41 ms après pour les codeuses SC et RV. Ainsi, il semble que l'ajustement plus fin LPC-parole se fasse plus au niveau de la tenue de la position manuelle (M2M3), la durée de la transition étant peu variable.

En résumé, nous trouvons chez ces trois codeuses un schéma temporel de coordination main-lèvres-son globalement assez similaire d'un sujet à l'autre. L'organisation LPC-parole trouvée chez ces trois codeuses est de plus similaire à celle que nous avions observée précédemment chez la codeuse GB durant la production de syllabes CV codées en LPC. Bien que le rythme syllabique et les durées d'intervalles soient différents de ceux observés chez la codeuse GB (de manière générale, le rythme est plus rapide et les durées sont plus courtes pour ces codeuses ; voir sections V.2.3 et V.3.3),

l'organisation temporelle des différents articulateurs de la LPC est maintenue. **Notre résultat majeur réside dans le *rendez-vous* temporel de la position manuelle vocalique** (et en même temps **de la configuration consonantique**, celle-ci étant superposée à la transition) **avec le début acoustique de la consonne et *l'anticipation* de la position manuelle LPC vocalique sur la cible labiale vocalique** (Figure 35).

VI.2.1.2. Comparaison interlocuteur

Afin de comparer les résultats des trois locutrices de manière objective (c'est-à-dire en éliminant les éventuelles fluctuations de rythme syllabique), nous les avons normalisés par la syllabe CV : pour chaque séquence, chaque intervalle en millisecondes est exprimé en pourcentage par rapport à la durée acoustique de la syllabe S_2 (sans unité, noté $\%_{rel}$; voir section V.4.1). Nous rappelons que ces durées de syllabes s'élevaient en moyenne à 252 ms (s= 41 ms) pour AM, à 253 ms (s= 45 ms) pour SC et à 258 ms (s= 56 ms) pour RV (Tableau 3) ; ces durées ne sont significativement pas différentes (ANOVA[14], F< 1 ; voir une représentation graphique de cette proximité sur le graphe en Annexe 3). Ces durées mènent à un rythme syllabique moyen de 4 Hz pour les sujets AM et SC et de 3,9 Hz pour RV. Les résultats moyens et les écarts-types obtenus (en $\%_{rel}$) pour chaque intervalle et chaque sujet sont présentés dans le Tableau 4 (voir également Figure 38). Pour comparer les trois codeuses nous avons effectué une ANOVA à un facteur pour chaque durée d'intervalle (les différences significatives sont indiquées par des astérisques dans le Tableau 4). Le lecteur intéressé pourra trouver en Annexe 3 (p. 22) les représentations graphiques, par des boîtes à moustaches, de ces données normalisées pour chaque intervalle et chaque sujet ainsi que les commentaires associés détaillés (incluant les résultats des analyses de variance et des comparaisons multiples a posteriori).

Ce qui ressort de cette comparaison en durées normalisées est le fait que nous trouvons dans le domaine syllabique, un schéma de coordination main-lèvres-son très similaire d'une codeuse à l'autre (voir les patrons de coordination schématisés sur la Figure 38 pour chaque codeuse) avec, certes, une part de variabilité en ce qui concerne les caractéristiques de codage LPC. De manière générale, l'anticipation de l'initiation du geste manuel sur le son (M1A1) caractérise le codage des trois sujets avec un timing relatif qui semble en moyenne ne pas varier d'un sujet à l'autre. En ce qui concerne la coordination temporelle des autres événements dans le domaine syllabique, nous trouvons une différence de la codeuse RV par rapport aux deux autres (révélée par les tests post-hoc, voir Annexe 3) mais le schéma général reste le même : la position de main LPC vocalique est atteinte en début de consonne acoustique (un peu plus tard dans la consonne pour RV), soit bien avant la réalisation de la

[14] Moyennes comparées également par le test non-paramétrique de Kruskal-Wallis indiquant une différence non significative (H_c= 0,46< H_t= 5,991 α= 5% ddl= 2).

cible labiale vocalique et la main repart ensuite vers la position suivante aux environs de la cible labiale (soit en même temps soit après).

Durées moyennes en $\%_{rel}$ (écart-type)	AM	SC	RV
M1M2 *	69 (15)	71 (17)	78 (21)
M2M3	58 (29)	64 (29)	65 (25)
M1A1	63 (28)	61 (29)	60 (27)
A1M2 *	6 (21)	10 (22)	18 (19)
M2L2 *	62 (21)	57 (20)	47 (23)
M3L2 *	4 (30)	-6 (23)	-18 (27)
M3M4 *	75 (19)	72 (19)	80 (23)

Tableau 4. Moyennes et écart-types des différents intervalles temporels obtenus pour les trois locutrices-codeuses, calculés en proportions par rapport à chaque durée de syllabe CV de chaque séquence (sans unité noté $\%_{rel}$). Les astérisques (*) près des noms d'intervalles indiquent les différences significatives (ANOVA) entre les trois sujets à $p < .01$.

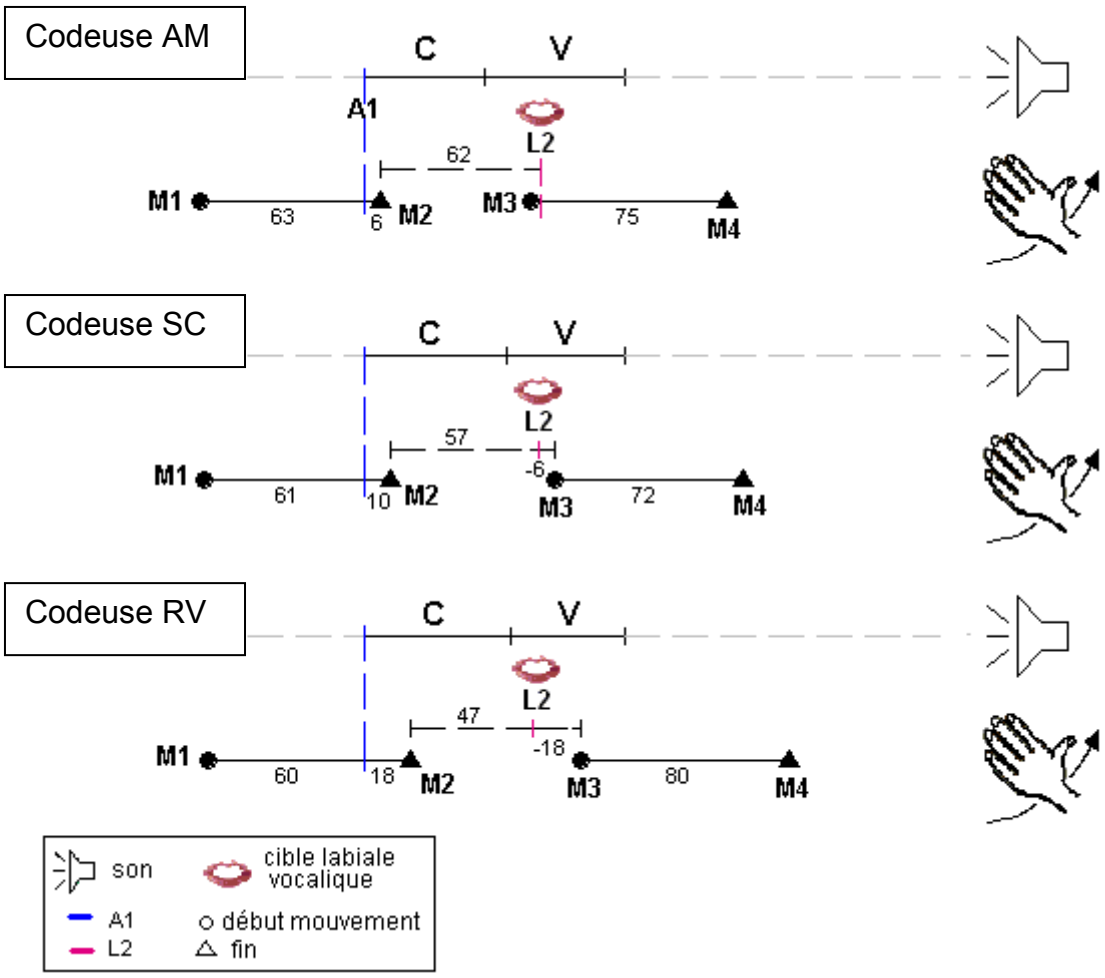


Figure 38. Schémas des patrons temporels de coordination main-lèvres-son dans le domaine syllabique pour chacune des codeuses AM, SC et RV. Les valeurs indiquées correspondent aux durées d'intervalles exprimées en pourcentages relatifs de la durée de la syllabe CV (la durée de la syllabe s'élève donc à $100\%_{rel}$).

Nous avons donc mis en évidence une très grande similarité dans le codage des sujets AM et SC ; la codeuse RV semble quant à elle coordonner son codage de manière un peu différente. L'examen des données pour le sujet RV indique que cette codeuse a tendance à exécuter des transitions manuelles qui durent en moyenne plus longtemps. L'initiation de son mouvement de main par rapport au début acoustique de la syllabe est semblable à celui des deux autres codeuses ; c'est l'atteinte de la position qui se fait plus tardivement après le début de la consonne. Il apparaît donc que son codage manuel est plus lent que celui des deux autres. De cette manière, la coordination temporelle mise en évidence chez les deux autres codeuses semble être légèrement « décalée » chez la codeuse RV : la main atteint sa position cible en début de consonne mais plus tard après le début acoustique, et donc de manière plus proche de la cible labiale. La tenue de cible n'étant pas significativement différente entre les trois codeuses, il s'ensuit que RV débute sa transition manuelle suivante plus tard après la réalisation de la cible vocalique aux lèvres. Notons que la cible vocalique labiale est réalisée en moyenne au même moment par rapport au début acoustique de la syllabe chez les trois codeuses (ANOVA non significative pour A1L2, $F < 1$) ; ce n'est donc pas la cible aux lèvres qui varie d'un sujet à l'autre. C'est bien la stratégie d'atteinte de position cible LPC qui semble différente chez la codeuse RV. Plus précisément, cette différence s'explique par le fait que RV effectue des transitions de main plus longues. Cette différence trouve certainement une explication dans la pratique quotidienne du code LPC par le sujet RV qui est moins intensive que celle des deux autres codeuses : RV a en effet l'habitude de coder pour des élèves au collège alors que AM et SC codent pour des élèves au lycée. Ainsi les deux codeuses AM et SC sont surentraînées et semblent avoir acquis toutes les deux une grande fluidité de codage par rapport à RV.

VI.2.2. Evolution des relations LPC-parole dans le domaine syllabique

Nous avons vu jusqu'à présent que les trois codeuses avaient certains points communs et certaines différences concernant les caractéristiques de codage LPC et nous avons mis en évidence un patron de coordination temporelle globalement stable pour la production du code LPC. Nous allons maintenant examiner plus précisément comment ce patron de coordinations évolue en fonction de la durée de la syllabe CV.

VI.2.2.1. Geste manuel

Examinons tout d'abord le geste LPC indépendamment de sa coordination avec les événements temporels de parole. Comment les durées de transition de main évoluent-elles dans le domaine de la syllabe ? Nous avons sur la Figure 39 le déploiement du geste de transition LPC en fonction de la durée de la syllabe CV pour chaque codeuse. Comme nous pouvons le remarquer, il ne semble pas y

avoir de lien entre la durée de la transition de main et la durée de la syllabe : les nuages de points sont en effet très dispersés. Le calcul des corrélations confirme cette observation : les coefficients sont tous non significatifs au seuil de 1% : r= 0,05 pour AM, r= 0,19 pour SC (significatif à 5% mais plus à 1%, p= 0,0146 ; notons que la part de variance expliquée est seulement de 3,7%) et r= 0,02 pour RV. Ainsi, pour les trois codeuses, les durées de transitions manuelles ne varient pas en fonction de la durée de la syllabe. Globalement, les transitions de main peuvent durer de 100 ms à 300 ms, indépendamment de la durée de la syllabe produite. Rappelons que nous avions montré précédemment que la durée de cette transition manuelle était statistiquement stable (section VI.2.1), ce qui ne signifie pas sans variance. Ici cette variance n'est pas régulièrement liée à la durée CV.

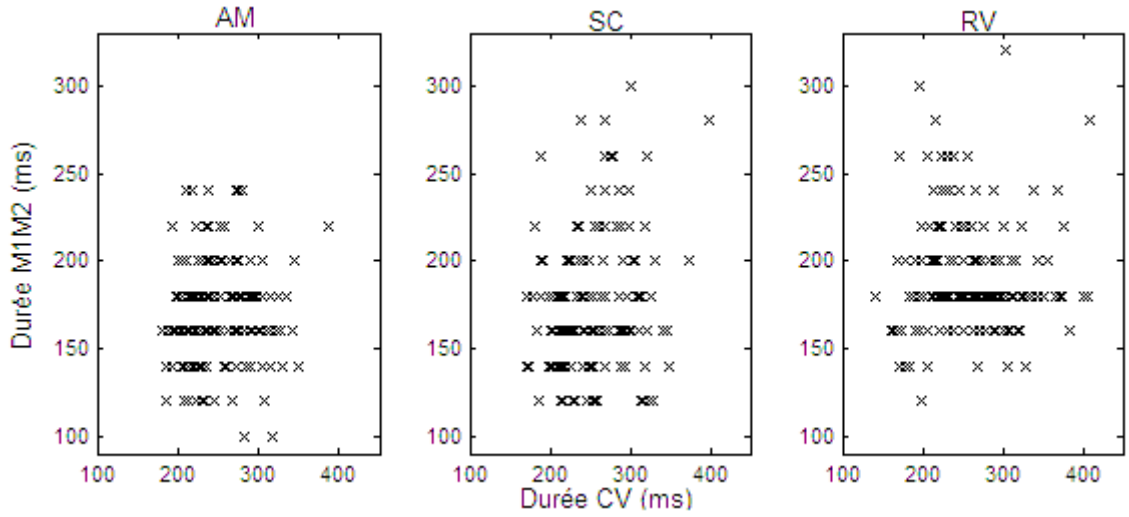


Figure 39. Evolution des durées de transitions manuelles M1M2 (ms) en fonction de la durée de la syllabe CV pour les trois codeuses.

VI.2.2.2. Coordination main-lèvres-son dans le domaine syllabique : à la recherche des invariants

Qu'en est-il maintenant de la coordination du geste de main avec la parole ? Comment le patron de coordination main-lèvres-son mis en évidence évolue-t-il dans le domaine syllabique ? Nous présentons sur la Figure 40, le positionnement relatif – algébrique (en ms) – des différents événements M1, M2 et L2, référencés par rapport au début acoustique de la consonne (A1) pour l'ensemble des séquences, en fonction de la durée de la syllabe CV (ms). Nous avons également calculé les équations des droites d'ajustement linéaire ainsi que les coefficients de corrélation (indiqués à côté de chaque graphe pour chaque événement). A première vue, ce qui ressort graphiquement est le fait que les trois événements temporels M1, M2 et L2 semblent déterminer trois « zones » bien distinctes dans le domaine syllabique : nous pouvons constater en effet que globalement les nuages de points pour chaque événement ne se mélangent pas. Nous retrouvons donc en fonction de la durée de la syllabe

CV, le schéma de coordination temporelle main-lèvres-son que nous avions mis en évidence précédemment : la main débute son geste de transition avant le début acoustique de la syllabe et atteint la position cible LPC en tout début de consonne, soit bien avant la cible labiale vocalique et cela quelle que soit la durée de la syllabe CV. Remarquons également que nous trouvons chez les trois codeuses un certain parallélisme des droites ajustées pour M1 et M2, soit une autre façon de voir que la durée de la transition manuelle M1M2 ne varie pas en fonction de la durée de la syllabe. Notons cependant que cette représentation n'est pas la meilleure façon pour examiner la préservation éventuelle, dite invariance, des relations de phases (Benoît & Abry, 1986 ; Gentner, 1987). Son utilisation a déjà donné lieu à débat dans la littérature (Barry, 1983 ; Munhall, 1985 ; Gentner, 1987 ; voir Sock, 1998 pour une revue) car elle souffre d'un artefact statistique lié à la corrélation « tout-partie » (*part-whole correlation* ; Benoît, 1986 ; le fait de calculer la corrélation entre un ensemble et ses parties peut engendrer de fortes corrélations ; voir par exemple, Tuller & Kelso, 1984). Remarquons en ce sens que dans nos données, tous les coefficients de corrélation sont significatifs (à $p < .01$) mais peu élevés en ce qui concerne le début (M1) et la fin (M2) de la transition manuelle. Notons que le positionnement relatif de la cible labiale L2 par rapport au début de la consonne A1 est bien corrélée à la durée de la syllabe pour les trois sujets ($r = 0,76$ pour AM, $r = 0,81$ pour SC et $r = 0,82$ pour RV, $p < .01$). Mais nous ne discuterons pas davantage ces données de timing relatif évaluées par une méthode entâchée d'un artefact statistique. Il était néanmoins nécessaire de les présenter pour alerter le lecteur sur la critique de cette approche.

La meilleure façon d'étudier le maintien des relations de phases quand la durée de la syllabe augmente est d'étudier l'évolution des proportions, par rapport à la durée totale de la syllabe, prises par chaque phase (Benoît & Abry, 1986 ; Gentner, 1987), en utilisant *le test de la proportion constante* (*Constant Proportion Test*) proposé par Gentner (1987), qui consiste à comparer la pente de la droite de régression avec 0. Cette statistique permet de tester le *modèle proportionnel de durée* (*proportional duration model*, Gentner, 1987 ; ce modèle correspond au concept de *Programme Moteur Généralisé* ou *PMG*, Schmidt, 1982, 1988) qui postule *l'invariance relative proportionnelle*, c'est-à-dire un maintien de proportions constantes (par rapport à la durée totale du geste) par les composantes d'un geste, malgré la variation de la durée totale du geste. Il a déjà été utilisé de nombreuses fois et a permis de démontrer que pour les activités motrices en général, le modèle proportionnel n'est pas une description adéquate, c'est-à-dire qu'il n'y a pas réellement d'invariance généralisée (voir Gentner, 1987 et Jeannerod, 1988 ; plus spécifiquement pour la parole, voir Benoît & Abry, 1986 ; Abry et al., 1990 ; Löfqvist, 1991 ; Sock, 1998).

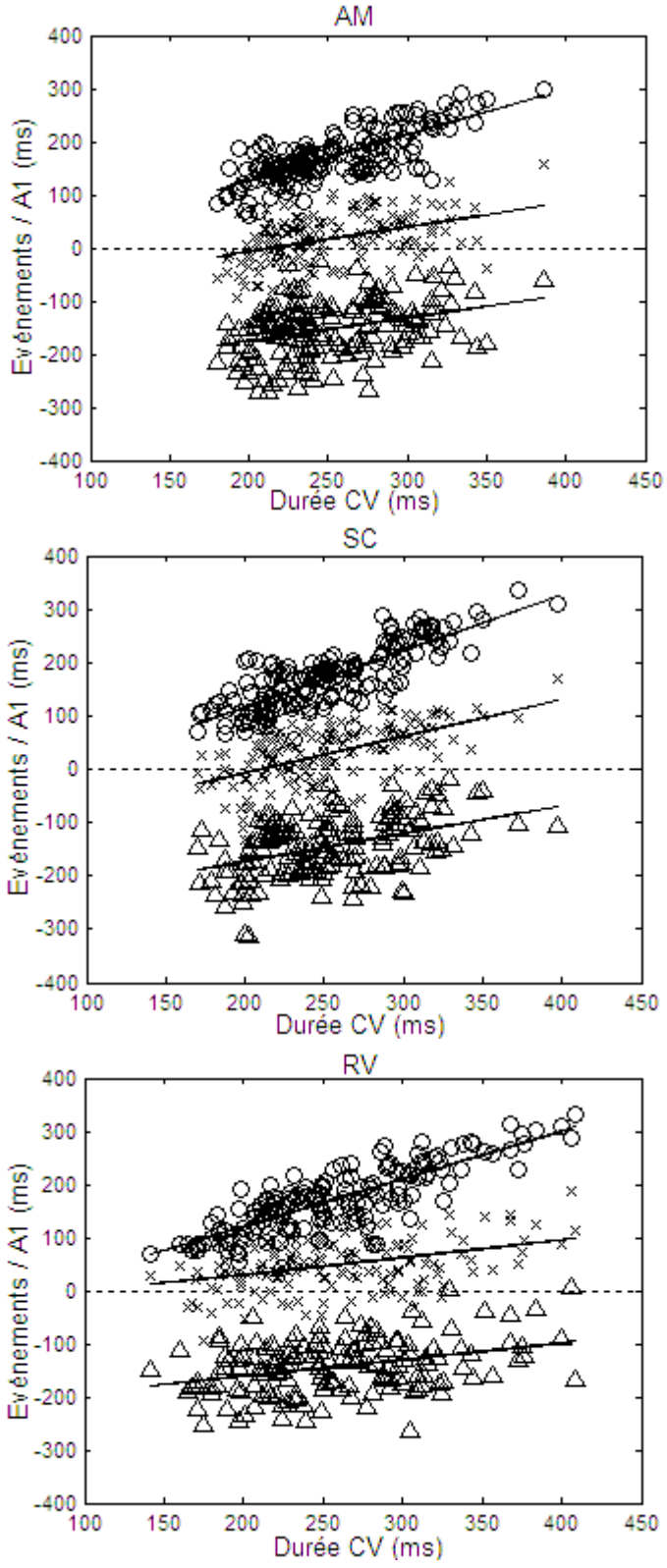
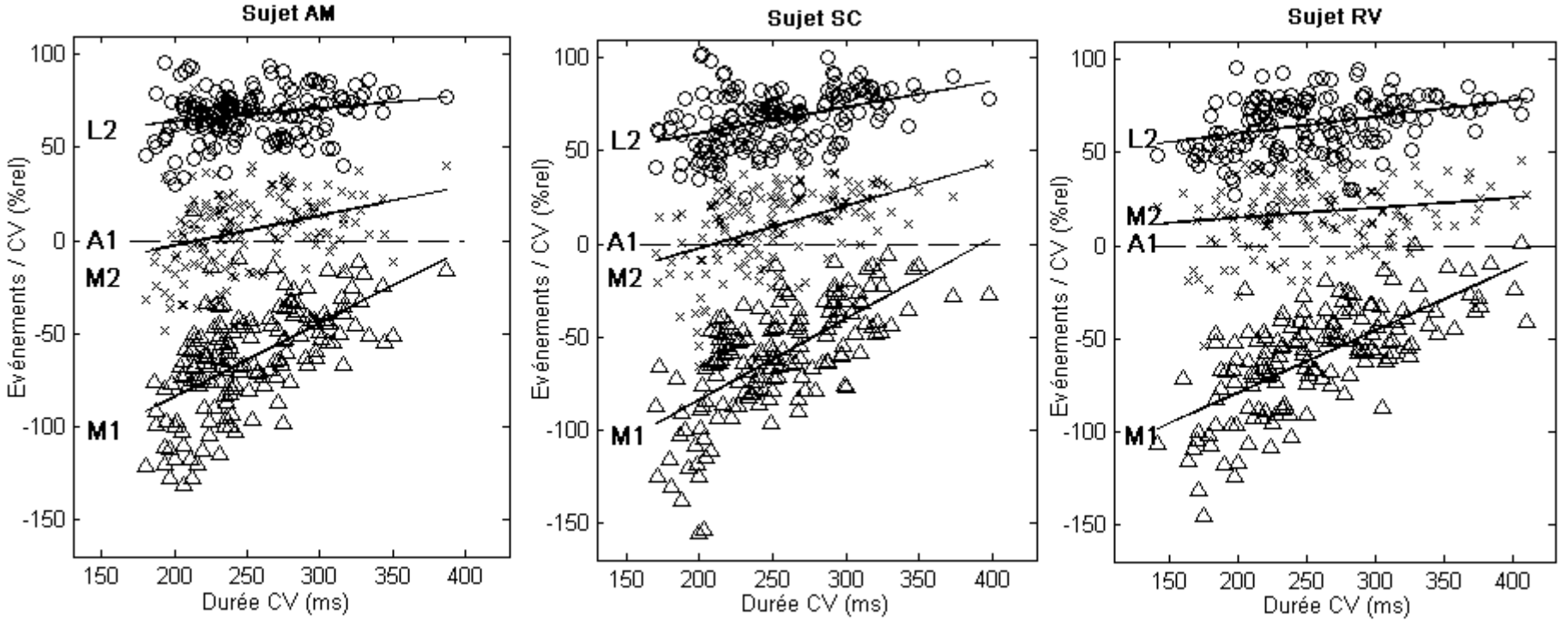


Figure 40. Organisation temporelle des événements manuels et orofaciaux durant la production de code LPC en fonction du cycle syllabique pour les trois codeuses AM, SC et RV : M1 est le début du geste de transition manuelle (représenté par des triangles), M2 est la fin de la transition (représenté par des croix) et L2 est la cible labiale vocalique (représenté par des cercles). Les événements temporels sont positionnés en relatif (ms) par rapport au début acoustique de la syllabe A1 (représenté par la droite en traits tiretés à 0 ms). Ces graphes complètent ceux de la Figure 41 pour le positionnement relatif proportionnel.



Equations des droites et coefficients de corrélations :

Sujet AM	Sujet SC	Sujet RV
Pour M1 : y= 0,4 x – 164 ; r= 0,59	Pour M1 : y= 0,44 x – 171 ; r= 0,67	Pour M1 : y= 0,33 x – 146 ; r= 0,69
Pour M2 : y= 0,16 x – 34 ; r= 0,31	Pour M2 : y= 0,23 x – 48 ; r= 0,46	Pour M2 : y= 0,05 x + 4 ; r= 0,16 (non sign.)
Pour L2 : y= 0,07 x + 50 ; r= 0,23	Pour L2 : y= 0,14 x + 31 ; r= 0,4	Pour L2 : y= 0,09 x + 42 ; r= 0,34

Figure 41. Organisation temporelle relative des événements manuels et orofaciaux durant la production de code LPC en fonction du cycle syllabique pour les trois codeuses AM, SC et RV : M1 est le début du geste de transition manuelle (représenté par des triangles), M2 est la fin de la transition (représenté par des croix) et L2 est la cible labiale vocalique (représenté par des cercles). Les événements temporels positionnés par rapport au début acoustique de la syllabe A1 sont exprimés en pourcentages de la durée de la syllabe CV ($\%_{rel}$). La droite en traits tiretés à $0\%_{rel}$ indique le début acoustique de la syllabe, soit l'événement A1. Les droites d'ajustement linéaire (au sens des moindres carrés) pour chacun des événements sont superposées sur chacun des graphiques. Les équations de droite et les coefficients de corrélations (Bravais-Pearson) sont également indiqués.

Qu'en est-il de notre coordination LPC-parole ? Est-ce que la coordination entre la main, les lèvres et le son suit un timing proportionnel dans le domaine syllabique ? Nous présentons nos données de coordination LPC-parole en timing proportionnel relatif à la durée de la syllabe acoustique CV dans la Figure 41 : comme précédemment, les événements temporels M1, M2 et L2 sont référencés par rapport à A1 (représenté par le zéro dans la figure) mais les intervalles résultants sont exprimés en pourcentages relatifs de la durée de la syllabe CV ($\%_{rel}$). Nous indiquons également les équations des droites d'ajustement linéaire (superposées sur chaque graphe pour chaque événement) ainsi que les coefficients de corrélation de Bravais-Pearson.

Globalement la durée de la syllabe peut varier de 150 ms à 400 ms environ. En ce qui concerne la relation temporelle entre le début du geste de transition de la main (M1) et le début acoustique de la syllabe (A1), nous pouvons voir que le début du geste de main se produit globalement entre -150%$_{rel}$ (pour les syllabes courtes) et 0%$_{rel}$ (pour les syllabes longues), ce qui signifie que la main peut initier son geste avec une avance de bien plus d'une syllabe (équivalent à 100%$_{rel}$) jusqu'à être quasiment synchrone avec le début acoustique de la consonne (0%$_{rel}$). Pour les trois codeuses, il apparaît clairement que plus la syllabe est longue, plus l'instant d'initiation du geste manuel se rapproche du début acoustique A1 de la syllabe en proportion relative à CV, c'est-à-dire que le geste va avoir tendance à moins anticiper en timing relatif par rapport au début de la consonne quand la syllabe s'allonge. Il est donc évident que la relation entre le début du geste de main et le début acoustique de la consonne dans le domaine de la syllabe n'est pas proportionnellement invariante ; nous aurions autrement obtenu des pentes de droites significativement non différentes de 0 (ce n'est pas le cas car toutes les pentes sont significativement différentes de 0 : t(AM)= 9,3 ddl= 159, t(SC)= 11,3 ddl= 157 et t(RV)= 11,9 ddl= 153). De plus, nous pouvons remarquer chez les trois codeuses un comportement relativement similaire (les trois pentes de droites ajustées sur les M1 ne sont pas significativement différentes pour les trois sujets : au risque α= 5%, t(AM-SC)= 0,63 ddl= 318 ; t(AM-RV)= 1,25 ddl= 314 ; t(SC-RV)= 2,12 ddl= 312 significatif à 5% mais pas au risque α= 1%).

En ce qui concerne la relation temporelle entre la fin du geste manuel de transition (M2) et le début acoustique de la syllabe (A1), nous observons cette fois des pentes de droite moins importantes (pentes à 0,16 pour AM, 0,23 pour SC et 0,05 pour RV). Plus précisément, les pentes pour AM et SC, bien que relativement faibles, sont significativement différentes de 0 (t(AM)= 4,1 ddl= 159 et t(SC)= 6,6 ddl= 157) alors que celle de RV ne l'est pas (au risque de 5%, t(RV)= 1,9755 ddl= 153 ; notons cependant que le t théorique est à 1,9756). Ainsi, les codeuses AM et SC semblent avoir un comportement similaire en fonction de la durée de la syllabe : globalement, pour des syllabes de durées inférieures à 250 ms, la position de main peut parfois être atteinte avant le début acoustique de

la consonne A1, puis avec l'allongement de la durée de la syllabe, la main atteint sa position après A1 (notons qu'un ajustement linéaire par parties semblerait plus approprié). La codeuse RV, quant à elle, a davantage tendance à atteindre la position de main après le début de la consonne de manière peu variable en fonction de la durée de la syllabe. Ainsi l'atteinte de la position cible LPC maintient une proportion constante avec l'allongement de la durée de la syllabe.

Comment comprendre la coordination de la transition manuelle avec le début acoustique de la consonne ? Nous avions vu précédemment que la durée absolue de la transition était statistiquement stable : quand l'anticipation du début du geste était importante (M1A1), la position était atteinte de manière très proche du début de la consonne (A1M2) et inversement (pour un rappel, voir Figure 37 p. 22). Nous avons ici cette relation en fonction de la durée de la syllabe acoustique CV. Pour des syllabes courtes, la main commence son geste avec une avance de plus d'une syllabe sur le début acoustique. Elle atteint sa position de manière synchrone avec le son, voire avec une légère avance sur le début acoustique de la consonne. Pour des syllabes plus longues, il apparaît que le début du geste de main se rapproche temporellement du début acoustique de la consonne. Dans ce cas, l'atteinte de la position cible LPC spatiale se fait plutôt après le début acoustique de la consonne. Il apparaît donc que la codeuse va beaucoup anticiper le geste manuel par rapport au son quand la syllabe qui va être produite est courte. Elle va anticiper dans une moindre mesure quand la syllabe qui va être produite est plus longue. Le sujet a donc tendance à anticiper davantage quand la « cible temporelle » – la première partie de la consonne – est petite, le point de rencontre temporelle entre la main et la consonne recquérant plus de précision. Il est important de rappeler, comme nous l'avons vu plus haut, que la durée de la transition manuelle ne varie pas avec la durée de la syllabe CV (Figure 39). C'est bien la coordination entre les deux qui va changer : en fonction de la durée de la syllabe qu'elle va produire, la codeuse va « décaler » temporellement plus ou moins le moment de l'initiation du geste de main de façon à atteindre la position cible LPC durant la consonne. La Figure 42 montre la distribution (sous forme de boîtes à moustaches) de l'intervalle A1M2 exprimé en pourcentage relatif de la durée de la consonne acoustique : les bornes à 0 et 100% représentent donc les frontières de la consonne (début et fin). Nous pouvons voir, pour la majorité des données, que l'atteinte de la position LPC se fait avant la fin de la consonne acoustique (c'est toujours le cas pour SC). Bien que les données soient assez dispersées, en moyenne, la position LPC est atteinte à 10% de la durée de la consonne pour AM, à 16% pour SC et à 33% pour RV (les moyennes sont représentées par des 'x' sur la figure), soit de manière très proche du début acoustique de la consonne.

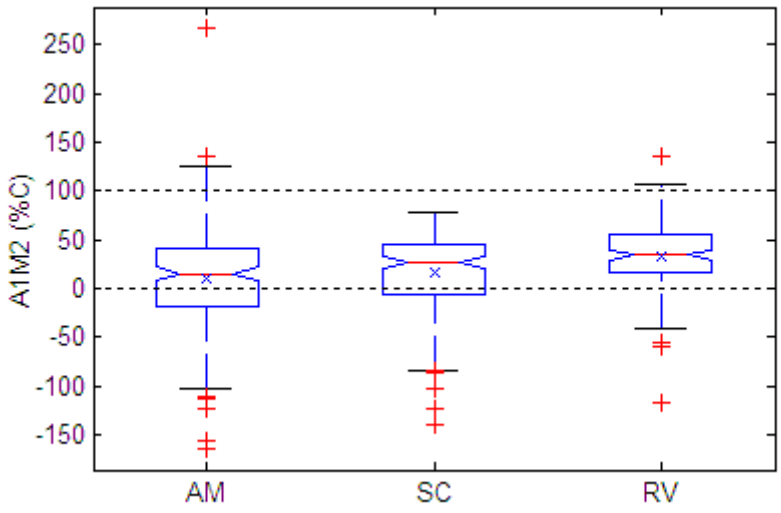


Figure 42. Boîtes à moustaches représentant la distribution de l'intervalle temporel A1M2 exprimé, pour chaque séquence, en pourcentage de la durée de la consonne acoustique, pour les trois codeuses AM, SC et RV. Chaque boîte indique la médiane, les 1er et 3ème quartiles, les valeurs frontières, les valeur hors normes ('+') et la moyenne ('x'). Les traits tiretés à 0 et à 100 % indiquent le début et la fin de la consonne acoustique.

VI.2.3. Variance et variabilité régulière (*lawful variability* ou invariance)

De manière générale, les données des trois codeuses ne valident pas le modèle proportionnel de durée (Gentner, 1987). A l'exception de l'atteinte de la position de main M2 pour la codeuse RV, toutes les phases de la coordination LPC-parole varient en effet en fonction de la durée de la syllabe CV. Comme d'autres auteurs qui ont étudié le timing proportionnel en parole (Abry et al., 1990 ; Löqfvist, 1991 ; Sock, 1998), nous ne trouvons donc pas d'invariance généralisée pour le contrôle moteur de la parole avec code LPC. Il est à noter par ailleurs que la plupart des activités motrices demandant de l'expertise ne sont pas décrites adéquatement par le modèle proportionnel, comme l'a démontré Gentner (1987) en réanalysant plusieurs études se réclamant de ce modèle (voir aussi Sock, 1998). En fait, il n'y aurait pas particulièrement de raison de penser que des durées proportionnelles soient maintenues constantes quand la durée totale augmente ; ce postulat sous-tendait la notion de Programme Moteur Généralisé (Schmidt, 1982, 1988). Gentner n'argumente pas contre cette notion mais propose plutôt d'étendre le concept : « […] The proportional duration model is only one possible version of a generalized motor program. The generalized motor program should be considered one component in a composite model of motor control with many other overlapping control mechanisms. Control of timing is determined at several levels in the perceptual-cognitive-motor system, and the nature and relative importance of these control levels can shift with skill acquisition and in response to the task environment. » (1987, p. 274-275) (pour une discussion sur ce point, voir aussi Heuer, 1991).

Nous avons mis en évidence en revanche une *invariance absolue* de la durée de la transition manuelle (voir section VI.2.1.1). Nous avons en effet observé que la durée du geste de transition est statistiquement stable chez nos trois codeuses (Figure 37 p. 22). Remarquons également que nous trouvons le même comportement chez la codeuse GB, aussi bien pour les durées de transitions que pour les formations de configurations de main (ceci est bien visible sur la Figure 27 p. 22 par exemple où nous trouvons un parrallélisme des distributions pour M1-M2 et D2-D1). Dans les recherches en contrôle moteur, une telle stabilité de la durée du mouvement a été parfois observée (Jeannerod, 1988). Viviani et Terzuolo (1980) par exemple ont observé des durées de mouvements de signatures maintenues constantes alors que leur taille variait largement (*space invariance*). Jeannerod (1984) a observé un même phénomène pour des mouvements de saisie d'objets situés à différentes distances. C'est le principe d'isochonie : le temps de mouvement tend à être constant pour des amplitudes différentes, grâce à une augmentation de la vitesse. Cette invariance absolue reste néanmoins assez rare et a été abandonnée au profit d'une invariance des lois caractérisant les mouvements humains. Ainsi, la plus connue, la loi de Fitts, établit une relation linéaire entre le temps de mouvement et l'indice de difficulté (pour une revue, voir Schmidt, 1988 et Meyer et al., 1990 ; pour une application, voir Jeannerod, 1988 ; voir aussi Gibet et al., 2004, pour l'intérêt de ces lois en animation de synthèse).

Nous avons donc observé chez nos codeuses une durée du geste de transition qui tend à être constante et qui peut expliquer le fait que nous ne trouvons pas d'invariance proportionnelle dans le domaine syllabique. Nous avons observé une stratégie de coordination temporelle très similaire d'un sujet à l'autre : les codeuses mettent en œuvre une stratégie de codage particulière qui vise à atteindre la position de main en début de consonne, la date d'initiation du geste dépendant de la durée de la syllabe à produire. Ces résultats mettent donc en évidence le fait que la production du code LPC est une activité complexe qui requiert une grande part de contrôle de la coordination de la main en relation avec la parole.

VI.3. Conclusion et discussion générale sur la production de syllabes CV codées

VI.3.1. Schéma général de coordination temporelle LPC-parole

En conclusion à cette étude, un même patron général de coordination temporelle main-lèvres-son, très similaire à celui de la codeuse GB, a été mis en évidence chez trois autres codeuses. Pour la production d'une syllabe CV codée, la main débute sa transition vers la position cible bien avant le début acoustique de la consonne, avec une avance moyenne de plus d'une demi-syllabe. La position cible LPC est atteinte durant le début acoustique de la consonne (durant sa première partie). La main

pointe donc sur la position bien avant la cible de la voyelle aux lèvres. Bien que nous ayons observé une certaine variabilité entre les codeuses, ce schéma de coordination temporelle dans la production de code LPC est respecté par les quatre codeuses. Le comportement *moyen* de codage LPC ainsi que l'examen des séquences particulières chez les quatre codeuses mettent en évidence l'anticipation de l'initiation du geste manuel sur le début acoustique de la syllabe et l'anticipation de la position manuelle sur la cible vocalique aux lèvres. En rassemblant toutes les études, nous pouvons donc résumer le schéma de coordination obtenu de la façon suivante.

En résumé, dans la production d'une syllabe CV codée, la main débute sa transition bien avant le début acoustique de la syllabe et atteint sa position cible sur le visage en début de consonne, donc bien avant la cible vocalique aux lèvres. La main reste en position durant toute la consonne, puis redémarre son geste vers la position suivante, dans la voyelle. En ce qui concerne la clé consonantique, elle est formée durant la transition de la main d'une position à une autre du visage (l'information consonantique est donc superposée à l'information vocalique) et est entièrement réalisée en début de consonne. Les deux informations vocaliques transmises à la fois par la position de main et la forme aux lèvres ne sont donc pas synchrones : *le geste de la main anticipe la réalisation de la cible labiale vocalique*. En revanche, la configuration consonantique de la main pointe la position du visage en synchronie avec la consonne.

Il apparaît donc que le code LPC vient se greffer complètement sur l'articulation naturelle de la parole. Comme pour d'autres types de gestes (pour un rappel, voir section III.2), nous avons observé en LPC une forte interdépendance main-parole au niveau de leur coordination. L'anticipation gestuelle – en considérant l'initiation du geste manuel par rapport au début de la parole – qui avait été observée pour les gestes co-verbaux est également mise en évidence ici pour des gestes LPC. On pourra même remarquer que cette anticipation est globalement du même ordre, le geste était initié moins d'une seconde avant le début de la parole (voir section III.2.2.2 p. 22). Il semblerait donc que cette avance temporelle de la main soit une règle générale dans les relations geste-parole, résultant peut-être d'un processus de compétition des deux activités motrices (voir section III.2.3). Cependant, le code LPC représente un système d'étude unique liant de manière aussi étroite la main à la parole. La fonction commune ne se situe pas au niveau sémantique mais à un niveau plus bas, au sens des traitements plus proches de l'*output*, au niveau phonologique. Le code manuel (la position et la forme de la main) est complètement déterminé par la parole *à venir*, car c'est bien une précédence de la main que nous trouvons. Les deux systèmes moteurs ayant des vitesses de conduction de l'influx nerveux différentes (Berthoz, 1997, p. 100), le cerveau anticipe sur le mouvement de la main ou retarde ceux de la parole par un système de temporisation. Il y a en effet en LPC une sorte de collaboration optimale de la main

et du système articulatoire de parole vers un même but phonologique, qui va être délivré de manière non synchrone (en ce qui concerne l'information vocalique). Il semble bien que l'anticipation manuelle que nous avons clairement mise en évidence soit le résultat d'une **stratégie de contrôle main-parole optimale**. Il y aurait une sorte de couplage fonctionnel entre les deux porteuses rythmiques LPC-parole, la main (l'ensemble poignet-main) et la mandibule, et les processus cognitifs centraux pourraient rendre compte de l'adaptabilité et de la flexibilité dans la pratique de ce code manuel, et plus généralement dans les performances motrices expertes (Summers, 1990).

VI.3.2. L'anticipation, résultat de la compatibilité des contrôles LPC-parole

De manière plus précise, quelle est donc notre proposition sur les contrôles des mouvements LPC-parole ? Nous pouvons faire un parallèle entre les différents types de contrôle moteur en parole seule et en LPC et ainsi mieux comprendre les relations qui se nouent entre les différentes composantes en jeu dans la *coproduction* LPC-parole. En parole, nous le rappelons, trois types de contrôle ont été proposés (Vilain, 2000 ; voir section II.2) : le contrôle de la mandibule qui génère le rythme syllabique, le contrôle postural global du conduit vocal pour la production de la voyelle et le contrôle local pour produire les contacts et pressions sur différentes parties du conduit vocal pour les consonnes. Öhman (1966, 1967) propose dans son modèle de coarticulation pour la parole, que le geste consonantique soit superposé sur le geste vocalique.

L'unité de codage étant la syllabe CV, nous pouvons trouver en LPC les trois composantes de la parole. Au cycle mandibulaire vient répondre le mouvement cyclique poignet-main permettant de passer d'une position vocalique à l'autre. Cet ensemble poignet-main est la *porteuse* du code LPC ; c'est elle qui lui donne son rythme. Par la resyllabisation systématique de la parole en suite de syllabes CV, le rythme manuel *entraîne* celui de la parole et la contraint ainsi à un ralentissement (son rythme naturel approchant les 6 Hz±1, Sorokin et al., 1980). La main étant vraisemblablement un articulateur plus lent que la mandibule dans le plan LPC, elle ne semble pas pouvoir suivre ses oscillations. Il est en effet connu que la parole avec LPC est généralement plus lente que la parole seule : c'est une observation notée par bon nombre de professionnels de la surdité et que Duchnowski et collaborateurs avaient également rapportée (2000). Et nous avons en effet mesuré chez nos codeuses des rythmes proches de 4 Hz pour des séquences syllabiques dans cette étude. Pour chacune de nos trois codeuses, des phrases avaient également été enregistrées avec et sans LPC (voir Annexe 4 pour plus de détails) : le rythme en parole naturelle était proche de 5 Hz (4,7 Hz pour AM, 5,1 Hz pour SC et 4,9 Hz pour RV) et était ralenti à moins de 4 Hz avec le code manuel (3,6 Hz pour AM, 3,7 Hz pour SC et 3,8 Hz pour RV). Notons cependant qu'au-delà d'un simple ralentissement, l'ajout du code LPC à la

parole ne perturbe pas l'oscillation mandibulaire naturelle ; c'est ce que montre une autre de nos études, variant la prosodie rythmique, non rapportée dans cette thèse (voir Annexe 5).

En ce qui concerne les deux autres composantes, le contrôle du geste porté pour la voyelle est un mouvement dirigé vers un but (celui de pointer la position cible correspondant à la voyelle) et il aboutit au placement *local* de la main près du visage : c'est donc la *composante locale*. L'atteinte de la position est souvent marquée par le contact de l'index ou du majeur sur le visage (à l'exception de la position côté qui, par principe, ne comporte pas de cible sur le visage). La formation de la clé implique une mise en forme globale de la main ou configuration des doigts pour coder la consonne : cette configuration est la *composante globale*.

Il apparaît donc que les deux types de contrôle en LPC pour les consonnes et les voyelles sont inversés par rapport à la parole. En LPC, la configuration de la main pour coder la consonne est le résultat d'un contrôle global de la mise en forme des doigts, alors qu'en parole la consonne est issue d'un contrôle local de la constriction. En ce qui concerne la voyelle, son codage est le résultat d'un contrôle local de la position de la main, alors qu'en parole, la mise en forme du conduit vocal pour la voyelle est issue d'un contrôle global de la posture des articulateurs dans le conduit vocal. Le code LPC est ainsi en quelque sorte par construction informationnelle (Cornett, 1967) le « monde à l'envers » du contrôle moteur du langage parlé.

En ce qui concerne la transmission des informations segmentales, nous trouvons en LPC un modèle comparable à celui d'Öhman. Le code LPC apparaît en effet comme une suite de transitions de main de position à position, définissant les voyelles, sur lesquelles se superposent les configurations digitales, définissant les consonnes. Les contrôles de ces deux composantes sont en phase, puisque, du fait de la resyllabisation de la parole en suites de syllabes CV, la main transmet simultanément l'information sur la consonne et sur la voyelle ; ce qui n'est pas forcément toujours le cas en parole (pour les consonnes en position de coda dans les structures CVC par exemple). De plus, le geste consonantique LPC, complètement superposé sur la transition de main, ne va jamais masquer le début du geste vocalique LPC, contrairement à ce qu'on peut observer en parole dans le cas de syllabes CV, où la consonne, bien qu'en phase avec la voyelle, va cacher le début du geste vocalique (l'abaissement de la langue pour le [a] dans [pa] est caché par la fermeture bilabiale du [p]). En effet, la superposition de la consonne sur le geste vocalique n'empêche pas, par principe, le maintien de la configuration globale du conduit vocal pour la voyelle, ou son évolution vers la voyelle suivante.

En ce qui concerne le phasage de la parole et du code LPC, nous avons montré pour le geste manuel vocalique LPC une synchronisation avec la consonne en parole et une anticipation temporelle sur la

cible articulatoire de la voyelle. La position LPC de la voyelle est donc en phase avec le début du geste consonantique de la parole. Il apparaît donc que **le phasage LPC-parole s'opère entre les deux composantes de contrôle moteur locales, soit le contrôle du mouvement pour le contact de la position articulatoire de la consonne (occlusion ou constriction) en parole et le contrôle du mouvement pour la position-contact de la main autour du visage en LPC**. Ces deux contacts locaux sont en effet quasi synchrones (*phase-locked contact control*). Ce phasage des deux contacts locaux explique que ce soit la main qui soit en avance sur les lèvres, de manière à assurer en début de syllabe articulatoire CV la présence de la position et de la clé. Ce phasage particulier *préférentiel*, résultat d'une compatibilité des composantes locales de contrôle, serait selon nous le reflet de la stratégie de contrôle optimal adoptée par les codeurs professionnels (voir d'autres comportements de *phase-lockings* dans les activités motrices, par exemple, Treffner & Peter, 2002). Le cerveau manœuvre vers une *cohérence* perceptive (Berthoz, 1997), qui se fait dans l'action par une synchronisation des deux composantes locales de contrôle moteur. La resyllabisation systématique en LPC de tout type de syllabes en suite de syllabes CV signifie que chaque geste LPC vocalique (position) sera en phase avec le début du geste consonantique de la parole, la transition vocalique en bouche, visible ou cachée, s'effectuant naturellement pendant cette consonne, avec un retard d'une bonne demi-syllabe sur la main vocalique. Le LPC est donc produit en relation étroite avec le système articulatoire de la parole, son organisation temporelle spécifique permettant de l'ancrer au mieux sur le contrôle de la faculté de la parole.

CHAPITRE VII.
Etude de l'anticipation coarticulatoire chez la codeuse GB

Attina V., Cathiard M.-A & Beautemps D. (2004).
L'ancrage de la main sur les lèvres :
Langue Française Parlée Complétée et anticipation vocalique.
XXVèmes Journées d'Etude sur la Parole, Fès, Maroc.

Cathiard M.-A., Attina V. & Alloatti D. (2003).
Labial Anticipation Behavior during Speech with and without Cued Speech.
15th International Congress of Phonetic Sciences,
Barcelona, Spain.

Dans ce chapitre, nous explorons l'organisation temporelle de la coarticulation anticipatoire augmentée de code LPC chez la codeuse GB. Jusqu'à présent, nous avons étudié le codage LPC et sa coordination avec la parole dans le cas classique de la syllabe CV ; rappelons qu'il s'agit du cas général puisque cette structure syllabique est la plus courante en parole (Blevins, 1995 ; Vallée et al., 2000) et qu'elle constitue de plus l'unité de base du codage LPC. Nous avons vu dans le CHAPITRE II qu'il existait des cas particuliers pour lesquels les gestes articulatoires anticipaient plus largement sur le son : en particulier, il a été montré pour des séquences $V_{\text{non arrondie}}C_nV_{\text{arrondie}}$ (C_n étant un groupe de plusieurs consonnes), que les lèvres peuvent débuter le geste vocalique d'arrondissement très tôt dans la séquence consonantique qui précède la voyelle (pour un rappel, voir section II.3). Ainsi, les lèvres s'adaptent en anticipant plus largement sur le son.

Dans le cas particulier où le code manuel LPC est co-produit avec la parole, il est intéressant de voir si le geste manuel va suivre, s'adapter ou résister à l'anticipation labiale. Dans un contexte où le geste labial est clairement en avance sur le son, comment le geste de la main se coordonne-t-il avec les formes labiales ? Nous avons montré précédemment pour la production de code LPC un ancrage de la main sur la parole, avec une forte interdépendance LPC-parole dans un contexte CV. Plus précisément, pour des séquences syllabiques CV, nous avons mis en évidence une anticipation systématique de la main sur les lèvres pour l'information vocalique, caractérisant le codage LPC de codeurs professionnels. Dès lors, plusieurs questions sont posées. Quelle influence la coproduction du code manuel va-t-elle avoir sur l'anticipation des gestes vocaliques labiaux ? Tout d'abord, l'anticipation labiale sera-t-elle préservée dans la parole codée ? Si oui, est-ce que l'anticipation manuelle sur la cible labiale vocalique, caractéristique des syllabes CV, va coexister avec l'anticipation labiale ? En d'autres termes, comment la main et la parole vont-elles s'organiser temporellement dans leur *coproduction* ?

Nous proposons donc une étude de l'anticipation vocalique par ses deux composantes visuelles : l'arrondissement et la hauteur. Nous présentons deux expériences explorant cette coordination temporelle. La première (expérience 4) étudie l'anticipation d'arrondissement et de hauteur dans des transitions de voyelle à voyelle au travers d'une pause prosodique silencieuse. Ce paradigme particulier, variant la longueur de la pause, teste spécifiquement l'établissement de la position articulatoire cible de la voyelle (son anticipation labiale dans notre cas) et a déjà été éprouvé pour ce genre d'études (Cathiard, 1994 ; Abry et al., 1996a). Notons que l'insertion de pauses en parole est une des stratégies utilisées en parole « claire » afin d'améliorer les conditions de communication (voir par exemple, Picheny et al., 1986, pour les sujets sourds). Le geste de base, de voyelle à voyelle, partant d'une voyelle étirée comme [i] à une voyelle arrondie protruse comme [y], peut ainsi être étudié

spécifiquement sans l'influence d'un geste consonantique superposé. La deuxième expérience (expérience 5), inspirée des études sur l'anticipation de Abry et Lallouache (1995a, 1995b), teste cette anticipation vocalique au travers d'une suite de consonnes variant en nombre (de 0 à 4 consonnes). Nous serons ainsi à même de pouvoir comparer nos résultats de parole coarticulée codée (le phénomène d'anticipation labiale) avec les prédictions des modèles classiques de l'anticipation (voir section II.3).

VII.1. Expérience 4 : anticipation vocalique d'arrondissement et de hauteur au travers d'une pause prosodique

VII.1.1. Corpus

Nous avons constitué un corpus nous permettant d'étudier le geste de voyelle à voyelle au travers d'une pause silencieuse plus ou moins longue. Nous étudions le geste labial de cette transition vocalique à partir de la configuration labiale étirée de la voyelle [i] à la configuration labiale arrondie de la voyelle [y] pour l'anticipation d'arrondissement et de la configuration labiale de la voyelle [i] à la configuration labiale de la voyelle [a] (qui implique une grande aperture labiale par un abaissement de la mandibule) pour l'anticipation de hauteur. Ces transitions vocaliques sont insérées dans la phrase porteuse[15] du type « T'as mis : UHI ise ? » [tami#yiiz] (ou bien « T'as mis : AHI ise ? » [tami#aiiz]), avec un temps de pause plus (#:) ou moins (#) long. La variation du temps de pause dans la première transition vocalique (de [i] à [y] dans [m**i#y**] et de [i] à [a] dans [m**i#a**]) nous permettra de maximiser la variabilité de l'anticipation du geste articulatoire (geste de constriction vs. geste d'aperture) pour la position cible de la voyelle. La fin en [iiz] garantit une fin de signal claire du point de vue du geste articulatoire et n'est pas gênante du point de vue de la clé LPC, car celle-ci ne sera pas étudiée.

Le codage LPC de ces deux phrases est illustré sur la Figure 43 et la Figure 44. Comme nous pouvons le voir sur ces figures, la transition [i#y] est codée par la main depuis la position « bouche » (correspondant au [i]) jusqu'à la position « cou » (correspondant au [y]), avec la configuration n°5 qui code les voyelles isolées (voir Figure 5), et la transition [i#a] est codée depuis la position « bouche » vers la position « côté » (qui code le [a]) avec cette même configuration. Inversement, les transitions [yi] et [ai] (qui ne seront pas étudiées) sont codées par un retour de la main vers la position « bouche ».

[15] Dans cette phrase, « UHI » ou « AHI » représente un nom propre et « ise » un pseudo-verbe à la 3ème personne du singulier

Figure 43. Codage LPC de la phrase « T'as mis : UHI ise ? »

Figure 44. Codage LPC de la phrase « T'as mis : AHI ise ? »

Nous enregistrons 12 répétitions de ces phrases dans chaque condition (pause courte - pause longue), disposées dans un ordre aléatoire. Au total, nous obtenons donc 24 séquences pour la dimension d'arrondissement et 24 séquences pour la dimension de hauteur. En condition contrôle, nous enregistrons quatre réalisations de la phrase « T'as mis : IHI ise ». Pour cette séquence, le geste reste en effet le même : les articulateurs ne changent pas de position mais restent sur le [i] (en LPC, ceci est codé par un léger mouvement avant-arrière de la main à la position « bouche »).

VII.1.2. Traitement des données

Pour l'acquisition et le traitement des données, nous avons procédé de la même manière que pour l'expérience 1 (voir section V.2.2, p. 22). Au niveau articulatoire, nous suivons l'aire aux lèvres qui nous donne un bon aperçu de l'anticipation articulatoire ; en effet, ce paramètre articulatoire pertinent pour l'arrondissement (Abry et al., 1980) est de plus fortement corrélé à l'aperture des lèvres (Abry & Boë, 1986 ; Robert-Ribes et al., 1998). De plus, dans une étude de la perception visuelle du geste d'aperture dans des transitions [i#a] comparables aux nôtres, Cathiard (1994) a mis en évidence la synchronie du début des gestes pour les mouvements d'aperture des lèvres et d'abaissement de la mâchoire ainsi que la quasi-superposition du décours temporel de ces gestes. Le suivi du décours de l'aire intérolabiale nous semble donc pertinent pour étudier l'anticipation de hauteur.

Nous analysons donc dans cette étude les paramètres suivants synchrones au cours du temps : (1) le décours de l'aire intérolabiale, avec une information toutes les 20 ms ; (2) les positions en *x* et en *y* de la pastille sur le dos de la main (échantillonnées à 50 Hz) (3) et le signal acoustique échantillonné à une fréquence de 22050 Hz. Pour notre analyse, nous prendrons en compte uniquement la coordonnée *y* de la pastille pour la transition [i#y] et la coordonnée *x* pour la transition [i#a]. En effet, pour les séquences contenant la voyelle [y], le passage de la position « bouche » à la position « cou » se caractérise par un déplacement de la main essentiellement dans l'axe vertical. Inversement, pour la

transition [i#a], nous avons choisi de ne traiter que la coordonnée *x* car le déplacement se fait de la position « bouche » à la position « côté » et implique un mouvement essentiellement dans l'axe horizontal.

Sur chacun des signaux, nous repérons les événements temporels suivants (grâce au profil d'accélération ; voir section V.2.2.2) illustrés sur la Figure 45 pour une séquence [tami#yiiz] : pour les lèvres, L1 est le démarrage du geste labial de constriction ou d'aperture (selon le corpus), L2 est l'atteinte de la cible vocalique du [y] ou du [a] (dans les transitions [i#y] et [i#a]) et L3 est le démarrage du geste labial pour former le [i] suivant (dans les transitions [yi] et [ai]) ; pour la main, M1 est le début de la transition manuelle LPC à partir de la position « bouche » vers la position « cou » pour le [y] ou la position « côté » pour le [a], M2 l'atteinte de la position LPC cible (dans les transitions [i#y] et [i#a]) et M3 est le début du geste manuel vers la position « bouche » suivante codant le [i] ; finalement, sur le signal acoustique, A1 indique le début acoustique des voyelles [y] ou [a]. Nous repérons également la fin acoustique de la voyelle [i] dans les transitions [i#y] et [i#a] afin de pouvoir mesurer la durée de la pause acoustique. La position manuelle au cours du temps est calculée dans le repère indiqué sur la Figure 46. Ainsi dans ce repère, une augmentation des valeurs en y sur la Figure 45 correspond à un éloignement vertical de la main par rapport au point de référence sur la lunette : c'est précisément ce qui se produit quand on passe de la position « bouche » à la position « cou » pour coder la transition de la voyelle [i] à la voyelle [y].

Pour les transitions testées [i#y] et [i#a], nous étudions les intervalles temporels suivants :

- IO est la durée de la pause silencieuse (mesurée depuis la fin de la structure formantique de [i] jusqu'au début de la structure formantique voisée de [y]) ;
- L1L2 est la durée de la phase de constriction pour [y] ou d'aperture pour [a] aux lèvres (équivalent du *Time-Falling* de Abry & Lallouache, 1995b) ;
- L1L3 est la durée totale du geste labial, incluant la phase de tenue (équivalent du *Time-Falling* + *Hold* de Abry & Lallouache, 1995b) ;
- M1L1 est l'intervalle entre le début du geste de main et le début du geste des lèvres pour la voyelle ;
- M2L2 est l'intervalle entre la cible manuelle et la cible labiale vocalique ;
- M1M2 est la durée de la transition de main de la position « bouche » pour le [i] vers la position cible, soit le « cou » pour le [y] ou le « côté » pour le [a] ;
- M1M3 est la durée totale de la clé LPC, incluant la phase de maintien de la main en position cible.

Chaque intervalle est obtenu par soustraction des valeurs temporelles d'une paire d'étiquettes ; par exemple, L1A1= A1 – L1 (ms). Ainsi pour un intervalle XY, une valeur positive indique que l'événement X se produit temporellement avant l'événement Y.

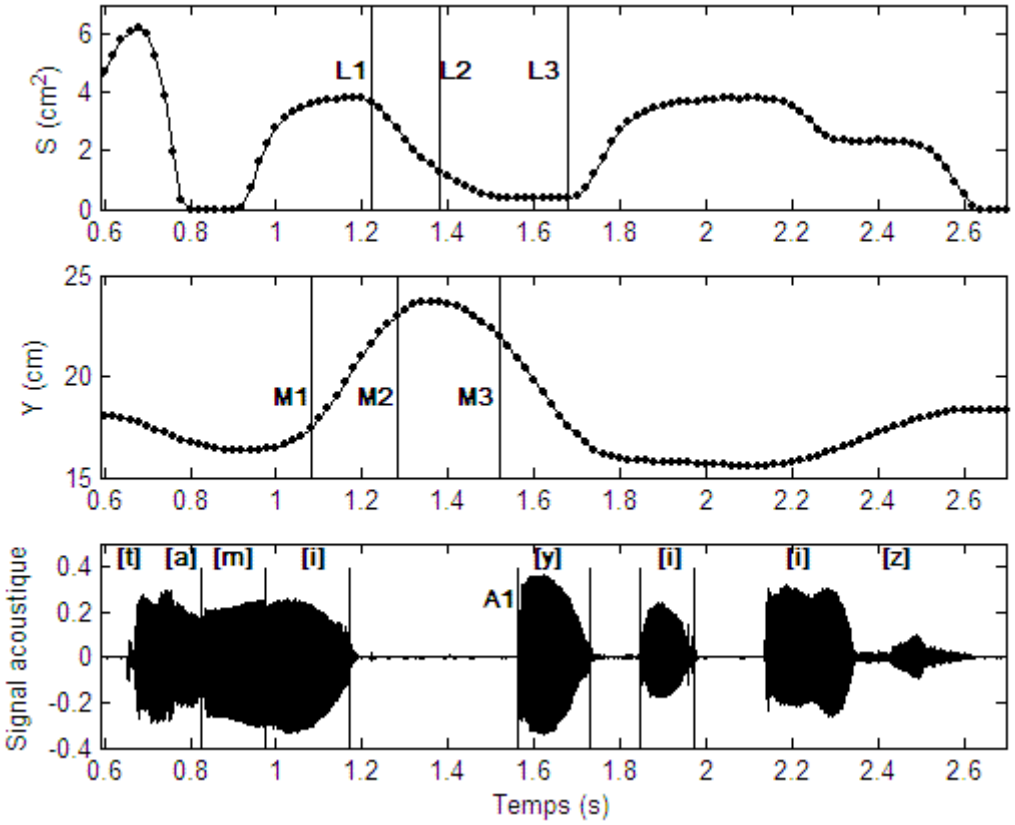


Figure 45. Tracé des différents signaux au cours du temps pour la séquence [tami#yiiz] (pause courte). De haut en bas : (1) décours de l'aire intérolabiale *S* (cm²) ; (2) position de la coordonnée *y* de la pastille sur le dos de la main (cm) ; (3) signal acoustique correspondant. Sur chacun des signaux sont superposés les événements temporels utilisés pour l'analyse. Afin de ne pas surcharger la figure, les intervalles temporels n'ont pas été indiqués ; ils se déduisent simplement grâce à la position des étiquettes.

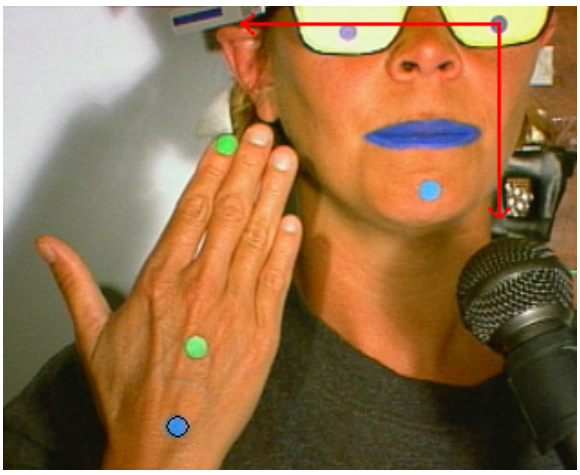


Figure 46. Photo de la locutrice-codeuse GB lors de l'enregistrement avec les positions des pastilles colorées sur la main et sur le visage et le repère utilisé.

VII.1.3. Résultats

Nous présenterons les résultats séparément pour les séquences contenant la voyelle [y] et celles contenant la voyelle [a].

VII.1.3.1. Anticipation d'arrondissement

Sur le plan articulatoire, pour l'ensemble des 24 réalisations de la séquence [tami#yiiz], toutes conditions de pause confondues, nous avons obtenu une valeur moyenne de l'aire intérolabiale de 2,69 cm^2 (s= 0,64 cm^2) pour les 24 cibles [i] et de 0,34 cm^2 (s= 0,06 cm^2) pour les 24 cibles [y], mesurées dans la transition [i#y]. Rappelons que ces valeurs d'aires sont classiques pour ce type de voyelle : la voyelle [i] est en effet caractérisée par un étirement important au niveau des lèvres et une grande aire intérolabiale alors que la voyelle [y] est caractérisée par une protrusion importante des lèvres et une petite aire intérolabiale.

Pour les transitions [i#y], la durée de la pause IO mesurée à partir du signal acoustique (depuis la fin du [i] jusqu'au début du [y]) est schématisée sur la Figure 47 pour les deux conditions de pause. L'intervalle de pause peut varier de 312 ms à 481 ms (m= 389 ms, s= 47 ms) en condition de petite pause et de 631 ms à 1452 ms (m= 961 ms, s= 239 ms) en condition de longue pause. La différence entre les deux conditions étant significative (|t|= 8,1 ddl= 22 p< .01), cela nous assure que la locutrice-codeuse GB a bien respecté la consigne. Il est à noter que ces durées importantes de pauses correspondent au rythme de parole assez lent de cette locutrice, que nous avions déjà noté lors des expériences 1 et 2. Dans le reste de nos analyses, nous ne distinguerons plus nos stimuli selon la condition de pause mais nous les étudierons en fonction de la durée de la pause IO.

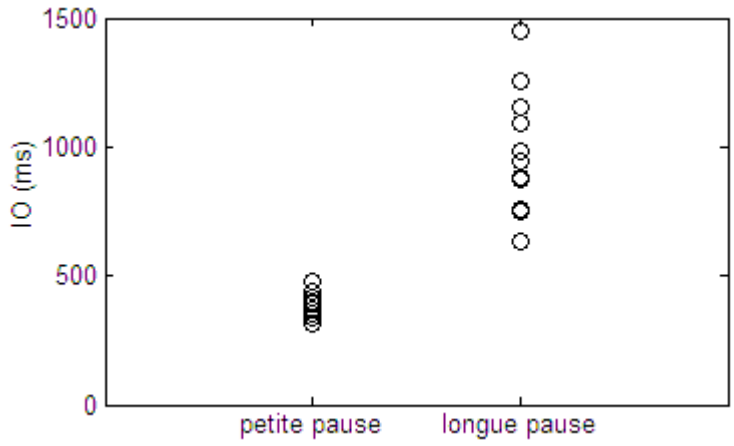


Figure 47. Durée de l'intervalle de pause IO pour l'ensemble des transitions [i#y]. Les séquences sont réparties en fonction de la consigne (petite pause – longue pause).

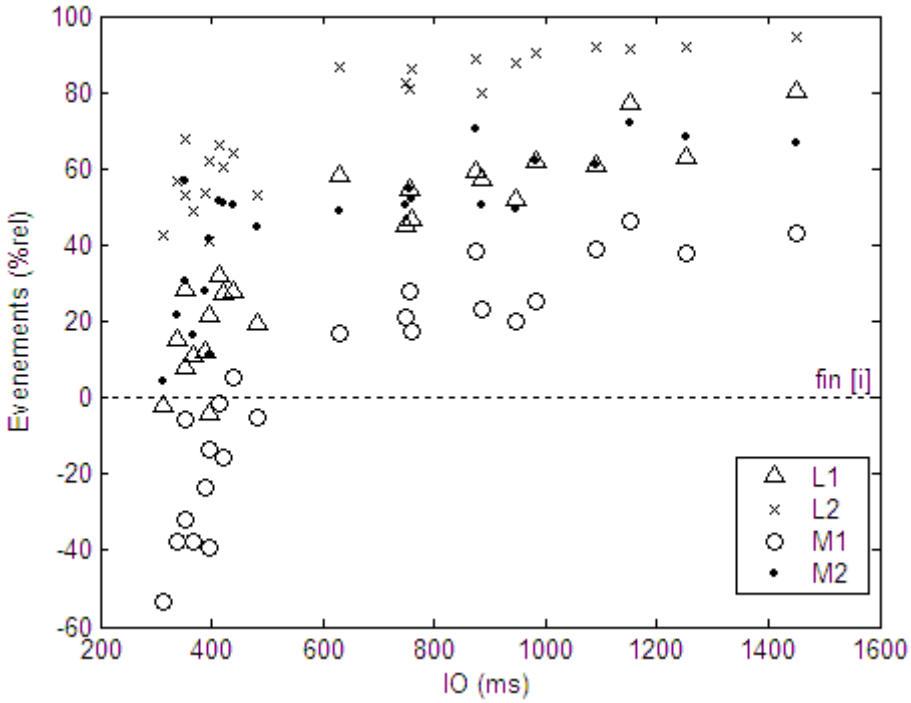


Figure 48. Organisation temporelle relative des événements manuels et labiaux dans la transition de [i] vers [y] pour la codeuse GB : M1 est le début du geste de transition manuelle, M2 est l'atteinte de la position du [y], L1 est le début du geste labial et L2 est la cible labiale vocalique. Les événements, repérés par rapport à la fin du [i], sont exprimés en pourcentages relatifs de la durée de l'intervalle de pause IO. 0% correspond donc à la fin du [i] et 100% indique le début du [y].

Nous avons sur la Figure 48 l'organisation temporelle relative des événements cinématiques manuels et labiaux en fonction de la longueur de la pause (IO), soit le phasage avec l'acoustique. Les différents événements sont référencés par rapport à la fin acoustique du [i] dans la transition [i#y] et ont été exprimés en pourcentages relatifs à la durée de IO. Nous pouvons voir sur cette figure que tous les événements se produisent avant le début acoustique du [y] (qui correspond à $100\%_{rel}$) ; ceci signifie clairement que, dans la transition [i#y] avec un intervalle de pause variable, les lèvres et la main anticipent toujours sur le son. Plus précisément, le début du geste manuel LPC est le plus en avance de tous les événements (les cercles représentant M1 sont les plus éloignés du début du [y]) : pour des pauses courtes (inférieures à 500 ms), la main peut initier sa transition avant même la fin acoustique du [i] (repérée à $0\%_{rel}$), soit bien avant la pause. Le positionnement relatif du départ de la main M1 (par rapport à la fin acoustique du [i]) augmente en fonction de la durée de la pause IO ; ainsi plus la locutrice-codeuse a de temps pour coder, plus elle va débuter son geste de main tardivement en timing relatif par rapport à la fin du [i], soit durant l'intervalle de pause. En ce qui concerne la fin du geste de main (M2) correspondant à l'atteinte de la position « cou » pour le [y], nous pouvons remarquer qu'elle se produit dans l'intervalle de pause et toujours avant la cible labiale (L2). Enfin, les lèvres arrivent en position cible pour la voyelle [y] (L2) avant son début acoustique. Cette avance est d'autant plus

importante en timing relatif que la pause est courte. Par rapport aux modèles de l'anticipation labiale, ces différents phasages indiquent que le geste labial peut démarrer dès la fin du [i], voire avant si la pause est courte, mais ne va pas forcément exploiter toute la durée de la pause lorsque celle-ci est longue.

La relation main-lèvres pour les débuts de gestes (M1L1) et les cibles (M2L2) est illustrée sur la Figure 49. Nous voyons clairement que plus l'intervalle de pause IO augmente, plus l'anticipation de la main sur les lèvres est importante aussi bien au niveau de l'initiation des gestes que des cibles manuelle et labiale (les coefficients de corrélations sont très forts : r= 0,87 pour M1L1 et r= 0,88 pour M2L2). Ceci indique que la main va profiter davantage du temps disponible pour anticiper sur les lèvres quand la durée de l'intervalle de pause est grande. Remarquons que dans tous les cas, la main est toujours en avance sur les lèvres (sur l'ensemble des séquences, les deux intervalles M1L1 et M2L2 sont toujours positifs).

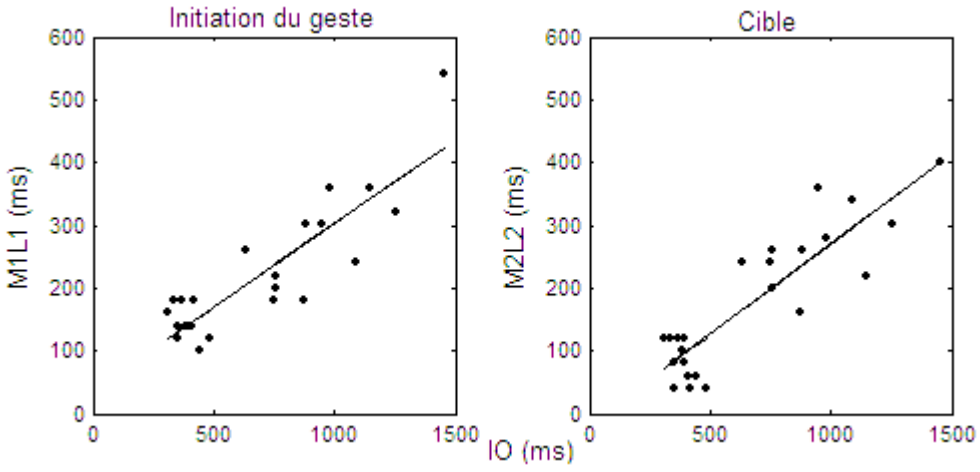


Figure 49. Coordination temporelle main-lèvres pour le début des gestes manuel et labial (M1L1, ms) et pour l'atteinte des cibles (M2L2, ms) en fonction de la durée de l'intervalle de pause IO (ms) dans les transitions [i#y]. Les valeurs positives de ces deux graphes indiquent que l'événement manuel se produit avant l'événement labial.

La Figure 50 nous donne les phases de mouvements labial et manuel en fonction de la durée de la pause ; pour le geste labial, nous considérons la durée de la phase de constriction (L1L2, *Time-Falling*) ainsi que le geste global d'arrondissement (L1L3, équivalent au *Time-Falling+Hold*) et pour le geste manuel, nous procédons de même en considérant la phase de transition M1M2 et la phase de transition + tenue (M1M3). Nous avons obtenu des corrélations toutes significatives (à $p < .01$) entre ces différentes phases et IO (voir les coefficients sur les graphes). De manière générale, le geste labial est plus bruité que le geste manuel, avec une variabilité observable surtout au niveau des pauses longues. Les valeurs de pente obtenues sont assez faibles (0,15 et 0,13) mais restent néanmoins significativement différentes de 0 et indiquent une relation linéaire positive entre les phases de

constriction et IO. Donc chez la codeuse GB, pour des transitions [i#y], la durée du geste d'arrondissement augmente linéairement en fonction de la durée de la pause, selon un coefficient d'expansion relativement bas pour la phase de constriction et de manière similaire pour le geste labial complet.

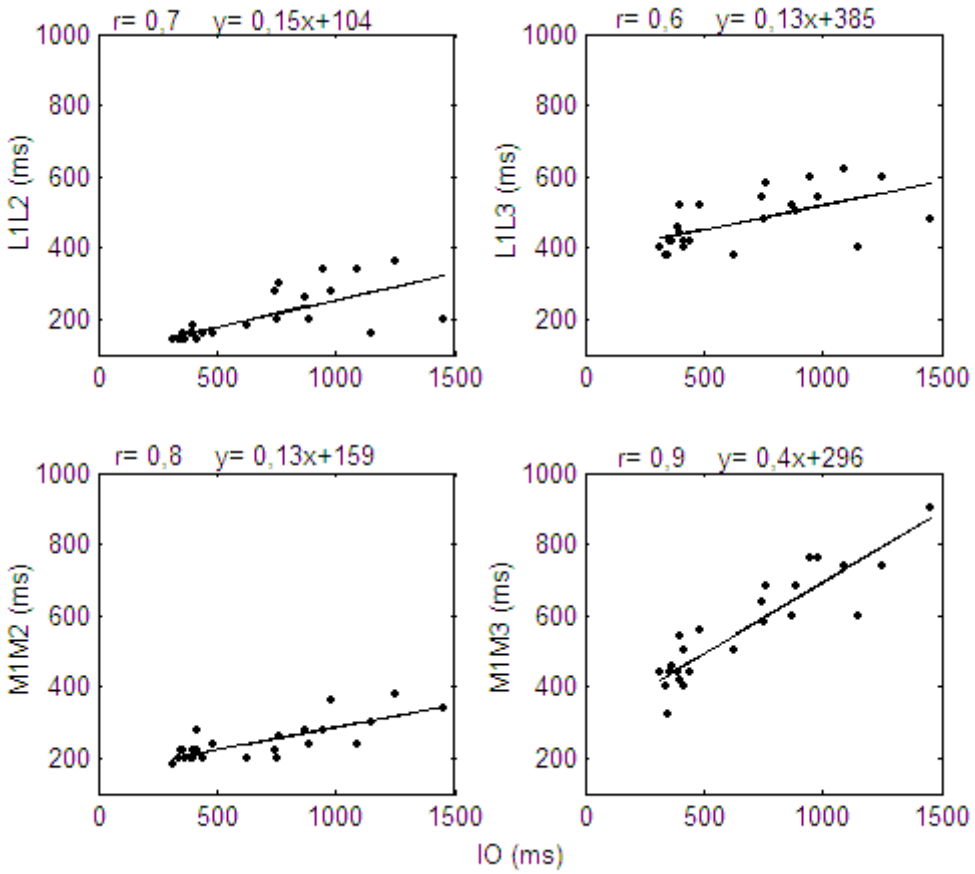


Figure 50. Corrélations entre l'intervalle de pause IO et les gestes labial et manuel pour la transition [i#y] : en fonction de IO, la phase de constriction L1L2, le geste global d'arrondissement L1L3, la transition de main M1M2 et la clé LPC globale M1M3. Au-dessus de chaque cadran sont indiqués les coefficients de corrélations (tous significatifs à p< .01) et les équations des droites de régression linéaire.

En ce qui concerne le geste manuel, il apparaît également que l'anticipation manuelle augmente linéairement en fonction de la durée de IO. Nous avons une croissance de la durée de transition (pente à 0,13 significative) qui indique que la codeuse va avoir tendance à effectuer une transition plus lente quand elle a plus de temps avec un coefficient d'expansion semblable à celui observé sur les lèvres. Contrairement à ce qui se passe pour les lèvres, on observe que la phase de tenue de la main est plus expansible que la phase de transition avec un coefficient d'expansion plus important (0,4) pour M1M3.

Nous obtenons donc chez cette locutrice des fonctions d'expansion très similaires à celles obtenues par Abry et al. (1996a) pour la modélisation du contrôle labial pour le geste vocalique d'arrondissement

avec des pauses intervocaliques plus ou moins longues : les coefficients d'expansion sont en effet très similaires (pour des pauses variant de 100 ms à 650 ms, ils avaient obtenu un coefficient d'expansion de 0,16). Nos données d'anticipation labiale peuvent donc être expliquées dans le cadre général du MEM (Abry & Lallouache, 1995a, 1995b).

VII.1.3.2. Anticipation de hauteur

Qu'en est-il de l'anticipation de hauteur mesurée sur le geste labial d'aperture dans des transitions [i#a], variant la longueur de la pause ? Un exemple de signaux analysés pour une séquence [tami#aiiz] avec petite pause est illustré sur la Figure 51 avec les événements temporels analysés.

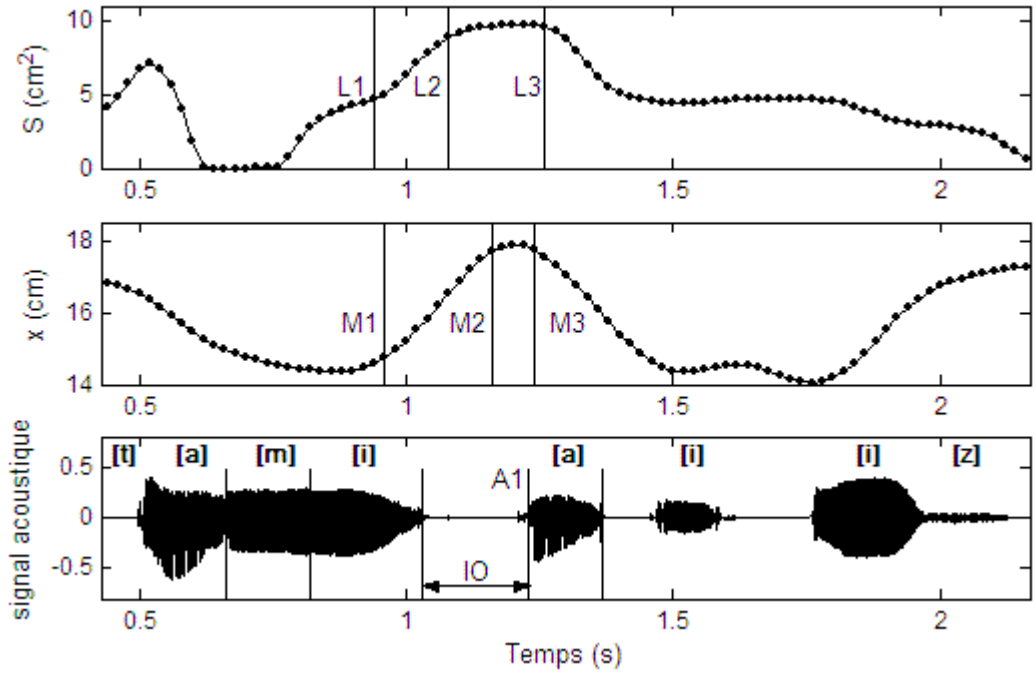


Figure 51. Tracé des signaux au cours du temps pour une séquence [tami#aiiz] avec petite pause. De haut en bas : (1) décours de l'aire intérolabiale *S* (cm²) ; (2) position de la coordonnée *x* de la pastille sur le dos de la main (cm) ; (3) signal acoustique correspondant. Sur chacun des signaux sont superposés les événements temporels utilisés pour l'analyse.

Sur le plan articulatoire, pour l'ensemble des 24 réalisations de la séquence [tami#aiiz], toutes conditions de pause confondues, nous avons obtenu une valeur moyenne de l'aire intérolabiale de 2,6 cm² (s= 0,6 cm²) pour les 24 cibles [i] et de 7 cm² (s= 1,1 cm²) pour les 24 cibles [a], mesurées dans la transition [i#a]. Nous retrouvons pour la cible [i] une valeur d'aire proche de celle obtenue dans le corpus précédent (2,7 cm²). La voyelle [a], quant à elle, est caractérisée par une grande aire aux lèvres due à l'aperture des lèvres et à l'abaissement important de la mâchoire.

Pour les séquences [i#a], la durée de la pause IO mesurée à partir du signal acoustique (voir Figure 52) peut varier de 182 ms à 400 ms (m= 260 ms, s= 66 ms) en condition de petite et de 452 ms à 1222

ms (m= 860 ms, s= 202,5 ms) en condition longue pause ; les deux conditions sont significativement différentes (|t|= 9,7 ddl= 22 p< .01) et certifient de la bonne application de la consigne par la codeuse. Remarquons dès à présent qu'elles sont en moyenne moins importantes en durées que les pauses du corpus précédent.

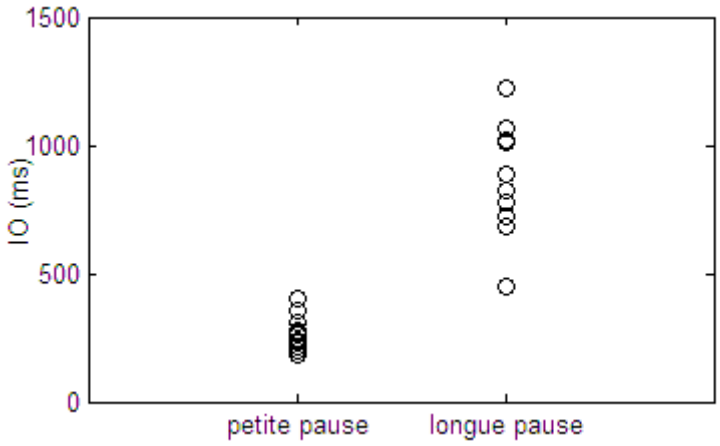


Figure 52. Durée de l'intervalle de pause IO pour l'ensemble des transitions [i#a]. Les séquences sont réparties en fonction de la condition (petite pause – longue pause).

En ce qui concerne l'organisation temporelle de ces événements en fonction de la durée de la pause IO, nous avons sur la Figure 53 le phasage des débuts et fins de gestes manuel et labial avec l'acoustique (les événements sont référencés par rapport à la fin acoustique du [i] dans la transition [i#a] et sont exprimés en pourcentages relatifs de la durée IO). Nous pouvons constater de manière assez remarquable une anticipation en pourcentage relatif assez importante sur le début acoustique de la voyelle [a], de tous les événements et cela pour toutes les réalisations (remarquons en particulier que, par rapport aux transitions [i#y], nous observons ici une anticipation relative plus importante de la cible labiale L2 ; Figure 48). Les débuts des gestes manuel et labial M1 et L1 peuvent se produire avant même la fin de la voyelle [i], donc durant sa production acoustique (c'est le cas pour les réalisations dont la durée de pause est inférieure à 600 ms). Ces initiations de geste semblent synchrones et affichent une claire superposition sur la figure, quelle que soit la durée de la pause ; plus précisément la Figure 54 montre une coordination temporelle entre ces deux événements qui ne varie pas en fonction de IO (r= -0,2 non significatif). En ce qui concerne les fins de geste, nous trouvons à ce niveau une dispersion plus importante à la fois pour la cible labiale du [a] et pour la cible manuelle. Ces deux événements sont dans tous les cas clairement en avance sur le son. De la même manière que pour les débuts de gestes, la relation main-lèvres au niveau des cibles ne varie cependant pas en fonction de la durée de la pause (r= 0,13 non significatif, Figure 54).

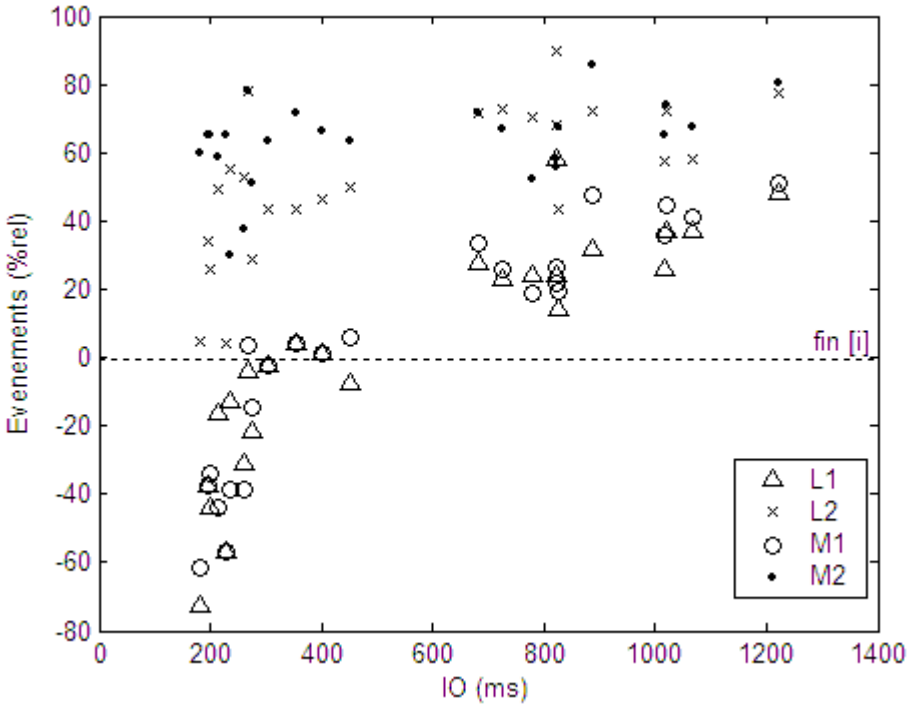


Figure 53. Organisation temporelle relative des événements manuels et labiaux dans la transition de [i] vers [a] pour la codeuse GB : M1 est le début du geste de transition manuelle, M2 est l'atteinte de la position du [a], L1 est le début du geste labial et L2 est la cible labiale vocalique. Les événements, repérés par rapport à la fin du [i], sont exprimés en pourcentages relatifs de la durée de l'intervalle de pause IO. 0% correspond donc à la fin du [i] et 100% indique le début du [a].

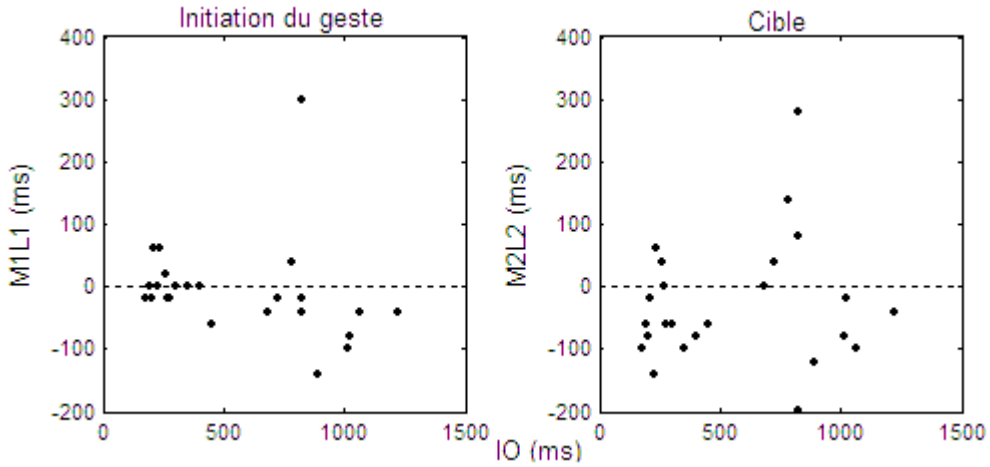


Figure 54. Coordination temporelle main-lèvres pour le début des gestes manuel et labial (M1L1, ms) et pour l'atteinte des cibles (M2L2, ms) en fonction de la durée de l'intervalle de pause IO (ms) dans les transitions [i#a]. Les valeurs positives de ces deux graphes indiquent que l'événement manuel se produit avant l'événement labial et inversement pour les valeurs négatives.

Nous trouvons donc un comportement manuel dans la coproduction avec le geste labial d'aperture bien différent de celui mis en évidence pour l'arrondissement. Il est important de préciser que, même si la main ne semble pas anticiper sur les lèvres, elle reste quand même largement en avance sur le début acoustique de la voyelle.

Comment pouvons-nous comprendre cette synchronisation avec les lèvres ? En ce qui concerne les débuts de gestes manuel et labial, leur synchronisation pourrait s'expliquer par le fait que la locutrice présente une anticipation labiale importante du début du geste : le geste d'aperture peut en effet être initié dans le [i] (pour des petites durées de IO) et dans tous les cas, durant la première partie de la pause acoustique (voir sur la Figure 53 le positionnement de L1). Ainsi la locutrice ne pourrait anticiper davantage le geste manuel sur les lèvres. En ce qui concerne l'atteinte des cibles labiale et manuelle, leur synchronisation pourrait s'expliquer par la position particulière qui code la voyelle [a], la position « côté ». En effet, pour cette voyelle le codeur va pointer sa configuration manuelle sur le côté du visage et non pas sur une partie bien précise de celui-ci comme il le fait quand il pointe pour [y] en allant toucher le cou au niveau de la pomme d'Adam. Ainsi il se peut que la main ait tendance à continuer sa course (puisqu'elle n'atteint pas un support fixe) en se synchronisant avec les lèvres pour ne s'arrêter qu'en même temps que la cible labiale. Nous pouvons voir sur la Figure 55 les ellipses de dispersion des positions cibles pour les deux types de transition [i#y] et [i#a]. La position « côté » du [a], qui n'a pas de support fixe sur le visage, est beaucoup plus dispersée que la position « cou » du [y].

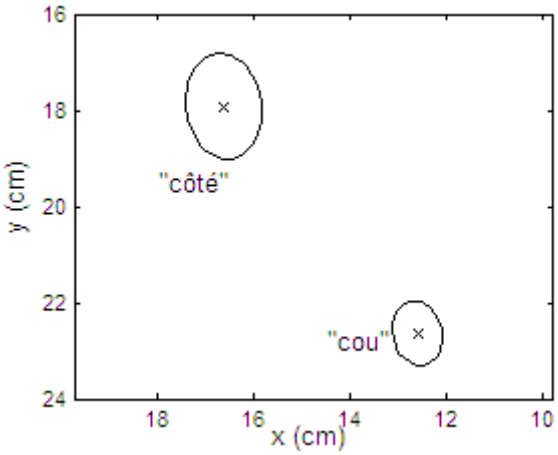


Figure 55. Ellipses de dispersion à un écart-type des positions cibles spatiales « côté » codant le [a] et « cou » codant le [y] pour la codeuse GB dans les transitions [i#y] et [i#a]. Les positions ont été calculées à partir de la position de la pastille du dos de la main (référencée dans le repère sur la lunette) au moment d'atteinte de position cible M2. La position moyenne de la pastille est indiquée dans chaque ellipse par une croix « x ».

En ce qui concerne l'évolution de la durée du geste d'aperture et de transition manuelle « bouche-côté » pour [i#a], nous avons tracé les phases des gestes (L1L2 pour la phase d'aperture, L1L3 pour le geste d'aperture complet incluant également la tenue de la cible labiale, M1M2, le geste de transition de main et M1M3, la clé manuelle complète incluant la phase de tenue en position cible) en fonction de l'intervalle de pause IO (voir Figure 56). Comme nous pouvions nous en douter (du fait de la synchronisation observée des deux gestes), les comportements labial et manuel sont très similaires.

Nous obtenons des durées de geste fortement corrélées à la durée de la pause, révélant donc assez peu de variabilité (voir les coefficients de corrélation sur les graphes). La durée des gestes des lèvres et de la main augmente linéairement en fonction du temps disponible selon un coefficient qui varie en fonction de la phase considérée. Nous trouvons pour le geste labial d'aperture une pente de 0,22 pour la phase de transition et une pente de 0,49 pour le geste complet indiquant que la stratégie de la locutrice va être de varier davantage la durée de la tenue de la cible labiale du [a]. En ce qui concerne le geste manuel, nous obtenons les mêmes tendances avec une phase de transition augmentant légèrement avec l'allongement de la pause tandis que la cible sera davantage tenue.

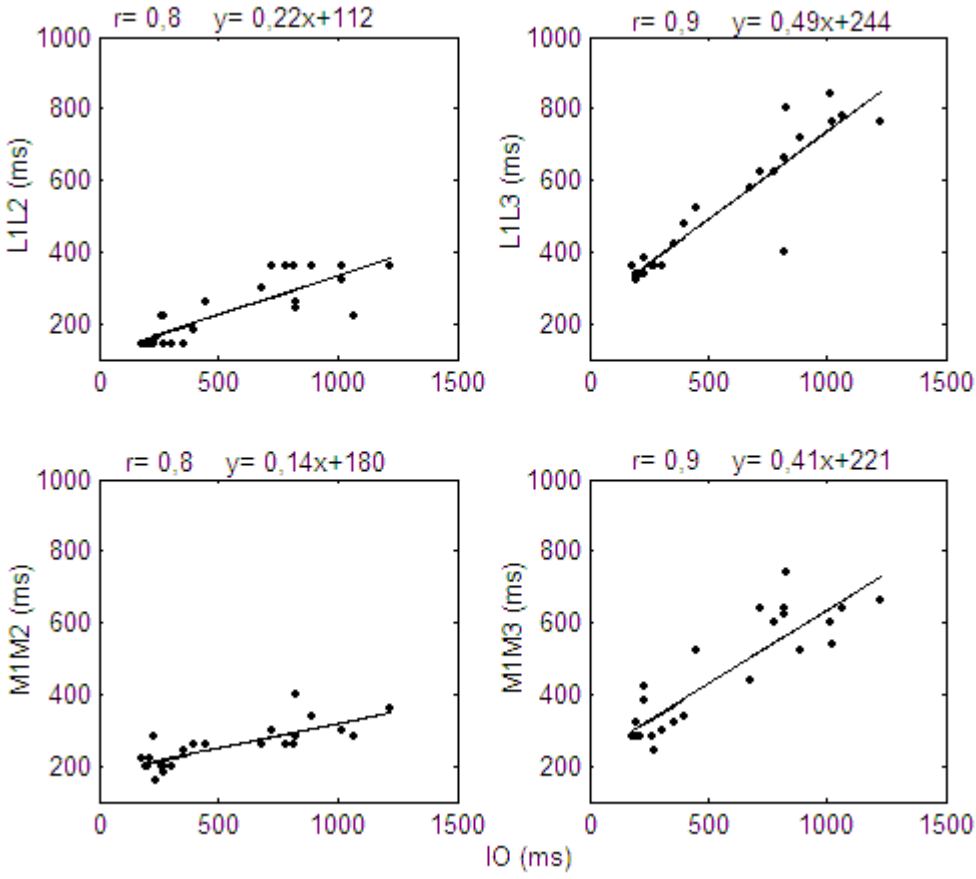


Figure 56. Corrélations entre l'intervalle de pause IO et les gestes labial et manuel dans la transition [i#a] : en fonction de IO, la phase de constriction L1L2, le geste global d'arrondissement L1L3, la transition de main M1M2 et la clé LPC globale M1M3. Au-dessus de chaque cadran sont indiqués les coefficients de corrélations (tous significatifs à p< .01) et les équations des droites de régression linéaire.

VII.1.4. Conclusion

En conclusion de cette étude sur l'anticipation d'arrondissement et de hauteur, nous avons observé pour les deux dimensions une anticipation claire des lèvres sur le son ; ainsi l'ajout du code manuel ne perturbe pas le comportement naturel de coarticulation de la parole. La durée du geste labial peut être

modélisée dans le cadre du MEM (Abry & Lallouache, 1996b) pour les deux dimensions vocaliques les plus visibles : elle a une expansion assez lente en fonction du temps disponible, avec un coefficient de 0,15 à 0,22 selon la dimension testée. En ce qui concerne le geste manuel LPC, nous avons confirmé l'anticipation de la cible manuelle sur la cible labiale, qui avait été obtenue pour des séquences CV, pour des transitions de voyelle à voyelle pour la dimension d'arrondissement : cette anticipation manuelle est préservée même dans le cas particulier où les lèvres sont fortement en avance sur le son. En ce qui concerne l'aperture, nous avons mis en évidence un patron de relations différent puisque nous observons une synchronie des gestes manuel et facial, que nous expliquons par la particularité de la position « côté » codant le [a] plutôt que par la spécificité du geste labial (pour une proposition des contraintes différentes sur le contrôle de ces deux gestes, voir la conclusion en fin de ce chapitre). Notons cependant que dans tous les cas, la main (ainsi que les lèvres) est bien en avance par rapport au début acoustique de la voyelle. La durée du mouvement LPC va également suivre une expansion linéaire en fonction de l'intervalle de pause : si la durée de la transition s'allonge, c'est surtout la durée du maintien de la main en position cible qui va être cruciale dans la coordination main-parole. Ainsi la main est clairement ancrée sur la parole : elle se coordonne avec le geste labial soit en anticipant, soit en se synchronisant avec lui.

VII.2. Expérience 5 : anticipation vocalique d'arrondissement au travers d'une séquence consonantique

Cette expérience teste l'anticipation du geste vocalique d'arrondissement pour la voyelle [y] au travers d'une suite de consonnes variant en nombre (de 0 à 4 consonnes) avec et sans code LPC. Ceci nous permettra de tester à la fois le geste labial et le geste manuel dans un contexte particulier. Pour les lèvres, il est connu qu'elles peuvent anticiper l'arrondissement de la voyelle plusieurs consonnes à l'avance (voir section II.3). Nous avons vu dans l'expérience précédente que l'arrondissement des lèvres était mis en place durant la pause silencieuse bien avant le début acoustique de la voyelle. En ce qui concerne le geste manuel, il est la plupart du temps en avance sur le geste labial : c'est le cas pour un codage CV classique mais aussi comme nous venons de le voir pour la dimension d'arrondissement testée lors de pauses acoustiques conséquentes. Que se passe-t-il quand une suite consonantique est insérée entre les voyelles ? Le principe de syllabification en CV du code LPC nécessite de coder les consonnes isolées avec la forme de main adéquate sur le « côté ». Ainsi pour des séquences incluant plus d'une consonne intervocalique, la transition manuelle d'une position à l'autre va être contrainte à un passage obligatoire en position « côté ». Que va-t-il advenir de l'anticipation de la main ? Le comportement labial va-t-il être modifié par l'ajout du code et de ses particularités ?

VII.2.1. Corpus

L'anticipation vocalique d'arrondissement est testée dans cette étude à l'aide de la transition vocalique de la voyelle étirée [i] vers la voyelle arrondie [y]. Cette transition est insérée dans la phrase porteuse du type « Ces deux Sc**ies u**tèrent ». Le geste d'arrondissement est étudié pour la transition simple [i#y], avec insertion d'une pause prosodique silencieuse intervocalique de longueur variable, et pour des transitions comportant jusqu'à quatre consonnes intervocaliques [ikssky]. Nous obtenons ainsi les transitions suivantes (voir la liste des phrases dans le Tableau 5) : [i#y], [iky], [isky], [iksky] et [ikssky]. Les séquences sont également enregistrées sans code LPC afin d'avoir un rythme naturel de parole. Nous voulions enregistrer un nombre égal de répétitions de chaque séquence avec et sans code LPC. Mais suite à des difficultés de notre codeuse à produire les séquences consonantiques complexes, ainsi qu'à des problèmes techniques de traitement d'images sur certaines séquences, nous avons finalement obtenu un nombre irrégulier de répétitions (voir Tableau 5) : au total, 25 transitions [iC_ny] codées et 15 transitions non codées.

Phrase	Transition étudiée	Nombre de séquences avec LPC	Nombre de séquences sans LPC
Ces deux scies utèrent	[i#y]	9	5
Ces deux scies kutèrent	[iky]	4	3
Ces deux sisses kutèrent	[isky]	3	2
Ces deux sixes kutèrent	[iksky]	7	3
Ces deux sixes skutèrent	[ikssky]	2	2
		Total= 25	Total= 15

Tableau 5. Liste des séquences enregistrées avec et sans code LPC pour les deux transitions vocaliques.

Le codage LPC de ces séquences implique une transition de main directe de la « bouche » vers le « cou » pour [i#y] et [iky] (la consonne [k] étant codée avec la voyelle) (Figure 57). Pour les autres séquences, les consonnes intermédiaires sont codées sur le « côté » (voir Figure 58 pour [sedøsiksskytɛʁ]).

Figure 57. Codage en LPC de la séquence « ces deux scies utèrent » [sedøsiytɛʁ].

Figure 58. Codage en LPC de la séquence [sikssky] extraite de la phrase « ces deux **sixes sku**tèrent ».

VII.2.2. Traitement des données

C'est de nouveau la codeuse GB qui a été enregistrée pour cette étude. Pour l'enregistrement, l'acquisition et le traitement des données, nous procédons comme précédemment (voir expérience 1 section V.2.2 et expérience 4 section VII.1.2).

Nous obtenons ainsi les signaux suivants synchrones au cours du temps : (1) le décours de l'aire intérolabiale, avec une information toutes les 20 ms ; (2) les positions en *x* et en *y* de la pastille sur le dos de la main (échantillonnées à 50 Hz) (3) et le signal acoustique échantillonné à une fréquence de 22050 Hz.

Les événements temporels suivants sont repérés (voir un exemple de signal sur la Figure 59 pour la séquence [sedøsiskytɛʁ]) : pour les lèvres, L1 est le démarrage du geste labial de constriction, L2 est l'atteinte de la cible vocalique du [y] et L3 est la fin de la tenue de l'arrondissement, soit le début du geste vocalique du [ɛ] ; pour la main, M1 est le début de la transition manuelle LPC vers la position « cou » pour le [y] (la position de départ pouvant être la « bouche » pour le [i] ou bien le « côté » dans les transitions à plus d'une consonne intervocalique), M2 l'atteinte de la position LPC cible (selon le profil des signaux de position en x et en y de la pastille du dos de la main, les repères manuels sont parfois étiquetés uniquement sur l'une des deux dimensions ; c'est le cas pour l'exemple présenté en Figure 59 où le déplacement horizontal est plus franc) ; finalement, sur le signal acoustique, A1 indique le début acoustique des voyelles [y]. Notons que, contrairement à l'étude précédente, nous n'avons pas repéré la fin de la tenue de la position manuelle (M3, soit le début de la transition vers le « menton » pour [ɛ]) car le déplacement de la pastille était peu marqué (Figure 59) et ne nous permettait pas de repérer fiablement cet événement. Nous repérons également la fin acoustique de la voyelle [i] dans la transition [i#y] afin de mesurer la durée de la pause acoustique ou de l'intervalle d'obstruence (IO). La position manuelle au cours du temps est calculée dans le repère indiqué sur la Figure 60. Ainsi dans ce repère, une diminution des valeurs en y correspond à un éloignement vertical de la main par rapport au point de référence sur la lunette et une diminution des valeurs en x correspond à un rapprochement horizontal de la main par rapport au point de référence (voir sur la Figure 59 le passage de la position « côté » au « cou » en x).

Les intervalles temporels suivants sont étudiés :

- IO est la durée de la pause ou selon le cas la durée de l'intervalle d'obstruence (déterminé à partir de la fin de la structure formantique du [i] et le début du [y] ou du [ø]) ;
- L1L2 est la durée de la phase de constriction ;
- L1L3 est la durée totale du geste de constriction, incluant la phase de tenue ;
- M1L1 est l'intervalle entre le début du geste de main et le début du geste des lèvres pour la voyelle ;
- M2L2 est l'intervalle entre la cible manuelle et la cible labiale vocalique ;
- M1M2 est la durée de la transition manuelle.

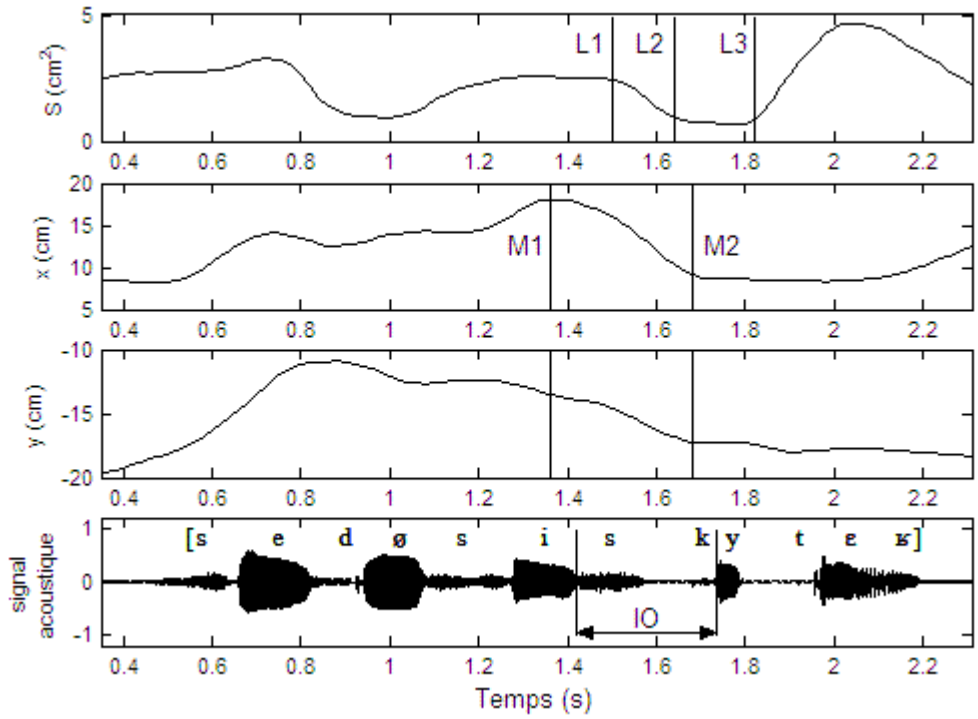


Figure 59. Tracé des différents signaux au cours du temps pour la séquence [sedøsiskytɛʁ] à deux consonnes. De haut en bas : (1) décours de l'aire intérolabiale *S* (cm²) ; (2) position de la coordonnée *x* de la pastille sur le dos de la main (cm) ; (3) position de la coordonnée *y* de la pastille sur le dos de la main (cm) et (4) signal acoustique correspondant. Sur chacun des signaux sont superposés les événements temporels utilisés pour l'analyse.

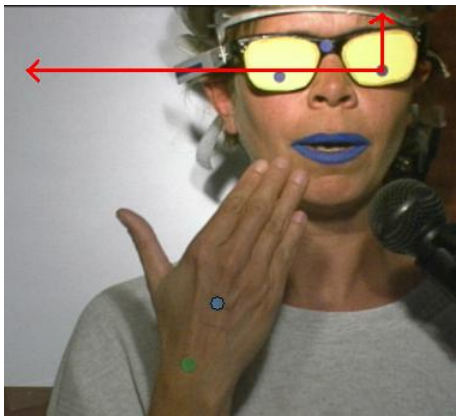


Figure 60. Photo de la locutrice-codeuse GB durant l'enregistrement avec indication de la pastille de la main et superposition du repère utilisé pour l'analyse des données.

VII.2.3. Résultats

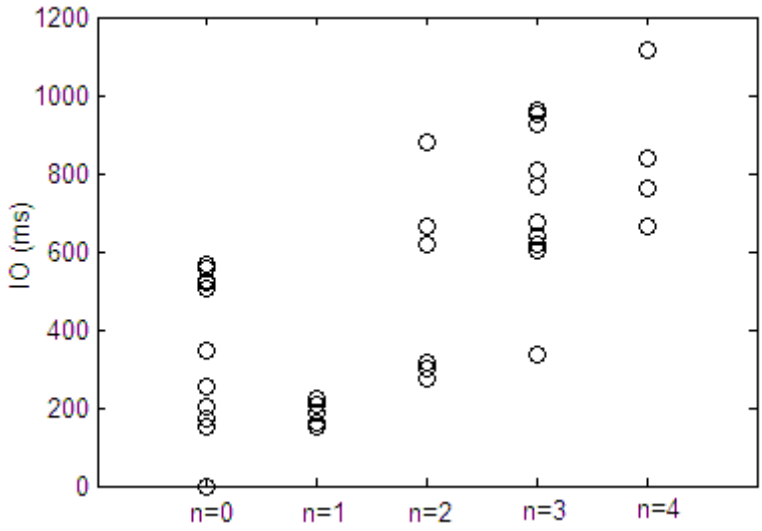


Figure 61. Durée de l'intervalle IO pour l'ensemble des séquences [iC$_n$y] (avec n= 0 à 4 consonnes).

Nous avons, sur la Figure 61, les durées d'intervalle IO selon le nombre de consonnes intervocaliques dans la transition de [i] vers [y] pour l'ensemble des 40 séquences de ce corpus (avec et sans code LPC). Pour les transitions simples [iy], nous avons au total trois séquences (deux avec code et une sans code) qui ont été réalisées sans pause entre les deux voyelles. L'intervalle de pause peut varier de 152 ms à 567 ms et l'intervalle d'obstruence de 149 ms (pour une séquence [iky]) à 1116 ms (pour une séquence [ikssky]). Nous pouvons observer une grande variabilité dans les données, en particulier pour les séquences à deux et trois consonnes intervocaliques ; ceci est lié au fait que la locutrice a eu un débit de parole assez lent et a parfois inséré des pauses dans les séquences consonantiques au niveau des jointures.

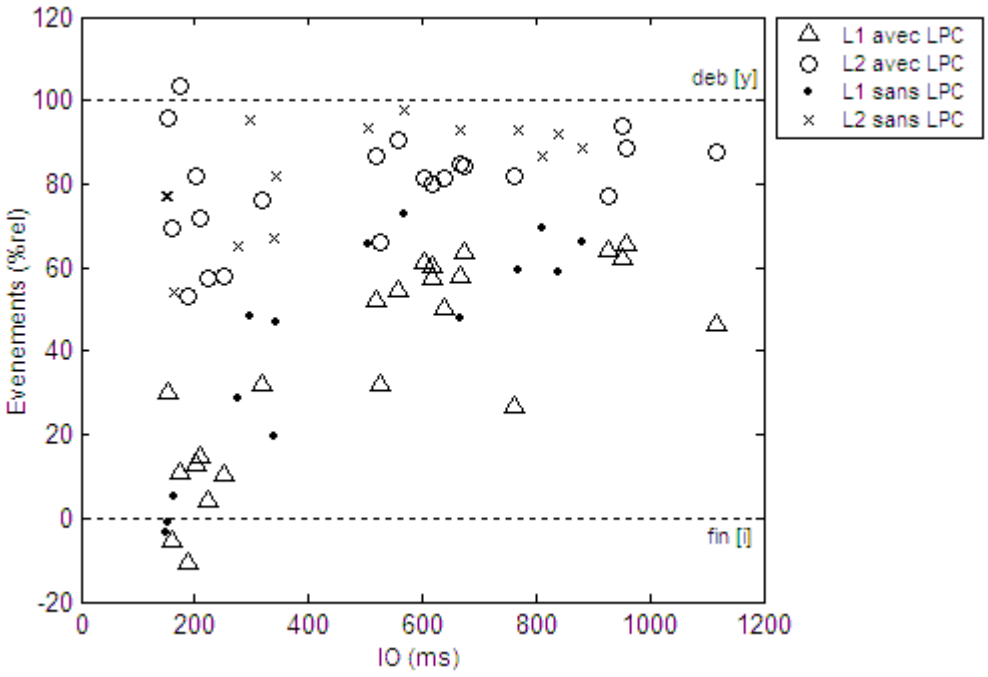


Figure 62. Organisation temporelle relative des événements labiaux dans la transition de [i] vers [y] pour les séquences avec et sans code LPC pour la codeuse GB : L1 est le début du geste de constriction et L2 l'atteinte de la cible labiale. Les événements sont référencés par rapport à la fin du [i] et sont exprimés en pourcentages relatifs de la durée de IO.

Pour l'ensemble des transitions [iC_ny] (avec n= 0 à 4 consonnes), nous pouvons voir sur la Figure 62, le phasage des événements labiaux L1 et L2, soit le début et la fin du geste de constriction, pour les séquences avec et sans code LPC, en fonction de la durée de l'intervalle IO (soit la durée de la pause silencieuse ou de l'intervalle d'obstruence). Ces événements sont référencés par rapport à la fin du [i] et sont exprimés en pourcentages relatifs de la base afin d'éviter le problème de la corrélation tout-partie (Benoît & Abry, 1986 ; représentation utilisée par Abry & Lallouache, 1995a). Les séquences [iy], ayant un intervalle IO nul (au nombre de trois), ne sont donc pas indiquées. Nous pouvons remarquer une grande similarité des séquences avec et sans LPC : en effet les deux jeux de données se recouvrent et indiquent des tendances similaires. Ainsi l'ajout du code LPC à la parole ne perturbe pas le geste naturel de constriction labiale dans les transitions de la voyelle [i] vers la voyelle [y] avec ou sans consonne intermédiaire. En ce qui concerne le timing relatif des événements, le geste de constriction labiale peut commencer avant même la fin de la structure formantique du [i] : c'est ce qui se produit pour quatre séquences avec un intervalle IO inférieur à 200 ms (il s'agit d'une séquence sans consonne et de trois séquences avec une consonne). Ce n'est pas toujours le cas car nous voyons que plus la durée de IO augmente, plus le geste labial commence tardivement dans IO par rapport à la fin du [i]. Ainsi le comportement de cette locutrice ne suit pas un modèle *Look-ahead* car la locutrice ne va pas forcément exploiter tout le temps disponible pour effectuer son geste labial. En ce

qui concerne la cible vocalique du [y], celle-ci va être mise en forme dans la majorité des séquences (sauf une) avant le début acoustique de la voyelle : la locutrice a tendance à anticiper davantage en timing relatif la cible labiale pour les séquences dont l'intervalle disponible est petit (IO inférieur à 400 ms).

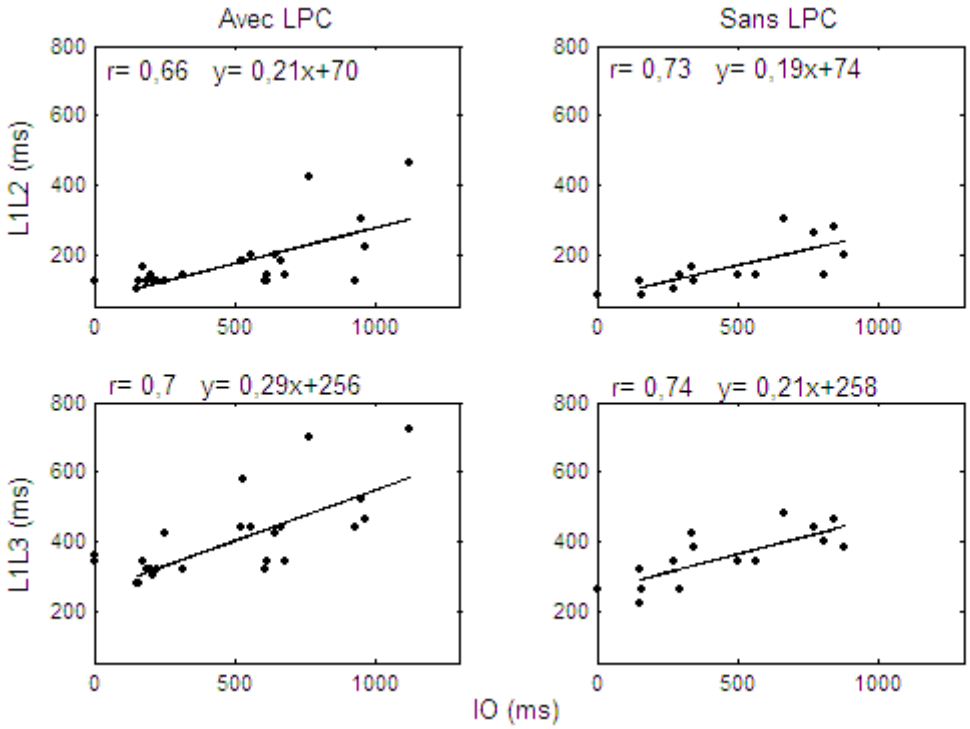


Figure 63. Phases du geste labial pour la voyelle [y] en fonction de l'intervalle IO pour les séquences avec (à gauche) et sans (à droite) code LPC : L1L2 est la phase de constriction (*Time-Falling*) et L1L3 la phase de constriction + la tenue de la cible labiale (*Time-Falling* + *Hold*). Les droites d'ajustement linéaire sont calculées sur toutes les données à l'exception de celles pour lesquelles IO= 0 ms. Les coefficients de corrélation et les équations des droites sont également indiqués.

Nous avons, sur la Figure 63, les phases du geste labial, soit, en adoptant les notations de Abry et Lallouache (1995b), L1L2 la phase de constriction ou *Time-Falling* et L1L3 le geste complet ou *Time-Falling+Hold*, en fonction de l'intervalle IO pour les séquences avec et sans code LPC. Comme nous pouvons le constater, les deux types de séquences donnent des gestes très similaires. De manière générale, les séquences avec LPC sont un peu plus bruitées que les séquences non codées (surtout pour les grandes valeurs de IO). Les droites d'ajustement linéaire (pour les séquences avec IO différent de zéro) sont superposées sur chaque graphe. Il apparaît que pour les deux phases L1L2 et L1L3, les pentes des droites ajustées pour les séquences avec et sans code LPC ne sont pas significativement différentes (au seuil α= 5%). Ainsi la locutrice GB semble bien adopter un même comportement labial pour le geste de constriction dans des transitions vocaliques de [i] à [y] (pouvant inclure jusqu'à quatre consonnes) avec ou sans code LPC. La durée du mouvement augmente de

manière linéaire en fonction de la durée de IO ; ainsi, nous pouvons modéliser le geste labial de constriction par le MEM (Abry & Lallouache, 1995b). Nous avons sur la Figure 64 le comportement labial général de la locutrice GB pour toutes les séquences mélangées. Les coefficients de corrélation sont significatifs en dépit d'une variabilité importante pour les valeurs les plus grandes de IO. Nous obtenons pour cette locutrice une fonction d'expansion pour le geste labial de constriction qui démarre à partir d'une constante d'exécution déterminée par les quelques séquences [iy] sans pause (sur les graphes, ces données correspondent aux séquences avec IO nul ; la durée du geste est en moyenne de 107 ms pour L1L2) et croît linéairement selon un coefficient de 0,2 en fonction de l'intervalle IO. Remarquons que la constante [iy] est proche de celle nécessaire pour effectuer les séquences [iky] à une consonne (en moyenne de 113 ms ; la valeur moyenne de IO pour ces séquences est de 182,5 ms ; voir en Annexe 6 le détail des séquences selon le nombre de consonnes intervocaliques). Notons de plus que le coefficient d'expansion obtenu dans ces données est proche de celui que nous avions déjà observé pour cette locutrice pour des transitions [i#y] avec pause plus ou moins longue (coefficient à 0,15). Il semble donc bien être propre à cette locutrice.

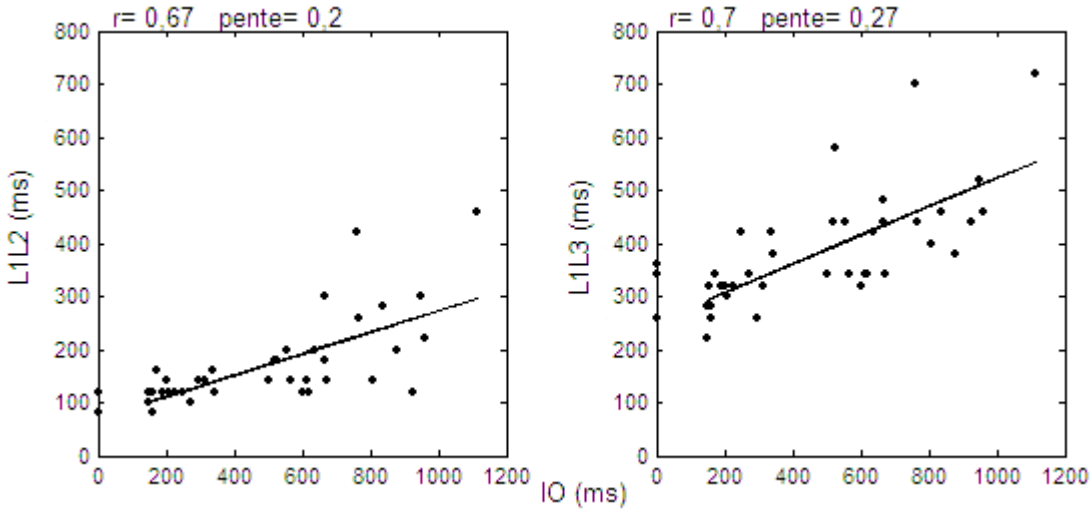


Figure 64. Phases de Time-Falling (L1L2) et de Time-Falling+Hold (L1L3) pour la transition de la voyelle [i] à la voyelle [y] en fonction de l'intervalle IO pour les séquences avec et sans LPC mélangées.

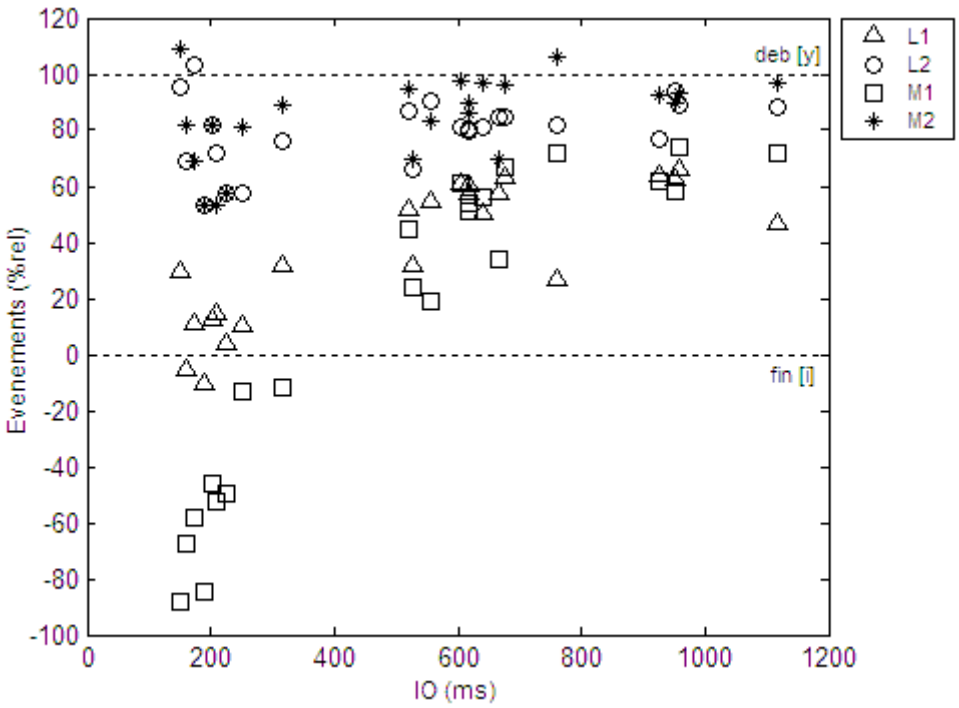


Figure 65. Organisation temporelle relative des événements labiaux et manuels dans la transition de [i] vers [y] pour les séquences avec code LPC pour la codeuse GB : L1 est le début du geste de constriction, L2 l'atteinte de la cible labiale, M1 le début du geste de transition manuelle vers le « cou » pour le [y] et M2 l'atteinte de la position cible LPC. Les événements sont référencés par rapport à la fin du [i] et sont exprimés en pourcentages relatifs de la durée de IO.

En ce qui concerne la main, nous avons l'organisation en timing relatif des événements manuels M1 et M2, soit le début et la fin de la transition manuelle pour coder la voyelle [y], dans le domaine de IO sur la Figure 65. Nous retrouvons les événements labiaux L1 et L2 correspondants pour les séquences avec code LPC. Nous pouvons remarquer un comportement manuel assez différent selon la longueur de l'intervalle IO disponible. En effet, le début du geste de main vers la position « cou » pour le [y] peut commencer largement durant la production acoustique de la voyelle précédente [i] pour les séquences dont l'intervalle IO est inférieur à 400 ms. Pour ces séquences, la main est également clairement en avance sur le début du geste labial (les carrés représentant les M1 sont tous en-dessous des triangles représentant les L1). Cette tendance s'atténue clairement pour les autres séquences ; nous voyons en effet que plus la durée de l'intervalle IO augmente, plus les M1 vont se superposer aux L1 jusqu'à être même en retard par rapport à eux pour les très grandes durées de IO. En ce qui concerne l'atteinte de la cible manuelle, nous observons pour toutes les séquences un comportement assez *inhabituel* : la position « cou » pour le [y] est atteinte quasiment en même temps que la cible labiale de la voyelle, en moyenne un peu avant son début acoustique (pour deux séquences cependant, la position de main est atteinte juste après le début de la structure formantique du [y]). Ce comportement s'explique sans

doute par la nécessité de maintenir les clés successives des consonnes en position « côté » avant de réaliser la syllabe CV porteuse de la voyelle [y].

Nous avons sur la Figure 66 le détail de cette relation main-lèvres, selon le nombre de consonnes intervocaliques, pour l'initiation des gestes manuel et labial (M1L1) et l'atteinte des cibles (M2L2) en fonction de l'intervalle IO. Nous retrouvons clairement la diminution de l'anticipation manuelle sur les lèvres quand IO augmente. En ce qui concerne l'initiation des gestes, la main devance les lèvres pour les séquences incluant jusqu'à deux ou trois consonnes (la frontière se situant dans les séquences à trois consonnes) ; au-delà, la main débute son geste de transition après le début du geste vocalique des lèvres. Rappelons que dans ce cas, les lèvres anticipent davantage sur le début acoustique de la voyelle. Ainsi, la main ne pourrait pas suivre une expansion du mouvement similaire à celle des lèvres. Le geste de constriction labiale peut en effet se déployer dans l'intervalle consonantique entre les deux voyelles [i] et [y] et nous avons vu que plus l'intervalle IO était grand, plus les lèvres profitaient de ce temps disponible pour débuter son mouvement. La main en revanche est contrainte par la structure du code LPC qui nécessite de coder les consonnes isolées (non CV) sur le « côté ». Il semble donc bien que ce soit ces consonnes intermédiaires qui limitent l'anticipation de la main jusqu'à inverser la tendance.

En ce qui concerne l'atteinte de la cible manuelle vocalique, la position de la main anticipe la cible aux lèvres dans une moindre mesure : son anticipation maximale est en effet de 100 ms pour une séquence [iy] sans pause mais aussi pour une séquence [isky] à deux consonnes intervocaliques dont l'intervalle d'obstruence est de 667 ms. Pour le reste, la main est plutôt synchrone et même en retard sur la cible labiale pour les séquences à plus de deux consonnes. Cela signifie que la main ne peut pas suivre la dynamique du geste labial dans les transitions vocaliques impliquant plusieurs consonnes intermédiaires ; même si elle démarre sa transition vers le « cou » en synchronie avec les lèvres, elle va atteindre la position en retard par rapport à la cible labiale. Ceci est certainement lié à la durée de la transition requise pour aller de la position « côté » à la position « cou ». La main qui est contrainte par la séquence consonantique ne peut pas tellement augmenter sa durée de mouvement comme lors des pauses acoustiques conséquentes dans l'étude précédente. Nous obtenons en effet des durées de transitions manuelles qui ne sont pas liées linéairement à l'intervalle IO ($r= 0,1$ non significatif au seuil $\alpha= 5\%$; voir Figure 67). Ainsi, comme pour les séquences CV chez les trois codeuses (section VI.2), les durées de transitions de main semblent ici assez constantes.

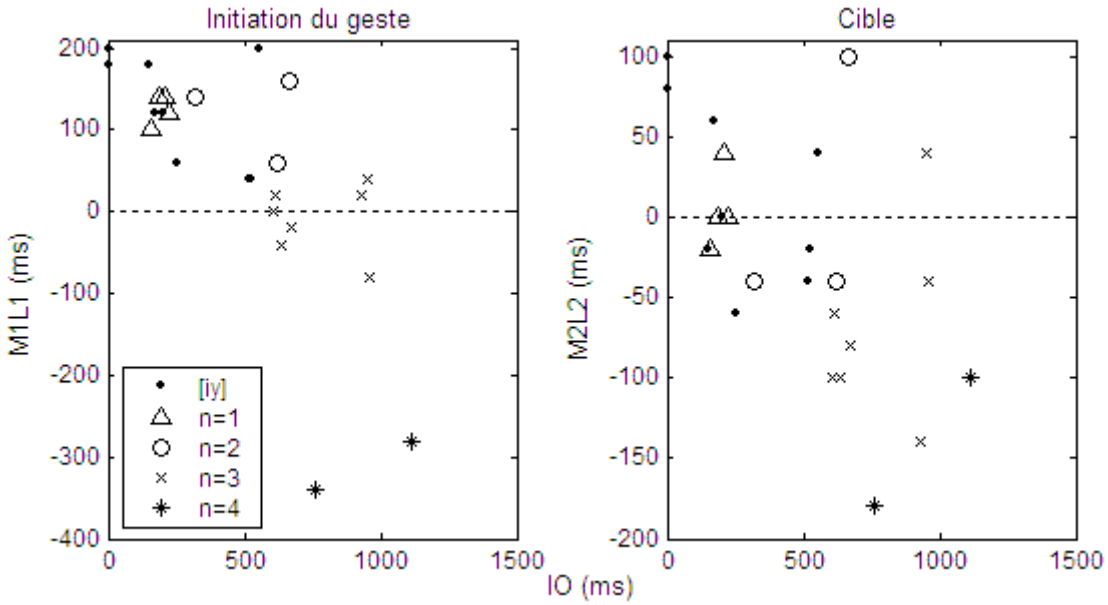


Figure 66. Coordination temporelle main-lèvres pour le début des gestes manuel et labial (M1L1) et pour l'atteinte des cibles (M2L2) en fonction de l'intervalle d'obstruence (ou de pause pour les séquences sans consonne) pour la locutrice-codeuse GB. Les séquences sont distinguées selon le nombre de consonne intervocalique (n= 0 à 4 consonnes) dans la transition [iC$_n$y]. Les valeurs positives de ces deux graphes indiquent que l'événement manuel se produit avant l'événement labial ; à l'inverse, les valeurs négatives indiquent que c'est l'événement labial qui se produit avant l'événement manuel.

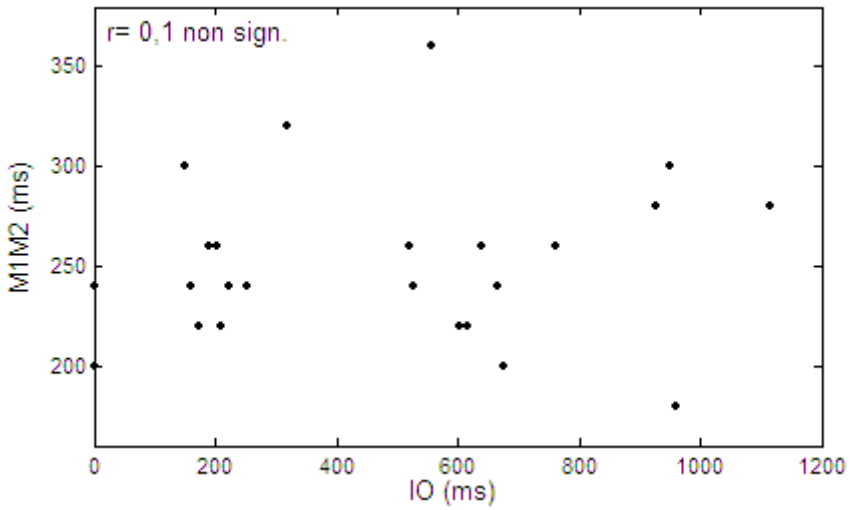


Figure 67. Durée de la transition manuelle M1M2 (ms) en fonction de l'intervalle IO (ms).

VII.2.4. Conclusion

En conclusion, nous avons montré dans cette étude sur le geste d'arrondissement, dans les transitions de la voyelle [i] à la voyelle [y] avec consonnes intervocaliques, que l'ajout du code LPC à la parole ne modifie pas (ou dans une moindre mesure) le comportement coarticulatoire labial naturel de la locutrice GB. Les lèvres anticipent le geste de constriction avant la production acoustique de la voyelle arrondie : le geste labial peut débuter dès la fin du [i] mais ne va pas forcément exploiter toute la

longueur de l'intervalle disponible pour le faire. Le comportement de cette locutrice ne suit donc ni le modèle *Time-Locked*, ni le modèle *Look-Ahead*. Il semble en revanche être adéquatement modélisé par le MEM de constriction (Abry & Lallouache, 1995b) : la durée du geste labial de constriction augmente en fonction de l'intervalle IO selon un coefficient d'expansion de 0,2. En ce qui concerne le geste manuel, il peut anticiper sur le geste labial pour les séquences incluant jusqu'à deux consonnes intervocaliques (soit une consonne seulement codée sur le « côté »). Au-delà, le nombre de consonnes à coder sur le « côté » contraint la main à se synchroniser avec les lèvres, voire même à prendre du retard sur elles quand la séquence est trop longue.

VII.3. Conclusion générale sur l'anticipation coarticulatoire

L'ensemble des deux études sur l'anticipation labiale au travers d'une pause prosodique silencieuse ou d'une séquence consonantique incluant jusqu'à quatre consonnes nous permet de conclure sur différents points.

L'anticipation labiale est préservée dans la parole codée. Ainsi l'ajout du code LPC ne modifie donc pas la coarticulation naturelle de la parole. On peut simplement remarquer que la coproduction parole-LPC contraint la parole à un ralentissement – ainsi que nous l'avions déjà remarqué pour des séquences syllabiques CV – ce qui mène à la production de longs intervalles IO par la locutrice GB. L'anticipation labiale augmente avec l'intervalle IO (de pause ou consonantique) et peut être adéquatement modélisée par le Modèle d'Expansion du Mouvement (MEM, Abry & Lallouache, 1991, 1995a, 1995b ; Abry et al, 1996a), avec un coefficient d'expansion du mouvement relativement constant chez cette locutrice d'un corpus à l'autre.

Rappelons que nous avions décidé d'explorer les deux dimensions d'arrondissement et d'aperture car ce sont les deux dimensions vocaliques visibles, pour le français, comme pour toutes les langues du monde. Il n'y a a priori aucune raison pour que le contrôle de ces deux dimensions soit le même et c'est bien d'ailleurs ce que nous trouvons. En effet, nous avons pu remarquer que le geste d'aperture pouvait démarrer très tôt, parfois nettement dans la voyelle [i] comme dans le cas de pauses courtes et dans tous les cas, dans la première partie de la pause acoustique (Figure 53 p. 22). Alors que nous avons observé, pour la dimension d'arrondissement, une anticipation moins précoce qui ne pénètre jamais la voyelle [i] mais se déploie dans la pause (Figure 48 p. 22). En ce qui concerne le geste manuel, on observe que la main s'adapte au timing de la parole, soit en étant synchrone avec les lèvres (cas de la dimension d'aperture), soit en étant en avance (dimension d'arrondissement) comme elle l'était déjà lors de séquences CV.

Quelles sont les différences essentielles entre ces deux gestes qui peuvent expliquer ces différences de contrôle ? Premièrement, du point de vue de la parole, l'essentiel de la mise en forme de l'aire aux lèvres pour le [a] est imputable au mouvement mandibulaire. Rappelons que cet articulateur joue à la fois un rôle pour l'ouverture du conduit vocal et un rôle de porteuse. Il est plutôt difficile de séparer la contribution des deux : sans faire appel aux variables intra-buccales (Vilain, 2000), on sait depuis Tuller et Kelso (1984) que l'essentiel de la contribution propre des lèvres est au mieux une petite élévation de la lèvre supérieure. Deuxièmement, en ce qui concerne le codage manuel, la cible du [a] (en position « côté ») ne vise pas au contact contrairement à [y]. Ces deux contraintes peuvent à notre avis très bien expliquer le comportement spécifique du geste de transition observé pour la dimension d'aperture. Des expériences mettant en évidence ces différences de contrôle, par exemple par perturbation de la mandibule et/ou du bras, restent bien évidemment à mener pour savoir laquelle de ces deux contraintes est la plus importante.

Dans le cas particulier de l'anticipation d'arrondissement avec une ou plusieurs consonnes, nous avons mis en évidence que l'anticipation manuelle sur la cible labiale vocalique – que nous avions mise en évidence dans le cas de syllabes CV – est bien présente tant que l'intervalle consonantique n'excède pas deux consonnes. Au-delà, la main peut se synchroniser, voire prendre du retard sur les lèvres, ce que nous expliquons par la nécessaire tenue de la main en position « côté » pour coder les consonnes à l'isolé. Ce n'est donc que sous la contrainte d'une position particulière de codage que la main se dégage du timing de la parole, mais sans néanmoins le perturber.

CHAPITRE VIII.

Expérience 6 : perception de syllabes CV codées

Cathiard M.-A., Attina V., Abry C. & Beautemps D. (2004).
La Langue française Parlée Complétée (LPC) : sa coproduction avec la parole et l'organisation temporelle de sa perception.
Revue PArole, n° 29/30/31, N° spécial :
« Handicap langagier et recherches cognitives : apports mutuels »,
J.-L. Nespoulous & J. Virbel (Eds.)

Le patron de coordinations qui ressort clairement de nos études sur la production du code LPC pour des syllabes CV est bien la synchronisation des deux informations consonantiques et l'anticipation de la position manuelle sur la cible labiale pour la voyelle. Cette anticipation temporelle est cruciale à prendre en compte pour aborder la question de l'intégration perceptive des informations manuelle et labiale. Il est en effet probable que le contrôle de la production de la parole avec LPC impose son organisation temporelle au traitement perceptif du code LPC. Le fait que la position manuelle vocalique soit nettement en avance sur la forme labiale vocalique peut laisser penser que la main propose, au sujet décodeur, par le biais de sa position, un premier ensemble de possibilités pour l'identification de la voyelle, possibilités qui seront restreintes à une solution unique lorsque la forme labiale sera donnée. La mise en forme de la main complètement réalisée en début de consonne pourra elle aussi selon le même système, mais probablement avec un délai moindre, pré-spécifier la consonne articulatoire. Dans les deux cas – identification de la consonne et identification de la voyelle –, l'idée force est que la main ne viendrait pas désambiguïser les formes labiales consonantiques et vocaliques, mais qu'elle proposerait par avance un sous-ensemble de formes possibles.

Comment la coordination main-lèvres est-elle traitée par le sujet sourd qui perçoit la parole codée ? Est-il capable de traiter l'information apportée par la main avant l'information portée par les lèvres ou attend-il pour intégrer les deux informations ? Autrement dit, un sujet sourd décodant la LPC va-t-il exploiter l'anticipation observée de la main sur les lèvres ? Si oui, on devrait observer une identification correcte de la position de la main plus précoce par rapport à l'identification correcte de la voyelle, et une identification quasi simultanée de la clé et de la consonne. Nous nous proposons de tester le décodage temporel de séquences syllabiques CV articulées et codées, c'est-à-dire la prise en compte perceptive de l'information manuelle et de l'information labiale dans son déroulement temporel, en utilisant un paradigme de *gating* (Grosjean, 1980, 1996), qui consiste à tronquer en différents points nos séquences. Cette technique nous permettra d'obtenir une identification pas à pas, au fur et à mesure du dévoilement progressif des informations manuelles et labiales.

Avant de présenter l'expérience perceptive proprement dite, nous allons dans un premier temps nous attarder sur l'analyse des séquences telles qu'elles ont été produites par la codeuse GB ; ceci nous permettra de confirmer le patron de coordinations main-lèvres-son déjà établi et nous donnera des repères temporels pour le découpage de nos séquences.

VIII.1. Corpus

Nous avons constitué un corpus de logatomes du type [my.ty.ma.CV.ma], avec C= [p, d, k, v] et V= [ø, ɛ̃, ɛ, ɔ] (par exemple, « mutumapeuma » [my.ty.ma.pø.ma]). Nous rappelons que ces consonnes sont codées en LPC par les configurations de main n°1 et 2 (voir Figure 68) et les voyelles par les positions « pommette » et « menton ».

configuration 1 [p, d, ʒ] | configuration 2 [k, v, z] | configuration 5 [m, t, f] | « pommette » [ø, ɛ̃] | « menton » [ɛ, ɔ, u] | « côté » [a, o, œ]

Figure 68. Clés consonantiques et positions codant les consonnes et voyelles utilisées dans le *corpus gating*.

Les phonèmes supplémentaires [ʒ] et [z] des clés 1 et 2 n'ont pas été retenus afin de ne pas multiplier le nombre de séquences qui seront testées dans l'expérience perceptive. La voyelle [u] de la position « menton » a également été éliminée en raison de son anticipation articulatoire trop importante qui serait facilement repérable visuellement (c'est également pour cette raison que nous n'avons pas retenu la position « cou » qui contenait le [y]). Ainsi, nous opposons pour la syllabe CV, deux clés consonantiques, assez proches au niveau de la configuration des doigts (de façon à ne pas deviner par avance la forme finale de la main), et deux positions spatiales, choisies en raison de leur clarté articulatoire correspondante (la position « bouche » a été éliminée principalement afin d'éviter tout problème ultérieur d'analyse de la forme labiale, la main pouvant occulter les contours). La syllabe cible est insérée dans le flux de parole, de façon à éviter toute influence de la prosodie finale, encerclée de part et d'autre par la syllabe [ma] : ainsi la main, en configuration ouverte (n°5), part de la position « côté » vers la position cible (« menton » ou « pommette ») tout en formant la clé correspondante (configuration 1 ou 2) et revient à la position « côté » en configuration ouverte (notons que la clé et la position du [ma] sont bien distinctes de celles testées pour la syllabe CV cible). Durant le test de perception, la syllabe cible CV sera de cette façon toujours précédée d'un même contexte. En combinant les différentes possibilités, nous obtenons donc 16 séquences différentes (Tableau 6), qui seront répétées trois fois par notre locutrice ; ce qui donnera un total de 48 séquences.

[my.ty.ma.dɛ.ma]	[my.ty.ma.kɛ.ma]	[my.ty.ma.pɛ.ma]	[my.ty.ma.vɛ.ma]
[my.ty.ma.dɛ̃.ma]	[my.ty.ma.kɛ̃.ma]	[my.ty.ma.pɛ̃.ma]	[my.ty.ma.vɛ̃.ma]
[my.ty.ma.dɔ.ma]	[my.ty.ma.kɔ.ma]	[my.ty.ma.pɔ.ma]	[my.ty.ma.vɔ.ma]
[my.ty.ma.dø.ma]	[my.ty.ma.kø.ma]	[my.ty.ma.pø.ma]	[my.ty.ma.vø.ma]

Tableau 6. Liste des logatomes de type [my.ty.ma.CV.ma] du *corpus gating.*

VIII.2. Etude préliminaire : coordination main-lèvres-son

VIII.2.1. Acquisition et traitement des données

La locutrice-codeuse GB a été enregistrée pour cette étude dans les mêmes conditions expérimentales que dans les expériences précédentes (voir expérience 1, section V.2.2 p. 22). La même méthode de traitement a été utilisée, nous fournissant ainsi pour chaque séquence, quatre signaux synchrones au cours du temps : le décours temporel de l'aire aux lèvres (échantillonné à 50 Hz), les positions horizontales et verticales de la pastille sur le dos de la main (échantillonnées à 50 Hz) et le signal acoustique (44100 Hz).

Les coordonnées en x et y de la pastille du dos de la main sont calculées par rapport au point de référence placé sur les lunettes que porte le sujet, dans le repère indiqué sur la Figure 69. De cette façon, lorsque la main effectue un mouvement vertical de bas en haut, les valeurs de y augmentent. Lorsque la main effectue un mouvement horizontal en direction du centre du visage, les valeurs de x diminuent. Nous pouvons voir un exemple des signaux analysés sur la Figure 70, pour laquelle la main passe de la position « côté » à la position « pommette » pour coder la syllable cible [pɛ̃] de la séquence [mytymapɛ̃ma]. On remarquera que dans ce cas, la variation en x est peu significative : ainsi pour l'étiquetage des signaux, c'est surtout la variation en y qui est prise en compte.

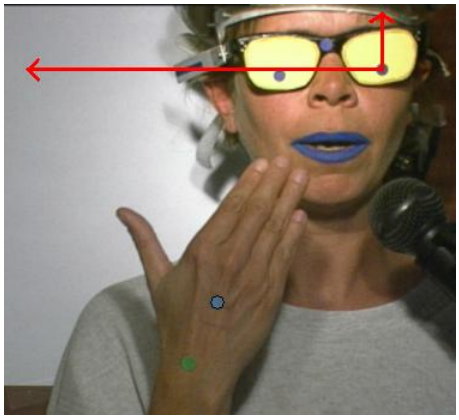


Figure 69. Photo de la locutrice-codeuse GB durant l'enregistrement avec indication de la pastille de la main et superposition du repère utilisé pour l'analyse des données.

Les événements temporels suivants ont été repérés (voir Figure 70) : pour chaque séquence, [mytymaCVma], A1 est le début acoustique de la consonne C, L1 est le début du geste labial pour former la voyelle V, L2 est la cible labiale de la voyelle, M1 est le début de la transition de la main vers la position codant la voyelle V (soit le « menton » ou la « pommette ») et M2 est l'atteinte de la position LPC cible, ces événements cinématiques étant repérés à l'aide du profil d'accélération (voir section V.2.2.2).

A partir de ces événements temporels, les caractéristiques de production étudiées sont les intervalles suivants : M1A1, entre le début du geste de main et le début de la consonne ; A1M2, entre le début acoustique de la consonne et l'atteinte de la position cible LPC ; M1L1, entre le début du geste de main et le début du geste labial vocalique ; M2L2, entre la cible LPC et la cible labiale de la voyelle et M3L2, entre le début du geste de main vers la position « côté » (transition de retour) et la cible de la voyelle. La durée de la syllabe cible CV est également calculée.

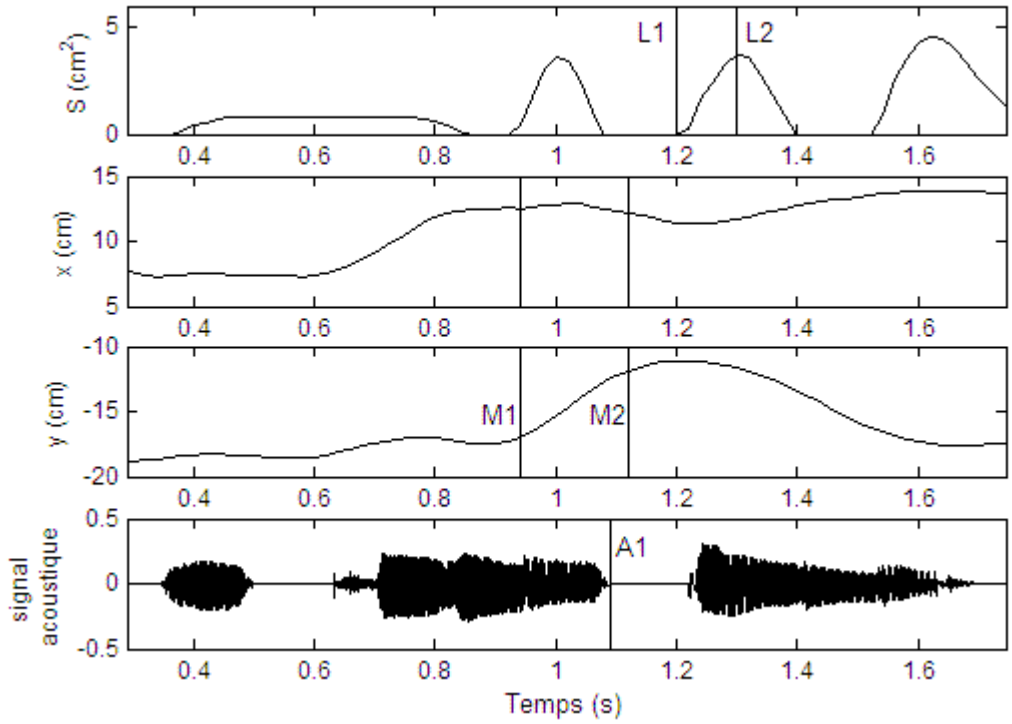


Figure 70. Tracé des différents signaux au cours du temps pour la séquence [mytymapɛ̃ma]. De haut en bas : (1) décours temporel de l'aire aux lèvres S, (2) position de la coordonnée x de la pastille sur le dos de la main, (3) position de la coordonnée y de la pastille et (4) signal acoustique correpondant. Sur chacun des signaux sont superposés les événements temporels repérés pour l'analyse de la coordination main-lèvres-son.

VIII.2.2. Résultats et conclusion

Nous avons observé une durée moyenne de 304 ms (s= 88 ms) pour la durée de la syllabe CV cible. En ce qui concerne la coordination main-son, nous avons obtenu une durée de 139 ms (s= 48 ms) pour M1A1 et de 59 ms (s= 44 ms) pour A1M2 : la main débute donc en moyenne son geste de transition bien avant le début acoustique de la consonne et atteint la position cible LPC durant la première partie de celle-ci (la durée de la consonne s'élevant en moyenne à 153 ms, s= 71). En ce qui concerne la coordination main-lèvres, nous avons obtenu une durée moyenne de 259 ms (s= 113 ms) pour M1L1, ce qui signifie que la main débute son mouvement avant celui des lèvres, et une moyenne de 155 ms (s= 112 ms) pour M2L2 indiquant que la position vocalique LPC est atteinte en avance par rapport à la

réalisation de la cible vocalique aux lèvres ; le geste de main anticipe donc clairement sur le geste labial pour la voyelle. Finalement, la valeur moyenne de -31 ms (s= 60 ms) pour M3L2 indique que la main repart vers la position « côté » un peu après la réalisation de la cible labiale, soit durant la production acoustique de la voyelle.

Sans strictement comparer les durées, nous pouvons tout de même tenter de rapprocher les résultats obtenus ici avec ceux des expériences précédentes, en particulier ceux de l'expérience 2 (section V.3.3.2 p. 22), obtenus chez la même codeuse pour des séquences syllabiques comparables, variant à la fois la position et la clé consonantique. Le rythme syllabique est assez proche dans ces deux expériences : la durée moyenne de syllabe était en effet de 316 ms. Concernant le patron de coordination, nous observons ici, en durées absolues, une anticipation manuelle qui semble moins importante par rapport au début acoustique de la consonne mais cependant du même ordre que celles observées chez les trois autres codeuses. La position vocalique LPC est atteinte dans tous les cas durant la première partie de la consonne acoustique. Par rapport aux lèvres, nous obtenons des durées très comparables, supérieures en moyenne à 150 ms. Nous avons donc une claire anticipation de la position de main sur la cible labiale vocalique. La main repart ensuite vers la position suivante durant la production acoustique de la voyelle, soit avant, soit après sa réalisation aux lèvres.

Nous avons donc obtenu sur ces données un patron de coordination main-lèvres-son très comparable à celui que nous avions déjà observé pour la même codeuse, mais aussi pour les trois autres codeuses AM, SC et RV. Les durées moyennes sont certes différentes, il n'en reste pas moins que le patron général est conservé : la main débute son mouvement de transition bien avant le son et atteint la position vocalique LPC en début de consonne, soit bien avant la formation de la voyelle aux lèvres. En ce qui concerne la formation de la clé consonantique, nous n'avons pas pu dans cette étude analyser les mouvements des doigts (le port d'un gant de données par la locutrice-codeuse aurait sans doute pu gêner la visibilité de la clé durant l'expérience perceptive) mais comme nous le constaterons dans le montage du test, la formation de la configuration manuelle est clairement incluse dans la transition de la main d'une position à une autre, la clé étant synchrone avec le début de la consonne.

VIII.3. Etude perceptive

Comment est traitée cette coordination main-lèvres par le sujet sourd qui perçoit la LPC ? Est-il capable de traiter l'information apportée par la main avant l'information portée par les lèvres ou attend-il forcément pour intégrer les deux informations ? Autrement dit, un sujet sourd décodant la LPC va-t-il exploiter l'anticipation observée de la main sur les lèvres ? Si oui, on devrait observer une identification

correcte de la position de la main plus précoce par rapport à l'identification correcte de la voyelle, et une identification simultanée de la clé et de la consonne.

Comme nous l'avions annoncé, nous utiliserons un paradigme de *gating* (Grosjean, 1980, 1996), en tronquant en différents points nos séquences. Ce paradigme consiste en un dévoilement progressif d'un stimulus au cours du temps : classiquement, le stimulus est montré depuis le début jusqu'au premier point de troncature, puis jusqu'au second et ainsi de suite jusqu'à montrer le stimulus complet. Il semble donc être un bon moyen de tester l'identification progressive d'un stimulus en fonction de l'évolution du décours temporel des paramètres le caractérisant. Ce paradigme a été utilisé dans diverses situations expérimentales (Grosjean, 1996, pour une revue) et il a déjà été éprouvé en perception de la parole (Warren & Marslen-Wilson, 1987, 1988 ; Cathiard, 1994 ; Munhall & Tohkura, 1998 ; Smits et al., 2003 ; Sock & Vaxelaire, 2004 ; de la Vaux & Massaro, 2004). Notons que les doutes qui avaient été émis sur cette méthode du *gating*, notamment par Ohala & Ohala (1995) et McQueen (1995), n'ont jamais concerné que l'accès lexical et ils ont d'ailleurs été dissipés depuis (U. Frauenfelder, communication personnelle). De plus, afin d'éviter les problèmes soulevés par l'utilisation classique de cette méthode (Grosjean, 1996), les séquences seront présentées en ordre aléatoire et non pas de manière continue.

Cette technique nous permettra donc d'obtenir une identification pas à pas, au fur et à mesure du dévoilement progressif des informations manuelles et labiales : ainsi, nous serons à même de voir si l'anticipation de la main est récupérée perceptivement et mise à profit pour l'identification d'un percept phonétique.

VIII.3.1. Montage et passation du test

VIII.3.1.1. Création des stimuli

Parmi les trois répétitions de nos 16 séquences, nous avons retenu un exemplaire de chaque. Pour ce choix, nous avons sélectionné les réalisations les plus représentatives, celles qui s'approchaient au mieux des valeurs moyennes quant à l'anticipation, tout en nous assurant de la bonne visibilité de la cible labiale et du mouvement de la main.

A partir des images numérisées puis détramées de chaque séquence (avec une trame toutes les 20 ms), et en nous aidant des événements temporels repérés sur les signaux de notre analyse en production, nous avons déterminé six points de troncature[16] dans le domaine de la syllabe CV pour

[16] Notons que notre pas de troncature n'est donc pas régulier. Il correspond plutôt à une unité d'identification. Cette procédure a déjà été utilisée pour des unités linguistiques (voir Grosjean, 1996).

chacune des 16 séquences (voir un exemple de ces images sur la Figure 71 et leurs repères correspondants sur les signaux sur la Figure 72). Plus précisément, les points 1, 5 et 6 correspondent aux événements temporels repérés sur les signaux pour l'analyse et les points 2, 3 et 4 ont été déterminés par une analyse visuelle des trames : ceci était nécessaire afin d'avoir des points de troncature nous permettant de tester l'identification de la configuration manuelle. Notons que ces points précèdent toujours le début acoustique de la consonne (voir Figure 72), ce qui confirme nos résultats sur la formation de la clé manuelle comme étant superposée sur la transition de main et synchrone avec le début acoustique de la consonne. Les séquences tronquées commencent toujours par le début de la phrase porteuse [mytyma] et se terminent à l'image correspondant à chaque point de troncature de la syllabe CV. Pour chaque séquence tronquée, le sujet devra identifier la syllabe CV.

Les points de troncature sélectionnés sont donc les suivants :

– Le 1er point de troncature correspond au début du mouvement de la main (trame correspondant à notre étiquette temporelle M1, récupérée à partir de l'analyse de données, voir Figure 72) pour la syllabe CV. Ce premier point correspondant au début du geste de la main pour coder cette syllabe, la séquence tronquée à ce point ne délivrera aucune information sur la nature de la syllabe CV : nous prévoyons, pour cette première troncature, que le sujet choisisse majoritairement la réponse [ma]. Nous aurons ainsi des courbes d'identification pour la cible CV démarrant quasiment à 0.
– Le 2ème point correspond au début de la formation de la configuration de la main. A ce point, la main n'est plus complètement ouverte (comme elle l'était pour coder la consonne [m] précédente) mais on ne peut pas encore identifier quelle sera la clé suivante. La trame retenue a été sélectionnée par analyse visuelle des images.
– Sur la trame au 3ème point de troncature, la clé est en cours de formation et la main est en train de se déplacer vers la position codant la voyelle V (soit la « pommette » ou le « menton »).
– Le 4ème point correspond à une trame où la clé et la position sont nettement visibles et identifiables. Notons que la main n'a pas encore vraiment atteint sa position (M2) mais s'en est clairement approchée ; la main se dirigeant soit vers la « pommette », soit vers le « menton », positions clairement distinctes spatialement, cela peut permettre une anticipation précoce de l'identification de la position. A cette étape, la forme labiale de la consonne est en formation (le point 4 précède toujours le début acoustique de la consonne).
– En 5ème point, nous avons pris le point L1 du signal de la trajectoire des lèvres correspondant au début du mouvement des lèvres vers la cible vocalique. La clé est totalement formée, la main a

atteint sa cible spatiale, la consonne est identifiable aux lèvres. Ainsi, il ne devrait plus y avoir de différence entre l'identification de la clé et de la consonne.

– Le 6ème point est la trame correspondant à L2, soit à l'atteinte de la cible labiale. A ce point, la syllabe devrait pouvoir être identifiée quelle que soit la séquence.

Le point 4 est donc critique en ce qui concerne l'anticipation des informations manuelles sur les informations labiales. Si les sujets récupèrent et utilisent l'information manuelle dans son décours temporel, nous devrions obtenir à ce point 4 une différence entre l'identification de la clé LPC (position et configuration) et l'identification des informations fournies par les indices labiaux. Dans le cas contraire, nous ne trouverons pas de différence à ce point, les deux types d'information étant intégrés ensemble au moment où les informations labiales seront disponibles, soit au point 5 pour la consonne et au point 6 pour la voyelle.

Figure 71. Exemple d'un découpage en six points de troncature de la syllabe [pɛ̃] pour la séquence [mytymapɛ̃ma] (de gauche à droite et de haut en bas). La main part de la position « côté » pour le [ma] et se dirige vers la « pommette », tout en formant la clé du [p], pour le [pɛ̃]. On voit clairement qu'au point 4, la clé est formée, la position est presque atteinte et le [p] n'est pas encore visible aux lèvres (il le sera au point 5).

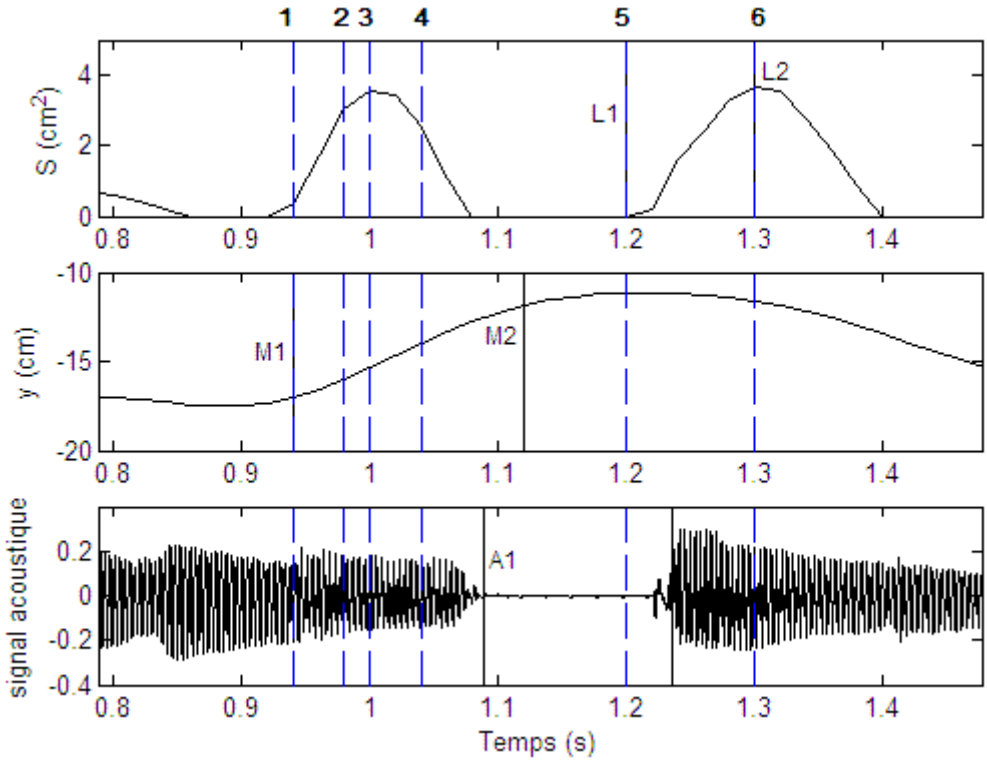


Figure 72. Tracé des signaux (aire aux lèvres S, position y de la pastille et signal acoustique) au cours du temps pour la portion [mapɛ̃ma] de la séquence [mytymapɛ̃ma] (voir Figure 70). Les barres verticales en traits pleins correspondent aux événements temporels repérés sur l'analyse des signaux. Les six lignes pointillées correspondent aux six points de troncature retenus pour l'expérience perceptive.

Nous avons donc 16 séquences coupées en six points de troncature, ce qui nous fait un total de 96 films. Nous avons réalisé ces films par le biais d'un programme Matlab créant des fichiers AVI à partir d'une série d'images : les films sont générés à 50 trames/sec (la taille des images étant de 640x465 pixels) et sont compressés à l'aide du codec Cinepak (nous avons bien entendu vérifié le rendu visuel et la qualité des images était très correcte).

VIII.3.1.2. Sujets et procédure

Seize sujets sourds profonds ont passé le test (voir Tableau 7 pour un détail) : treize adolescents âgés de 11 à 19 ans (âge moyen : 15,5 ans) et trois adultes (de 35, 25 et 24 ans). Ils avaient tous des parents entendants qui codent. Les treize adolescents bénéficiaient de codage régulier en milieu scolaire. L'adulte de 35 ans lit parfaitement sur les lèvres et décode le LPC lors de réunions professionnelles. Les deux autres adultes ne bénéficient plus de code depuis plusieurs années (de 3 ans à 7 ans) mais lisent parfaitement sur les lèvres. Ces jeunes ont été rencontrés soit dans leur milieu scolaire soit lors d'une journée organisée par l'association URAPEDA réunissant des sourds et leur famille ou encore par le biais de l'Institut des Jeunes Sourds de Cognin (Savoie). Une anamnèse

(Tableau 7 ; voir aussi Annexe 9) a été réalisée auprès des parents pour chaque jeune : la grande majorité de ces anamnèses a été contrôlée par l'orthophoniste.

Chaque sujet lisait, avant le test, une consigne écrite qui décrivait la tâche et la manière de donner la réponse (voir Annexe 8) ; au besoin, des précisions orales codées étaient données. L'expérimentatrice restait auprès du sujet pendant la phase de familiarisation avec l'interface afin de vérifier que la consigne avait été bien comprise. Puis le sujet était laissé seul pendant la passation.

Durant le test, les sujets sont assis à une distance confortable de l'écran d'un ordinateur portable (écran 14 pouces ; résolution 1024x768 pixels). Les séquences sont présentées en vision seule (nous n'avons pas intégré le son afin de garantir une identification uniquement visuelle, les restes auditifs pouvant varier d'un sujet à l'autre). Le test est géré par une interface utilisateur Matlab (voir Annexe 7), qui présente les séquences filmées, en ordre aléatoire et différent pour chaque sujet (les séquences apparaissent aléatoirement pendant le test afin d'éviter d'éventuels effets d'apprentissage ou de stratégie de décodage ; Grosjean, 1996) et enregistre les réponses données. A la fin de chaque séquence tronquée, une fenêtre noire apparaît à la place de l'image, de façon à ne pas laisser au sujet la possibilité de développer un raisonnement sur la dernière image de la séquence et ainsi de deviner la syllabe présentée. Après visualisation d'une séquence, le sujet doit indiquer la consonne et la voyelle perçues en cliquant sur les boutons réponses à sa disposition ([m, p, v, k, d] et [a, ɔ, ø, ɛ, ɛ̃]). Lorsqu'il valide sa réponse, le film suivant est déroulé. Les sujets peuvent sélectionner la syllabe [ma] lorsqu'ils n'ont perçu que cette syllabe (rappelons que la séquence [mytyma] précède la syllabe cible ; si les sujets répondent [ma], cela signifie qu'ils n'ont pas encore identifié la syllabe cible qui suit) ; on s'attend donc à ce que cette réponse soit sélectionnée majoritairement pour les séquences tronquées au premier point. Une phase de familiarisation avec six exemples de séquences était proposée avant la passation complète du test.

Sujet	Sexe	Age (au moment du test)	Type de surdité	Cause	Aide auditive (âge)	Type de communication	Age début exposition LPC	Fréquence décodage par semaine	Aptitude LL	Précoce /Tardif
AC	F	24	Profonde bilatérale	?	port appareils puis arrêt	LL+LPC	2 ans et demi	pendant études, 20h/sem. Depuis 3 ans, pas de code	++	P
CJ	M	12	Profonde (dépistée à 2 ans)	Génét.	Contours	LPC	2 ans et demi	Plusieurs heures + soir	++	P
GV	M	15	Profonde	Génét.	Implanté (vers 8 ans)	LPC + LSF (car LPC n'a pas marché au début)	7 ans	15h + soir	++	T
JH	F	17	Profonde (dépistée à 10mois)	Génét.	Contours	LPC	10 mois	Ecole + soir + sœur sourde	++	P
LF	F	17	Profonde (dépistée à 3 ans)	?	Contours aux 2 oreilles (3 ans)	LL + un peu de LPC	15 ans (surtout à l'école)	30h (au lycée)	++	T
LM	M	15	Profonde	?	Contours (avant 1 an)	LPC	environ 18 mois	20h + soir	+++	P
LV	F	11	Profonde	?		LPC	4 ans (mais surtout depuis 9 ans)	7h30 + soir		T
MD	F	16	Profonde	?	Appareils (11 mois)	LPC + LL	1-2 ans	14h + maison	+++	P
MD2	F	17	Profonde (dépistée à 9mois)	Génét.	Contours (1 an)	LPC (+ LSF au collège)	1 an et demi	école + soir	+++	P
MH	F	19	Profonde	?	Appareils (2 ans)	LPC + LL (+ un peu de LSF)	2 ans (maison+2-3 fois/sem. en institut)	13h (au lycée)	+++	P
MP	F	12	Profonde	Génét.	Contours	LPC (mère+frère sourd)	3 ans	15h + soir	++	P
ND	M	25	Très sévère bilatérale (dépistée à 18 mois)	Accident médical	Appareils aux 2 oreilles (2 ans)	LL (très peu de LPC)	3-4 ans	Ne décode plus depuis plusieurs années	++	T
NS	F	35	Profonde (dépistée à 2 ans)	?	Contours puis implant vers 30 ans	LL		1 fois / mois	+++	T
SG	F	17	Profonde bilatérale	virus durant grossesse	Port d'appareils puis arrêt	LL+LPC (+très peu de LSF)	8 mois	35h (jusqu'au lycée) puis 2h/sem.	+	P
ST	F	13	Profonde	?		LPC	8 ans	15h + soir		T
TC	M	15	Profonde	?	Appareils (très peu portés)	un peu de LPC + LSF (bilingue)	(5 ans) surtout depuis 11 ans	15h	++	T

Tableau 7. Anamnèse des sujets.

VIII.3.2. Résultats

Nous présenterons les résultats généraux pour tous les sujets confondus, afin de dégager des tendances perceptives générales (nous donnons cependant les résultats individuels en Annexe 9).

A l'issue du test, nous obtenons pour chaque sujet une matrice d'identification des 96 stimuli présentés. Nous avons calculé, pour la cible CV, tous sujets confondus, les pourcentages d'identification correcte de la consonne, de la voyelle et de la syllabe. Nous avons aussi calculé les scores d'identification de la syllabe [ma] précédant la cible CV.

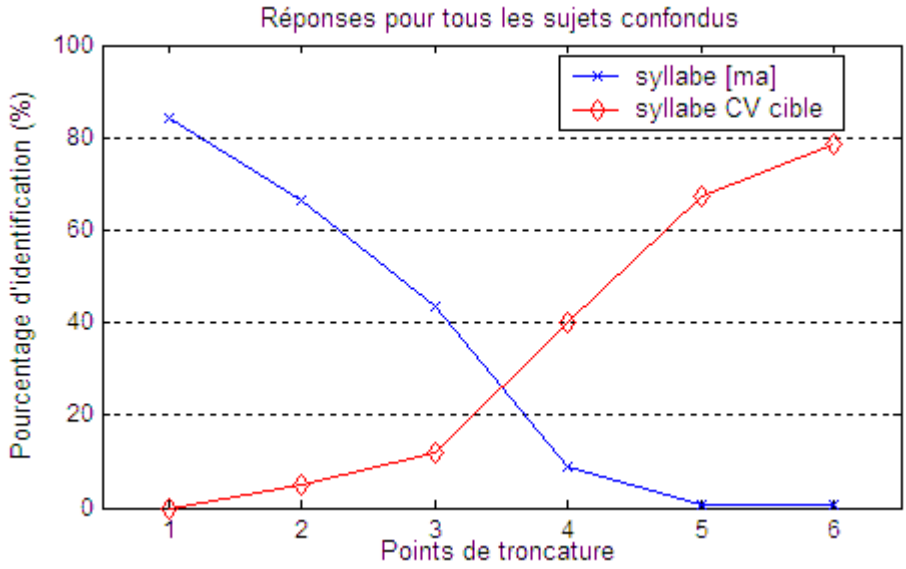


Figure 73. Scores moyens d'identification de la syllabe [ma] et de la syllabe CV cible obtenus aux différents points de troncature pour les 16 sujets confondus.

Nous présentons sur la Figure 73 les courbes d'identifications moyennes obtenues pour la syllabe [ma] et pour la syllabe CV cible à chaque point de troncature pour l'ensemble des sujets. La courbe de la syllabe cible, passant de 0% à 78,5%, traduit l'identification progressive de la syllabe CV au fur et à mesure du dévoilement des informations. Inversement, en ce qui concerne l'identification de [ma], nous voyons qu'au premier point de troncature, les sujets ne pouvant rien identifier de la syllabe CV cible, ont répondu en majorité [ma] (à 84%), comme attendu. Puis au fur et à mesure que le sujet a accès a davantage d'information, la courbe d'identification du [ma] décroît jusqu'à un seuil de nullité (0,8% au dernier point).

Afin d'analyser la prise en compte des informations manuelle et labiale dans les réponses des sujets, nous avons séparé la contribution spécifique de chaque information. En ce qui concerne l'information manuelle, à partir des scores sur la voyelle, nous avons calculé les pourcentages d'identification correcte de la position manuelle (la position « pommette » correspondant aux voyelles [ø, ɛ̃] et

« menton » aux voyelles [ɛ, ɔ]) et à partir des scores sur la consonne, les pourcentages d'identification de la clé manuelle (configurations 1 pour [p, d] et 2 pour [k, v]). De la même façon, nous avons extrait la contribution de l'information labiale, en calculant le pourcentage d'identification correcte des voyelles classées selon le critère d'arrondissement aux lèvres ([ɔ, ø] arrondies et [ɛ, ɛ̃] étirées) et celui des consonnes classées selon leur niveau de labialité ([p, v] labiales et [k, d] non labiales). Ainsi, lors de la présentation de la voyelle [ø] par exemple (qui correspond à la position « pommette » pour la main et à une voyelle arrondie en ce qui concerne les lèvres), si le sujet répond [ɛ̃], sa réponse est comptabilisée comme une réponse fausse en ce qui concerne l'identification du trait d'arrondissement, mais correcte du point de vue de la position (puisque la voyelle [ɛ̃] correspond également à la position « pommette »). Cette réponse signifierait que le sujet a identifié correctement l'information manuelle de position, mais s'est trompé concernant l'information labiale. En revanche, si pour cette voyelle [ø], le sujet répond [ɔ], sa réponse est dans ce cas correcte du point de vue de la labialité ([ɔ] et [ø] sont en effet toutes les deux arrondies) mais fausse en ce qui concerne la position de main (la voyelle [ɔ] étant codée par la position « cou »). Nous présentons sur la Figure 74 les pourcentages moyens d'identification correcte de la consonne et de la voyelle, avec en superposition, les contributions manuelles (clé et position) et labiales (consonnes et voyelles regroupées selon le trait de labialité) pour tous les sujets confondus. Nous pouvons remarquer au moins à partir du point 4 de troncature que la courbe d'identification correcte de la voyelle et celle de l'information vocalique labiale (voyelle/labialité) sont presque superposées, de même que la courbe d'identification de la consonne et celle de l'information consonantique labiale (consonne/labialité). On peut en conclure que c'est l'identification de l'information labiale (vocalique et consonantique) qui pose le plus de difficultés aux sujets.

Afin d'étudier statistiquement la contribution des informations manuelles et labiales, nous avons réalisé une analyse de variance à deux facteurs intra-sujets : « troncature » (6 points) et « identification » (4 modalités : position, clé, consonne/labialité, voyelle/labialité). Le facteur « troncature » a un effet significatif ($F(5, 75)= 173$, $p< .01$), de même que le facteur « identification » ($F(3, 45)= 38$, $p< .01$). L'interaction « troncature »x« identification » est également significative ($F(15, 225)= 21$, $p< .01$). Un test post-hoc de comparaison par paires (test de Sheffé) nous a permis de mettre en évidence qu'au point 4 de troncature[17], les identifications de la clé et de la consonne aux lèvres sont significativement différentes ($p< .01$) ainsi que les identifications de la position et de la voyelle aux lèvres ($p< .01$)

[17] Les scores aux trois premiers points de troncature étant relativement faibles (tous inférieurs à 50%), nous ne donnerons pas le détail des comparaisons entre les différentes identifications à ces points.

(l'identification de la voyelle est également différente de celle de la clé et de celle de la consonne). Au point 5 de troncature, l'identification de la voyelle est différente de toutes les autres ($p < .01$). Ainsi, à ce point, les identifications de la consonne et de la clé manuelle ne sont plus significativement différentes. Finalement au point 6 de troncature, nous ne trouvons plus de différence significative entre les différentes identifications.

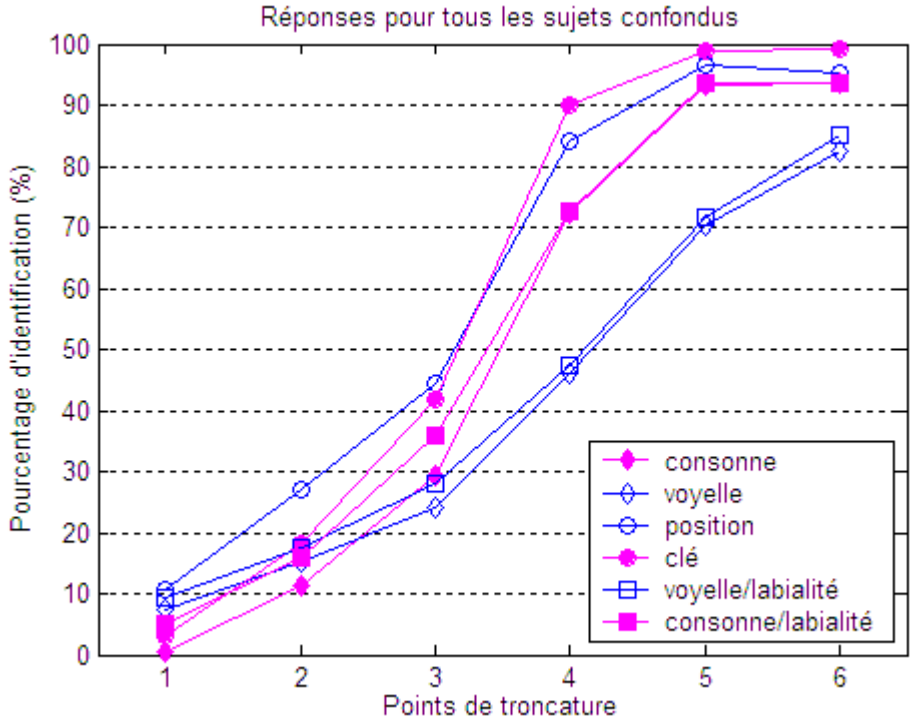


Figure 74. Scores moyens d'identification pour la consonne, la voyelle, la consonne classée selon son niveau de labialité, la voyelle classée selon son arrondissement, la clé et la position manuelle obtenus aux différents points de troncature pour les 16 sujets confondus.

A partir de cette analyse statistique, nous pouvons donc dégager les points suivants :

– Au premier point de troncature, tous les scores sont très faibles (inférieurs à 10%), les sujets répondant majoritairement [ma]. Aux points 2 et 3, les identifications [ma] diminuent et ce sont les scores pour l'identification de la position qui croissent les plus rapidement, suivis de ceux de l'identification de la clé. Au point 3, tous les scores restent néanmoins en-dessous des 50% ; ce qui signifie que les sujets commencent à récupérer des informations mais celles-ci ne leur permettent pas encore d'identifier les stimuli.

– Au point 4, c'est-à-dire au moment où la clé manuelle est visible, la position de main est presque atteinte et la consonne commence à être formée aux lèvres, nous avons un maximum de différences entre les différentes courbes d'identification. Toujours en-dessous des 50%, l'identification de la voyelle aux lèvres est la plus faible de toutes (47%). Elle est clairement moins

bonne que l'identification de la position de main qui atteint un pourcentage élevé : 84%. De même, l'identification de la clé manuelle atteint un score d'identification correcte de 90%. A ce point, l'identification de la consonne aux lèvres commence à augmenter mais elle n'est encore que de 73% et reste significativement différente de l'identification de la clé. Ainsi, nous obtenons au point 4, une meilleure identification des informations manuelles LPC sur la consonne et sur la voyelle par rapport à celle donnée par les lèvres, ce qui signifie que les sujets ont commencé à traiter les informations manuelles alors que les informations labiales sur cette syllabe ne sont pas encore complètement disponibles.

- Au point 5, l'information labiale sur la consonne est disponible et nous ne trouvons plus de différence entre l'identification de la clé manuelle et celle de la consonne aux lèvres (respectivement de 99% et de 94%). L'identification de la position de main (96,5%) reste toujours statistiquement meilleure que celle de la voyelle aux lèvres (71%).
- Finalement au point 6, c'est-à-dire au moment où la cible vocalique est réalisée aux lèvres, l'identification du visème vocalique atteint un score de 85%, non différent de celui de la position manuelle. On ne s'étonnera pas que les scores globaux pour la voyelle (à 82%) mais aussi pour la syllabe (78,5%) ne soient pas maximaux au point 6 puisque nous n'avons pas montré toute la tenue de la voyelle, après son climax (rappelons que ce point 6 de troncature est déterminé par l'événement L2 sur le décours de l'aire intérolabiale, soit le début de l'établissement de la cible vocalique).

VIII.3.3. Discussion

VIII.3.3.1. Le geste de la main est bien perçu en avance des lèvres

La différence de scores obtenue au point 4 entre l'identification de l'information vocalique manuelle (position) et l'identification de l'information vocalique labiale (voyelle/labialité) indique que la position de la main, donnant l'information vocalique, est nettement identifiée en avance de la voyelle labiale. De même, on constate, à ce même point, une identification meilleure de la clé (consonne manuelle) que de la consonne labiale. Les sujets exploitent ainsi dès que possible (c'est-à-dire dès qu'elles sont fournies par le codeur) les informations de la main.

Pour mieux comprendre encore nos résultats, il nous semble important d'examiner en détail nos scores d'identification. Commençons tout d'abord par les identifications des informations sur la consonne données par la main et les lèvres. La Figure 75 présente les identifications de chaque clé et de chaque groupe labial consonantique pour tous les sujets confondus et la Figure 76 présente l'évolution de l'identification de chaque consonne [p, d, k, v]. Nous pouvons constater que chaque clé manuelle est

aussi bien identifiée, au fur et à mesure de son dévoilement temporel. Ainsi le score moyen de 90% d'identification des clés au point 4 (Figure 74) provient d'une identification également réussie des deux clés. En revanche, pour les lèvres, les consonnes [p, v] sont globalement mieux identifiées que les consonnes [d, k]. En particulier, au point 4, c'est-à-dire au moment où la forme labiale de la consonne n'est pas encore complètement formée, nous avons une claire différence entre les deux groupes : l'identification correcte des consonnes [p, v] s'élève à 91% alors qu'elle n'est qu'à 55% pour [d, k]. Le score moyen de 73% d'identification pour la consonne aux lèvres est ainsi partagé entre une bonne identification des consonnes bilabiale et labiodentale [p, v] (à 91%) et une identification plus incertaine des consonnes [k, d] (55%). On sait que ces deux dernières consonnes sont non marquées aux lèvres puisque leur lieu d'articulation est intra-buccal (respectivement alvéolaire et vélaire), et qu'elles appartiennent au même visème (Gentil, 1981). Mais n'oublions pas que nos sujets n'ont à ce point 4, aucune difficulté à identifier la clé manuelle : ayant vu la clé, il leur reste donc à choisir entre [p] et [d] ou entre [v] et [k]. Quelle est donc la raison de cette performance de nos sujets à mieux identifier les consonnes les plus marquées aux lèvres au point 4 ?

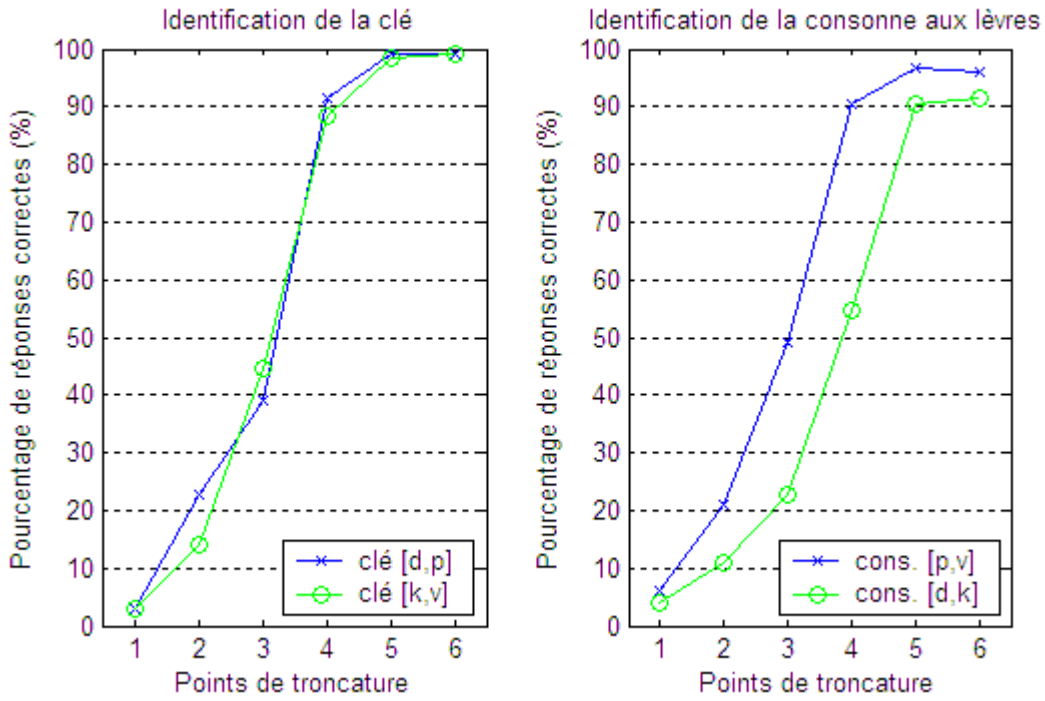


Figure 75. Résultats d'identification pour chaque clé (à gauche) et pour chaque groupe de consonnes aux lèvres (à droite) obtenus aux différents points de troncature pour tous les sujets confondus.

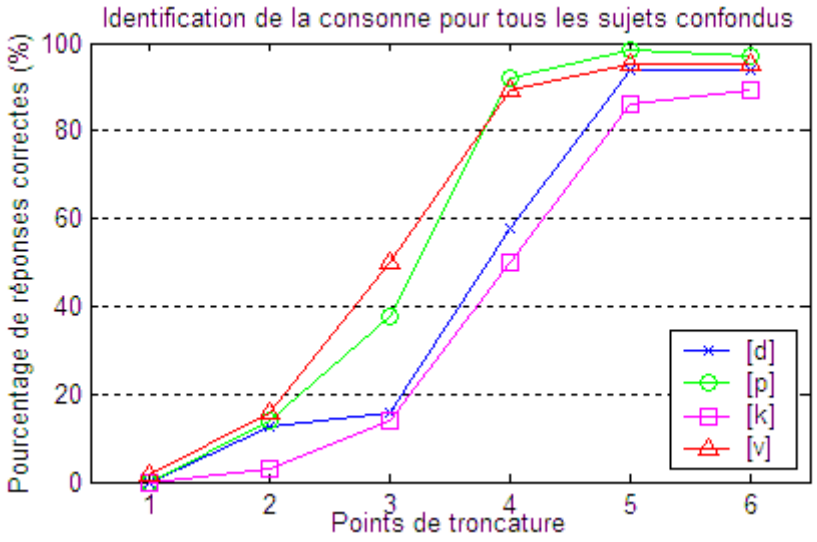


Figure 76. Résultats globaux (main-lèvres) d'identification pour la consonne obtenus aux différents points de troncature pour tous les sujets confondus.

La Figure 77 nous présente les décours d'aires aux lèvres pour les séquences [mapɛ̃], [mapø] et [madɛ̃] avec les six points de troncature. Remarquons sur ces signaux que le point 4 de troncature (correspondant sur chaque séquence vidéo à l'image où la clé et la position sont nettement visibles, voir section VIII.3.1.1) est toujours situé dans la portion de signal après le maximum d'ouverture intérolabiale du [a], autrement dit dans la phase de diminution de l'aire aux lèvres en vue de réaliser la consonne. On peut noter que cette diminution sera importante et rapide pour [p] (et [v]) jusqu'à complète fermeture des lèvres, et plus faible et moins rapide pour [d] (et [k]). C'est sans doute cette amorce de mouvement distinctif qui a été bien récupérée par nos sujets dès le point 4 pour les deux consonnes [p] et [v]. Notons aussi que notre description des lèvres limitée au décours de l'aire intérolabiale ne reflète sans doute pas toutes les variations visibles des mouvements labiaux au cours de la production de ces consonnes, en particulier en ce qui concerne les variations du contour labial externe, ou encore les informations de profondeur qui ont pu concourir à l'identification plus efficace des consonnes labiales.

Au point 5, date à laquelle la consonne est entièrement réalisée aux lèvres, nous obtenons, pour les informations labiales, des taux d'identification correcte très proches de 91% pour [d, k] et 97% pour [p, v] (Figure 75). Combinés aux informations manuelles, ces scores donnent des pourcentages globaux d'identification de 98% pour [p], 95% pour [v], 94% pour [d] et 86% pour [k] (Figure 76). Ces scores d'identification suivent le niveau de visibilité articulatoire bien connu des consonnes, de la bilabiale [p] à la postérieure [k] (Gentil, 1981).

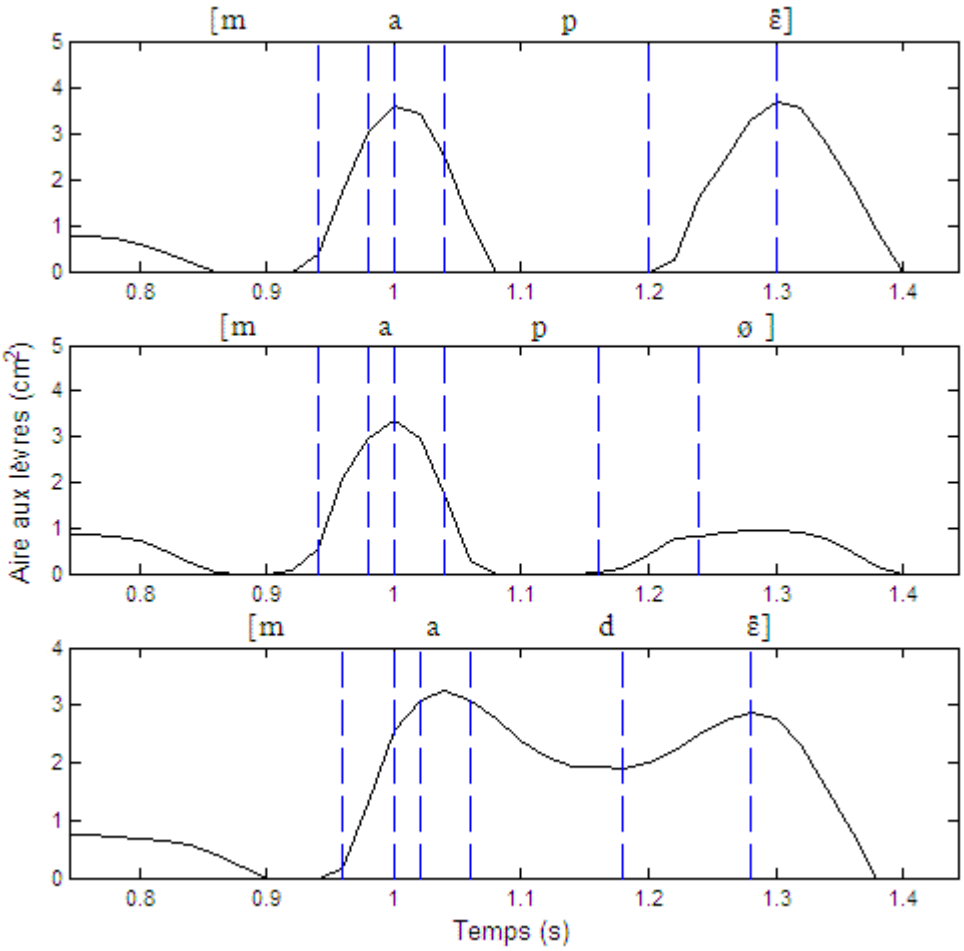


Figure 77. Décours d'aires aux lèvres au cours du temps pour les séquences [mapɛ̃], [mapø] et [madɛ̃] (de haut en bas) avec sur chaque graphe les six points de troncature en traits tiretés.

En ce qui concerne l'information vocalique, nous présentons sur la Figure 78 les identifications de chaque position de main (« pommette » et « menton ») et de chaque groupe de voyelle aux lèvres (arrondies et non arrondies) et sur la Figure 79 les courbes d'identification de chaque voyelle [ɛ, ø, ɛ̃, ɔ] pour l'ensemble des sujets. Nous avions obtenu au point 4 une meilleure identification de la position de main (84%) par rapport à la voyelle aux lèvres (47%) (voir Figure 74). Nous voyons ici (Figure 78) que les deux positions manuelles, « pommette » et « menton », sont globalement identifiées de manière égale (au point 4, elles le sont à plus de 80%), alors que l'identification de l'information vocalique labiale montre une franche supériorité pour les voyelles non arrondies ([ɛ, ɛ̃] à 80% et [ø, ɔ] à 14%). Puisque la position est identifiée correctement, le sujet n'a donc plus qu'à choisir entre deux voyelles, l'arrondie ou la non-arrondie. Comment expliquer les scores très différenciés que nous obtenons ?

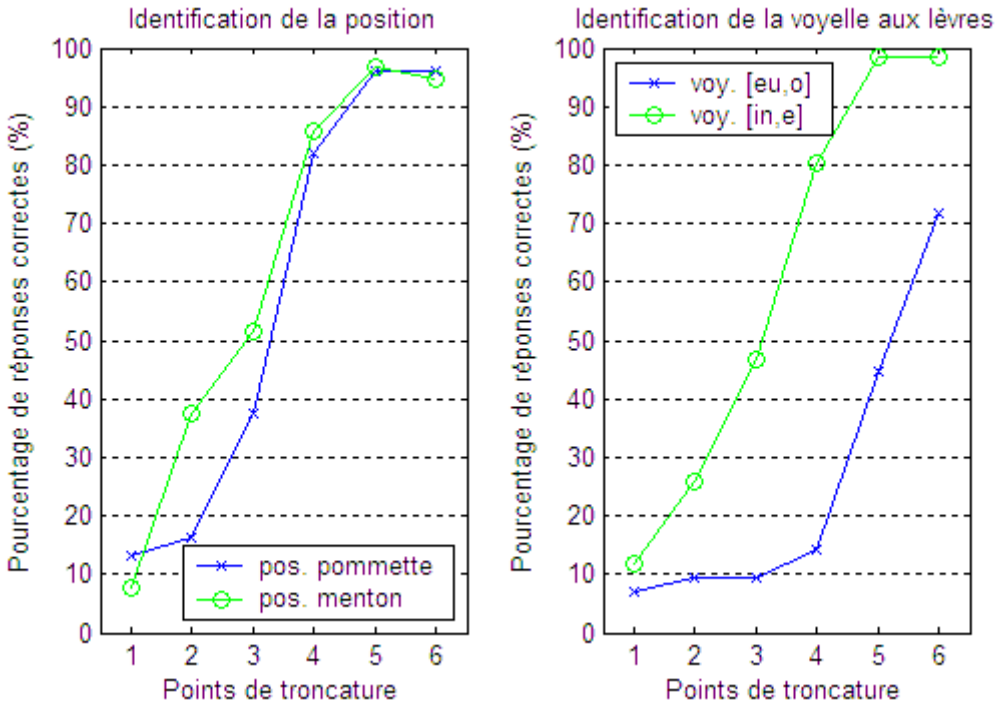


Figure 78. Résultats d'identification pour chaque position de main (à gauche) et pour chaque groupe de voyelles labiales (à droite) obtenus aux différents points de troncature pour tous les sujets confondus. La légende [e, eu, in, o] correspond aux voyelles [ɛ, ø, ɛ̃, ɔ].

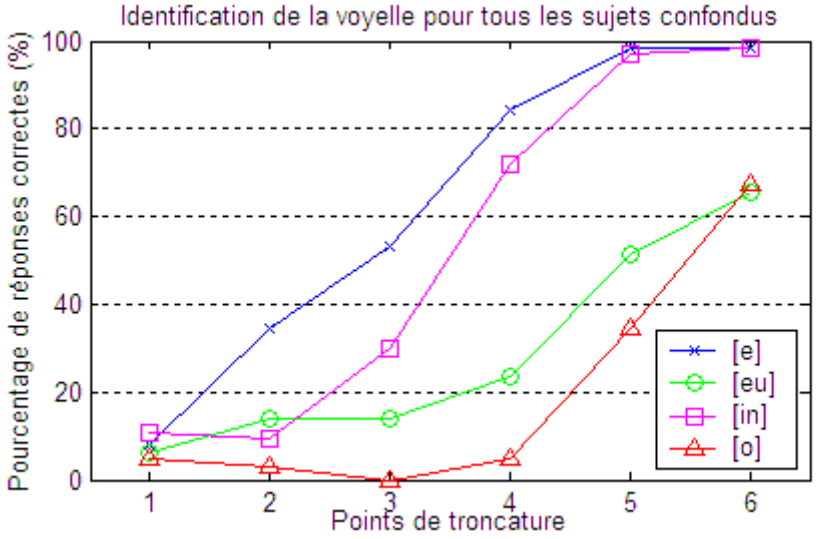


Figure 79. Résultats globaux (main-lèvres) d'identification pour la voyelle obtenus aux différents points de troncature pour tous les sujets confondus. La légende [e, eu, in, o] correspond aux voyelles [ɛ, ø, ɛ̃, ɔ].

Revenons à notre Figure 77 : nous pouvons voir que, quelle que soit la séquence, il sera bien difficile d'identifier au point 4, la voyelle à venir après la consonne. Que voit le sujet ? Une forme labiale dont l'aire aux lèvres est encore bien importante, que le sujet peut au mieux identifier comme correspondant à une voyelle étirée mais en aucun cas comme une voyelle arrondie. Dans la situation de choix forcé à laquelle il est contraint, il ne peut que faire ce choix de voyelle non arrondie (qui correspond en fait à la

voyelle [a] de la syllabe précédente). Les bons scores de [ɛ] et [ɛ̃] prouvent simplement qu'il a bien tenu compte de l'information manuelle vocalique donnée par la position. On notera que les identifications correctes des voyelles arrondies restent longtemps difficiles à établir, puisqu'au point 6 de notre gating, les scores avoisinent seulement 72% (score néanmoins nettement supérieur au hasard). A ce point 6, l'arrondissement est certes établi mais il manque la phase de tenue de la voyelle qui permettrait une meilleure identification.

Ainsi, l'analyse de ces résultats en fonction des voyelles nous montre que : (1) nos sujets intègrent bien à chaque étape de gating, les informations manuelles et labiales dont ils disposent : ils sont en effet capables de les composer de manière cohérente, pour peu qu'on lise les résultats perceptifs à la lumière du décours articulatoire des signaux ; (2) l'information sur la position vocalique manuelle est bien récupérée précocement, dès qu'elle est disponible, soit aux alentours du début acoustique de la syllabe cible CV, tandis que l'identification correcte de la voyelle produite ne pourra être effectuée qu'aux alentours de cette voyelle. Cette expérience de gating nous permet donc de confirmer que c'est bien l'information manuelle qui est identifiée avant l'information labiale.

VIII.3.3.2. Sujets « précoces » versus « tardifs »

Les recherches en LPC ont mis en évidence une différence régulière au niveau des performances dans différentes tâches perceptives et cognitives entre les sujets qui ont été exposés précocement au code LPC (avant l'âge de 3 ans) et ceux qui en ont bénéficié plus tardivement (sections I.6 et I.7 ; Leybaert & Alegria, 2003 ; Colin, 2004). Bien qu'à l'origine notre étude ne visait pas à différencier nos sujets en deux groupes, il nous a paru néanmoins intéressant de séparer, sur la base de nos anamnèses, nos sujets, selon qu'ils avaient été exposés précocement (9 sujets) ou tardivement au LPC (7 sujets) (notés respectivement P et T dans le Tableau 7) : les LPC précoces regroupent les sujets qui ont bénéficié de code LPC à la fois à la maison et à l'école et qui ont été exposés à cette méthode avant l'âge de 3 ans.

Nous présentons sur la Figure 80 les courbes d'identification pour les deux groupes de sujets (pour chaque groupe, nous avons en haut sur la figure, les courbes pour la syllabe [ma], la syllabe cible, la consonne et la voyelle, et en bas, les courbes des contributions manuelles et labiales). Graphiquement, nous pouvons remarquer que, jusqu'au 5ème point de troncature, les deux groupes semblent présenter un patron d'identification très similaire. C'est seulement au point 6 de troncature qu'il semblerait y avoir quelques différences entre les deux groupes : les sujets LPC tardifs ont en effet en moyenne des scores légèrement moins bons que les sujets LPC précoces, surtout en ce qui concerne les identifications globales de la consonne (98% contre 88%), de la voyelle (87,5% contre 76%) et donc de la syllabe (87% contre 68%).

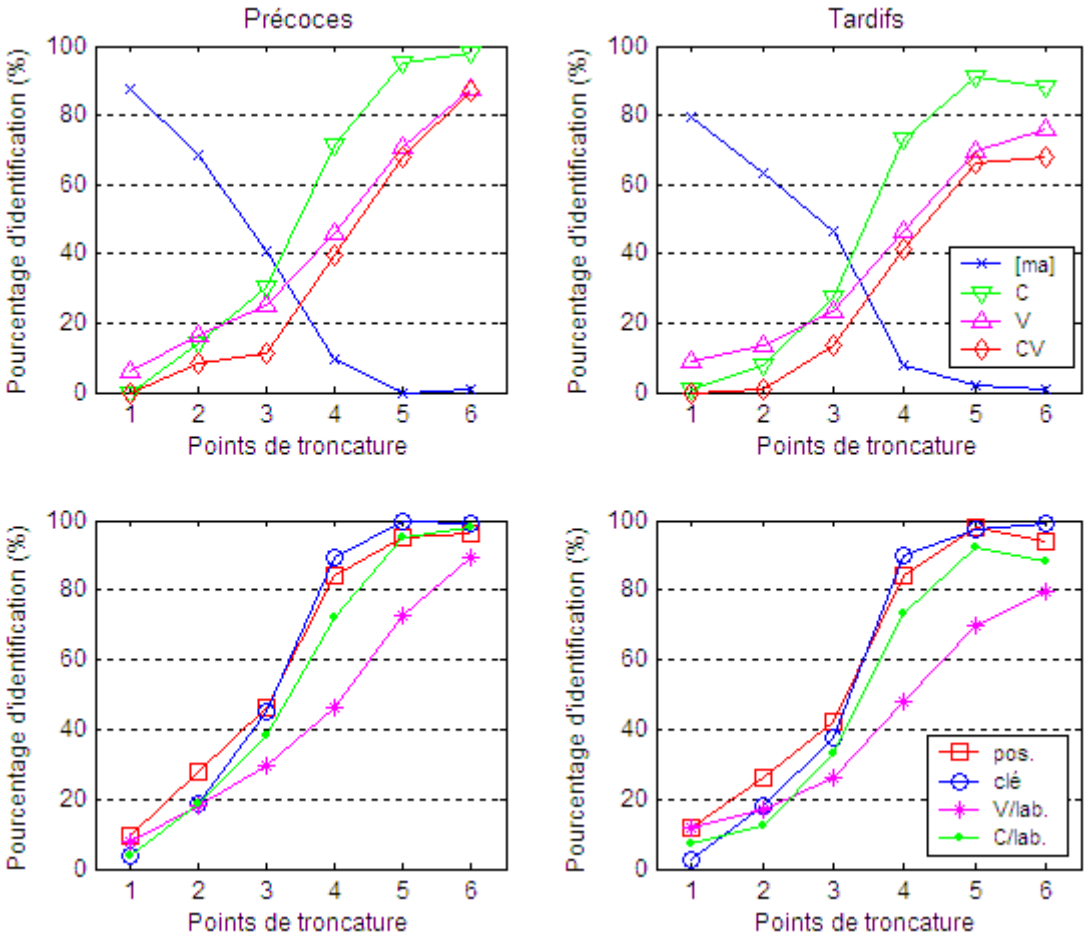


Figure 80. Résultats d'identification obtenus aux différents points de troncature pour les deux groupes de sujets LPC précoces (à gauche) et tardifs (à droite), pour la consonne (C), la voyelle (V), la syllabe (CV) et la syllabe [ma] (en haut) et pour la position manuelle, la clé manuelle, la voyelle (voyelle/labialité, classée selon son arrondissement) et la consonne (consonne/labialité, classée selon son niveau de labialité) (en bas).

Afin de comparer statistiquement les résultats des deux groupes, nous avons réalisé une ANOVA à deux facteurs intra-sujets (« troncature » avec 6 points et « identification » avec 4 modalités, consonne/labialité, voyelle/labialité, position et clé) et un facteur inter-sujet (« exposition » à deux modalités, précoce vs. tardive). Les résultats montrent que les seuls effets significatifs sont ceux des facteurs troncature ($F(5, 70)= 162$ $p< .01$) et identification ($F(3, 42)= 35{,}6$ $p< .01$), et de leur interaction (« troncature-identification » $F(15, 210)= 20{,}3$ $p< .01$). Tous les autres effets sont non significatifs ($F< 1$). Nous ne trouvons ainsi pas de différence significative entre les deux groupes de sujets, la tendance des LPC tardifs à moins bien identifier les informations fournies par les lèvres n'étant pas significative.

VIII.3.4. Conclusion

Nous avons retrouvé une fois de plus, dans cette étude, le patron temporel de coordination LPC-parole, soit le rendez-vous des contacts (celui de la position de main et de la consonne articulatoire),

entrainant ainsi une anticipation de la main dans la production du codage LPC par rapport au mouvement des lèvres. Notre test de gating montre que nos sujets sourds tirent profit de cette organisation temporelle spécifique en récupérant perceptivement cette anticipation de la main par rapport aux lèvres : les sourds décodeurs récupèrent et utilisent l'information donnée par le geste manuel en avance de l'information labiale correspondante. En effet, leurs performances indiquent un décodage progressif avec, dans un premier temps, identification par les informations manuelles d'un groupe de consonnes et voyelles possibles, suivi du choix d'un percept correct, une fois que les informations labiales pour la consonne et la voyelle sont disponibles. Il se pourrait donc bien finalement que la main « prépare le terrain » par avance pour l'identification visuelle de la parole codée. La main qui se dirige vers une position propose donc un premier ensemble de voyelles, le bon candidat étant finalement identifié par les lèvres.

En outre, il ressort de notre analyse post-hoc, en distinguant les sujets LPC précoces et tardifs, que ce patron d'identification est le même pour les deux groupes. La coordination temporelle LPC-parole est aussi bien récupérée par les sourds ayant bénéficié précocement de code que par ceux qui ont été exposés plus tardivement et de manière moins intensive. Nous avons juste observé une petite tendance des LPC tardifs à être moins bons labio-lecteurs. Il se peut cependant que notre classement en deux groupes souffre de la composition même de notre population, avec des sujets peut-être trop hétérogènes (selon des variables d'âge, d'exposition au code, etc.).

Ces résultats ne sont qu'un premier pas dans la compréhension de l'intégration perceptive des deux informations visuelles de la LPC. Mais d'ores et déjà, ils proposent pour la production-perception de cette invention tardive de communication, une organisation qui permet de l'ancrer au mieux sur le contrôle de la faculté de la parole. Les sourds perçoivent et utilisent l'anticipation de la main ; de cette manière, ils récupèreraient les contrôles de la coproduction parole-LPC.

Discussion générale et Conclusion

Afin d'améliorer la perception de la parole par les sourds, Orin Cornett a inventé le code LPC (ou Cued Speech, 1967) comme un complément manuel aux formes labiales ambiguës. S'inspirant largement de la théorie de l'information (Shannon & Weaver, 1949) pour constituer son système, ce physicien de formation a donné ainsi la possibilité aux codeurs de transmettre visuellement aux sourds toute l'information phonologique par une association optimale de la main – sa forme et sa position près du visage – et des lèvres. Son objectif a été atteint puisque l'utilisation de ce code manuel « artificiel » s'est révélée être très efficace pour la perception d'une parole complète, ayant des effets remarquables sur la qualité des représentations phonologiques développées chez les enfants sourds. En particulier, les LPC précoces (ceux qui ont été exposés à cette méthode de manière intensive et avant l'âge de trois ans) bénéficient d'une telle habileté phonologique leur permettant l'accès à la lecture et l'écriture qu'ils rivalisent sans problème particulier avec les enfants entendants. Ceci est maintenant bien établi : nous disposons en effet d'une grande littérature sur le sujet, avec en particulier, les avancées de Jacqueline Leybaert et de ses collègues à Bruxelles (*Laboratoire de Psychologie Expérimentale*, *LAPSE*, de l'*Université Libre de Bruxelles*). Alors que les études sur la perception et la mémoire du LPC ont largement avancé, les premiers essais de synthèse du *Cued Speech* ont dû utiliser des heuristiques, ne disposant d'aucune donnée sur le timing du contrôle manuel par rapport au mouvement des lèvres. Cet état de fait traduit un retard des études sur la production du Cued Speech, un retard tel que cette thèse représente la première étude sur la production du code LPC.

Nos recherches ont montré que l'ajout du code LPC à la parole ne modifie pas sa coarticulation naturelle. On peut simplement remarquer un ralentissement de la parole, comme cela avait été noté anecdotiquement par les utilisateurs de ce code (voir aussi Duchnowski et al., 2000). Nous avons mis en évidence une certaine automatisation du codage manuel : la durée de la transition de main d'une position à l'autre du visage est en effet stable pour un codage classique de syllabes CV et peut avoir une expansion qui reste relativement faible dans certains cas particuliers où la main a largement le temps d'exécuter son mouvement. Comment cette mise en rythme syllabique brachio-manuelle se coordonne-t-elle avec la parole ? Dans la coproduction parole-LPC, la main vient littéralement « se greffer » sur la configuration spatio-temporelle oro-faciale : le code manuel est ainsi complètement *ancré* sur la parole. Le codage LPC apparaît comme une activité experte complexe, dans laquelle la main et la parole sont coordonnées temporellement d'une manière très spécifique. Nous avons en effet mis en évidence, chez quatre codeuses expertes, ce patron de coordinations pour la production du

code LPC : pour une syllabe CV, la main débute sa transition avant le début acoustique de la syllabe visant l'atteinte de la position LPC cible en synchronie avec la consonne, soit bien avant la cible vocalique labiale. Cette coordination particulière est complètement liée à la parole puisque nous avons mis en évidence une dépendance du début d'initiation du geste vis-à-vis de la durée de la syllabe à produire, ce qui se traduit par un glissement de la transition de main dans une fenêtre temporelle autour du début de la tenue acoustique de la consonne. La main reste en position durant la consonne et repart ensuite vers la position suivante durant la production de la voyelle. La configuration de main se superpose à la transition de position à position. En résumé, l'information vocalique transmise par la main (sa position) anticipe celle transmise par les lèvres (la cible labiale vocalique), tandis que la configuration consonantique de la main pointe la position du visage en synchronie avec la consonne. Cette organisation temporelle spécifique des deux systèmes moteurs, manuel et articulatoire, est selon nous le résultat de la *compatibilité* des types de contrôle en jeu, ce qui se traduit par le phasage des deux composantes de contrôle moteur *locales,* soit le contrôle du mouvement visant à produire le contact de la position articulatoire de la consonne (occlusion ou constriction) en parole, avec le contrôle visant à atteindre la position-contact de la main près du visage en LPC. En outre, il semble que le contrôle de la production de la parole avec LPC impose son organisation temporelle au traitement perceptif de ce code. Les sourds LPC ont montré qu'ils étaient capables de récupérer par la vision la coordination temporelle main-lèvres, soit le phasage spécifique des deux systèmes moteurs, et de mettre à profit l'avance de la main dans le traitement perceptif de la parole codée. Les cibles récupérées – la position articulatoire de la consonne et la clé manuelle (sa configuration et sa position) – sont en effet identifiées avant la cible vocalique aux lèvres. Notre étude de *gating* a montré que les sourds intègrent les informations *aussi tôt* qu'elles leur sont fournies dans leurs décours temporels : autrement dit, ils exploitent le plus avantageusement possible l'anticipation manuelle. Cette organisation en production comme en perception nous laisse penser que les sourds, par le biais de la modalité visuelle, récupèrent bel et bien les contrôles de la parole codée en LPC. Cette récupération des contrôles pour précisément contrôler la perception est typiquement un des deux principes de la *Théorie de la Perception pour le Contrôle de l'Action*, la PACT (Schwartz et al., 2002b), comme d'ailleurs de la *Théorie Motrice de la Perception de la Parole* (TMPP). Le versant des contraintes perceptives, l'autre principe de la PACT – en l'occurrence les contraintes sur les contrastes de positions et de configurations de la LPC – est directement intégré dans cette théorie, pas dans la TMPP même classiquement révisée (Liberman & Mattingly, 1985).

Nous allons maintenant situer ces principaux résultats et leurs implications selon trois perspectives différentes : la modélisation psycholinguistique de la production de la parole codée ; la perception de

ce code en insistant particulièrement sur les mécanismes d'intégration ; et enfin l'intérêt de nos recherches pour le développement des technologies cognitives du handicap.

Vers un modèle de génération de parole codée

En ce qui concerne la production de la parole, ainsi que nous l'avons déjà rappelé, Levelt (1989) a proposé le modèle *Speaking* (voir aussi, Wheeldon & Levelt, 1995) dans lequel la représentation phonologique des mots est, à un stade donné du traitement, syllabifiée. Le locuteur aurait accès à un *syllabaire* – un répertoire mental de gestes syllabiques – qui permettrait de récupérer les gestes articulatoires correspondant aux syllabes phonologiques (Levelt & Wheeldon, 1994). De Ruiter (2000), s'appuyant sur ce modèle a proposé d'ajouter une composante pour la production des gestes co-verbaux, le modèle *Sketch*. Comment la modalité LPC peut-elle se situer par rapport à ces deux modèles ?

Lors de l'apprentissage du codage LPC, il est souvent recommandé d'apprendre à coder avec la main non dominante, de manière à pouvoir utiliser l'autre main plus facilement pour effectuer des tâches parallèles. Dans la pratique, tous les codeurs ne suivent pas forcément cette recommandation car ils peuvent être plus performants avec la main dominante. Quoiqu'il en soit, les codeurs utilisent souvent la main libre pour exécuter naturellement des gestes co-verbaux, en particulier des gestes de pointage (pour désigner la personne qui est en train de parler par exemple). Nous rapportons ici les témoignages de professionnels et de nos codeuses. Il parait donc possible d'envisager d'augmenter le modèle « Speaking + Sketch » (Levelt, 1989 ; de Ruiter, 2000) de la production de parole et de gestes par une composante LPC. En effet, les gestes de main du code LPC ne devraient a priori pas empêcher la production de gestes co-verbaux (du moins en partie). De plus, il est inutile de rappeler que le code LPC est un système syllabique, cela devrait donc faciliter l'intégration d'un module spécifique pour la LPC dans la mesure où le modèle de Levelt donne un rôle tout particulier à la syllabe.

Levelt (1989) explique le processus de génération de mots dans la parole en plusieurs étapes cognitives et propose un modèle à modules encapsulés correspondant à différents niveaux de traitement (pour un rappel, voir Figure 16 p. 22) : la conceptualisation de ce que le locuteur veut dire qui est gérée par le *conceptualizer*, la formulation de l'intention du locuteur avec les bons mots gérée par le *formulator* et l'articulation, étape durant laquelle le locuteur produit le mot et qui est gérée par l'*articulator*. L'étape de formulation est elle-même divisée en trois sous-processus : la sélection lexicale

durant laquelle le locuteur récupère les informations sémantiques et syntaxiques d'un mot, l'encodage phonologique qui donne la forme phonologique du mot, l'encodage phonétique qui associe à chaque syllabe du mot un geste articulatoire correspondant. Nos recherches ont montré que dans la coproduction parole-LPC, la coordination des différents articulateurs est une coordination finement contrôlée qui dépend complètement de la parole qui va être produite. Le séquencement sériel de l'information phonologique et syllabique (*serial ordering*, MacNeilage, 1970 ; Meyer, 1990, 1991) est délivré d'une façon régulière (*smooth*) et coarticulée à la fois par les articulateurs de la parole et par le code manuel.

Contrairement aux gestes co-verbaux qui ont une origine commune avec la parole au niveau de la conceptualisation de l'intention du locuteur (de Ruiter, 2000), la composante LPC ne peut entrer en jeu qu'une fois que les consonnes et les voyelles à produire sont spécifiées : la forme de main ainsi que sa position près du visage seront en effet déterminées par elles. Ce n'est donc qu'à partir de l'étape d'encodage phonologique (dans le formulator) que le module LPC est susceptible de venir se greffer (voir une schématisation de cette étape et des suivantes sur la Figure 81), une fois que le lexème, la forme phonologique sonore du mot, a été récupéré. A partir de là, nous pouvons formuler différentes hypothèses.

La première hypothèse fait intervenir le module LPC dès l'étape d'encodage phonologique. Durant cette étape, le lexème subit une mise en forme prosodique et la forme phonologique du mot est syllabifiée. A la manière du lexème qui a une composition segmentale et une structure métrique, on peut imaginer qu'il en est de même pour les composantes LPC. Chaque segment du lexème aurait un équivalent en clé manuelle LPC, c'est-à-dire les positions sur le visage pour les voyelles (cinq au total) et les configurations de main (huit) pour les consonnes. On aurait ainsi un module qui viendrait directement se greffer sur la composition segmentale du lexème, avec en entrée les consonnes et les voyelles et en sortie l'identité des configurations et des positions de main correspondantes. On disposerait ainsi d'une suite de clés manuelles (ou du moins une représentation de ces clés) avant même l'association des composantes segmentales et métriques. Cette intervention précoce nécessiterait d'avoir en plus une composante qui serait dédiée à l'information métrique et qui ferait le lien entre la représentation de ces clés et leur association en unité syllabique CV. Il semble cependant assez difficile de dissocier en LPC la composition segmentale (la forme de main et sa position) de la structure métrique (l'association des deux). Cela nécessiterait en effet une multiplication de sous-modules, qui devraient à chaque étape de l'encodage phonologique, effectuer les mêmes traitements pour le code manuel. Cette hypothèse est donc peu plausible ; il semble plus pertinent de faire

intervenir le module LPC une fois que la forme syllabique phonologique du mot est créée, soit en sortie de l'étape d'encodage phonologique.

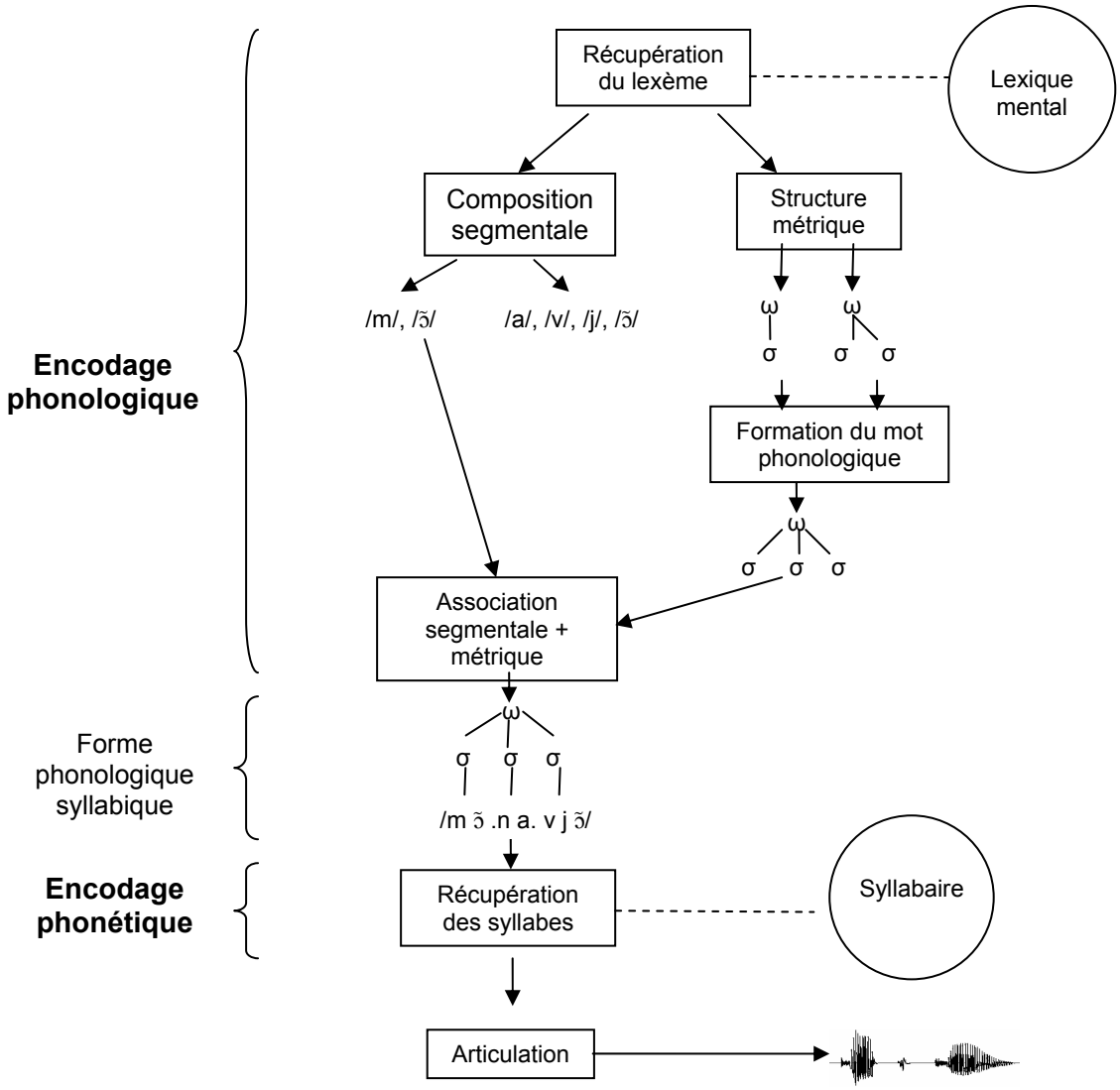


Figure 81. Détail de l'étape d'encodage phonologique dans le modèle « Speaking » + encodage phonétique et articulation pour « mon avion » (adapté de Levelt & Wheeldon, 1994).

Notre deuxième hypothèse propose donc que le module LPC intervienne en sortie de l'étape d'encodage phonologique, en parallèle à l'étape d'encodage phonétique, durant laquelle les gestes syllabiques articulatoires sont récupérés dans le syllabaire (voir une proposition schématisée sur la Figure 82). Le module LPC disposerait donc en entrée de la forme phonologique syllabique du mot. Rappelons que cette forme syllabique spécifie également les liaisons entre les mots ; ce qui est tout à fait pertinent en LPC. Cette forme doit être d'abord resyllabifiée en syllabes CV (pour aboutir à une forme phonologique syllabique CV du mot). A partir de là, le module LPC aurait accès à une sorte de *LPCaire*, un stock de programmes moteurs fournissant les clés LPC (soit une paire configuration-position de la main) correspondant à chaque syllabe CV du mot. Le plan LPC serait ensuite envoyé au

module de contrôle moteur du geste LPC, de la même façon que le plan phonétique est envoyé au module d'articulation.

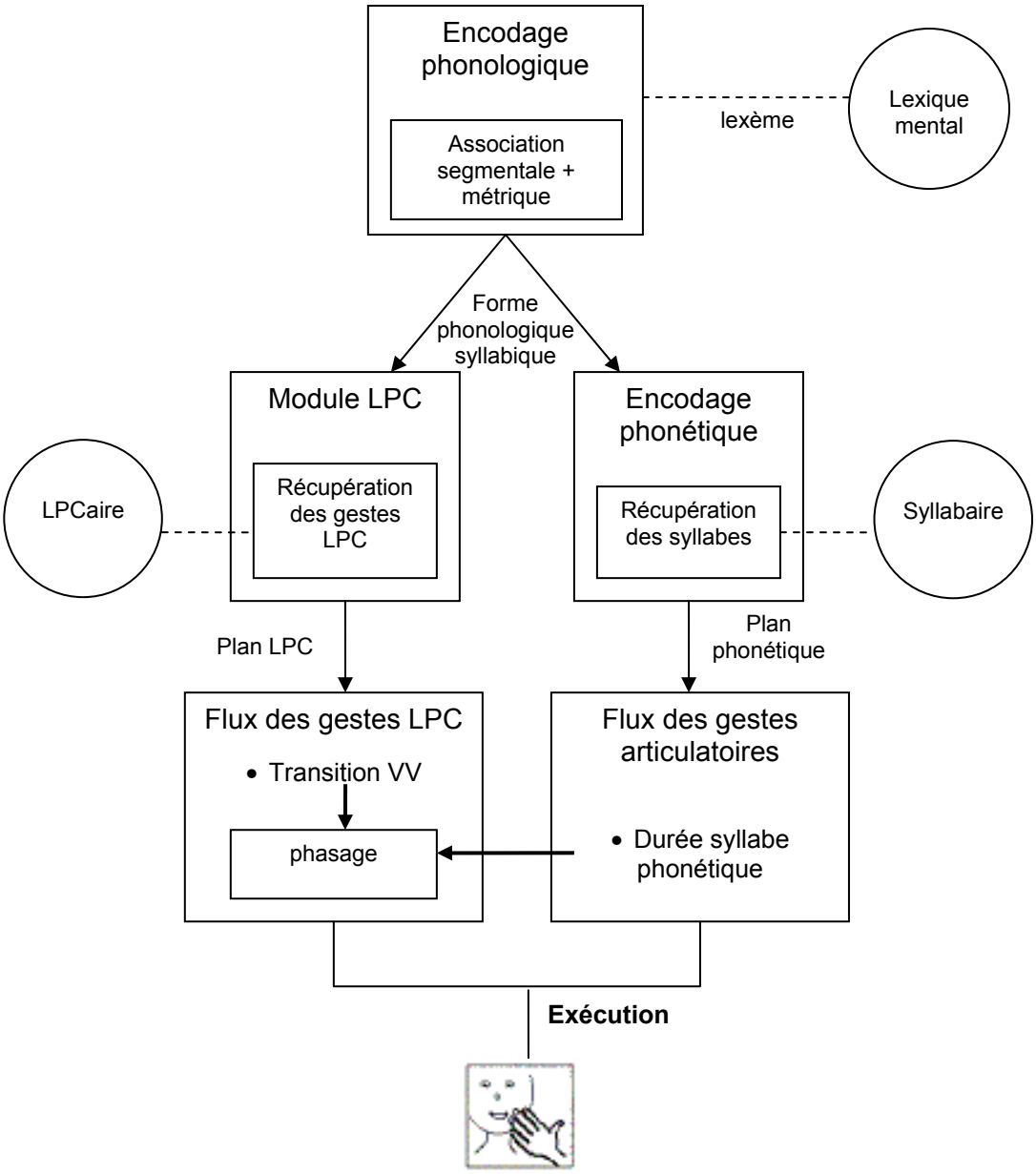


Figure 82. Proposition d'une composante LPC dans le modèle « Speaking ».

Pourquoi proposons-nous que ce module LPC fonctionne en parallèle à l'étape d'encodage phonétique ? Ce parallélisme garantit avant tout une récupération rapide des programmes moteurs LPC chez les codeurs professionnels, ce qui ne va pas perturber la planification de la parole, mais peut simplement la ralentir (le ralentissement du rythme de parole observé dans le codage LPC pourrait provenir en effet d'une compétition sur des ressources communes de traitement). Nous aurions dans le cas contraire pu penser que le module LPC vienne se greffer avant l'encodage phonétique ; dans cette hypothèse, une forme phonologique syllabique CV aurait été envoyée au syllabaire et nous aurions eu

en sortie une articulation de parole « hachée », qui suivrait un découpage CV. Ce n'est pas le cas pour les codeurs expérimentés, comme nous avons pu le montrer dans nos recherches : la coarticulation naturelle de la parole n'est en effet pas perturbée par l'ajout du code manuel. Cette hypothèse serait plutôt à tester dans le cas de codeurs en cours d'apprentissage.

Comment se fait la coordination entre le plan phonétique pour la parole et le plan LPC ? Nous avons mis en évidence une coordination temporelle très spécifique entre le geste de main et la parole : il ressort en particulier de ces études en production que l'organisation du code manuel et de la parole suit une certaine *hiérarchie temporelle*, quel que soit le contexte de coarticulation : le geste de transition de la main, qui est relativement incompressible, se déploie en avance par rapport au son pour atteindre sa position finale aux alentours du début acoustique de la consonne. Ainsi la position de main est finalement toujours atteinte avant que la cible de la voyelle ne soit formée aux lèvres. Notre hypothèse est que cette organisation temporelle particulière résulte du phasage des composantes de contrôle moteur locales en LPC et en parole : la position de main et la consonne articulatoire se rencontrent temporellement dans une *fenêtre de compatibilité motrice des contacts* en fonction de la durée de la syllabe phonétique. Ainsi nous proposons une interaction entre le contrôle moteur en LPC et le module d'articulation en parole : les deux programmes moteurs vont par ce biais pouvoir se coordonner, avant l'exécution. Par rapport aux gestes co-verbaux, la relation LPC-parole s'inscrirait ainsi plutôt dans une théorie balistique ; évidemment, le rôle potentiel d'un feedback durant l'exécution motrice ne doit pas être négligé, permettant de maintenir un but perceptif par un allongement de la tenue de la cible manuelle par exemple.

Cette proposition reste bien entendu une hypothèse de travail. D'autres expériences restent à mener, particulièrement sur les structures syllabiques complexes du type $C(C_n)V(C_m)$, en portant une attention particulière au phasage des gestes manuels avec la parole, ce qui nous permettra de mieux comprendre les étapes de planification de parole codée. Nous pourrons en retour émettre des hypothèses sur la façon dont la structure syllabique de la langue est perçue par les sourds LPC, ceux qui ont acquis par « imprégnation visuelle » la *phonologie* de la langue. En effet, la syllabe apparaît également comme une unité fondamentale dans le domaine de la perception de la parole (pour une revue, voir Dumay et al., 2002) ; supplantant le phonème, elle est soit proposée simplement comme unité de segmentation (Cutler & Norris, 1988 ; Content et al., 2001a), soit comme une unité de classification, constituant ainsi un intermédiaire mental entre la perception du signal et le lexique mental (le signal de parole serait recodé et catégorisé en unités syllabiques avant même l'accès au lexique ; Mehler et al., 1981 ; Mehler et al., 1990). Le codage LPC donnant à la fois une information sur la voyelle et sur la consonne, c'est un code dont la pratique peut apporter des informations sur cette

question. Rappelons qu'Alegria et al. (1992, 2005) avaient déjà mis en évidence l'influence de la main dans la perception de parole codée : ils avaient observé en particulier que les sourds LPC précoces avaient tendance à percevoir des syllabes supplémentaires quand la structure du mot ne respectait pas la forme *canonique* CV (*CS errors*, c'est-à-dire des erreurs liées à la structure du LPC). Quoi qu'il en soit la sensibilité des sourds LPC à la structure syllabique simple et universelle CV ne semble pas devoir être mise en question.

Vers un modèle d'intégration parole-LPC

En dehors des études de Alegria et collaborateurs (Alegria et al., 1992, 1999 ; Alegria & Lechat, 2005), très peu de choses sont connues sur la façon dont les informations manuelles et labiales sont intégrées par les sourds décodeurs de LPC. Ces auteurs ont proposé deux modèles possibles pour l'intégration main-lèvres en LPC :

1) un modèle hiérarchique, selon une approche de type résolution de problèmes, dans lequel l'information de lecture labiale, qui serait première, fournirait le corps de l'information phonologique et l'information manuelle, plus tardive et optionnelle, permettrait de résoudre les ambiguïtés restantes ;

2) un modèle phonémique, reposant sur une intégration des informations manuelles et labiales, comparable à l'intégration des informations auditives et visuelles en parole (Summerfield, 1987).

A propos de ces deux modèles, Leybaert & Alegria (2003) soulignent que le premier modèle s'apparentant à une conception « lip gestures first, cues next » pourrait être approprié pour rendre compte des performances des enfants élevés tardivement avec la LPC mais ne semblerait pas convenir pour les sourds LPC précoces, selon l'argument suivant : « It is possible that sometimes the lip gesture disambiguate the hand gestures, while sometimes the reverse occurs. If this speculation is true, it points toward a more integrated model of C[ued] S[peech] perception than a simple "lip gestures first, cues next", at least for experienced CS receivers. » (Leybaert & Alegria, 2003, p. 264). Ils proposent ainsi, pour les LPC précoces, que les informations manuelles et labiales, pouvant alternativement se désambiguïser, seraient combinées ensemble en une sorte de percept phonémique unique.

Nos recherches sur la production du codage LPC ont montré que les deux modalités, la main et la parole, suivent une organisation temporelle spécifique au niveau des informations qu'elles délivrent : alors que l'information consonantique donnée par la configuration de main est en-phase avec celle donnée par la parole, l'information vocalique délivrée par la position de main anticipe clairement la cible labiale. Notre étude par gating visuel montre que le contrôle de la production du code LPC impose son organisation temporelle au traitement perceptif de ce code. La position de main est en effet temporellement identifiée correctement avant la cible labiale de la voyelle, au moment où elle est fournie par les codeurs : donc au niveau temporel, l'identification du phonème vocalique se fait progressivement par le décodage de l'information manuelle puis de l'information labiale. En ce qui concerne la récupération de cette position vocalique anticipée, nous l'avons observée aussi bien chez nos sujets LPC classés en précoces ou en tardifs ; ce qui signifie que temporellement, ils sont les uns et les autres autant capables d'intégrer l'information manuelle dès qu'elle est donnée, avec les informations labiales disponibles au même moment. L'analyse des réponses de nos sujets, éclairée par le décours temporel des signaux articulatoires et manuels, nous a en effet permis de voir que les sujets répondent en tenant compte à la fois des informations manuelle et labiale. Ainsi nous montrons que l'information labiale n'est clairement pas première.

En accord avec la proposition de Leybaert et Alegria (2003), nous pensons qu'un modèle phonémique, reposant sur une véritable intégration des informations manuelles et labiales (« a more integrated model of CS perception ») semble mieux correspondre au processus d'intégration main-lèvres chez les sourds LPC. Cette description reste cependant assez générale et ne définit pas un modèle particulier. Quelle serait donc notre proposition sur l'intégration des deux flux visuels[18] ? Dans l'état actuel de nos résultats, il nous manque en fait plus d'une expérience pour traiter de cette question de l'intégration avec des arguments aussi fournis que ceux qui sont donnés pour l'intégration audio-visuelle. Nous pouvons néanmoins tenter de la situer par rapport aux quatre modèles « cardinaux » possibles proposés par Schwartz et al. (1998).

1) Le modèle d'identification directe pourrait vraiment être rejeté si nous disposions de données montrant que les sujets peuvent faire état de leur perception d'une discordance pour des stimuli main-lèvres non-concordants tout en les intégrant.

[18] La modalité auditive, récupérée notamment par le biais de l'implant cochléaire, ne doit pas être omise. Cependant, l'utilisation du code LPC par des enfants implantés est encore relativement récente et nous disposons de très peu de données sur cette pratique. Ainsi, nous nous limiterons dans cette proposition aux deux informations visuelles délivrées par la main et par les lèvres.

2) Le modèle d'identification séparée qui décode en premier lieu, à partir des deux informations manuelle et labiale, un ensemble de phonèmes ou de traits phonétiques (type VPAM en audiovisuel), qu'il va ensuite fusionner pour déterminer un phonème unique, correspondrait assez bien à un modèle de type résolution de problème tel que Alegria et al. (1992) le proposaient pour les enfants LPC tardifs, mais aussi à leur modèle phonémique, la différence résidant dans le type de traitement, qui peut être sériel suivant un modèle dit « maître-esclave » (avec les lèvres en modalité dominante ; Andre-Obrecht et al., 1997) ou parallèle, et dans la fiabilité portée sur chaque information (ainsi dans le modèle phonémique, on pourrait par exemple augmenter le poids de l'information manuelle selon le degré de saillance de l'information labiale, comme l'ont observé par Alegria & Lechat, 2005). De manière schématique, on pourrait envisager, qu'à partir de la forme labiale et de l'information manuelle, deux sous-ensembles de phonèmes ou de traits visuels soient extraits et qu'ils soient ensuite fusionnés pour aboutir à un phonème unique. Par exemple, pour un [u], le sujet voit une forme labiale arrondie et la main en position « menton ». Chaque composante serait décodée séparément : la forme labiale arrondie fournirait le visème [y, u, ø, ɔ̃, o] et la position de main indiquerait le sous-ensemble [ɛ, u, ɔ]. Les deux groupes de phonèmes seraient ensuite comparés et le [u] final serait finalement identifié. A l'heure actuelle, nous disposons de trop peu de données en LPC pour rejeter ce modèle. Notons qu'il faudrait néanmoins introduire un système de temporisation des flux, à la manière du modèle multi-stream (Luettin & Dupont, 1998), par exemple, qui resynchronise les deux flux indépendants par des rendez-vous temporels contraints. Les décodeurs des informations labiale et manuelle pourraient de cette façon se resynchroniser par les contacts locaux de position vocalique manuelle et de consonne articulée aux lèvres.

3) Le modèle de recodage dans une modalité dominante n'est à l'évidence pas viable même si l'on prend pour modalité dominante la labio-lecture. Le code, de par sa construction même, désambiguïse les formes labiales identiques. Le recodage de l'information manuelle en forme labiale reviendrait à annuler cet avantage. En effet, dans le cas d'un [m] par exemple, le sujet voit la main formant la configuration 1 (main grande ouverte) associée à une occlusion bilabiale. L'information manuelle serait alors recodée en une forme labiale fermée (pour [m]), en une forme labio-dentale (pour [f]) et en une forme ouverte (pour [t]) de manière équiprobable. La fusion des deux informations donnerait ainsi une plus grande probabilité à la forme labiale fermée. Mais cette forme fermée est toujours hautement ambiguë puisqu'elle correspond aux trois phonèmes [m], [p] et [b] formant le visème bilabial.

4) Enfin, le modèle de recodage dans un espace des commandes motrices a pu naturellement poser question à ces auteurs (Alegria et al., 1992, 1999 ; Leybaert & Lechat, 2001) : parfaitement au fait de la Théorie Motrice de la Perception de la Parole, ils ont pensé que la LPC posait un véritable problème à celle-ci : « We do not think that our deaf participants processed their perception of C[ued] S[peech] in a code derived from articulatory features of speech » (Leybaert & Lechat, 2001, p. 960). En effet, on voit mal a priori comment on pourrait recoder les commandes ou traits LPC en commande articulatoires. Notre préférence va cependant bien vers ce type de modèle. Nous allons voir comment.

Le fait que nous ayons découvert que, dans leurs performances LPC, les sujets anticipaient systématiquement la commande de contact vocalique pour la faire coïncider au maximum avec la commande de contact articulatoire consonantique, ouvre en effet la possibilité de trouver un espace de commande commun pour la bouche et la main. Nous présentons sur la Figure 83 cet espace commun de manière phonologique simplifiée, dans une approche de phonologie dite non-linéaire (inspirée dès ces débuts de la phonologie *autosegmentale*, Goldsmith, 1976, 1990). Les consonnes et les voyelles peuvent être représentés par des commandes ou traits, de posture et de contact (classiquement des traits dits de mode en phonétique) : la consonne articulatoire est caractérisée par un trait de contact (sur différentes parties du conduit vocal), tout comme la voyelle LPC (qui vise des contacts main-visage en différentes positions), alors que la voyelle articulatoire est caractérisée par un trait de posture (mise en forme globale du conduit vocal), comme l'est la consonne LPC. Ces traits, situés sur des lignes (paliers ou tires) parallèles distinctes, peuvent être projetés sur les unités d'une ligne du temps (unités de timing), dit *squelette*. Ainsi la lecture des commandes pour les consonnes et les voyelles en parole et LPC peut se faire le long de l'axe du temps : le trait de contact pour la voyelle LPC, qui est alors sur la même ligne auto-segmentale que le trait de contact pour le visème de la consonne, se produit sur la même unité de timing, de même que le trait de posture de la consonne LPC. Ainsi ces trois commandes sont quasi synchrones, précédant la commande de posture du visème vocalique qui tombe sur l'unité de timing suivante. Notre proposition est en somme que les informations manuelle et labiale sont projetées dans cet espace de commande commun, puis fusionnées, tout en tenant compte de la coordination temporelle, avant d'aboutir à un percept unique.

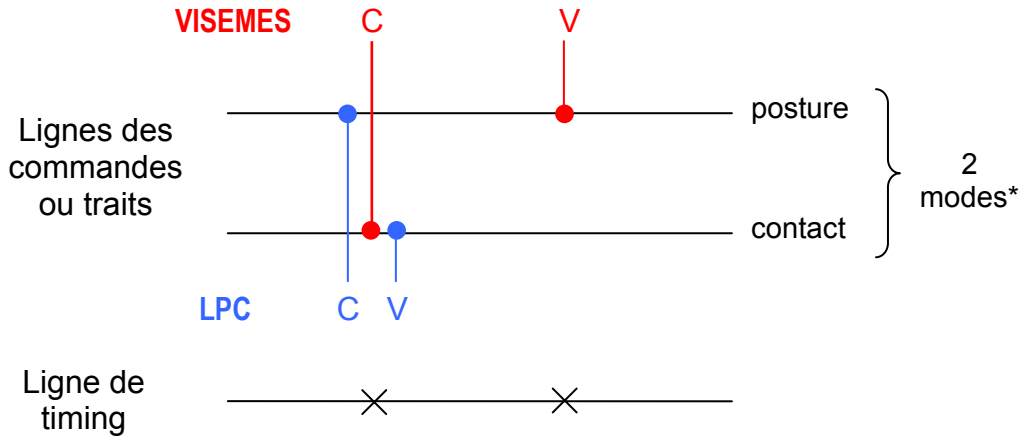


Figure 83. Schéma simplifié d'une représentation en phonologie non-linéaire d'un espace de commande commun (cas prototypique de la syllabe CV) à une étape très proche de l'output du modèle ; * la ligne de lieux buccaux et faciaux n'est pas montrée, de même que d'autres lignes de mode.

Notons que cette proposition est clairement différente d'une représentation phonologique purement visuelle, complètement dissociée de la parole, telle que celle proposée par Fleetwood et Metzger (1998). Elle s'inscrit plutôt dans une théorie motrice de la perception de la parole, ou mieux dans la PACT (Schwartz et al., 2002), qui lie la perception, le contrôle de l'action et la phonologie. L'étude plus spécifique du format de représentation du code LPC, ainsi que celle du développement de ce codage chez les sourds, nous permettraient d'apporter des arguments en faveur de ces théories. Ainsi, la question de l'intégration des informations manuelles et faciales en LPC reste ouverte et nécessite encore de nombreuses expériences.

Amélioration des technologies cognitives du handicap

Les nouvelles technologies ont permis depuis quelques années de créer des êtres virtuels et de les animer comme agents communicants. Nous avons même maintenant, avec le développement important de la réalité virtuelle, la possibilité de nous immerger dans des mondes virtuels et d'interagir avec des personnages de synthèse dans un environnement totalement fictif. Il est largement admis que le caractère naturel de ces personnages artificiels facilite l'interaction de l'homme avec la machine. Dans ce contexte, l'étude du mouvement n'est plus seulement une avancée théorique sur la connaissance de l'humain, mais représente un champ porteur pour améliorer les avatars de synthèse, notamment le réalisme prenant en compte les caractéristiques du mouvement humain. Le développement important des techniques d'acquisition, avec en particulier la multiplication des

systèmes de *motion capture* ces dernières années, a permis de mettre au point de véritables clones de synthèse. Différentes techniques sont mises en œuvre pour calquer le naturel : certains dotent par exemple leur système d'un modèle de contrôle intégrant les lois du mouvement humain (Gibet & Marteau, 1995 ; Gibet et al., 2004). Dans notre laboratoire, plusieurs têtes parlantes audiovisuelles, intégrant les propriétés multisensorielles de la parole, ont été clonées à partir de différents locuteurs. Ces clones virtuels sont constitués d'un ensemble de modèles (entre autres des modèles articulatoires établis à partir de données articulatoires obtenues par cinéradiographie et par Imagerie par Résonance Magnétique ; notons, de plus, qu'un modèle de visage, doté d'une texture vidéo-réaliste, complète l'apparence de la tête parlante) qui représentent les mécanismes articulatoires sous-jacents à la production de la parole (Revéret et al., 2000 ; Elisei et al., 2001 ; Bailly et al., 2003 ; Odisio & Bailly, 2004). Ces systèmes ouvrent la voie à de nombreuses applications : synthèse de parole multimodale à partir de texte, téléconférence et télécommunication, interaction homme-machine, réalité augmentée, etc. On comprend bien l'intérêt potentiel de tels systèmes pour la réhabilitation des personnes handicapées ; ce champ novateur suscite déjà un intérêt accru auprès de nombreux chercheurs, industriels et utilisateurs.

Concernant le code LPC, la meilleure compréhension de sa production par les codeurs expérimentés, ainsi que celle des relations production-perception qui se forment naturellement chez les utilisateurs de ce code manuel, sont indispensables pour mettre au point des technologies du handicap adaptées aux utilisateurs sourds décodeurs. Les études de simulation de Bratakos et al. (1998) avaient déjà pointé l'importance d'un codage synthétique simulant le mouvement naturel pour les systèmes de synthèse de Cued Speech. Duchnowski et al. (2000) avaient, dans cette optique, adopté des règles heuristiques pour le pilotage de la main synthétique pour améliorer leur système : en particulier, la main était avancée de 100 ms par rapport au son émis. Nous avons pu démontrer que cette avance arbitraire du MIT pour l'anglais correspondait en fait de la pratique réelle des codeurs adultes expérimentés et était mise à profit par la perception qu'en font les sourds.

Ainsi, dans le cadre d'un programme cognitique (action « Réhabilitation et remédiation », « Tête parlante audiovisuelle virtuelle : réalité augmentée et Langage Parlé Complété pour la réhabilitation des déficients auditifs », sous la responsabilité de Denis Beautemps), le clone PB de notre laboratoire (Beautemps et al., 2001 ; Badin et al., 2002) a été dotée d'une main synthétique (qui consiste en un ensemble de photographies de main formant les clés de la LPC) pilotée par les règles temporelles que nous avons mises en évidence dans nos études (Figure 84 ; pour plus de détails, voir Attina et al., 2004 ; voir aussi Celle, 2002).

Figure 84. Image de la tête parlante audiovisuelle PB augmentée de la modalité LPC.

Dans le cadre d'un autre projet piloté par notre laboratoire (projet RNRT ARTUS, « Animation Réaliste par Tatouage audiovisuel à l'Usage des Sourds », sous la responsabilité de Gérard Bailly), ce sont la main et le visage d'une codeuse qui ont été clonés par une technique de motion capture (Gibert et al., 2005), permettant ainsi de disposer d'un codeur virtuel 3D, qui vise à donner la possibilité aux sourds de remplacer le télétexte par ce personnage. Enfin, très récemment, un autre projet ambitieux dans le domaine de la télécommunication téléphonique (Projet TELMA, « Téléphonie à l'usage des Malentendants », sous la responsabilité de Denis Beautemps), vise à enrichir un terminal téléphonique de fonctionnalités audio-visuelles (débruitage, transcodage LPC-parole) pour améliorer les conditions de communication des sourds, mais aussi des personnes âgées.

La LPC : un codage phonologique incorporé

Lors de son invention par Cornett, la LPC a été conçue comme un codage artificiel venant se greffer sur un code naturel, le code phonologique. Comment ce codage manuel est-il devenu un codage phonologique incorporé ? On sait que le code orthographique a pu venir se greffer sur le code phonologique. Mais il n'y a rien de différent dans le contrôle grapho-moteur des lettres qui représentent les consonnes par rapport à celles qui représentent les voyelles. En témoigne l'évolution de l'alphabet depuis son invention dans les langues sémitiques, dont les racines consonantiques ont été utilisées pour noter les voyelles d'une langue comme le grec où elles devaient être précisées (et finalement nos voyelles alphabétiques a, e, i, o, u). Quoi qu'il en soit, s'il y a effectivement, pour ces deux catégories, quelques rares cas de dissociation offerts par la neuropsychologie (Cubelli, 1991, dysgraphie qui se ramène en fait à une dissociation phonologique, voir Caramazza et al., 2000), ce n'est pas au niveau des représentations motrices. Les comportements en LPC qu'il nous a été donné d'observer nous ont révélé une compatibilité inattendue entre les contrôles des consonnes de la parole et des voyelles de la LPC. Entre les différentes possibilités de coordination temporelle parole-LPC qui ont été testées, le

« bricolage » pour la synthèse réalisé par Duchnowski et collègues – qui ont obtenu empiriquement une amélioration pour la solution main en avance sur la face – s'est révélé dans notre étude sur la production comme étant tout simplement le comportement naturel de nos codeuses ; ce dont les sujets sourds testés ont aussi tiré le meilleur parti temporel en perception.

Le cœur de l'explication que nous avons donnée de ce phénomène d'optimisation d'un code, à l'origine artificiel, réside dans l'*embodiement* ou incorporation de la main et de la face dans un espace de contrôle neural, somatotopiquement suffisamment proche pour les deux articulateurs, voire commun : c'est celui du contrôle des contacts des articulateurs du langage dans l'espace sensorimoteur orofacial. Cette hypothèse est formulée de manière suffisamment contrainte pour pouvoir être précisément testée par imagerie cérébrale. Cette hypothèse est d'autre part non seulement formulée en termes spatiaux (somatotopiques) mais aussi en termes temporels, donc testables en chronométrie des signaux cérébraux : nous avons en effet mis en évidence une régularité statistiquement invariante du *stroke* manuel, dont le phasage avec le contact consonantique dans la parole dépend de la durée de l'unité rythmique de base qu'est la syllabe.

Les progrès au niveau du dépistage de plus en plus précoce de la surdité et le développement spectaculaire des implants cochléaires ces dernières années n'ont pas entraîné l'abandon de cette méthode manuelle. Le codage phonologique LPC reste une méthode nécessaire au bon développement du langage en pré- et post-implantation (Calmels et al., 2003 ; Le Normand & Berger, 2003). Il permet de faciliter l'accès à la modalité auditive. Même si l'enfant sourd semble avoir recouvré la totalité de son audition avec l'implant, les spécialistes de la surdité sont les premiers à reconnaître la nécessité de continuer à utiliser la LPC avec ces enfants, afin de garantir une bonne acquisition de la langue (voir notamment les témoignages de nombreux professionnels durant les journées d'étude de l'ALPC). La généralisation de l'implant ne marque donc pas la fin du code LPC ; au contraire, il est important de connaître au mieux ce système afin de permettre une réussite complète de la réhabilitation de la surdité de l'enfant.

Nos travaux sur la production-perception du LPC, démontrant que la perception vient se couler dans le déroulé temporel de la coordination anticipatoire main-visage, viennent tout naturellement s'insérer dans les mises en évidence de plus en plus nombreuses d'une « pratique mentale » qui de fait améliore la perception. Lorsque les actions motrices observées entreront en résonance – via le système des neurones dits « miroirs » de l'équipe de Giacomo Rizzolatti à Parme – avec les représentations motrices du sujet qui possède ces habiletés, le sujet percevant le code LPC aura d'autant plus de facilités qu'il aura développé lui-même par imitation un codage spontané. Cette

hypothèse pratique est elle-même tout autant testable que l'hypothèse plus théorique de la relation phonologique motrice que nous avons défendue.

Références bibliographiques

Abry C., Boë L.-J., Corsi P., Descout R., Gentil M. & Graillot P. (1980). *Labialité et phonétique. Données fondamentales et études expérimentales sur la géométrie et la motricité labiales*. Publications de l'Université des Langues et Lettres de Grenoble.

Abry C. & Boë L.-J. (1986). « Laws » for lips. *Speech Communication*, 5, pp. 97-104.

Abry C., Orliaguet J.-P. & Sock R. (1990). Patterns of speech phasing. Their robustness in the production of a timed linguistic task: single vs. double (abutted) consonants in French. *Cahiers de Psychologie Cognitive (European Bulletin of Cognitive Psychology)*, 10 (3), pp. 269-288.

Abry C. & Lallouache M. T. (1991). Audibility and stability of articulatory movements. Deciphering two experiments on anticipatory rounding in French. *Proceedings of the 12th International Congress of Phonetic Sciences*, 1, pp. 220-225.

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

Abry C. & Lallouache M. T. (1995b). Modeling lip constriction anticipatory behaviour for rounding in French with the MEM (Movement Expansion Model). *Proceedings of the 13th International Congress of Phonetic Sciences*, 4, pp. 152-155.

Abry C., Cathiard M.-A., El Abed R., Lallouache M. T., Leroy M. C., Perrier P., Poveda F. & Savariaux C. (1996a). Silent speech production: Anticipatory behaviour for 2 out of the 3 main vowel gestures/features while pausing. *Proceedings of the 1st ESCA Tutorial and Research Workshop on Speech Production Modeling & 4th Speech Production Seminar*, Autrans, France, pp. 101-104.

Abry C., Lallouache M. T. & Cathiard M.-A. (1996b). How can coarticulation models account for speech sensitivity to audio-visual desynchronization? In: Stork D. & Hennecke M. (Eds.), *Speechreading by Humans and Machines* (pp. 247-255), Springer-Verlag, Berlin.

Abry C. & Perrier P. (1996). Le contrôle des mouvements audibles et visibles dans la parole. In H. Méloni (Coordinateur), *Fondements et perspectives en traitement automatique de la parole*, AUF, Manuels universitaires.

Abry C., Stefanuto M., Vilain A. & Laboissière R. (2001). Que nous apprennent les "tan, tan" du Tan de Broca sur l'hypothèse d'une syllabe émergeant du babillage ? In : Keller D., Durafour J.P., Bonnot J.F.P. & Sock R. (Eds.), *Percevoir : monde et langage. Invariance et variabilité du sens vécu* (pp. 241-260). Sprimont : Mardaga.

Abry C., Stefanuto M., Vilain A. & Laboissière R. (2002). What can the utterance 'Tan, Tan' of Broca's patient Leborgne tell us about the hypothesis of an emergent 'babble-syllable' downloaded by SMA? In: Durand J & Laks B. (Eds.), *Phonetics, Phonology and Cognition* (pp. 226-243). Oxford University Press, Oxford.

Alegria J., Leybaert J., Charlier B. & Hage C. (1992). On the origin of phonological representations in the deaf: hearing lips and hands. In: Alegria J., Holender D., Morais J. & Radeau M. (Eds.), *Analytic Approaches to Human Cognition* (pp. 107-132). Amsterdam, North-Holland.

Alegria J., Charlier B. L. & Mattys S. (1999). The role of lip-reading and Cued Speech in the processing of phonological information in french-educated deaf children. *European Journal of Cognitive Psychology*, 11 (4), pp. 451-472.

Alegria J. & Lechat J. (2005). Phonological processing in deaf children: When lipreading and cues are incongruent. *Journal of Deaf Studies and Deaf Education*, 10 (2), pp. 122-133.

Alibali M. W., Heath D. C. & Myers H. J. (2001a). Effects of visibility between speaker and listener on gesture production: some gestures are meant to be seen. *Journal of Memory and Language*, 44, pp. 169-188.

Alibali M. W., Kita S., Bigelow L. J., Wolfman C. M. & Klein S. M. (2001b). Gesture plays a role in thinking for speaking. In: Cavé C., Guaïtella I & Santi S. (Eds.), *Oralité et gestualité. Interactions et comportements multimodaux dans la communication* (pp. 407-410). L'Harmattan, Paris.

Andre-Obrecht R., Jacob B. & Parlangeau N. (1997). Audiovisual speech recognition and segmental master-slave HMM. *Proceedings of AVSP'97*, pp. 49-52.

Audouy M. (2000). *Logiciel de traitement d'images vidéo pour la détermination de mouvements des lèvres*. Projet de fin d'études, option génie logiciel, ENSIMA Grenoble.

Badin P., Motoki K., Miki N., Ritterhaus D. & Lallouache M.-T. (1994). Some geometric and acoustic properties of the lip horn. *Journal of Acoustical Society of Japan (E)*, 15 (4), pp. 243-253.

Badin P., Bailly G., Reveret L., Baciu M., Segerbarth C. & Savariaux C. (2002). Three dimensional articulatory modelling of tongue, lips and face, based on MRI and video images. *Journal of Phonetics*, 30 (3), pp. 533–553.

Bailly G., Bérar M., Elisei F. & Odisio M. (2003). Audiovisual speech synthesis. *International Journal of Speech Technology*, 6, pp. 331-346.

Barry W. J. (1983). Some problems of interarticulator phasing as an index of temporal regularity in speech. *Journal of Experimental Psychology: Human Perception and Performance*, 9, pp. 826-828.

Beaupré W. J. (1997). *Gaining Cued Speech Proficiency. A manual for parents, teachers, and clinicians*. Web edition, http://www.uri.edu/comm_service/cued_speech/gcsphome.html.

Beautemps D., Badin P. & Bailly G. (2001). Linear degrees of freedom in speech production: Analysis of cineradio and labio-films data for a reference subject, and articulatory–acoustic modelling. *Journal of the Acoustical Society of America,* 109 (5), pp. 2165–2180.

Beautemps D., Cathiard M.-A. & Le Borgne Y. (2003). Benefit of audiovisual presentation in close shadowing task. *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona, 3-9 August, pp. 841-844.

Bell A. G. (1895). L'art subtil de la lecture sur les lèvres. (Traduction de Dupont & Legrand, 1976). *Bulletin d'Audiophonologie*, 6 (5), pp. 1-21.

Bell-Berti F. & Harris K. S. (1981). A temporal model of speech production. *Phonetica*, 38, pp. 9-20.

Bell-Berti F. & Harris K. S. (1982). Temporal patterns of coarticulation: Lip rounding. *The Journal of Acoustical Society of America*, 71 (2), pp. 449-454.

Benguérel A.-P & Cowan H. A. (1974). Coarticulation of upper lip protrusion in French. *Phonetica*, 30, pp. 41-55.

Benguérel A.-P. & Pichora-Fuller M. K. (1982). Coarticulation effects in lipreading. *Journal of Speech and Hearing Research*, 25, p. 600-607.

Benoît C. (1986). Note on the use of correlations in speech timing. *Journal of the Acoustical Society of America*, 80(6), pp. 1846-1849.

Benoît C. & Abry C. (1986). Vowel-consonant timing across speakers. *Proceedings of the 12th International Congress on Acoustics*, Toronto, Canada, A6-1.

Benoît C., Mohamadi T. & Kandel S. (1994). Effects of phonetic context on audio-visual intelligibility of French. *Journal of Speech and Hearing Research*, 37, pp. 1195-1203.

Bernstein L. E., Demorest M. E. & Tucker P. E. (1998). What makes a good speechreader? First you have to find one. In R. Campbell, B. Dodd & D. Burnham (Eds.*), Hearing by eye (II): The psychology of speechreading and auditory-visual speech* (pp. 211-228). East Sussex, UK: Psychology Press.

Bernstein L. E., Demorest M. E. & Tucker P. E. (2000). Speech perception without hearing. *Perception & Psychophysics*, 62(2), pp. 233-252.

Bernstein L. E. & Auer T. Jr. (2003). Speech perception and spoken word recognition. In: M. Marschark & P. E. Spencer (Eds), *Oxford Handbook of Deaf studies, language, and education*, (pp. 379-391), Oxford University Press.

Berthier V., Abry C. & Lallouache M. T. (1991). Coordination du geste et de la parole dans la production d'un instrument traditionnel. *Proceedings of the XIIth International Congress of Phonetic Sciences*, Aix-en-Provence, France, pp. 34 -37.

Berthoz A. (1997). *Le sens du mouvement*, Editions Odile Jacob.

Bladon A. & Al-Bamerni A. (1982). One stage and two-stage temporal patterns of velar coarticulation. *The Journal of Acoustical Society of America*, 72, S104(A).

Blevins J. (1995). The syllable in Phonological Theory. In: Goldsmith J. A. (Ed.), *The Handbook of Phonological Theory* (pp. 206-244). Blackwell Publishers, Oxford.

Blondel M. & Chauvin-Payan C. (2002). Syntaxe, rythme et gestualité dans les enfantines orales et signées. *Revue de linguistique et de didactique des langues*, 26, pp. 99-124.

Bonnot J.-F. P. (1990a). Production de la parole et coarticulation. Une analyse critique des principaux modèles. *Travaux de l'Institut de Phonétique de Strasbourg*, 20.

Bonnot J.-F. P. (1990b). Organisation temporelle des événements moteurs. Communication au Séminaire d'ouverture *Cognition, Perception et Action en Communication parlée*, Groupe francophone de la Communication parlée de la Société française d'acoustique.

Bonnot J.-F. P. & Keller D. (2004). Anticipation dans la parole, « bases articulatoires » et modèles phonétiques : un aperçu historique, In: Sock R. & Vaxelaire B. (Eds.), *L'anticipation à l'horizon du présent* (pp. 239-252). Sprimont, Mardaga.

Boyce S. E., Krakow R. A., Bell-Berti F. & Gelfer C. E. (1990). Converging sources of evidence for dissecting articulatory movements into core gestures. *Journal of Phonetics*, 18, pp. 173-188.

Boyer J. (2001). Iconicité et gestes. In: Cavé C., Guaïtella I & Santi S. (Eds.), *Oralité et gestualité. Interactions et comportements multimodaux dans la communication* (pp. 215-218). L'Harmattan, Paris.

Boyer J., Di Cristo A. & Guaïtella I. (2001). Rôle de la voix et des gestes dans la focalisation. In: Cavé C., Guaïtella I & Santi S. (Eds.), *Oralité et gestualité. Interactions et comportements multimodaux dans la communication* (pp. 459-463). L'Harmattan, Paris.

Braida L. D., Uchanski R. M. & Delhorne L. A. (1995). Speechreading aids based on automatic speech recognition: Prospects for the automatic generation of cued speech. *Journal of Acoustical Society of America*, 97(5), May 1995, p. 3308.

Braida L. D. (1991). Crossmodal integration in the identification of consonant segments. *Quaterly Journal of Experimental Psychology*, 43 (A-3), pp. 647-677.

Braida L. D., Duchnowski P., Lum D. S., Sexton M. G. & Bratakos M. S. (1997). Development of aids to speechreading based on automatic production of cued speech. *Journal of Acoustical Society of America*, 102(5), November 1997, pp. 3133-3134.

Bratakos M. S., Duchnowski P. & Braida L. D. (1998). Toward the automatic generation of Cued Speech. *Cued Speech Journal*, 6, pp. 1-37.

Browman C. P. & Goldstein L. (1990). Gestural specification using dynamically-defined articulatory structures. *Journal of Phonetics*, 18, pp. 299-320.

Browman C. P. & Goldstein L. (2000). Competing constraints on intergestural coordination and self-organization of phonological structures. *Bulletin de la Communication Parlée*, 5, pp. 25-34.

Butterworth B. L. & Beattie G. W. (1978). Gesture and silence as indicators of planning in speech. In: Campbell R. N. & Smith P. T. (Eds.), *Recent advances in the psychology of language 4: Formal and experimental approaches* (pp. 347-360). London: Plenum.

Butterworth B. L. & Hadar U. (1989). Gesture, speech, and computational stages: A reply to McNeill. *Psychological Review*, 96 (1), pp. 168-174.

Calmels M. N., Cochard N., Deguine O. & Fraysse B. (2003). L'implant cochléaire chez l'enfant : bilan, technique chirurgicale et résultats à long terme. *Actes du Colloque International ACFOS IV*, 8-10 Novembre 2002, pp. 105-109.

Cathiard M.-A. (1988-1989). La perception visuelle de la parole : Aperçu de l'état des connaissances. *Bulletin de l'Institut de Phonétique de Grenoble*, 17-18, pp. 109-193.

Cathiard M.-A. (1994). *La perception visuelle de l'anticipation des gestes vocaliques : cohérence des événements audibles et visibles dans le flux de la parole*. Thèse de Doctorat de Psychologie Cognitive, UFR SHS, Université Pierre Mendès France (Grenoble 2).

Cathiard M.-A., Schwartz J.-L. & Abry C. (2001). Asking a naive question about the McGurk effect: Why does audio [b] give more [d] percepts with visual [g] than with visual [d]? *Proceedings of AVSP'2001*, Aalborg, Denmark, pp. 138-142.

Cathiard M.-A. (2003). Interactions audiovisuelles. *Actes des Journées d'Etude de l'ALPC*, 40, Nantes.

Cathiard M.-A., Attina V. & Alloatti D. (2003). Labial Anticipation Behavior during Speech with and without Cued Speech. *Proceedings of the 15th ICPHS*, 4-10 August 2003, Barcelona, Spain, pp. 1939-1942.

Caramazza A., Chialant D., Capasso R. & Miceli G. (2000). Separable processing of consonants and vowels. *Nature*, 403, pp. 428-430.

Celle B. (2002). *Démonstrateur 2D de séquences consonne-voyelle pour le Langage Parlé Complété*. Rapport de stage d'IUT, IUT2, Grenoble.

Chafcouloff M., Guaïtella I. & Boyer J. (2001). Etude de la coordination temporelle des événements vocaux et gestuels dans l'exécution du haka par les rugby-men néo-zélandais, les "All Blacks". In: Cavé C., Guaïtella I & Santi S. (Eds.), *Oralité et gestualité. Interactions et comportements multimodaux dans la communication* (pp. 503-509). L'Harmattan, Paris.

Charlier B. L. & Leybaert J. (2000). The rhyming skills of deaf children educated with phonetically augmented speechreading. *The Quaterly Journal of Experimental Psychology*, 53A (2), pp. 349-375.

Chauvin-Payan C. (1999a). *Comptines, formulettes et jeux enfantins dans les Alpes occidentales (région Rhône-Alpes, Suisse romande et Val d'Aoste). Etude gestuelle, rythmique et verbale.* Thèse de Doctorat de 3ème cycle en Sciences du Langage, Centre de Dialectologie, Université Stendhal, Grenoble.

Chauvin-Payan C. (1999b). « Quand Fanny… » Autour d'un jeu de tape-mains. *Le Monde alpin et rhodanien*, 4ème trimestre 1999, pp. 55-68.

Chistovich L. A. (1980). Auditory processing of Speech. *Language and Speech*, 23, pp. 67-72.

Cholin J., Schiller N. O. & Levelt W. J. M. (2004). The preparation of syllables in speech production. *Journal of Memory and Language*, 50, pp. 47-61.

Clarke B. R. & Ling D. (1976). The effects of using Cued Speech: A follow-up study. *Volta Review*, 78, pp. 23-34.

Colin C. (2001). *Etude comportementale et électrophysiologique des processus impliqués dans l'effet McGurk et dans l'effet de ventriloquie*. Thèse de Doctorat, Université libre de Bruxelles, Belgique.

Colin C. & Radeau M. (2003). Les illusions McGurk dans la parole : 25 ans de recherches. *L'Année Psychologique*, 103 (3), pp. 497-542.

Colin S. (2004). *Développement des habiletés phonologiques précoces et apprentissage de la lecture et de l'écriture chez l'enfant sourd : apport du Langage Parlé Complété (LPC)*. Thèse de Doctorat de 3ème cycle, Institut de psychologie, Université Lumière Lyon 2.

Content A., Kearns R. K., & Frauenfelder U. H. (2001). Boundaries versus onsets in syllabic segmentation. *Journal of Memory and Language*, 45, pp. 177–199.

Cornett R. O. (1967). Cued Speech. *American Annals of the Deaf*, 112, pp. 3-13.

Cornett R. O. (1982). Le Cued Speech. In Francis Destombes (Ed.), *Aides manuelles à la lecture labiale et perspectives d'aides automatiques* (pp. 5-15). Centre scientifique IBM-France, Paris.

Cornett R. O. (1988). Cued Speech, manual complement to lipreading, for visual reception of spoken language. Principles, practice and prospects for automation. *Acta Oto-Rhino-Laryngologica Belgica*, 42(3), pp. 375-384.

Cornett R. O. (1994). Adapting Cued Speech to additional languages. *Cued Speech Journal V*, pp. 19-29.

Cubelli R. (1991). A selective deficit for writing vowels in acquired dysgraphia. *Nature*, 353, pp. 258-260.

Cutler A. & Norris D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance,* 14, pp. 113–121.

Davis C. & Kim J. (1998). Repeating and remembering foreign language words: Does seing help? *Proceedings of the International Conference on Audio-Visual Speech Processing,* Terrigal, Australia, 4-7 Dec., pp. 121-125.

Davis B. & MacNeilage P. (1995). The articulatory basis of babbling. *Journal of Speech and Hearing Research*, 38 (4), pp. 1199-1211.

Denes P. B. (1963). On the Statistics of Spoken English. *The Journal of Acoustical Society of America*, 35 (6), pp. 892-904.

Destombes F. (1982) Le projet VIDVOX. In Francis Destombes (Ed.), *Aides manuelles à la lecture labiale et perspectives d'aides automatiques* (pp. 35-36). Centre scientifique IBM-France, Paris.

Dittman A. T. (1972). The body movement-speech rhythm relashionship as a cue to speech encoding. In: Siegman A. W. & Pope B. (Eds.), *Studies in dyadic communication* (pp. 135-152). New York: Pergamon Press.

Dodd B. (1976). The phonological systems of deaf children. *Journal of Speech and Hearing Disorders*, 41, pp. 185-198.

Dodd B. (1987). Lip-reading, phonological coding and deafness. In: Dodd B. & Campbell R. (Eds.), *Hearing by Eye: The pyschology of lip-reading* (pp. 177-190). Lawrence Erlbaum Associates, Hillsdale NJ.

Dodd B. & Campbell R., editors (1987). *Hearing by Eye: The pyschology of lip-reading*. Lawrence Erlbaum Associates, Hillsdale NJ.

Duchnowski P, Braida L. D., Bratakos M., Lum D., Sexton M. & Krause J. (1998a). A speechreading aid based on phonetic ASR. *5th International Conference on Spoken Language Processing*, Sydney, Australia, 7, pp. 3289-3292.

Duchnowski P, Braida L. D., Lum D., Sexton M., Krause J. & Banthia S. (1998b). Automatic generation of Cued Speech for the deaf: Status and outlook. *International Conference Auditory-Visual Speech Processing*, Terrigal, Australia, pp. 161-166.

Duchnowski P., Lum D., Krause J., Sexton M., Bratakos M. & Braida L. D. (2000).Development of speechreading supplements based on automatic speech recognition. *IEEE Transactions on Biomedical Engineering*, 47 (4), pp. 487-496.

Dumay N., Frauenfelder U. H. & Content A. (2002). The Role of the Syllable in Lexical Segmentation in French: Word-Spotting Data. *Brain and Language,* 81, pp. 144-161.

Elisei F., Odisio M., Bailly G. & Badin P. (2001). Creating and controlling video-realistic talking heads. *Auditory-Visual Speech Processing Workshop*, Scheelsminde, Denmark, pp. 90-97.

Erber N. P. (1969). Interaction of audition and vision in the recognition of oral speech stimuli. *Journal of Speech and Hearing Research*, 12, pp. 423-425.

Erber N. P. (1972). Auditory, visual, and auditory-visual recognition of consonants by children with normal and impaired hearing. *Journal of Speech and Hearing Research*, 15, pp. 407-412.

Erber N. P., Sachs R. M. & DeFilippo C. L. (1979). Labiometrics I: Analysis of articulatory dynamics in relation to perception of vowels through lipreading. *The Journal of Acoustical Society of America*, 65, Suppl. N°1, S136.

Faraco M. (2000). Gestes et fonctionnalité dans une séquence narrative en langue seconde. *TIPA*, 19, pp. 63-71.

Farnetani E. (1997). Coarticulation and connected speech processes. In: Hardcastle W. J. & Laver J. (Eds.), *The handbook of phonetic sciences* (pp. 371-404). Blackwell Publishers, Oxford, UK.

Farnetani E. (1999). Labial coarticulation. In: Hardcastle W. J. & Hewlett N. (Eds.), *Coarticulation theory, data and techniques* (pp. 144-163), Cambridge University Press.

Fisher C. G. (1968). Confusions among visually perceived consonants. *Journal of Speech and Hearing Research*, 11, pp. 796-804.

Ferbach-Hecker V., Vaxelaire B., Cathiard M.-A., Savariaux C. & Sock R. (2001). How lip protrusion expansion influences auditory perceptual extent: Probing into the Movement Expansion Model. In: Cavé C., Guaïtella I & Santi S. (Eds.), *Oralité et gestualité. Interactions et comportements multimodaux dans la communication* (pp. 450-456). L'Harmattan, Paris.

Ferrand L., Segui J. & Grainger J. (1996). Masked priming of word and picture naming: the role of syllabic units. *Journal of Memory and language*, 35, pp. 708-723.

Ferrand L., Segui J. & Humphreys G. W. (1997). The syllable's role in word naming. *Memory and Cognition*, 25, pp. 458-470.

Ferrand L. & Segui J. (1998). The syllable's role in speech production: Are syllables chunks, schemas, or both? *Psychonomic Bulletin & Review*, 5 (2), pp. 253-258.

Feyereisen P. (1997). The competition between gesture and speech production in dual-task paradigms. *Journal of Memory and Language*, 36, pp. 13-33.

Fleetwood E. & Metzger M. (1998). *Cued language structure: an analysis of Cued American English based on linguistic principles*. Silver Spring, Maryland: Calliope Press.

Fodor J. A. (1983). *The modularity of mind: An essay on faculty psychology*. Cambridge, MA: MIT Press.

Fowler C. A. (1980). Coarticulation and theories of extrinsic timing. *Journal of Phonetics*, 8, pp. 113-133.

Fowler C. A. & Saltzman E. (1993). Coordination and coarticulation in speech production. *Language and Speech*, 36, pp. 171-195.

Furuyama N., McNeill D. & Park-Doob M. (2002). Is speech-gesture production ballistic or interactive? Papier présenté à *the Congress of the International Society for Gesture Studies*, Austin, TX, USA.

Gentil M. (1981). *Etude de la perception de la parole : Lecture labiale et sosies labiaux*. Paris : IBM France, 48 p.

Gentner D. R. (1987). Timing of skilled motor performance: Tests of the proportional duration model. *Psychological Review*, 94(2), pp. 255-276.

Gibert G., Bailly G., Beautemps D., Elisei F. & Brun, R. (2005). Analysis and synthesis of the 3D movements of the head, face and hands of a speech cuer. *Journal of the Acoustical Association of America*, 118 (2), pp. 1144-1153.

Gibet S. & Marteau P. F. (1995). Modèle sensori-moteur pour le contrôle et la commande de mouvement de bras. *Intellectica*, 2, pp. 233-251.

Gibet S., Kamp J.-F. & Poirier F. (2004). Gesture analysis: Invariant laws in movement. In A. Camurri & G. Volpe (Eds.), *GW 2003, Lecture Notes in Artificial Intelligence 2915* (pp. 1-9). Springer-Verlag, Berlin, Heidelberg.

Goldin-Meadow S. (1999). The role of gesture in communication and thinking. *Trends in Cognitive Sciences*, 3 (11), pp. 419-429.

Goldin-Meadow S., Alibali M. W. & Church R. B. (1993). Transitions in concept acquisition: using the hand to read the mind. *Psychological Review*, 100, pp. 279-297.

Goldin-Meadow S. & Momeni Sandhofer C. (1999). Gestures convey substantive information about a child's thoughts to ordinary listeners. *Developmental Science*, 2 (1), pp. 67-74.

Goldsmith J. A. (1976). *Autosegmental Phonology*, thèse de doctorat, Massachusetts Institute of Technology, Cambridge. Publiée en 1979, New-York, Garland Press.

Goldsmith J. A. (1990). *Autosegmental and Metrical Phonology*, Oxford, Blackwell.

Green K. P. (1998). The use of auditory and visual information during phonetic processing: implications for theories of speech perception. In: Campbell R., Dodd B. & Burnham D. (Eds.), *Hearing by*

eye, II. Perspectives and directions in research on audiovisual aspects of language processing (pp. 3-25). Hove (UK): Psychology Press.

Gregory J. F. (1987). An investigation of speechreading with and without Cued Speech. *American Annals of the Deaf*, 132 (6), pp. 393-398.

Grosjean F. (1980). Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics*, 28, pp. 267-283.

Grosjean F. (1996). Gating. *Language and Cognitive Processes*, 11 (6), pp. 597-604.

Hack Z. C. & Erber N. P. (1982). Auditory, visual and auditory-visual perception of vowels by hearing-impaired children. *Journal of Speech and Hearing Research*, 25, pp. 100-107.

Hadar U., Steiner T. J. & Clifford Rose F. (1984). The relationships between head movements and speech dysfluencies. *Language and Speech*, 27, pp. 333-342.

Hage C., Alegria J. & Périer O. (1990). Cued Speech and language acquisition: with specifics related to grammatical gender. *Cued Speech Journal*, 4, pp. 39-46.

Hage C., Alegria J. & Périer O. (1991). Cued Speech and language acquisition: The case of grammatical gender morpho-phonology. In: Martin D. S. (Ed.), *Advances in cognition, education and deafness*. Washington, DC: Gallaudet University Press.

Hardcastle W. J. & Hewlett N. (1999*). Coarticulation. Theory, Data and Techniques*. Cambridge University Press.

Hays W. L. (1988). *Statistics for the social sciences*. New York: Holt, Rinehart and Winston.

Henke W. L. (1967). Preliminaries to speech synthesis based on an articulatory model. *Proceedings of IEEE Boston Speech Conference*, pp. 170-171.

Heuer H. (1991). Invariant relative timing in motor-program theory. In Fagard J. & Wolff P.H. (Eds.), *The development of timing control and temporal organization in coordinated action* (pp. 37-68). Amsterdam: North-Holland.

Holender D. (1980). Interference between a vocal and a manual response to the same stimulus. In: Stelmach G. E. & Requin J. (Eds.), *Tutorials in Motor Behavior* (pp. 421-431). Amsterdam: North-Holland.

Jackson P. L., Montgomery A. A. & Binnie C. A. (1976). Perceptual dimensions underlying vowel lipreading performance. *Journal of Speech and Hearing Research*, 19, pp. 796-812.

Jackson P. L. (1988). The theoretical minimal unit for visual speech perception: Visemes and coarticulation. *The Volta Review*, 90 (5), pp. 99-115.

Jeannerod M. (1984). The timing of natural prehension movements. *Journal of Motor Behavior*, 16, pp. 235-254.

Jeannerod M. (1988). *The neural and behavioural organization of goal-directed movements*. Oxford Psychology series n°15, Clarendon Press Oxford.

Keele S. W., Cohen A. & Ivry R. (1990). Models of speech production. In Jeannerod M. (Ed.), *Attention and Performance XIII* (pp. 77-110), Lawrence Erlbaum Associates, Hillsdale (NJ), Hove and London.

Kelly S. D., Barr D., Church R. B., & Lynch K. (1999). Offering a hand to pragmatic understanding: The role of speech and gesture in comprehension and memory. *Journal of Memory and Language,* 40, pp. 577-592

Kendon A. (1980). Gesticulation and speech: Two aspects of the process of utterance. In: Key M. R. (Ed.), *The relation between verbal and nonverbal communication* (pp. 207-227). Mouton, The Hague.

Kendon A. (1997). Gesture. *Annual Review of Anthopology*, 26, pp. 109-128.

Kita S., van Gijn I. & van der Hulst H. (1998). Movement Phases in signs and co-speech gestures, and their transcription by human coders. In: Wachsmuth I & Fröhlich M. (Eds.), *Gesture and sign language in human-computer interaction*, International Gesture Workshop 1997, Bielefeld, Germany, September 17-19. *Lecture Notes in Artificial Intelligence*, 1371, (pp.23-35). Berlin: Springer-Verlag.

Krauss R. M. & Hadar U. (2001). The role of speech-related arm/hand gestures in word retrieval. In: Campbell R. & Messing L. (Eds.), *Gesture, Speech and Sign* (pp. 93-116). Oxford: Oxford University Press.

Kricos P. B. & Lesner S. A. (1982). Differences in visual intelligibility across talkers. *The Volta Review*, 84, pp. 219-225.

Lallouache M. T. (1991). *Un poste visage-Parole couleur. Acquisition et traitement automatique des contours des lèvres*. Thèse de doctorat, INP Grenoble.

LaSasso C., Crain K. & Leybaert J. (2003). Rhyme generation in deaf students: The effect of exposure to Cued Speech. *Journal of Deaf Studies and Deaf Education*, 8 (3), pp. 250-270.

Le Normand M.-T. & Berger B. (2003). Acquisition du langage chez l'enfant sourd porteur d'un implant cochléaire. *Actes du Colloque International ACFOS IV*, 8-10 Novembre 2002, pp. 148-156.

Lesner S. A. & Kricos P. B. (1981). Visual vowel and diphthong perception across speakers. *Journal of the Academy of Rehabilitative Audiology*, 14, pp. 252-258.

Lee B. S. (1950). Effects of delayed speech feedback. *The Journal of Acoustical Society of America*, 22, pp. 824-826.

Levelt W. J. M. (1989). *Speaking. From intention to articulation*. MIT Press, Cambridge, Massachusetts.

Levelt W. J. M. (1994). On the skill of speaking. *Proceedings of International Congress on Speech and Language Processing*, Yokohama, pp. 2253-2258.

Levelt W. J. M., Richardson G. & La Heij W. (1985). Pointing and voicing in deictic expressions. *Journal of Memory and Language*, 24, pp. 133-164.

Levelt W. J. M. & Wheeldon L. (1994). Do speakers have access to a mental syllabary? *Cognition*, 50, pp. 239-269.

Leybaert J. (1996). La lecture chez l'enfant sourd : l'apport du Langage Parlé Complété. *Revue Française de Linguistique Appliquée*, Paris, AFLA, 1, pp. 81-94.

Leybaert J. (1998). Phonological representations in deaf children: The importance of early linguistic experience. *Scandinavian Journal of Psychology*, 39, pp. 169-173.

Leybaert J., Alegria J., Hage C. & Charlier B. (1998). The effect of exposure to phonetically augmented lipspeech in the prelingual deaf. In: Dodd B., Campbell R. & Burnham D. (Eds.), *Hearing By Eye II*, *Advances in the psychology of speechreading and auditory visual speech* (pp. 283-301), Hove (UK): Psychology Press.

Leybaert J. & Charlier B. (1996). Visual speech in the head: the effect of cued speech on rhyming, remembering and spelling. *Journal of Deaf Studies and Deaf Education*, 1, pp. 234-248.

Leybaert J. (2000). Phonology acquired through the eyes and speeling in deaf children. *Journal of Experimental Child Psychology*, 75, pp. 291-318.

Leybaert J. & Lechat J. (2001). Phonological similarity effects in memory for serial order of cued speech. *Journal of Speech, Language and Hearing Research*, 44, pp. 949-963.

Leybaert J. & Alegria J. (2003). The Role of Cued Speech in Language Development of Deaf Children. In: Marschark M. & Spencer P. E. (Eds), *Oxford Handbook of Deaf Studies, Language, and Education* (pp. 261-274). Oxford University Press.

Leybaert J. & Van Reybroeck M. (2004). L'évaluation de la conscience phonologique et des mécanismes de production écrite de mots : Que peuvent nous apprendre les enfants sourds et les enfants dysphasiques? In: Metz-Lutz M.-N., Demont E., Seegmuller C., de Agostini M. & Bruneau N. (Eds.), *Développement cognitif et trouble des apprentissages : évaluer, comprendre, rééduquer et prendre en charge* (pp. 193-218), Solal, Marseille.

Liberman A. M. (1982). On finding that speech is special. *American Psychologist*, 37 (2), pp. 148-167. (Reprinted In: *Handbook of Cognitive Neuroscience*, ed by Michael S. Gazzaniga. (1984) Plenum Press: New York, pp. 169-197.)

Liberman A. M. & Mattingly I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21, pp. 1-36.

Ling D. & Clarke B. R. (1975). Cued Speech: An evaluative study. *American Annals of the Deaf*, 120, pp. 480-488.

Löfqvist A. (1991). Proportional timing in speech motor control. *Journal of Phonetics*, 19, pp. 343-350.

Luettin J. & Dupont S. (1998). Continuous audio-visual speech recognition. *Proceeding of 5th European Conference on Computer Vision (ECCV-98)*, volume II of *Lecture Notes in Computer Science*, (pp. 657–673), Springer Verlag.

MacDonald J. & McGurk H. (1978). Visual influences on speech perception processes. *Perception and Psychophysics*, 24, pp. 253-257.

MacLeod A. & Summerfield Q. (1987). Quantifying the contribution of vision to speech perception in noise. *British Journal of Audiology*, 21, pp. 131-141.

MacLeod A. & Summerfield Q. (1990). A procedure for measuring auditory and audio-visual speech-reception thresholds for sentences in noise: Rationale, evaluation and recommendations for use. *British Journal of Audiology*, 24, pp. 29-43.

MacNeilage P. (1970). Motor control of serial ordering of speech. *Psychological Review*, 77 (3), pp. 182-196.

MacNeilage P. (1998). The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences*, 21, pp. 499-546.

Marchal A. & Farnetani E. (1993) (Guest Eds.). Special issue on coarticulation, *Language and Speech*, 36, parts 2 & 3.

Massaro D. W. (1987a). *Speech perception by ear and eye: A paradigm for psychological inquiry.* Hillsdale, NJ: Erlbaum.

Massaro D. W. (1987b). Speech perception by ear and eye. In: Dodd B. & Campbell R. (Eds.), *Hearing by eye: the psychology of lipreading* (pp. 53-83). Lawrence Erlbaum Associates, London.

Massaro D. W. (2001). Perceiving the many modalities of spoken language: Theories and data. In: Cavé C., Guaïtella I & Santi S. (Eds.), *Oralité et gestualité. Interactions et comportements multimodaux dans la communication* (pp. 34-37). L'Harmattan, Paris.

Maunoury B. (2004). Formation au métier de codeur : information sur le projet de mise en place de la licence professionnelle. *Communication pendant la Rencontre nationale des codeurs en LPC*, Lyon, 14-15 Mai 2004.

McClave E. (1998). Pitch and manual gestures. *Journal of Psycholinguistic Research*, 1, pp. 69-89.

McGurk H. & MacDonald J. (1976). Hearing lips and seeing voices. *Nature*, 264, pp. 746-748.

McNeill D. (1985). So you think gestures are nonverbal. *Psychological Review,* 92, pp. 350-371.

McNeill D. (1992). *Hand and mind. What gestures reveal about thought*. The University of Chicago Press.

McNeill D., Levy E. T. & Pedelty L. L. (1990). Speech and gesture. In: Hammond G. R. (Ed.), *Advances in psychology: Cerebral control of speech and limb movements* (pp. 203-256). Amsterdam: Elsevier/North Holland Publishers.

McNeill D., Quek F., McCullough K.-E., Duncan S., Furuyama N., Bryll R. & Ansari R. (2001). Catchments, prosody and discourse. In: Cavé C., Guaïtella I & Santi S. (Eds.), *Oralité et gestualité. Interactions et comportements multimodaux dans la communication* (pp. 474-481). L'Harmattan, Paris.

McQueen J. M. (1995). Processing versus representation: Comments on Ohala and Ohala. In: Connell B. & Arvaniti A. (Eds.), *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV* (pp. 61-67). Cambridge: Cambridge University Press.

Mehler J., Dommergues J. Y., Frauenfelder U. H., & Segui J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior,* 20**,** pp. 298–305.

Mehler J., Dupoux E., & Segui J. (1990). Constraining models of lexical access: The onset of word recognition. In: Altmann G. T. M. (Ed.), *Cognitive models of speech processing* (pp. 236–262). Cambridge, MA: MIT Press.

Meyer A. S. (1990). The time course of phonological encoding in language production: The encoding of successive syllables of a word. *Journal of Memory and Language*, 29, pp. 524-545.

Meyer A. S. (1991). The time course of phonological encoding in language production: The phonological encoding inside a syllable. *Journal of Memory and Language*, 30, pp. 69-89.

Meyer A. S. & Schriefers H. (1991). Phonological facilitation in picture-word interference experiments: Effects of stimulus onset asynchrony and types of interfering stimuli. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 17, pp. 1145-1160.

Meyer D. E., Smith J. E. K., Kornblum S., Abrams R. A. & Wright C. E. (1990). Speed-accuracy tradeoffs in aimed movements: Toward a theory of rapid voluntary action. In: Jeannerod M. (Ed.), *Attention and Performance XIII, Motor representation and Control* (pp. 173-226). Hillsdale, NJ: Lawrence Erlbaum.

Montgomery A. A. & Jackson P. L. (1983). Physical characteristics of the lips underlying vowel lipreading performance. *The Journal of Acoustical Society of America*, 73(6), pp. 2134-2144.

Montgomery A. A., Walden B. E. & Prosek R. A. (1987). Effects of consonantal context on vowel lipreading. *Journal of Speech and Hearing Research*, 30, pp. 50-59.

Morrel-Samuels P. & Krauss R. M. (1992). Word familiarity predicts temporal asynchrony of hand gestures and speech. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 18 (3), pp. 615-622.

Munhall K. G. (1985). An examination of intra-articulator relative timing. *The Journal of Acoustical Society of America*, 78 (5), pp. 1548-1553.

Munhall K. G. & Tohkura Y. (1998). Audiovisual gating and the time course of speech perception. *The Journal of Acoustical Society of America*, 104 (1), pp. 530-539.

National Cued Speech Association of America (NCSA) (1994). Guidelines on the mechanics of cueing. *Cued Speech Journal V*, pp. 73-80.

Nicholls G. & Ling D. (1982). Cued Speech and the reception of spoken language. *Journal of Speech and Hearing Research*, 25, pp. 262-269.

O'Donoghue G. (2003). Les defies poses par l'incidence du dépistage néonatal systématique de la surdité et l'implantation cochléaire pédiatrique. *Actes du Colloque International ACFOS IV*, 8-10 Novembre 2002, pp. 116-119.

O'Shaughnessy D. (2000). *Speech Communications. Human and machine*. IEEE Press, Piscataway, NJ, USA.

Odisio M. & Bailly G. (2004). Shape and appearance models of talking faces for model-based tracking. *Speech Communication*, 44, pp. 63-82.

Ohala J. J. & Ohala M. (1995). Speech perception and lexical representation: The role of vowel nasalization in Hindi and English. In: Connell B. & Arvaniti A. (Eds.), *Phonology and Phonetic Evidence. Papers in Laboratory Phonology IV* (pp. 41-60). Cambridge: Cambridge University Press.

Öhman S. E. G. (1966). Coarticulation in VCVs utterances: Spectrographic measurements. *The Journal of Acoustical Society of America*, 39, pp. 151-168.

Öhman S. E. G. (1967). Numerical model of coarticulation. *The Journal of Acoustical Society of America*, 41 (2), pp. 310-320.

Owens E. & Blazek B. (1985). Visemes observed by hearing-impaired and normal hearing adult viewers. *Journal of Speech and Hearing Research*, 28, pp. 381-393.

Özyürek A. (2001). What do speech-gesture mismatches reveal about speech and gesture integration? A comparison between English and Turkish. In: Cavé C., Guaïtella I & Santi S. (Eds.), *Oralité et gestualité. Interactions et comportements multimodaux dans la communication* (pp. 576-581). L'Harmattan, Paris.

Pashler H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin*, 116 (2), pp. 220-244.

Périer O., Charlier B., Hage C. & Alegria J. (1990). Evaluation of the effects of prolonged Cued Speech practice upon the reception of spoken language. *Cued Speech Journal*, 4, pp. 47-59. Egalement apparu dans: Taylor I. G. (Ed.), *The education of the deaf: Current perspectives*, 1, 1987 (pp. 616-628), London: Croom Helm.

Perkell J. S. (1986). Coarticulation strategies: Preliminary implications of a detailed analysis of lower lip protrusion movements. *Speech Communication*, 5, pp. 47-68.

Perkell J. S. (1990). Testing theories of speech production: Implications of some detailed analyses of variable articulatory data. In: Hardcastle W. J. & Marchal A. (Eds.), *Speech Production and speech modelling* (pp. 263-288). Dordrecht/ Boston/ London : Kluwer Academic Publishers.

Perkell J. S. & Chiang C. (1986). Preliminary support for a "hybrid model" of anticipatory coarticulation. *Proceedings of the 12th International Congress of Acoustics*, A3-6.

Perkell J. S. & Matthies M. L. (1992). Temporal measures of anticipatory labial coarticulation for the vowel /u/: Within- and cross-subject variability. *The Journal of Acoustical Society of America*, 91 (5), pp. 2911-2925.

Perkell J. S., Zandipour M., Matthies M. L. & Lane H. (2002). Economy of effort in different speaking conditions. I. A preliminary study of intersubject differences and modeling issues, *Journal of Acoustical Society of America*, 112, pp. 1627-1641.

Perrier P., Payan Y. & Marret R. (2004). Modéliser le physique pour comprendre le contrôle : le cas de l'anticipation en production de parole. In : Sock R. & Vaxelaire B. (Eds.), *L'anticipation à l'horizon du présent* (pp. 159-177), Mardaga, Belgique.

Picheny M. A., Durlach N. I. & Braida L. D. (1986). Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. *Journal of Speech and Hearing Research*, 29, pp. 434-446.

Pietrosemoli L., Mora E. & Blondet M. A. (2001). Synchronisation des mouvements des mains et de la ligne de fréquence fondamentale lors des incises en espagnol parlé. In: Cavé C., Guaïtella I & Santi S. (Eds.), *Oralité et gestualité. Interactions et comportements multimodaux dans la communication* (pp. 492-495). L'Harmattan, Paris.

Recasens D. (1984) Vowel-to-vowel coarticulation in Catalan VCV sequences. *Journal of the Acoustical Society of America*, 76, pp. 1624-1635.

Recasens D. (1999). Acoustic analysis. In: Hardcastle W. J. & Hewlett N. (Eds*), Coarticulation. Theory, Data and Techniques* (pp. 322-336). Cambridge University Press.

Reisberg D., McLean J. & Goldfield A. (1987). Easy to hear but hard to understand: A lipreading advantage with intact auditory stimuli. In: Dodd B. & Campbell R. (Eds.), *Hearing by eye: The psychology of lipreading* (pp. 97-113). Lawrence Erlbaum Associates, London.

Revéret L., Bailly G. & Badin P. (2000). MOTHER: a new generation of talking heads providing a flexible articulatory control for video-realistic speech animation. *International Conference on Speech and Language Processing*, Beijing, China, pp. 755-758.

Robert-Ribes J. (1995). *Modèles d'intégration audiovisuelle de signaux linguistiques : de la perception humaine à la reconnaissance automatique des voyelles*. Thèse de Doctorat, INPG, Spécialité Signal, Image, Parole.

Robert-Ribes J., Schwartz J.-L., Lallouache M. T. & Escudier P. (1998). Complementarity and synergy in bimodal speech: Auditory, visual and audio-visual identification of French oral vowels in noise. *The Journal of Acoustical Society of America*, 103 (6), pp. 3677-3689.

Rossato S., Badin P. & Bouaouni F. (2003). Velar movements in French: An articulatory and acoustical analysis of coarticulation. Proceedings of the 15th International Congress on Phonetic Sciences, Barcelona, Spain, pp. 3141-3144.

de Ruiter J. P. & Wilkins D. P. (1998). The synchronisation of gesture and speech in Dutch and Arrernte (an australian aboriginal language): a cross-cultural comparison. In: Santi G. et al. (Eds.), *Oralité et Gestualité : Communication multimodale, Interaction* (pp. 603-607). Paris : L'Harmattan.

de Ruiter J. P. (2000). The production of gesture and speech. In: McNeill D. (Ed.), *Language and gesture* (pp. 284-311), Cambridge University Press.

Sato M., Schwartz J.-L., Cathiard M.-A., Abry C. & Loevenbruck H. (2002). Intrasyllabic articulatory control constraints in verbal working memory. *Proceedings of the VIIth International Congress of Speech and Language Processes*, Sept. 16-20, Denver, USA, pp. 669-672.

Sato M. (2004). *Représentations verbales multistables en mémoire de travail : vers une perception active des unités de parole*. Thèse de Doctorat de Sciences Cognitives, Institut National Polytechnique de Grenoble, Institut de la Communication Parlée.

Shannon C. E. & Weaver W. (1949). *The mathematical theory of communication*. Urbana IL: University of Illinois Press.

Scheffé H. (1959). *The Analysis of Variance*. Wiley J. & Sons (Ed.), Wiley Publication in Mathematical Statistics (New-York).

Schmidt R. A. (1982). *Motor control and learning*. Champaign, IL: Human Kinetics Publishers.

Schmidt R. A. (1988). *Motor control and learning: A behavioural emphasis*. Champaign, IL: Human Kinetics Publishers.

Schwartz J.-L., Beautemps D., Arrouas Y. & Escudier P. (1992). Auditory analysis of speech gestures. In: Schouten M.E.H. (Ed.), *The Auditory Processing of Speech. From Sounds to Words* (pp. 239-252). Berlin: Mouton de Gruyter.

Schwartz J.-L., Robert-Ribes J. & Escudier P. (1998). Ten years after Summerfield: a taxonomy of models for audio-visual fusion in speech perception. In: Dodd B., Campbell R. & Burnham D. (Eds.), *Hearing By Eye II*, *Advances in the psychology of speechreading and auditory visual speech* (pp. 85-108), Hove (UK): Psychology Press.

Schwartz J.-L. (2001). Une théorie de la perception pour le contrôle de l'action. In : Keller D., Durafour J.P., Bonnot J.F.P. & Sock R. (Eds.), *Percevoir : monde et langage. Invariance et variabilité du sens vécu* (pp. 261-270). Sprimont : Mardaga.

Schwartz J.-L., Teissier P., Escudier P. (2002a). La parole multimodale: deux ou trois sens valent mieux qu'un. In: Mariani, J. J. (Ed.), *Traitement automatique du langage parlé 2 : reconnaissance de la parole* (pp. 141–178). Hermes, Paris.

Schwartz J.-L., Abry C., Boë L.-J. & Cathiard M.-A. (2002b). Phonology in a theory of Perception-for-Action-Control. In: Durand J. & Laks B. (Eds.), *Phonetics, Phonology and Cognition* (pp. 254-280), Oxford University Press.

Schwartz J.-L. (2004). La parole multisensorielle : Plaidoyer, problèmes et perspectives. Actes des *XXVèmes Journées d'Etude sur la Parole*, Fès, Maroc, 19-22 Avril, pp. xi-xvii.

Segui J. & Ferrand L. (2000). *Leçons de parole*. Odile Jacob (Ed.), Paris.

Serniclaes W. (2005). Formation CNRS aux statistiques multivariées, 19-22 Avril.

Sevald C. A., Dell G. S. & Cole J. S. (1995). Syllable structure in speech production: Are syllables chunks or schemas? *Journal of Memory and Language*, 34, pp. 807-820.

Seyfeddinipur M. & Kita S. (2001). Gestures and dysfluencies in speech. In: Cavé C., Guaïtella I & Santi S. (Eds.), *Oralité et gestualité. Interactions et comportements multimodaux dans la communication* (pp. 266-270). L'Harmattan, Paris.

Smits R., Warner N., McQueen J. M. & Cutler A. (2003). Unfolding of phonetic information over time: A database of Dutch diphone perception. *Journal of the Acoustical Society of America*, 113, pp. 563-574.

Sock R. (1998). *Organisation temporelle en production de la parole. Emergence de catégories sensori-motrices phonétiques*. Thèse de Doctorat d'Etat en Lettres et Sciences Humaines, Linguistique, spécialité Phonétique expérimentale et motrice, Institut de la Communication Parlée, Université Stendhal, Grenoble.

Sock R. & Vaxelaire B. (2004). Le diable perceptif dans les détails sensori-moteurs anticipatoires. In : Sock R. & Vaxelaire B. (Eds.), *L'anticipation à l'horizon du présent* (pp. 141-157), Mardaga, Belgique.

Sorokin V. N., Gay T. & Ewan W. G. (1980). Some biomechanical correlates of the jaw movements, *Journal of the Acoustical Society of America*, Suppl. 1, 68, S 32.

Stevens K. N. & House A. S. (1955). Development of quantitative description of vowel articulation. *The Journal of Acoustical Society of America*, 27 (3), pp. 484-493.

Stevens K. N. (1989). On the quantal nature of speech. *Journal of Phonetics,* 17, pp. 3-46.

Sumby W. H. & Pollack I. (1954). Visual contribution to speech intelligibility in noise. *The Journal of Acoustical Society of America*, 26 (2), pp. 212-215.

Summerfield Q. (1979). Use of visual information for phonetic perception. *Phonetica*, 36, pp. 314-331.

Summerfield Q. (1983). Audio-visual speech perception, lipreading and artificial stimulation. In: M. E. Lutman & M. P. Haggard (Eds.), *Hearing Science and Hearing Disorders* (pp. 131-182). Academic Press: London.

Summerfield Q. (1987). Some preliminaries to a comprehensive account of audio-visual speech perception. In: Dodd B. & Campbell R. (Eds.), *Hearing by eye: the psychology of lipreading* (pp. 3-51). Lawrence Erlbaum Associates, London.

Summerfield Q. & McGrath M. (1984). Detection and resolution of audio-visual incompatibility in the perception of vowels. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 36 (1-A), pp. 51-74.

Summers J. J. (1990). Temporal constraints on concurrent task performance. In: Hammond G. E. (Ed.), *Cerebral control of speech and limb movements* (pp. 661-680). Elsevier Science Publishers B. V., North-Holland.

Treffner P. & Peter M. (2002). Intentional and attentional dynamics of speech-hand coordination. *Human Movement Science*, 21, pp. 641-697.

Tseva A. & Cathiard M.-A. (1990). Paroles vues : la dimension d'arrondissement dans l'identification visuelle des voyelles du français. Actes du 1er Congrès Français d'Acoustique, *Colloque de Physique*, Colloque C2, suppl. 2, 51, pp. 507-510.

Tuller B. & Kelso J. A. S. (1984). The timing of articulatory gestures: Evidence for relational invariants. *Journal of the Acoustical Society of America*, 76 (4), pp. 1030-1036.

Uchanski R. M., Delhorne L. A., Dix A. K., Braida L. D., Reed C. M. & Durlach N. I. (1994). Automatic speech recognition to aid the hearing impaired: Prospects for the automatic generation of cued speech. *Journal of Rehabilitation Research and Development*, 31(1), pp. 20-41.

Vallée N., Boë L.-J., Maddieson I. & Rousset I. (2000). Des lexiques aux syllabes des langues du monde. Typologies et structures. *XXIIIèmes Journées d'Etude sur la Parole*, Aussois, France, pp. 93-96.

de la Vaux S. K. & Massaro D. W. (2004). Audiovisual speech gating: examining information and information processing. *Cognitive Process*, 5, pp. 106-112.

Vilaclara G. (1988). Speech processing to aid the profoundly deaf. *Journal of Acoustical Society of America*, Suppl. 1, 84, S41.

Vilain A. (2000). *Apports de la modélisation des degrees de liberté articulatoires à l'étude de la coarticulation et du développement de la parole.* Thèse de Doctorat de 3ème cycle, discipline Sciences du Langage, Institut de la Communication Parlée, Université Stendhal, Grenoble.

Vilain A., Abry C. & Badin P. (2000). Coproduction strategies in french VCVs: Confronting Öhman's model with adult and developmental articulatory data. *Proceedings of the 5th Seminar on Speech Production. Models and Data*, Kloster Seon, Bavaria, pp. 81-84.

Vilain A., Abry C., Badin P. & Brosda S. (1999). From idiosynchratic pure frames to variegated babbling: Evidence from articulatory modelling. *Proceedings of the International Congress of Phonetic Sciences*, pp. 2497-2500.

Viviani P. & Terzuolo C. (1980). Space-time invariance in learned motor skills. In: Stelmach G.E. & Requin J. (Eds.), *Tutorials in motor behaviour* (pp. 525-536). Amsterdam: North-Holland.

Walden B. E., Prosek R. A., Montgomery A. A., Scherr C. K. & Jones C. J. (1977). Effects of training on the visual recognition of consonants. *Journal of Speech and Hearing Research*, 20, pp. 130-145.

Walden B. E., Erdman S. A., Montgomery A. A., Schwartz D. M. & Prosek R. A. (1981). Some effects of training on speech recognition by hearing-impaired adults. *Journal of Speech and Hearing Research*, 24, pp. 207-216.

Warren P. & Marslen-Wilson W. (1987). Continuous uptake of acoustic cues in spoken word recognition. *Perception and Psychophysics*, 41, pp. 262-275.

Warren P. & Marslen-Wilson W. (1988). Cues to lexical choice: Discriminating place and voice. *Perception and Psychophysics*, 43, pp. 21-30.

Whalen D. H. (1990). Coarticulation is largely planned. *Journal of Phonetics*, 18, pp. 3-35.

Wheeldon L. R. & Levelt W. J. M. (1995). Monitoring the time course of phonological encoding. *Journal of Memory and Language*, 34, pp. 311–334.

Winter D. A. (1990). *Biomechanics and motor control of human movement*. Wiley-Interscience Publication, John Wiley & Sons, Inc., USA.

Woodward M. F. & Barber C. G. (1960). Phoneme perception in lipreading. *Journal of Speech and Hearing Research*, 3(3), pp. 212-222.

Wouts W. (1982). L'AKA. In : Destombes F. (Ed.), *Aides manuelles à la lecture labiale et perspectives d'aides automatiques* (pp. 16-29). Centre scientifique IBM-France, Paris.

Wozniack V. D. & Jackson P. L. (1979). Visual vowel and diphthong perception from two horizontal viewing angles. *Journal of Speech and Hearing Research*, 22, pp. 355-365.

Yakel D. A., Rosenblum L. D. & Fortier M. A. (2000). Effects of talker variability on speechreading. *Perception & Psychophysics*, 62 (7), pp. 1405-1412.

Liste des publications

Articles et chapitres

Attina V., Cathiard M.-A. & Beautemps D. (accepté). Temporal measures of hand and speech coordination during French Cued Speech production. *Lecture Notes in Artificial Intelligence, LNAI/LNCS*, Springer Verlag.

Attina V., Beautemps D., Cathiard M.-A. & Odisio M. (2004). A pilot study of temporal organization in Cued Speech production of French syllables: Rules for a Cued Speech synthesizer. *Speech Communication*, vol. 44, pp. 197-214.

Beautemps D., Cathiard M.-A., Attina V., Savariaux C. & Arnal A. (à paraître). Temporal organization of Cued Speech production. In Festschrift Christian Benoît, G. Bailly, P. Perrier & E. Vatiokis-Bateson, (Eds), *Audiovisual Speech Processing*: MIT Press.

Cathiard M.-A., Attina V., Abry C. & Beautemps D. (2004 - à paraître). La Langue française Parlée Complétée (LPC) : sa coproduction avec la parole et l'organisation temporelle de sa perception. *Revue PArole*, n° 29/30/31, N° spécial : « Handicap langagier et recherches cognitives : apports mutuels », J.-L. Nespoulous & J. Virbel (Eds.).

Conférences invitées

Attina V. (2005). Organisation temporelle de la perception du code LPC. *Journées d'Etude de l'ALPC,* Paris, 21-22 Mai 2005.

Attina V. (2004). Cued Speech production: giving a hand to speech acoustics. *CFA/DAGA'04*, Strasbourg, France, 22-25 Mars 2004.

Actes de congrès avec comité de lecture

Attina V., Cathiard M.-A. & Beautemps D. (2005). French Cued Speech production : giving a hand to speech. *2e Congrès de l'International Society for Gesture Studies (ISGS): Interacting Bodies-Corps en Interaction*, Juin 2005.

Attina V., Cathiard M.-A. & Beautemps D. (2005). Temporal measures of hand and speech coordination during French Cued Speech production. *The 6th International Workshop on Gesture in Human-Computer Interaction and Simulation (GW'2005)*, Mai 2005.

Attina V., Cathiard M.-A & Beautemps D. (2004). L'ancrage de la main sur les lèvres : Langue Française Parlée Complétée et anticipation vocalique. *XXVèmes Journées d'Etude sur la Parole*, Fès, Maroc, 19-22 Avril.

Cathiard M.-A, Bouaouni F., Attina V. & Beautemps D. (2004). Etude perceptive du décours de l'information manuo-faciale en Langue Française Parlée Complétée. *XXVèmes Journées d'Etude sur la Parole*, Fès, Maroc, 19-22 Avril.

Attina V., Beautemps D., Cathiard M.-A. & Odisio M. (2003). Toward an audiovisual synthesizer for Cued Speech: Rules for CV French syllables. *Proceedings of AVSP*, St Jorioz, pp. 227-232.

Attina V., Beautemps D. & Cathiard M.-A. (2003). Temporal motor organization of Cued Speech gestures in the French language. *Proceedings of the 15th ICPHS*, 4-10 August 2003, Barcelona, Spain, pp. 1935-1938.

Cathiard M.-A., Attina V. & Alloatti D. (2003). Labial Anticipation Behavior during Speech with and without Cued Speech. *Proceedings of the 15th ICPHS*, 4-10 August 2003, Barcelona, Spain, pp. 1939-1942.

Attina V., Beautemps D. & Cathiard M.-A. (2003). Temporal organization of French Cued Speech production. *Actes de COG-Speech OMLL* (*Origin of Man, Language and Languages*) *Conference. Vocalise to localise: a missing piece in the puzzling route towards language,* Grenoble, France, January 30-31.

Attina V., Beautemps D. & Cathiard M.-A. (2002). Coordination of hand and orofacial movements for CV sequences in french Cued Speech. *Proceedings of ICSLP*, Denver, Colorado, pp. 1945-1948.

Attina V., Cathiard M.-A. & Beautemps D. (2002). Controling anticipatory behavior for rounding in french Cued Speech. *Proceedings of ICSLP*, Denver, Colorado, pp. 1949-1952.

Attina V., Beautemps D. & Cathiard M.-A. (2002). Organisation spatio-temporelle main-lèvres-son de séquences CV en Langage Parlé Complété. *XXIVèmes Journées d'Etude sur la Parole*, Nancy, 24-27 Juin 2002, pp. 241-244.

Attina V., Cathiard M.-A. & Beautemps D. (2002). Contrôle de l'anticipation vocalique d'arrondissement en Langage Parlé Complété. *XXIVèmes Journées d'Etude sur la Parole*, Nancy, 24-27 Juin 2002, pp. 161-164.

Colloques sans acte

Attina V., Cathiard M.-A. & Beautemps D. (2002). Parole audio-visuelle et Langage Parlé Complété : Etude des coordinations main-lèvres-son. *Journées ACFOS*, Paris, 8-10 Novembre 2002.

Attina V., Cathiard M.-A. & Beautemps D. (2002). Parole audio-visuelle et Langage Parlé Complété : Etude des coordinations main-lèvres-son. *Journée scientifique du PRASC*, Archamps, 8 Mars 2002.

Table des figures

Figure 1. Architectures de base pour la fusion audiovisuelle proposées par Schwartz et al. (1998) : (a) identification directe, (b) identification séparée, (c) recodage dans la modalité dominante et (d) recodage dans la modalité motrice. Figure tirée de Schwartz et al., 2002a. 10

Figure 2. Taxinomie des modèles de fusion. Figure tirée de Schwartz et al., 2002a. 11

Figure 3. Clés manuelles du Cued Speech conçu par Cornett pour l'anglais-américain 17

Figure 4. En haut : codage de la syllabe [pi] (la configuration n°1 est utilisée pour coder la consonne [p] en pointant la « bouche », position codant la voyelle [i], voir Figure 3) ; en bas, codage du mot « left » impliquant la même clé à la suite. 18

Figure 5. Clés manuelles de la LPC pour le français. Adaptation du Cued Speech. 22

Figure 6. Codage manuel du mot « structure » en LPC suivant la décomposition syllabique C.C.CV.C.CV.C 22

Figure 7. Exemple d'expression codée avec et sans liaison, « mon avion ». A gauche, le [n] de liaison est codé par la configuration appropriée avec la main en position « côté » pour la voyelle [a] associée à cette consonne. 22

Figure 8. Schématisation du modèle d'Öhman montrant l'évolution dans le temps de l'aire aux lèvres pour un geste de voyelle à voyelle (transition de la voyelle [y] vers la voyelle [i] ; à gauche) et pour un geste vocalique avec consonne surimposée (transition [ybi] ; à droite) (figure tirée de Cathiard, 2003). 22

Figure 9. Contours sagittaux pour la production de [u], [b], [u] dans [ubu]. Figure tirée de Vilain, 2000. 22

Figure 10. Contours sagittaux pour la production de [a], [b], [a] dans [aba]. Figure tirée de Vilain, 2000. 22

Figure 11. Représentation schématique des différentes prédictions des trois modèles d'anticipation du geste de protrusion labiale pour une séquence du type [iC_ny] (signal acoustique schématisé en haut). Chaque tracé représente la protrusion labiale de la lèvre supérieure en fonction du temps. Les points noirs indicés de v=0^+ correspondent au début du geste de protrusion, déterminé au moment où la vitesse de la courbe passe à 0. Le point noir indicé de γMax indique le moment du pic d'accélération, qui détermine le début de la $2^{ème}$ phase du geste de protrusion dans le modèle hybride. (Figure tirée de Cathiard, 1994). 22

Figure 12. MEM de protrusion en relation avec les modèles Look-Ahead et Time-Locked : représentation de la durée du mouvement (MT) en fonction de l'intervalle d'obstruence IO. La durée de 140 ms désigne la durée minimale incompressible du geste de protrusion (pour une durée de IO de 0 et 100 ms). La pente de la droite représente le coefficient d'expansion du mouvement. Ce coefficient est propre au locuteur (ici, cas de trois locuteurs : Annie, Jean-Luc et Benny). Figure tirée de Abry et al., 1996a. 22

Figure 13. Signal acoustique et évolution temporelle de l'aire aux lèvres pour la séquence [sedøsikstkyltɛʁ]. Les événements temporels suivants sont repérés : (1) correspond au maximum de [i], (2) au 90%aire.on, (3) à 10%aire.on, (4) au minimum de [y], (5) à 10%aire.off, (6) à la fin acoustique du [i] et (7) au début acoustique du [y]. D'après C. Abry, publié dans Cathiard et al., 2003. 22

Figure 14. MEM de constriction : représentation de la phase de Time-Falling+Hold en fonction de la durée de l'intervalle d'obstruence pour quatre locuteurs. Figure tirée de Abry et al., 1996a. 22

Figure 15. Décours temporel du cycle de volée en correspondance avec le signal acoustique de parole. En haut, portion de signal acoustique [sa] extraite de la formulette d'incantation. En bas, évolution au cours du temps de l'angle entre la phalange du sujet et la lame du couteau avec les trois phases du cycle du geste de volée. Sur le signal acoustique est repéré le moment où se produit la percussion (coup), soit durant la consonne. Figure tirée de Berthier et al., 1991. 22

Figure 16. Les étapes de la production de mots dans le modèle « Speaking ». Figure adaptée de Levelt, 1994. 22

Figure 17. Détail de l'étape d'encodage phonologique dans le modèle « Speaking » + encodage phonétique et articulation. Adaptée de Levelt & Wheeldon, 1994 (voir également Segui & Ferrand, 2000). 22

Figure 18. Le Sketch Model. Figure adaptée de de Ruiter, 2000. 22

Figure 19. Photo de la locutrice-codeuse lors de l'enregistrement avec les positions des pastilles colorées sur la main et sur le visage. 22

Figure 20. Tracé des différents signaux au cours du temps pour la portion [tatuta] extraite de la séquence [tatatuta]. De haut en bas : (1) décours temporel de l'aire aux lèvres S (cm^2), avec en dessous (2) le profil d'accélération correspondant ; (3) tracé de la position de la coordonnée x de la pastille sur le dos de la main (cm) avec (4) son profil d'accélération ; (5) tracé de la position de la coordonnée y de la pastille sur le dos de la main (cm) avec (6) son profil d'accélération ; (7) signal acoustique correspondant. Les signaux en traits pointillés correspondent aux signaux bruts et les signaux en traits pleins correspondent aux signaux filtrés que nous utilisons pour calculer le profil d'accélération nécessaire à l'étiquetage (voir texte). 22

Figure 21. Positionement temporel des événements M1 (début du geste manuel), M2 (fin du geste manuel), L2 (cible labiale vocalique) et M3 (début du geste manuel vers la position suivante) par rapport à A1 (repéré par 0), début acoustique de la syllabe, pour les différents types de séquences (avec et sans consonne). Les items sont ordonnés dans l'ordre croissant de M1 (ce qui correspond à une courbe de fréquence cumulée avec ici la variable en y et le rang en x). 22

Figure 22. Schéma de la coordination temporelle main-lèvres-son obtenue dans l'expérience 1 pour les deux conditions pour la codeuse GB. En haut, schéma temporel des résultats de la condition VCV pour la production de syllabe CV. En bas, schéma temporel des résultats de la condition V-V. 22

Figure 23. Code de la séquence [mabama]. Toute la séquence est codée en position de main sur le « côté ». Durant la séquence, la configuration consonantique est changée pour le codage de la consonne [b] (passage de la configuration n°5 à la n°4, puis retour à la n°5) : ceci est caractérisé par un geste d'effacement puis de réapparition du pouce. 22

Figure 24. Code de la séquence [mabuma]. La séquence implique à la fois un changement de clé consonantique pour la consonne [b] (disparition du pouce) et un déplacement de la main (de la position « côté » vers la position « menton »). 22

Figure 25. Photo de la locutrice-codeuse portant le gant de données et positions des pastilles de couleur utilisées pour le suivi des mouvements. Le repère référentiel utilisé est tracé en superposition sur la photo. 22

Figure 26. Tracé des différents signaux pour la séquence [mabema]. De haut en bas : (1) décours temporel de l'aire aux lèvres (cm^2) ; (2) position de la coordonnée x de la pastille sur le dos du gant au cours du temps (cm) ; (3) position de la coordonnée y de la pastille sur le dos du gant au cours du temps (cm) ; (4) données brutes issues du capteur du gant (capteur sur le pouce) (en pointillés, données non filtrées ; en trait plein, données filtrées pour le calcul de

l'accélération), (5) avec en-dessous le profil d'accélération correspondant ; (6) signal acoustique. Sur chacun des signaux sont superposés les étiquettes et les intervalles étudiés (voir texte). 22

Figure 27. Positionnement temporel des événements suivants par rapport à A1 (repéré par 0), début acoustique de la consonne, pour les deux conditions avec transitions de main (corpus 2, à droite) et sans transition de main (corpus 1, à gauche) : D1 (début de la configuration des doigts), D2 (fin de la formation de la configuration), M1 (début de la transition de main), M2 (atteinte de la position cible) et L2 (atteinte de cible labiale). Les items sont ordonnés dans l'ordre croissant de D1 pour le corpus 1 et de M1 pour le corpus 2. 22

Figure 28. Schéma temporel de coordination entre les doigts, les lèvres, le son et éventuellement la main résumant les résultats de l'expérience 2 dans les deux conditions pour la codeuse GB. En haut, schéma résumé de l'étude 1 des changements de configuration consonantique sans transition manuelle. En bas, schéma résumé de l'étude 2 des changements de configurations avec transitions manuelles. 22

Figure 29. Boîtes à moustaches représentant la distribution de chaque intervalle temporel dans le cycle syllabique pour la codeuse GB dans les expériences 1 et 2. Chaque boîte indique la médiane, les 1er et 3ème quartiles, les valeurs frontières, les valeur hors normes ('+') et la moyenne ('x'). 22

Figure 30. Boîtes à moustaches indiquant la distribution de chaque intervalle temporel pour l'étude de la configuration digitale relativement au cycle syllabique pour la codeuse GB dans l'expérience 2 (corpus 2). 22

Figure 31. Graphique en bâtons superposés représentant les durées (en $\%_{rel}$) des formations de configurations digitales (foncé) superposées sur les durées totales de transitions de main (clair) pour toutes les séquences. Les données sont triées dans l'ordre croissant de la durée M1M2. 22

Figure 32. Schéma de coordination temporelle de la main, des doigts et des lèvres en relation avec le son durant la production d'une syllabe CV codée en LPC chez le sujet GB dans le domaine syllabique. Les intervalles indiqués sur la figure sont extraits des expériences 1 et 2 (voir également le Tableau 2) : les durées sont exprimées en pourcentages relatifs à la durée de la syllabe acoustique CV (unité : $\%_{rel}$). 22

Figure 33. Photo de la locutrice-codeuse SC avec superposition du repère utilisé pour le calcul de la position de la pastille colorée sur le dos de la main. 22

Figure 34. Tracé des différents signaux pour la portion [ma.be.ma] extraite de la séquence [ma.ma.be.ma] de la locutrice-codeuse SC. De haut en bas : (1) décours temporel de l'aire aux lèvres S (cm^2) ; (2) position de la coordonnée x de la pastille sur le dos de la main au cours du temps (cm) ; (3) position de la coordonnée y de la pastille au cours du temps (cm) ; (4) signal acoustique correspondant. Sur chacun des signaux sont superposés les événements temporels (en gras) et les intervalles étudiés. 22

Figure 35. Schémas des patrons temporels de coordination main-lèvres-son pour chacune des codeuses AM, SC et RV. 22

Figure 36. Positionnement temporel (en ms) des événements M1 (début du geste manuel), M2 (fin du geste manuel), L2 (cible labiale vocalique) et M3 (début du geste manuel vers la position suivante) par rapport au début acoustique de la consonne A1, pour les trois codeuses AM, SC et RV. Pour faciliter la lecture des graphiques, les items sont triés dans l'ordre croissant en fonction de M1. 22

Figure 37. Evolution de l'intervalle A1M2 en fonction de l'intervalle M1A1 (ms) pour les trois codeuses AM, SC et RV. Sur chacun des graphes sont indiqués les coefficients de

corrélation (tous significatifs à p< .01) et les équations de droites ajustées linéairement sur les données. 22

Figure 38. Schémas des patrons temporels de coordination main-lèvres-son dans le domaine syllabique pour chacune des codeuses AM, SC et RV. Les valeurs indiquées correspondent aux durées d'intervalles exprimées en pourcentages relatifs de la durée de la syllabe CV (la durée de la syllabe s'élève donc à 100%$_{rel}$). 22

Figure 39. Evolution des durées de transitions manuelles M1M2 (ms) en fonction de la durée de la syllabe CV pour les trois codeuses. 22

Figure 40. Organisation temporelle des événements manuels et orofaciaux durant la production de code LPC en fonction du cycle syllabique pour les trois codeuses AM, SC et RV : M1 est le début du geste de transition manuelle (représenté par des triangles), M2 est la fin de la transition (représenté par des croix) et L2 est la cible labiale vocalique (représenté par des cercles). Les événements temporels sont positionnés en relatif (ms) par rapport au début acoustique de la syllabe A1 (représenté par la droite en traits tiretés à 0 ms). Ces graphes complètent ceux de la Figure 41 pour le positionnement relatif proportionnel. 22

Figure 41. Organisation temporelle relative des événements manuels et orofaciaux durant la production de code LPC en fonction du cycle syllabique pour les trois codeuses AM, SC et RV : M1 est le début du geste de transition manuelle (représenté par des triangles), M2 est la fin de la transition (représenté par des croix) et L2 est la cible labiale vocalique (représenté par des cercles). Les événements temporels positionnés par rapport au début acoustique de la syllabe A1 sont exprimés en pourcentages de la durée de la syllabe CV (%$_{rel}$). La droite en traits tiretés à 0%$_{rel}$ indique le début acoustique de la syllabe, soit l'événement A1. Les droites d'ajustement linéaire (au sens des moindres carrés) pour chacun des événements sont superposées sur chacun des graphiques. Les équations de droite et les coefficients de corrélations (Bravais-Pearson) sont également indiqués. 22

Figure 42. Boîtes à moustaches représentant la distribution de l'intervalle temporel A1M2 exprimé, pour chaque séquence, en pourcentage de la durée de la consonne acoustique, pour les trois codeuses AM, SC et RV. Chaque boîte indique la médiane, les 1er et 3ème quartiles, les valeurs frontières, les valeur hors normes ('+') et la moyenne ('x'). Les traits tiretés à 0 et à 100 % indiquent le début et la fin de la consonne acoustique. 22

Figure 43. Codage LPC de la phrase « T'as mis : UHI ise ? » 22

Figure 44. Codage LPC de la phrase « T'as mis : AHI ise ? » 22

Figure 45. Tracé des différents signaux au cours du temps pour la séquence [tami#yiiz] (pause courte). De haut en bas : (1) décours de l'aire intérolabiale *S* (cm^2) ; (2) position de la coordonnée *y* de la pastille sur le dos de la main (cm) ; (3) signal acoustique correspondant. Sur chacun des signaux sont superposés les événements temporels utilisés pour l'analyse. Afin de ne pas surcharger la figure, les intervalles temporels n'ont pas été indiqués ; ils se déduisent simplement grâce à la position des étiquettes. 22

Figure 46. Photo de la locutrice-codeuse GB lors de l'enregistrement avec les positions des pastilles colorées sur la main et sur le visage et le repère utilisé. 22

Figure 47. Durée de l'intervalle de pause IO pour l'ensemble des transitions [i#y]. Les séquences sont réparties en fonction de la consigne (petite pause – longue pause). 22

Figure 48. Organisation temporelle relative des événements manuels et labiaux dans la transition de [i] vers [y] pour la codeuse GB : M1 est le début du geste de transition manuelle, M2 est l'atteinte de la position du [y], L1 est le début du geste labial et L2 est la cible labiale vocalique. Les événements, repérés par rapport à la fin du [i], sont exprimés en pourcentages relatifs de la durée de l'intervalle de pause IO. 0% correspond donc à la fin du [i] et 100% indique le début du [y]. 22

Figure 49. Coordination temporelle main-lèvres pour le début des gestes manuel et labial (M1L1, ms) et pour l'atteinte des cibles (M2L2, ms) en fonction de la durée de l'intervalle de pause IO (ms) dans les transitions [i#y]. Les valeurs positives de ces deux graphes indiquent que l'événement manuel se produit avant l'événement labial. 22

Figure 50. Corrélations entre l'intervalle de pause IO et les gestes labial et manuel pour la transition [i#y] : en fonction de IO, la phase de constriction L1L2, le geste global d'arrondissement L1L3, la transition de main M1M2 et la clé LPC globale M1M3. Au-dessus de chaque cadran sont indiqués les coefficients de corrélations (tous significatifs à p< .01) et les équations des droites de régression linéaire. 22

Figure 51. Tracé des signaux au cours du temps pour une séquence [tami#aiiz] avec petite pause. De haut en bas : (1) décours de l'aire intérolabiale *S* (cm^2) ; (2) position de la coordonnée *x* de la pastille sur le dos de la main (cm) ; (3) signal acoustique correspondant. Sur chacun des signaux sont superposés les événements temporels utilisés pour l'analyse. 22

Figure 52. Durée de l'intervalle de pause IO pour l'ensemble des transitions [i#a]. Les séquences sont réparties en fonction de la condition (petite pause – longue pause). 22

Figure 53. Organisation temporelle relative des événements manuels et labiaux dans la transition de [i] vers [a] pour la codeuse GB : M1 est le début du geste de transition manuelle, M2 est l'atteinte de la position du [a], L1 est le début du geste labial et L2 est la cible labiale vocalique. Les événements, repérés par rapport à la fin du [i], sont exprimés en pourcentages relatifs de la durée de l'intervalle de pause IO. 0% correspond donc à la fin du [i] et 100% indique le début du [a]. 22

Figure 54. Coordination temporelle main-lèvres pour le début des gestes manuel et labial (M1L1, ms) et pour l'atteinte des cibles (M2L2, ms) en fonction de la durée de l'intervalle de pause IO (ms) dans les transitions [i#a]. Les valeurs positives de ces deux graphes indiquent que l'événement manuel se produit avant l'événement labial et inversement pour les valeurs négatives. 22

Figure 55. Ellipses de dispersion à un écart-type des positions cibles spatiales « côté » codant le [a] et « cou » codant le [y] pour la codeuse GB dans les transitions [i#y] et [i#a]. Les positions ont été calculées à partir de la position de la pastille du dos de la main (référencée dans le repère sur la lunette) au moment d'atteinte de position cible M2. La position moyenne de la pastille est indiquée dans chaque ellipse par une croix « x ». 22

Figure 56. Corrélations entre l'intervalle de pause IO et les gestes labial et manuel dans la transition [i#a] : en fonction de IO, la phase de constriction L1L2, le geste global d'arrondissement L1L3, la transition de main M1M2 et la clé LPC globale M1M3. Au-dessus de chaque cadran sont indiqués les coefficients de corrélations (tous significatifs à p< .01) et les équations des droites de régression linéaire. 22

Figure 57. Codage en LPC de la séquence « ces deux scies utèrent » [sedøsiytɛʁ]. 22

Figure 58. Codage en LPC de la séquence [sikssky] extraite de la phrase « ces deux **sixes sku**tèrent ». 22

Figure 59. Tracé des différents signaux au cours du temps pour la séquence [sedøsiskytɛʁ] à deux consonnes. De haut en bas : (1) décours de l'aire intérolabiale *S* (cm^2) ; (2) position de la coordonnée *x* de la pastille sur le dos de la main (cm) ; (3) position de la coordonnée *y* de la pastille sur le dos de la main (cm) et (4) signal acoustique correspondant. Sur chacun des signaux sont superposés les événements temporels utilisés pour l'analyse. 22

Figure 60. Photo de la locutrice-codeuse GB durant l'enregistrement avec indication de la pastille de la main et superposition du repère utilisé pour l'analyse des données. 22

Figure 61. Durée de l'intervalle IO pour l'ensemble des séquences [iC$_n$y] (avec n= 0 à 4 consonnes). 22

Figure 62. Organisation temporelle relative des événements labiaux dans la transition de [i] vers [y] pour les séquences avec et sans code LPC pour la codeuse GB : L1 est le début du geste de constriction et L2 l'atteinte de la cible labiale. Les événements sont référencés par rapport à la fin du [i] et sont exprimés en pourcentages relatifs de la durée de IO. 22

Figure 63. Phases du geste labial pour la voyelle [y] en fonction de l'intervalle IO pour les séquences avec (à gauche) et sans (à droite) code LPC : L1L2 est la phase de constriction (*Time-Falling*) et L1L3 la phase de constriction + la tenue de la cible labiale (*Time-Falling* + *Hold*). Les droites d'ajustement linéaire sont calculées sur toutes les données à l'exception de celles pour lesquelles IO= 0 ms. Les coefficients de corrélation et les équations des droites sont également indiqués. 22

Figure 64. Phases de Time-Falling (L1L2) et de Time-Falling+Hold (L1L3) pour la transition de la voyelle [i] à la voyelle [y] en fonction de l'intervalle IO pour les séquences avec et sans LPC mélangées. 22

Figure 65. Organisation temporelle relative des événements labiaux et manuels dans la transition de [i] vers [y] pour les séquences avec code LPC pour la codeuse GB : L1 est le début du geste de constriction, L2 l'atteinte de la cible labiale, M1 le début du geste de transition manuelle vers le « cou » pour le [y] et M2 l'atteinte de la position cible LPC. Les événements sont référencés par rapport à la fin du [i] et sont exprimés en pourcentages relatifs de la durée de IO. 22

Figure 66. Coordination temporelle main-lèvres pour le début des gestes manuel et labial (M1L1) et pour l'atteinte des cibles (M2L2) en fonction de l'intervalle d'obstruence (ou de pause pour les séquences sans consonne) pour la locutrice-codeuse GB. Les séquences sont distinguées selon le nombre de consonne intervocalique (n= 0 à 4 consonnes) dans la transition [iC$_n$y]. Les valeurs positives de ces deux graphes indiquent que l'événement manuel se produit avant l'événement labial ; à l'inverse, les valeurs négatives indiquent que c'est l'événement labial qui se produit avant l'événement manuel. 22

Figure 67. Durée de la transition manuelle M1M2 (ms) en fonction de l'intervalle IO (ms). 22

Figure 68. Clés consonantiques et positions codant les consonnes et voyelles utilisées dans le *corpus gating*. 22

Figure 69. Photo de la locutrice-codeuse GB durant l'enregistrement avec indication de la pastille de la main et superposition du repère utilisé pour l'analyse des données. 22

Figure 70. Tracé des différents signaux au cours du temps pour la séquence [mytymapɛ̃ma]. De haut en bas : (1) décours temporel de l'aire aux lèvres S, (2) position de la coordonnée x de la pastille sur le dos de la main, (3) position de la coordonnée y de la pastille et (4) signal acoustique correpondant. Sur chacun des signaux sont superposés les événements temporels repérés pour l'analyse de la coordination main-lèvres-son. 22

Figure 71. Exemple d'un découpage en six points de troncature de la syllabe [pɛ̃] pour la séquence [mytymapɛ̃ma] (de gauche à droite et de haut en bas). La main part de la position « côté » pour le [ma] et se dirige vers la « pommette », tout en formant la clé du [p], pour le [pɛ̃]. On voit clairement qu'au point 4, la clé est formée, la position est presque atteinte et le [p] n'est pas encore visible aux lèvres (il le sera au point 5). 22

Figure 72. Tracé des signaux (aire aux lèvres S, position y de la pastille et signal acoustique) au cours du temps pour la portion [mapɛ̃ma] de la séquence [mytymapɛ̃ma] (voir Figure 70). Les barres verticales en traits pleins correspondent aux événements temporels repérés sur

l'analyse des signaux. Les six lignes pointillées correspondent aux six points de troncature retenus pour l'expérience perceptive. 22

Figure 73. Scores moyens d'identification de la syllabe [ma] et de la syllabe CV cible obtenus aux différents points de troncature pour les 16 sujets confondus. 22

Figure 74. Scores moyens d'identification pour la consonne, la voyelle, la consonne classée selon son niveau de labialité, la voyelle classée selon son arrondissement, la clé et la position manuelle obtenus aux différents points de troncature pour les 16 sujets confondus. 22

Figure 75. Résultats d'identification pour chaque clé (à gauche) et pour chaque groupe de consonnes aux lèvres (à droite) obtenus aux différents points de troncature pour tous les sujets confondus. 22

Figure 76. Résultats globaux (main-lèvres) d'identification pour la consonne obtenus aux différents points de troncature pour tous les sujets confondus. 22

Figure 77. Décours d'aires aux lèvres au cours du temps pour les séquences [mapɛ̃], [mapø] et [madɛ̃] (de haut en bas) avec sur chaque graphe les six points de troncature en traits tiretés. 22

Figure 78. Résultats d'identification pour chaque position de main (à gauche) et pour chaque groupe de voyelles labiales (à droite) obtenus aux différents points de troncature pour tous les sujets confondus. La légende [e, eu, in, o] correspond aux voyelles [ɛ, ø, ɛ̃, ɔ]. 22

Figure 79. Résultats globaux (main-lèvres) d'identification pour la voyelle obtenus aux différents points de troncature pour tous les sujets confondus. La légende [e, eu, in, o] correspond aux voyelles [ɛ, ø, ɛ̃, ɔ]. 22

Figure 80. Résultats d'identification obtenus aux différents points de troncature pour les deux groupes de sujets LPC précoces (à gauche) et tardifs (à droite), pour la consonne (C), la voyelle (V), la syllabe (CV) et la syllabe [ma] (en haut) et pour la position manuelle, la clé manuelle, la voyelle (voyelle/labialité, classée selon son arrondissement) et la consonne (consonne/labialité, classée selon son niveau de labialité) (en bas). 22

Figure 81. Détail de l'étape d'encodage phonologique dans le modèle « Speaking » + encodage phonétique et articulation pour « mon avion » (adapté de Levelt & Wheeldon, 1994). 22

Figure 82. Proposition d'une composante LPC dans le modèle « Speaking ». 22

Figure 83. Schéma simplifié d'une représentation en phonologie non-linéaire d'un espace de commande commun (cas prototypique de la syllabe CV) à une étape très proche de l'output du modèle ; * la ligne de lieux buccaux et faciaux n'est pas montrée, de même que d'autres lignes de mode. 22

Figure 84. Image de la tête parlante audiovisuelle PB augmentée de la modalité LPC. 22

Index des tableaux

Tableau 1. Détail des scores d'identification obtenus dans les quatre expériences selon les différentes conditions (voir texte). Le calcul du bénéfice est donné par : 100(score_ACS - score_SA)/(score_MCS - score_SA). Tableau tiré de Duchnowski et al., 2000, p. 493. 22

Tableau 2. Moyennes et écart-types ($\%_{rel}$) des intervalles temporels exprimés en fonction du cycle syllabique CV pour les expériences 1 et 2. 22

Tableau 3. Durées moyennes et écarts-types (en ms) des différents intervalles temporels obtenus pour les trois locutrices-codeuses AM, SC et RV. 22

Tableau 4. Moyennes et écart-types des différents intervalles temporels obtenus pour les trois locutrices-codeuses, calculés en proportions par rapport à chaque durée de syllabe CV de chaque séquence (sans unité noté $\%_{rel}$). Les astérisques (*) près des noms d'intervalles indiquent les différences significatives (ANOVA) entre les trois sujets à $p < .01$. 22

Tableau 5. Liste des séquences enregistrées avec et sans code LPC pour les deux transitions vocaliques. 22

Tableau 6. Liste des logatomes de type [my.ty.ma.CV.ma] du *corpus gating.* 22

Tableau 7. Anamnèse des sujets. 22

Table des Annexes

Annexe 1. Liste complète des séquences du corpus de l'expérience 3 (voir p. 89) (symboles phonétiques). 22

Annexe 2. Photos des locutrices-codeuses avec superposition des repères utilisés pour le traitement des données 22

Annexe 3. Comparaison normalisée pour les trois codeuses AM, SC et RV 22

Annexe 4. Rythmes de parole pour les trois codeuses 22

Annexe 5. Etude de la prosodie rythmique 22

Annexe 6. Phases du geste labial de constriction pour les transitions [iC_ny] (n= 0 à 4 consonnes) 22

Annexe 7. Interface Matlab pour l'expérience perceptive 22

Annexe 8. Consigne donnée aux sujets pendant l'expérience perceptive 22

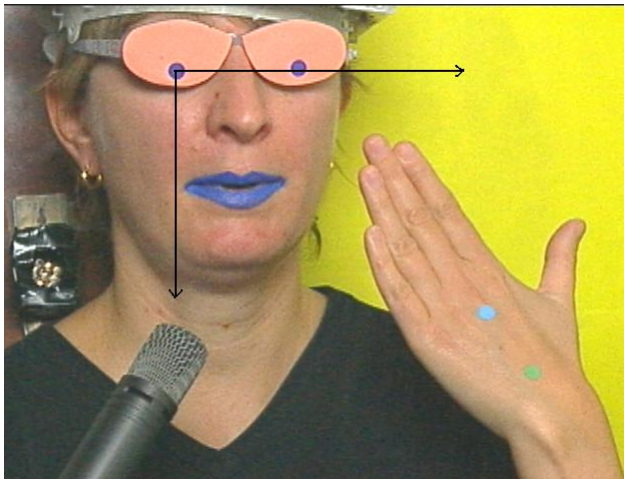
Annexe 9. Résultats de l'expérience perceptive pour les 16 sujets testés 22

Annexes

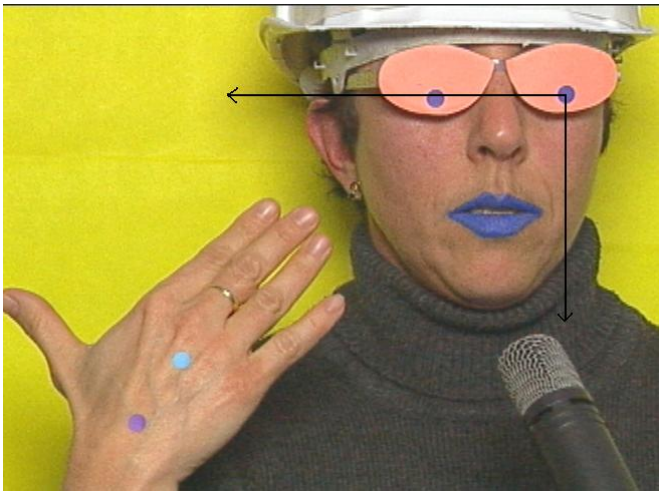
Annexe 1. Liste complète des séquences du corpus de l'expérience 3 (voir p. 22) (symboles phonétiques).

[mamapija] [mamapuja] [mamapøja] [mamapeja]	[mimipaji] [mimipuji] [mimipøji] [mimipeji]	[mumupaju] [mumupiju] [mumupøju] [mumupeju]	[mømøpajø] [mømøpijø] [mømøpujø] [mømøpejø]	[memepaje] [memepije] [memepuje] [memepøje]
[mamajipa] [mamajupa] [mamajøpa] [mamajepa]	[mimijapi] [mimijupi] [mimijøpi] [mimijepi]	[mumujapu] [mumujipu] [mumujøpu] [mumujepu]	[mømøjapø] [mømøjipø] [mømøjupø] [mømøjepø]	[memejape] [memejipe] [memejupe] [memejøpe]
[mamasila] [mamasula] [mamasøla] [mamasela]	[mimisali] [mimisuli] [mimisøli] [mimiseli]	[mumusalu] [mumusilu] [mumusølu] [mumuselu]	[mømøsalø] [mømøsilø] [mømøsulø] [mømøselø]	[memesale] [memesile] [memesule] [memesøle]
[mamalisa] [mamalusa] [mamaløsa] [mamalesa]	[mimilasi] [mimilusi] [mimiløsi] [mimilesa]	[mumulasu] [mumulisu] [mumuløsu] [mumulesu]	[mømølasø] [mømølisø] [mømølusø] [mømølesø]	[memelase] [memelise] [memeluse] [memeløse]
[mamaviga] [mamavuga] [mamavøga] [mamavega]	[mimivagi] [mimivugi] [mimivøgi] [mimivegi]	[mumuvagu] [mumuvigu] [mumuvøgu] [mumuvegu]	[mømøvagø] [mømøvigø] [mømøvugø] [mømøvegø]	[memevage] [memevige] [memevuge] [memevøge]
[mamagiva] [mamaguva] [mamagøva] [mamageva]	[mimigavi] [mimiguvi] [mimigøvi] [mimigevi]	[mumugavu] [mumugivu] [mumugøvu] [mumugevu]	[mømøgavø] [mømøgivø] [mømøguvø] [mømøgevø]	[memegave] [memegive] [memeguve] [memegøve]
[mamabima] [mamabuma] [mamabøma] [mamabema]	[mimibami] [mimibumi] [mimibømi] [mimibemi]	[mumubamu] [mumubimu] [mumubømu] [mumubemu]	[mømøbamø] [mømøbimø] [mømøbumø] [mømøbemø]	[memebame] [memebime] [memebume] [memebøme]
[babamiba] [babamuba] [babamøba] [babameba]	[bibimabi] [bibimubi] [bibimøbi] [bibimebi]	[bubumabu] [bubumibu] [bubumøbu] [bubumebu]	[bøbømabø] [bøbømibø] [bøbømubø] [bøbømebø]	[bebemabe] [bebemibe] [bebemube] [bebemøbe]

Annexe 2. Photos des locutrices-codeuses avec superposition des repères utilisés pour le traitement des données



Codeuse AM



Codeuse RV

Annexe 3. Comparaison normalisée pour les trois codeuses AM, SC et RV

Cette annexe détaille le Tableau 4 de la section VI.2.1.2. Nous pouvons résumer graphiquement l'ensemble des distributions pour chaque intervalle et chaque sujet par des boîtes à moustaches (voir section Annexe 1.V.4.1 pour un rappel des caractéristiques de ces diagrammes) présentées dans la figure suivante (les durées de syllabes en millisecondes sont également indiquées et nous avons indiqué les moyennes par des 'x'). Pour chaque intervalle, les boîtes à moustaches des trois codeuses sont présentées côte à côte de manière à faciliter la comparaison graphique des distributions.

En ce qui concerne le rythme manuel, nous pouvons remarquer sur la figure, une grande similarité des boîtes à moustaches représentant les différentes durées de transitions de main pour les trois codeuses (M1M2 et M3M4). Très grossièrement, nous retrouvons chez les trois sujets 50% des données totales entre 60%$_{rel}$ et 80%$_{rel}$ (plutôt 90%$_{rel}$ pour la codeuse RV) ; ce qui signifie que globalement, les transitions durent bien plus longtemps qu'une demi-syllabe. Nous remarquons une grande variabilité dans les durées révélée par l'étendue très large des moustaches de chaque boîte : les durées de transitions manuelles peuvent varier d'une trentaine de %$_{rel}$ à plus de 100%$_{rel}$ (elles peuvent aller jusqu'à 140%$_{rel}$ pour la codeuse RV, soit en durée absolue, bien plus que la durée d'une syllabe). Parmi les trois codeuses, il semblerait que RV ait tendance à effectuer des transitions plus longues dans le domaine de la syllabe, en démontrant plus de variabilité que les deux autres (les données pour cette codeuse sont en effet plus dispersées ; notamment, les écarts interquartiles de ses boîtes sont plus grands). Cette observation est confirmée statistiquement par une ANOVA[19] à un facteur sur toutes les durées de transitions manuelles : les durées se révèlent être en effet significativement différentes ($F(5, 949)= 7,79$ $p< .01$). Les comparaisons multiples a posteriori par le test de Scheffé montrent que c'est la durée de transition M3M4 de la codeuse RV qui diffère de la durée M1M2 de la codeuse AM ($p< .01$). En ce qui concerne la durée de la tenue de main en position cible LPC (M2M3), nous pouvons remarquer des distributions relativement proches entre les trois sujets, avec une plus grande variabilité dans les durées pour la codeuse AM. Les moyennes s'élèvent à 58%$_{rel}$ pour AM, 64%$_{rel}$ pour SC et 65%$_{rel}$ pour RV : ces moyennes ne sont pas statistiquement différentes (ANOVA à un facteur, $F(2, 474)= 2,7$ au risque $\alpha= 5\%$, $p= 0,07$). Cette similarité de résultats pour les tenues de cibles LPC

[19] Nous avons également utilisé le test non paramétrique de Kruskal-Wallis pour comparer les différentes moyennes, indiquant une différence significative ($H_c= 26,2 > H_t= 11,07$ pour $\alpha= 5\%$ ddl= 5, $p< .01$).

suggère donc une certaine constance du rythme manuel chez les trois codeuses dans le domaine de la syllabe.

En ce qui concerne la coordination des ces transitions avec le son, nous pouvons remarquer une remarquable similarité des distributions des trois sujets pour l'intervalle M1A1, représentant le début du geste manuel par rapport au début acoustique de la consonne. Pour les trois sujets, la moitié des données totales positionnées dans la partie centrale des distributions se trouve largement entre 40%$_{rel}$ et 80%$_{rel}$, ce qui démontre une assez grande variabilité intra-locuteur. Les durées sont de manière générale assez dispersées pour chaque sujet, mais sont distribuées normalement et de façon assez similaire entre les sujets. L'intervalle M1A1 s'élève en moyenne à 60-63%$_{rel}$ pour les trois sujets : une analyse de variance à un facteur montre que ces moyennes ne sont pas significativement différentes (au risque α= 5%, F< 1). Ainsi les trois locutrices-codeuses semblent avoir un même comportement en ce qui concerne l'initiation du geste de transition manuelle dans le domaine syllabique : celui-ci est initié en avance par rapport au son, avec une avance d'une bonne demi-syllabe.

En ce qui concerne la fin de la transition manuelle, elle se produit aux environs du début acoustique de la consonne pour les trois codeuses (A1M2). Globalement, nous pouvons remarquer une certaine similarité des distributions des codeuses AM et SC, alors que la codeuse RV semble différer (sur le graphe, la position de la boîte à moustaches pour RV est globalement plus haute que les deux autres). Cette différence graphique est confirmée statistiquement par une ANOVA à un facteur sur ces durées d'intervalles ; les moyennes, qui s'élèvent à 6%$_{rel}$ pour AM, 10%$_{rel}$ pour SC et 18%$_{rel}$ pour RV, sont significativement différentes ($F(2, 474)= 14,9$ $p< .01$). Les tests post-hoc (Scheffé) montrent que c'est bien la codeuse RV qui diffère des deux autres ($p< .01$) : elle a en effet tendance à atteindre la position cible LPC légèrement plus tard relativement au début acoustique de la consonne. Nous pouvons cependant remarquer pour les trois codeuses que le 1er quartile de chaque distribution est positionné autour de 0 (à -8%$_{rel}$ pour AM, -3%$_{rel}$ pour SC et 8%$_{rel}$ pour RV), ce qui indique que la majorité des données de chaque distribution se trouve dans les valeurs positives : la position de main est donc atteinte en majorité après le début acoustique de la consonne. Pour la codeuse AM, la partie centrale de la distribution est vraiment positionnée autour de 0 : cette codeuse aurait donc tendance à atteindre la position cible LPC en quasi-synchronie avec le début acoustique de la consonne. Pour la codeuse SC, nous trouvons la moitié des données positionnées dans la partie centrale de la distribution entre -3%$_{rel}$ et 14%$_{rel}$: ce résultat indique que dans la plupart des cas, la position manuelle est atteinte en tout début de consonne. Enfin, pour la codeuse RV, la moitié des données sont concentrées entre 8%$_{rel}$ et 32%$_{rel}$: pour cette codeuse, il semble donc y avoir un léger retard de la position cible LPC sur le début

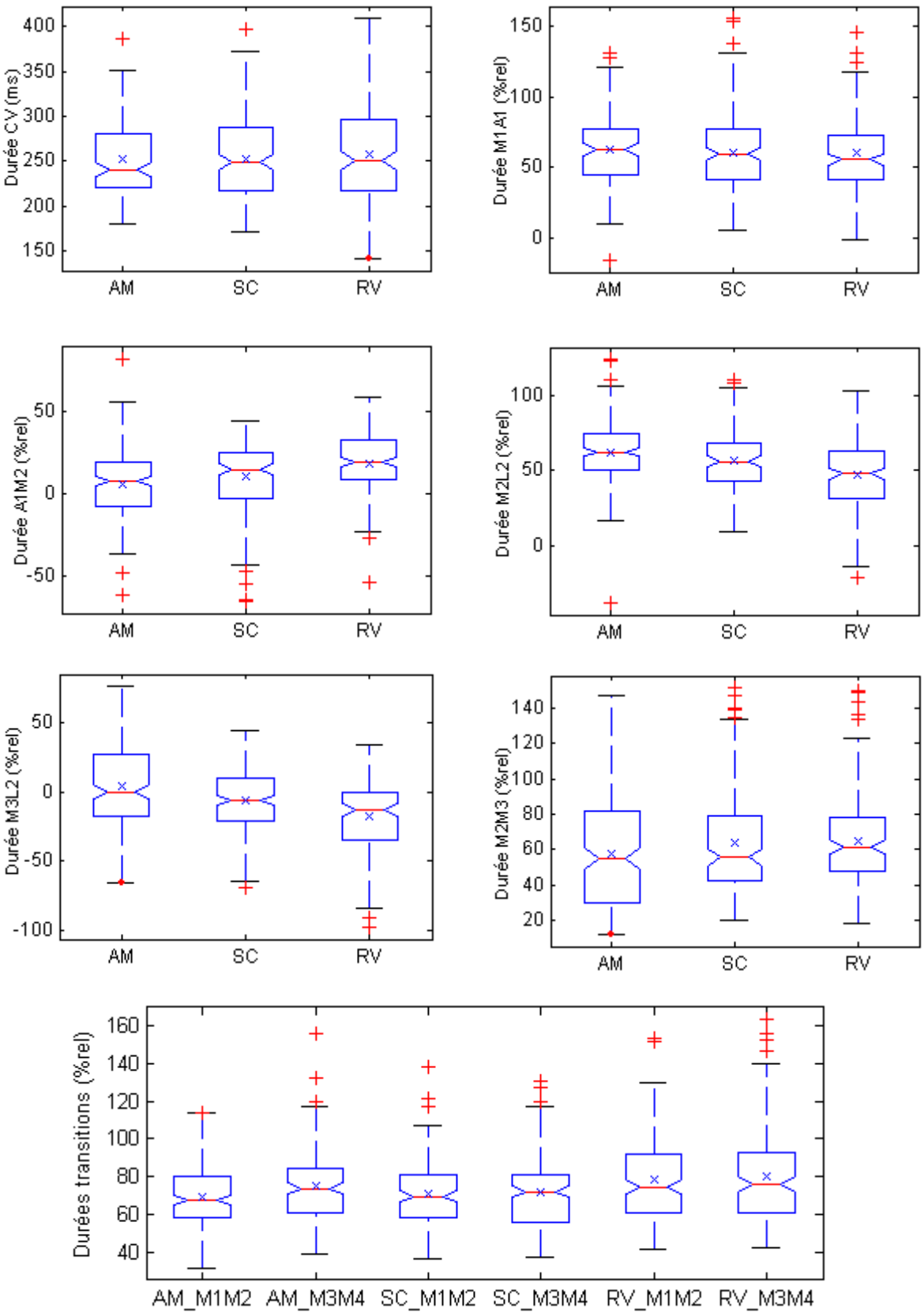
acoustique de la consonne par rapport aux autres codeuses. La main atteind cependant sa position cible dans la première partie de la consonne acoustique.

La position manuelle est donc atteinte avant la réalisation de la cible labiale vocalique (M2L2). Pour les codeuses AM et SC, c'est toujours le cas ; on peut voir en effet que les durées d'intervalle M2L2 sont toutes distribuées dans les valeurs positives (à l'exception d'une valeur hors norme pour AM). La codeuse RV montre à ce niveau un peu plus de variabilité : la boîte à moustaches représentant la distribution de cet intervalle pour cette codeuse semble un peu différente des deux autres. Elle produit quelques séquences où la position de main est atteinte après la cible labiale (la moustache inférieure de la boîte s'étend jusqu'à -14%$_{rel}$). Les données centrales pour l'intervalle M2L2 sont positionnées entre 51%$_{rel}$ et 74%$_{rel}$ pour la codeuse AM, entre 43%$_{rel}$ et 68%$_{rel}$ pour la codeuse SC et entre 32%$_{rel}$ et 63%$_{rel}$ pour la codeuse RV et en moyenne, la position manuelle est atteinte en avance de 62%$_{rel}$ pour AM, 57%$_{rel}$ pour SC et 47%$_{rel}$ pour RV par rapport à la cible labiale. Ceci indique que dans tous les cas, dans la grande majorité des séquences, la position manuelle vocalique est atteinte avant la cible labiale vocalique, avec une avance équivalent à plus d'une demi-syllabe pour AM et SC (la codeuse RV a quant à elle en moyenne une demi-syllabe d'avance). Au niveau statistique, une ANOVA à un facteur montre que les durées M2L2 sont significativement différentes d'un sujet à l'autre ($F(2, 474)= 18,7$ $p< .01$). Les tests post-hoc confirment nos observations : c'est bien la codeuse RV qui diffère des deux autres ($p< .01$).

En ce qui concerne le début du geste manuel vers la position suivante, nous pouvons clairement voir qu'il se produit aux environs de la réalisation de la cible labiale vocalique (M3L2). Plus précisément, pour la codeuse AM, nous pouvons remarquer que les données sont distribuées autour de 0 : la médiane s'élève à 0%$_{rel}$, ce qui signifie que la moitié des données est négative et l'autre moitié positive. La moyenne s'élève à 4%$_{rel}$: il y a donc une plus grande dispersion des données dans les valeurs positives. Ces résultats indiquent que la codeuse AM débute la transition manuelle suivante autour de la réalisation de la cible labiale, avec une légère tendance à partir en avance. En ce qui concerne la codeuse SC, nous retrouvons également une répartition des données globalement autour de 0 (plus précisément centrée à -6%$_{rel}$). Cette codeuse a donc un comportement similaire à celui de AM en ce qui concerne la coordination du début du geste manuel suivant et de la cible labiale, avec toutefois une tendance à initier son geste après la réalisation de la cible de la voyelle aux lèvres. La codeuse RV quant à elle, accentue cette tendance : en effet, le 3ème quartile est positionné à 0%$_{rel}$, ce qui indique clairement que 75% des données sont négatives et donc que la transition de main vers la position suivante est initiée dans la majorité des cas après la cible labiale vocalique. Ces différences observées

entre les trois sujets sont confirmées statistiquement par une ANOVA[20] significative ($F(2, 474)= 24,8$ $p< .01$). Des comparaisons multiples a posteriori montrent que c'est de nouveau la codeuse RV qui diffère des deux autres : elle semble initier la transition manuelle suivante plus tard par rapport à la réalisation de la cible de la voyelle aux lèvres.

[20] Notons que le test non-paramétrique de Kruskal-Wallis a été également utilisé car les conditions d'application de l'ANOVA n'étaient pas vraiment respectées : une différence significative a également été obtenue ($H_c= 38,8 > H_t= 5,991$ pour $\alpha= 5\%$ et ddl= 2).



Boîtes à moustaches indiquant la distribution de chaque intervalle temporel normalisé dans le domaine CV pour les trois sujets AM, SC et RV. Chaque graphique indique la médiane, les 1er et 3ème quartiles, les valeurs frontières, les valeur hors normes ('+') et la moyenne ('x').

Annexe 4. Rythmes de parole pour les trois codeuses

Des phrases avec et sans code LPC ont été enregistrées par les trois codeuses AM, SC et RV à la suite de chacun des enregistrements de corpus CV. Ces phrases consistaient en un texte de présentation (appris par cœur) : « Bonjour je suis codeuse en Langage Parlé Complété. Avec les gestes de ma main, en plus du mouvement de mes lèvres, un malentendant, connaissant cette méthode, peut comprendre tout ce que je dis ». Les phrases sans code ont été répétées une seule fois alors que celles avec code ont été répétées une ou deux fois.

Les signaux acoustiques ont été segmentés et les durées de syllabes de type CV ont été calculées. Les résultats sont indiqués dans le tableau ci-dessous :

	AM	SC	RV
Sans code LPC			
Moyenne (Ecart-type)	211 ms (57 ms)	196 ms (66 ms)	203 ms (62 ms)
Avec code LPC			
Moyenne (Ecart-type)	280 ms (89 ms)	273 ms (78 ms)	264 ms (82 ms)
Moyenne (Ecart-type)	282 ms (72 ms)		262 ms (81 ms)

Annexe 5. Etude de la prosodie rythmique

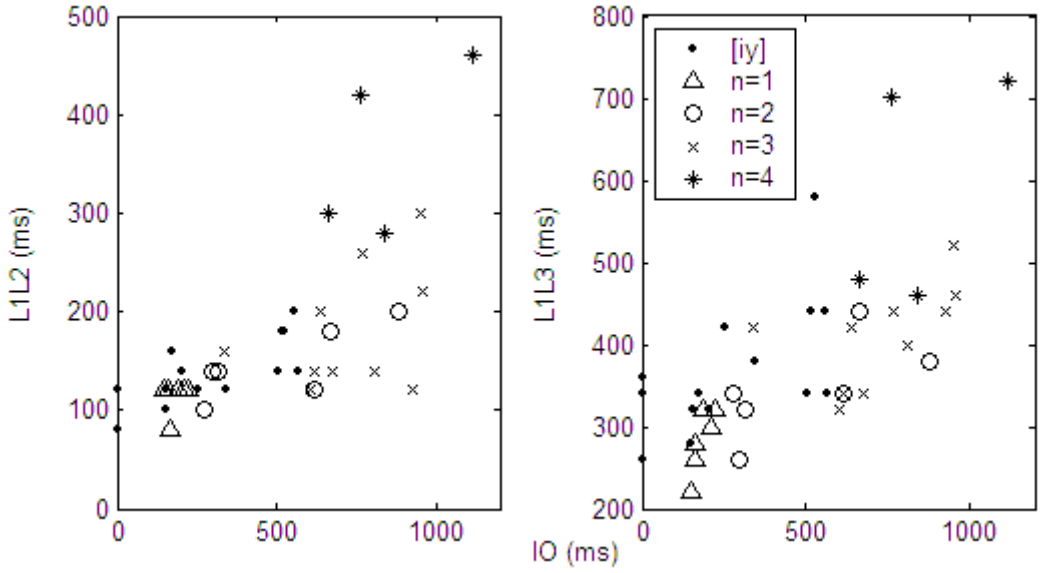
Nous avons effectué un enregistrement, à partir d'une autre codeuse diplômée EH, qui n'a pas pu être exploité quant à l'analyse des coordinations main-lèvres-son, suite à des problèmes techniques. Nous avons cependant pu analyser les signaux acoustiques résultants. Cette étude visait à analyser les différents rythmes de parole codée en fonction des variations de vitesse d'élocution. Nous avons enregistré les phrases d'une comptine (« Une poule sur un mur, […] etc. ») produites avec LPC, variant la prosodie rythmique (rythmé versus non rythmé ; on s'attend à ce que la marque du rythme ralentisse la parole) et le débit de parole (vitesse normale versus rapide) (un groupe contrôle de phrases sans LPC, rythmées et non rythmées à vitesse normale, a également été enregistré). A partir de la segmentation du signal acoustique, des durées de syllabes CV ont été calculées. Les résultats sont présentés dans le tableau ci-dessous en fonction des différentes conditions.

Conditions	Moyenne (ms)	Ecart-type (ms)	Rythme syllabique (Hz)
Sans LPC non rythmé	212	91	4,7
Avec LPC non rythmé vitesse rapide	230	68	4,3
Avec LPC rythmé vitesse rapide	236	63	4,24
Avec LPC non rythmé vitesse normale	240	68	4,17
Sans LPC rythmé	277	97	3,6
Avec LPC rythmé vitesse normale	310	89	3,2

Durées moyennes (ms) et écart-types des syllabes CV extraites de la comptine dans les différentes conditions rythmiques pour la codeuse EH.

Les résultats montrent qu'en moyenne les comptines « sans LPC non rythmées », c'est-à-dire approchant une élocution normale de parole, sont les plus rapides (4,7 Hz). L'ajout d'une prosodie rythmique, marquant davantage les frontières syllabiques, ralentit la parole à un rythme de 3,6 Hz (« sans LPC rythmé »). En ce qui concerne les phrases codées, nous observons la même hiérarchie (les phrases non rythmées sont plus rapides) avec cependant un rythme syllabique plus petit qu'en parole non codée. Nous pouvons de plus noter que la parole codée en « vitesse rapide » reste plus lente que la parole naturelle. Ainsi l'ajout du code LPC ralentit le rythme de parole mais ne perturbe pas son organisation naturelle.

Annexe 6. Phases du geste labial de constriction pour les transitions [iC$_n$y] (n= 0 à 4 consonnes)



Phases du geste labial en fonction de l'intervalle IO pour les transitions [iC$_n$y]. Il s'agit de la même figure que la Figure 64 avec en plus le détail des séquences selon le nombre de consonnes n variant de 0 à 4.

Annexe 7. Interface Matlab pour l'expérience perceptive

Présentation de l'interface du test d'identification des syllabes CV par *gating* visuel. Chaque séquence filmée est présentée au sujet au centre de l'écran. Une fois que la séquence est terminée, une fenêtre noire apparaît (à la place de l'image) et est maintenue tant que le sujet n'a pas donné sa réponse (il doit cliquer sur les boutons correspondants), c'est-à-dire une réponse sur la consonne et la voyelle identifiées (soit « m », « d », « p », « v », « k » et « a », « eu », « e », « o », « in », ces dernières correspondant à [ø], [ɛ], [ɔ] et [ɛ̃]). Pour valider ses réponses, le sujet clique ensuite sur le bouton « film suivant » (les réponses sont stockées dans un vecteur résultat) ; la séquence suivante (en ordre aléatoire) est alors présentée. A la fin des 96 séquences, les réponses du sujet pour chaque séquence, ainsi que l'ordre d'apparition des séquences, sont stockées dans un fichier matlab (.mat).

Annexe 8. Consigne donnée aux sujets pendant l'expérience perceptive

Consigne pour le test de perception visuelle du LPC

Vous allez voir le visage d'une personne s'exprimant en Langage Parlé Complété. Pour les besoins de l'enregistrement, elle a dû porter des lunettes, donc vous ne verrez pas ses yeux.

Les phrases qu'elle va coder n'ont pas de sens et commencent toutes par « mutuma » suivi d'une syllabe que vous devrez identifier. Cette dernière syllabe est formée d'une consonne C et d'une voyelle V : par exemple, « mutumado ».

En fait, vous ne verrez pas forcément toute cette syllabe car nous avons le plus souvent coupé les films plus tôt.
Vous devez, après avoir vu chaque film, indiquer la dernière syllabe que vous aurez pu identifier en cliquant sur l'une des consonnes suivantes [d, p, k, v] et sur l'une des voyelles [in, eu, o, e].
Ces voyelles correspondent à :
[in] de « pain »
[eu] de « peu »
[o] de « bol »
[e] de « mère »

Dans le cas où la séquence serait si courte que l'identification de cette syllabe ne serait pas possible, vous répondrez par [m] et [a] (« ma ») : ce qui voudra dire que vous n'avez identifié que « mutuma ».

Après avoir choisi une consonne et une voyelle, cliquez sur le bouton « film suivant », pour observer le prochain film.

Les premiers films que vous allez voir sont là pour vous familiariser avec le test et les réponses à donner.

Bonne chance !!!

Annexe 9. Résultats de l'expérience perceptive pour les 16 sujets testés

POURCENTAGES D'IDENTIFICATION CORRECTE (%) POUR TOUS LES SUJETS CONFONDUS

Tous les sujets (n= 16)	Point 1	Point 2	Point 3	Point 4	Point 5	Point 6
[ma]	83,98	66,41	43,36	8,98	0,78	0,78
consonne	0,39	11,33	29,30	72,27	93,36	93,75
voyelle	7,42	15,23	24,22	46,09	70,31	82,42
syllabe	0	5,08	12,11	40,23	67,19	78,52
clé	3,13	18,36	41,80	89,84	98,83	99,22
position	10,55	26,95	44,53	83,98	96,48	95,31
consonne/labialité	5,08	16,02	35,94	72,66	93,75	93,75
voyelle/labialité	9,38	17,58	28,13	47,27	71,48	85,16

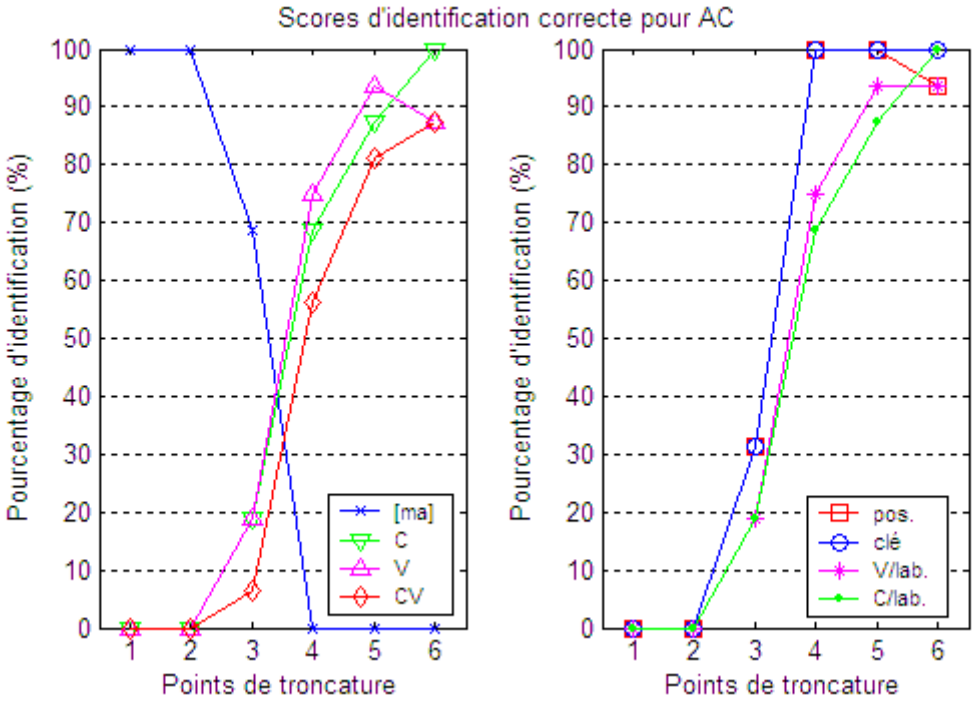
Précoces (n= 9)	Point 1	Point 2	Point 3	Point 4	Point 5	Point 6
[ma]	87,50	68,75	40,97	9,72	0	0,69
consonne	0	13,89	30,56	71,53	95,14	97,92
voyelle	6,25	16,67	25	45,83	70,83	87,50
syllabe	0	8,33	11,11	39,58	68,06	86,81
clé	3,47	18,75	45,14	89,58	100	99,31
position	9,72	27,78	46,53	84,03	95,14	96,53
consonne/labialité	3,47	18,75	38,19	72,22	95,14	97,92
voyelle/labialité	7,64	18,06	29,86	46,53	72,92	89,58

Tardifs (n= 7)	Point 1	Point 2	Point 3	Point 4	Point 5	Point 6
[ma]	79,46	63,39	46,43	8,04	1,79	0,89
consonne	0,89	8,04	27,68	73,21	91,07	88,39
voyelle	8,93	13,39	23,21	46,43	69,64	75,89
syllabe	0	0,89	13,39	41,07	66,07	67,86
clé	2,68	17,86	37,50	90,18	97,32	99,11
position	11,61	25,89	41,96	83,93	98,21	93,75
consonne/labialité	7,14	12,50	33,04	73,21	91,96	88,39
voyelle/labialité	11,61	16,96	25,89	48,21	69,64	79,46

RESULTATS INDIVIDUELS

Sujet AC

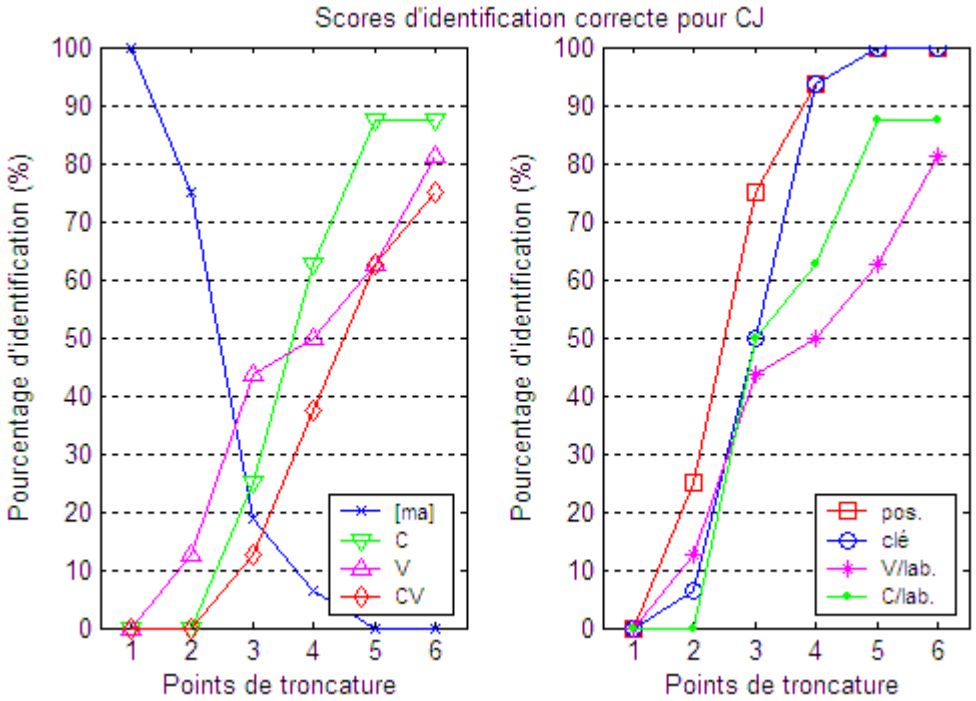
AC a 24 ans. Elle a une surdité profonde bilatérale de cause inconnue qui a été dépistée à l'âge de 2 ans et demi. Elle a porté des appareils quand elle était petite mais a arrêté car ils ne lui servaient à rien. Ses parents sont entendants et font de la LPC et de la lecture labiale avec elle. Elle a du code LPC depuis l'âge de 2 ans et demi. Elle a en surtout eu à la maison au début. En primaire, elle en avait 2 heures par semaine en classe de CM2 et avait du soutien scolaire codé 5 heures par semaine. Au collège et au lycée, 20 heures de cours étaient codées par semaine. En BTS, elle a bénéficié de 20 à 30 heures de codage par semaine. Elle travaille depuis 3 ans et ne fait que de la lecture labiale (elle a un peu de code avec ses proches le week-end). Ses capacités en lecture labiale sont bonnes.



%	1	2	3	4	5	6
[ma]	100	100	68,75	0	0	0
consonne	0	0	18,75	68,75	87,5	100
voyelle	0	0	18,75	75	93,75	87,5
syllabe	0	0	6,25	56,25	81,25	87,5
clé	0	0	31,25	100	100	100
position	0	0	31,25	100	100	93,75
consonne/labialité	0	0	18,75	68,75	87,50	100
voyelle/labialité	0	0	18,75	75	93,75	93,75

Sujet CJ

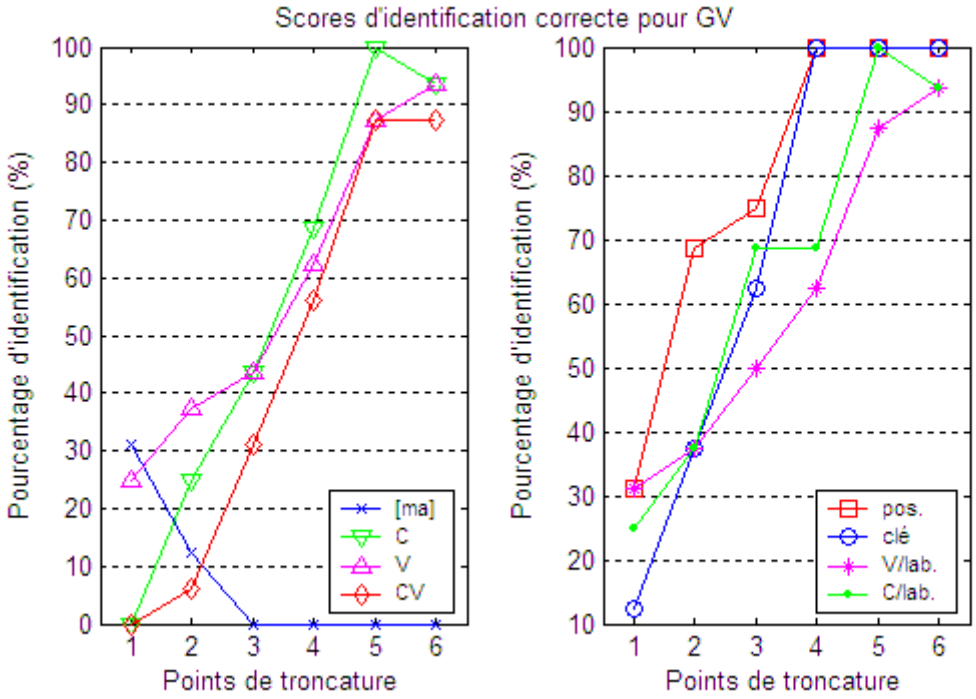
CJ a 12 ans, il est sourd profond de naissance du fait d'une malformation de l'oreille interne (sa surdité a été dépistée à l'âge de deux ans). Ses parents, qui sont entendants, codent, ainsi que son environnement familial (tante, cousine…). CJ est exposé au LPC depuis l'âge de 2 ans et demi. Il effectue plusieurs heures de décodage par jour à l'école et le soir avec ses parents. Il porte des appareils avec contours.



%	1	2	3	4	5	6
[ma]	100	75	18,75	6,25	0	0
consonne	0	0	25	62,50	87,50	87,50
voyelle	0	12,50	43,75	50	62,50	81,25
syllabe	0	0	12,50	37,50	62,50	75
clé	0	6,25	50	93,75	100	100
position	0	25	75	93,75	100	100
consonne/labialité	0	0	50	62,50	87,50	87,50
voyelle/labialité	0	12,50	43,75	50	62,50	81,25

Sujet GV

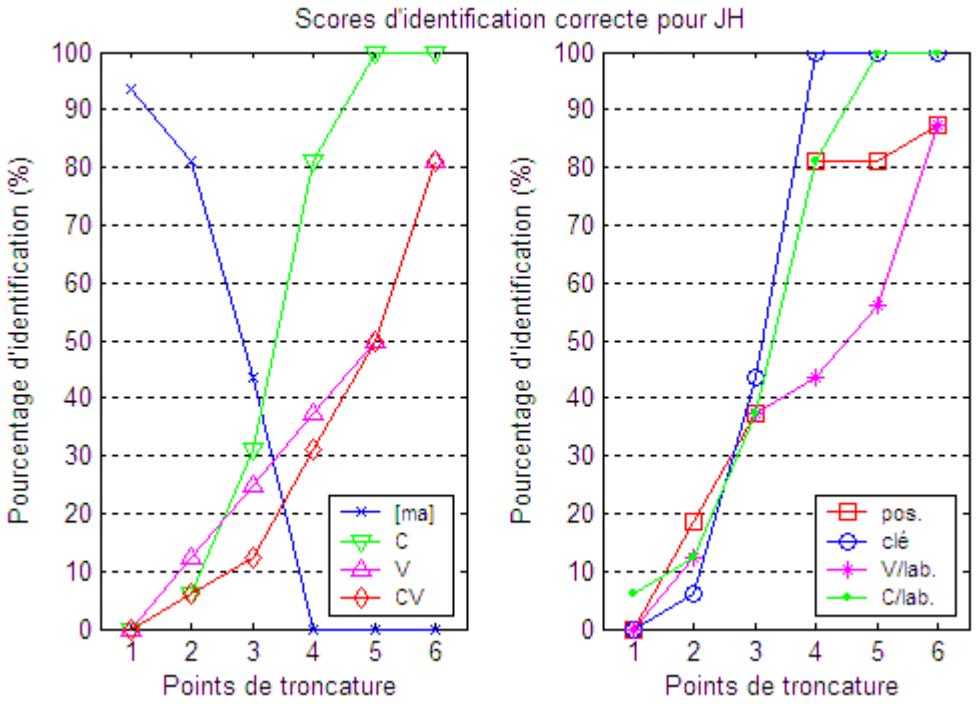
GV a 15 ans, il est sourd profond de naissance. Ses parents sont entendants, codent et signent avec lui. Il décode vraiment depuis qu'il a 7 ans. Il est exposé au LPC environ trois heures par jour au collège et le soir avec ses parents. Il est implanté depuis l'âge de 8 ans.



%	1	2	3	4	5	6
[ma]	31,25	12,50	0	0	0	0
consonne	0	25	43,75	68,75	100	93,75
voyelle	25	37,50	43,75	62,50	87,50	93,75
syllabe	0	6,25	31,25	56,25	87,50	87,50
clé	12,50	37,50	62,50	100	100	100
position	31,25	68,75	75	100	100	100
consonne/labialité	25	37,50	68,75	68,75	100	93,75
voyelle/labialité	31,25	37,50	50	62,50	87,50	93,75

Sujet JH

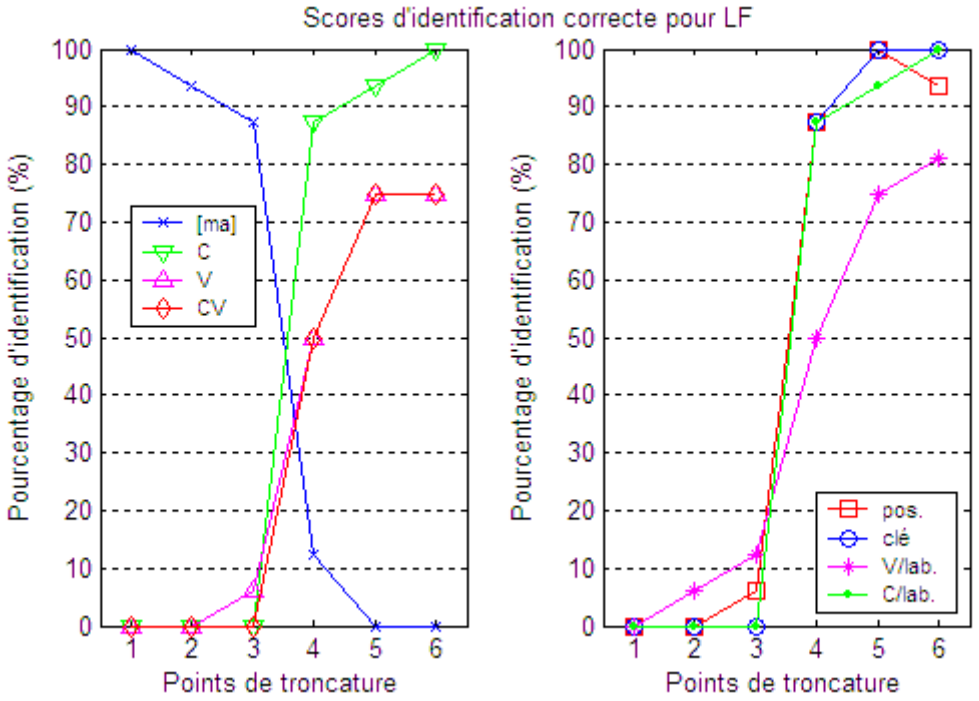
J.H. a 17 ans. Elle est atteinte de surdité profonde qui a été dépistée à l'âge de 10 mois. Ses parents sont entendants et lui codent en LPC depuis le dépistage de sa surdité. Elle a donc été exposée au LPC très précocement. Sa fréquence de décodage par jour est importante : décodage à l'école et à la maison le soir avec ses parents et sa sœur aînée qui est également sourde. Elle porte des appareils avec contours.



%	1	2	3	4	5	6
[ma]	93,75	81,25	43,75	0	0	0
consonne	0	6,25	31,25	81,25	100	100
voyelle	0	12,50	25	37,50	50	81,25
syllabe	0	6,25	12,50	31,25	50	81,25
clé	0	6,25	43,75	100	100	100
position	0	18,75	37,50	81,25	81,25	87,50
consonne/labialité	6,25	12,50	37,50	81,25	100	100
voyelle/labialité	0	12,50	37,50	43,75	56,25	87,50

Sujet LF

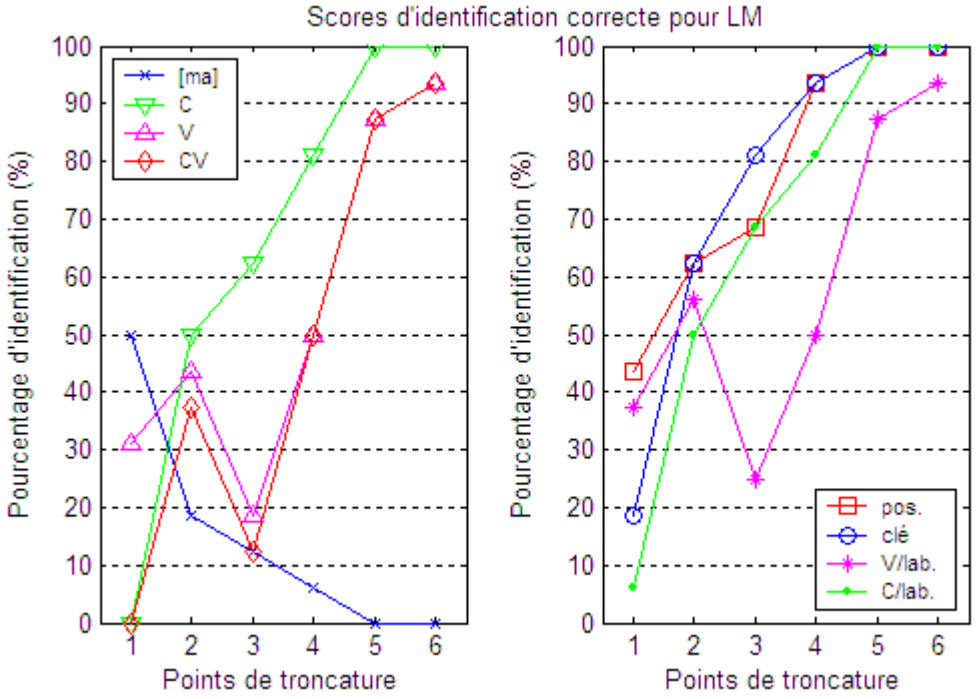
LF a 17 ans. Elle a une surdité profonde de cause inconnue qui a été dépistée à l'âge de 3 ans. Elle porte des contours aux deux oreilles depuis. Elle communique avec ses parents qui sont entendants surtout par le biais de la lecture labiale ; ils font très peu de LPC. Elle décode du LPC depuis l'âge de 15 ans (2001 en 3ème) et surtout à l'école. Depuis, elle bénéficie de 6 à 8 heures de cours codés au lycée (par jour). Ses capacités en lecture labiale sont bonnes.



%	1	2	3	4	5	6
[ma]	100	93,75	87,5	12,5	0	0
consonne	0	0	0	87,5	93,75	100
voyelle	0	0	6,25	50	75	75
syllabe	0	0	0	50	75	75
clé	0	0	0	87,5	100	100
position	0	0	6,25	87,5	100	93,75
consonne/labialité	0	0	0	87,50	93,75	100
voyelle/labialité	0	6,25	12,50	50	75	81,25

Sujet LM

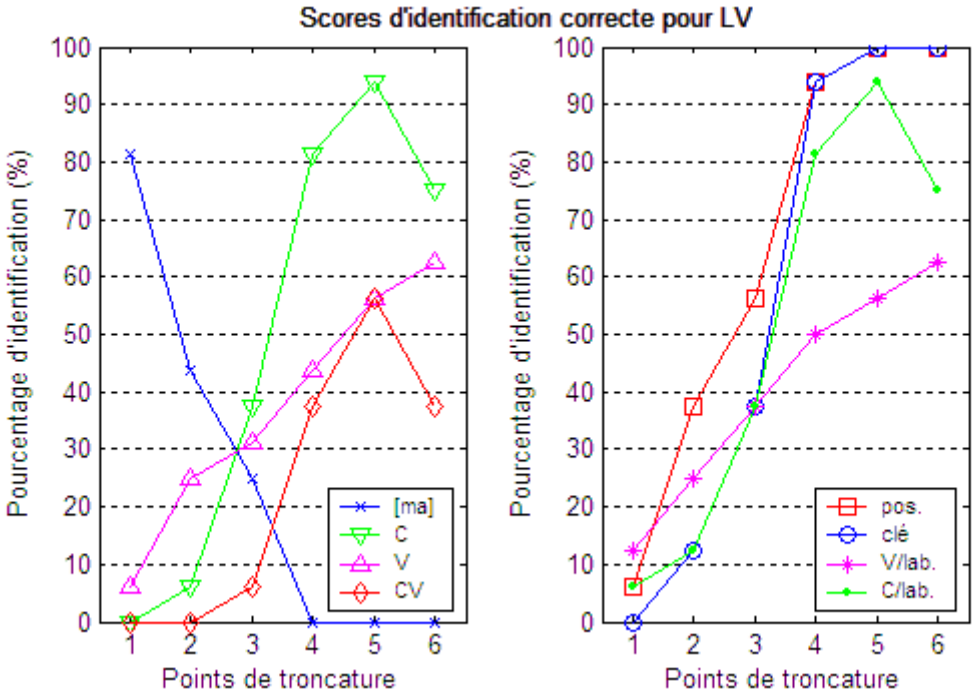
LM a 15 ans. Il est sourd profond de naissance. Ses parents sont entendants et pratiquent le code LPC. LM bénéficie de code depuis l'âge de 18 mois environ. Il décode tous les jours à l'école (environ 4 heures par jour) et le soir avec ses parents. Sa capacité à lire sur les lèvres est très bonne. Il est appareillé avec contours depuis l'âge de 1 an.



%	1	2	3	4	5	6
[ma]	50	18,75	12,50	6,25	0	0
consonne	0	50	62,50	81,25	100	100
voyelle	31,25	43,75	18,75	50	87,50	93,75
syllabe	0	37,50	12,50	50	87,50	93,75
clé	18,75	62,50	81,25	93,75	100	100
position	43,75	62,50	68,75	93,75	100	100
consonne/labialité	6,25	50	68,75	81,25	100	100
voyelle/labialité	37,50	56,25	25	50	87,50	93,75

Sujet LV

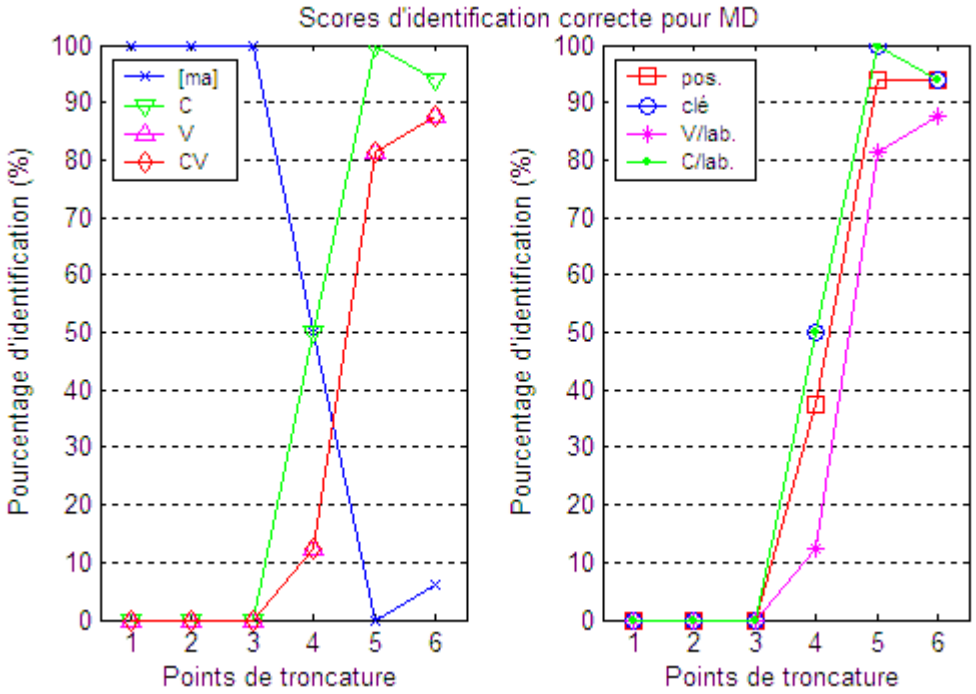
LV a 11 ans. Elle est sourde profonde de naissance et sa mère entendante code depuis que LV a 4 ans mais celle-ci ne décode vraiment que depuis deux ans. Fréquence de codage par jour : 1h30 à l'école et le soir avec sa famille.



%	1	2	3	4	5	6
[ma]	81,25	43,75	25	0	0	0
consonne	0	6,25	37,50	81,25	93,75	75
voyelle	6,25	25	31,25	43,75	56,25	62,50
syllabe	0	0	6,25	37,50	56,25	37,50
clé	0	12,50	37,50	93,75	100	100
position	6,25	37,50	56,25	93,75	100	100
consonne/labialité	6,25	12,50	37,50	81,25	93,75	75
voyelle/labialité	12,50	25	37,50	50	56,25	62,50

Sujet MD

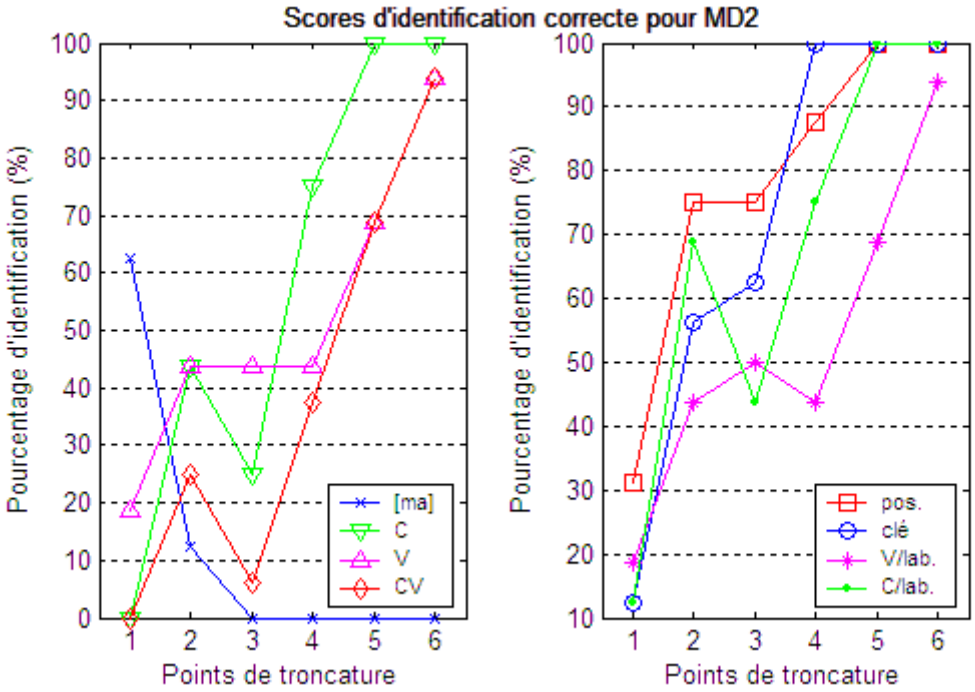
MD a 16 ans. Elle est sourde profonde de naissance de cause inconnue. Elle a été appareillée depuis le dépistage de sa surdité à l'âge de 11 mois. Ses parents sont entendants et communiquent avec elle en LPC et par la lecture labiale. Elle a été exposée au LPC depuis l'âge de 1-2 ans. Actuellement, elle a 14 heures par semaine de cours codés et quelques heures à la maison. Elle a une très bonne capacité à lire sur les lèvres.



%	1	2	3	4	5	6
[ma]	100	100	100	50	0	6,25
consonne	0	0	0	50	100	93,75
voyelle	0	0	0	12,5	81,25	87,5
syllabe	0	0	0	12,5	81,25	87,5
clé	0	0	0	50	100	93,75
position	0	0	0	37,5	93,75	93,75
consonne/labialité	0	0	0	50	100	93,75
voyelle/labialité	0	0	0	12,50	81,25	87,50

Sujet MD2

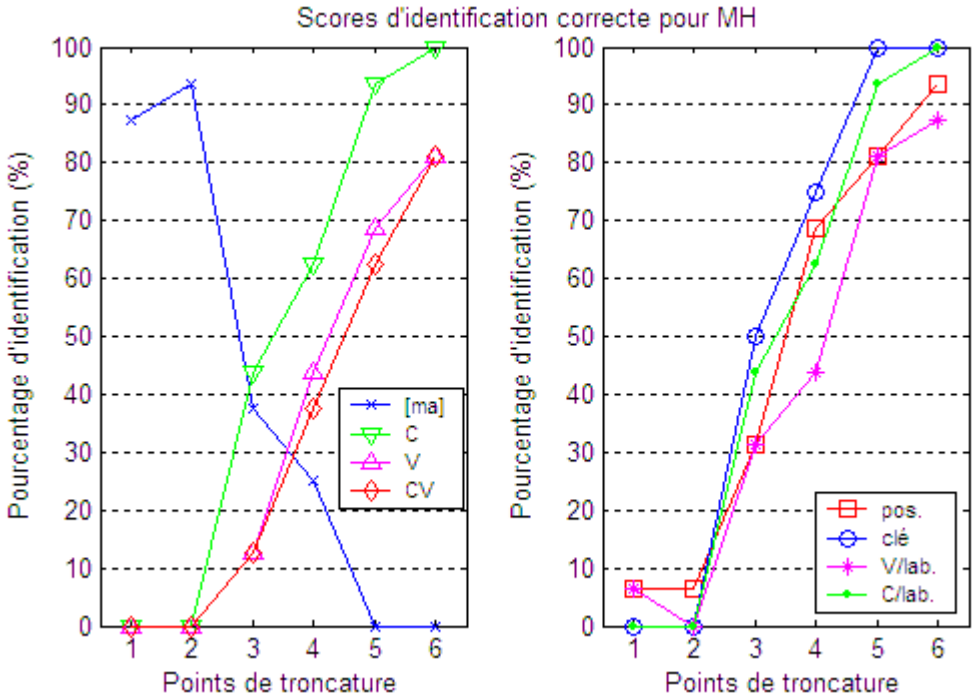
MD2 a 17 ans, elle est atteinte de surdité profonde (dépistée à l'âge de 9 mois). Ses parents qui sont entendants codent, sa maman est notamment devenue codeuse professionnelle. Elle a commencé à bénéficier de code à 1 an et demi. Fréquence de décodage par jour : toute la journée au lycée, et le soir avec ses parents. Elle est appareillée avec contours depuis l'âge de 1 an. Elle lit très bien sur les lèvres.



%	1	2	3	4	5	6
[ma]	62,50	12,50	0	0	0	0
consonne	0	43,75	25	75	100	100
voyelle	18,75	43,75	43,75	43,75	68,75	93,75
syllabe	0	25	6,25	37,50	68,75	93,75
clé	12,50	56,25	62,50	100	100	100
position	31,25	75	75	87,50	100	100
consonne/labialité	12,50	68,75	43,75	75	100	100
voyelle/labialité	18,75	43,75	50	43,75	68,75	93,75

Sujet MH

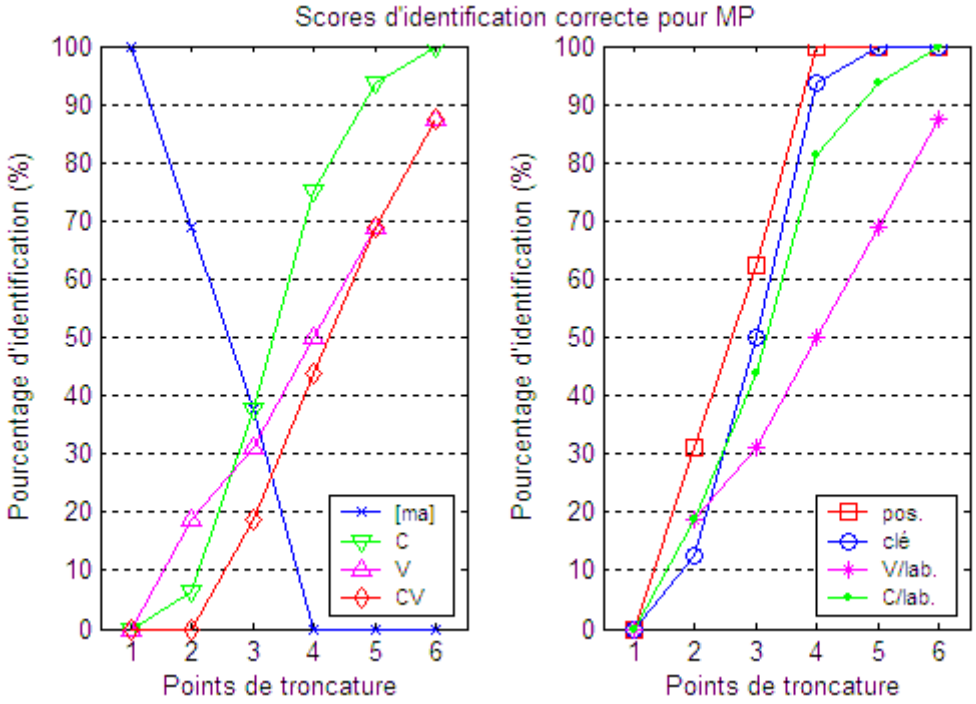
MH a 19 ans. Elle a une surdité profonde de naissance de cause inconnue qui a été détectée à l'âge de 2 ans. Elle est appareillée depuis. Ses parents entendants communiquent avec elle en utilisant surtout le LPC et la lecture labiale et aussi un peu de LSF. Elle a été exposée au LPC depuis l'âge de 2 ans surtout à la maison et 2 ou 3 fois par semaine à l'institut. Actuellement, elle a 13 heures de cours codés par semaine au lycée. Sa capacité à lire sur les lèvres est bonne.



%	1	2	3	4	5	6
[ma]	87,5	93,75	37,5	25	0	0
consonne	0	0	43,75	62,5	93,75	100
voyelle	0	0	12,5	43,75	68,75	81,25
syllabe	0	0	12,5	37,5	62,5	81,25
clé	0	0	50	75	100	100
position	6,25	6,25	31,25	68,75	81,25	93,75
consonne/labialité	0	0	43,75	62,50	93,75	100
voyelle/labialité	6,25	0	31,25	43,75	81,25	87,50

Sujet MP

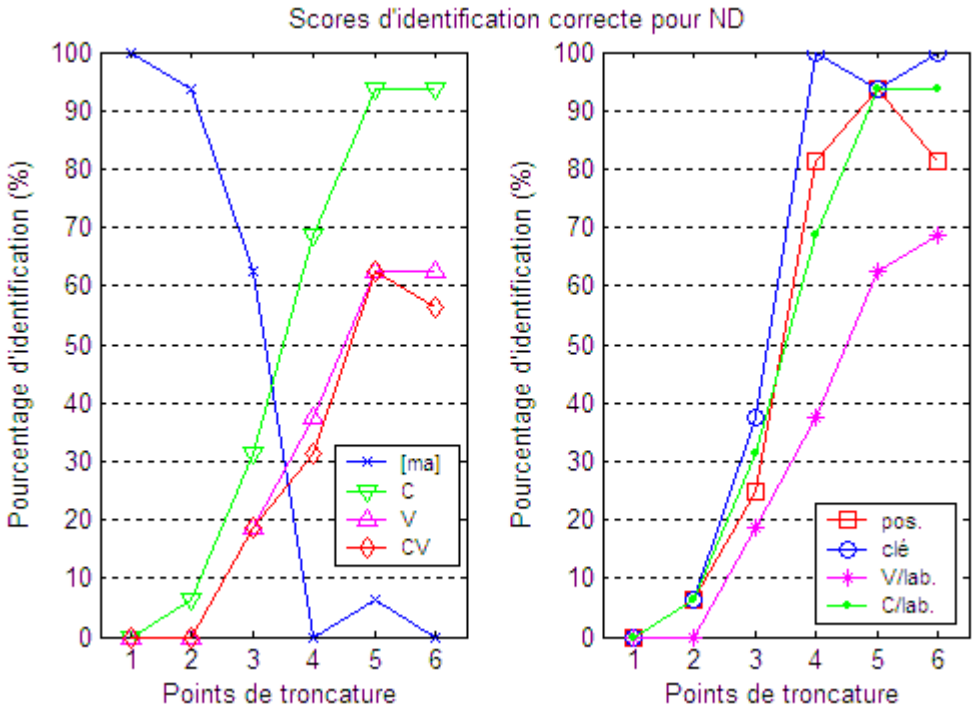
M.P a 12 ans, elle est sourde profonde de naissance. Ses parents sont entendants et sa mère code en LPC. Elle bénéficie de code depuis la maternelle (3 ans). Fréquence de décodage : 3 heures par jours au collège et le soir avec sa famille et son frère (qui est sourd et qui code également). Elle lit bien sur les lèvres.



%	1	2	3	4	5	6
[ma]	100	68,75	37,50	0	0	0
consonne	0	6,25	37,50	75	93,75	100
voyelle	0	18,75	31,25	50	68,75	87,50
syllabe	0	0	18,75	43,75	68,75	87,50
clé	0	12,50	50	93,75	100	100
position	0	31,25	62,50	100	100	100
consonne/labialité	0	18,75	43,75	81,25	93,75	100
voyelle/labialité	0	18,75	31,25	50	68,75	87,50

Sujet ND

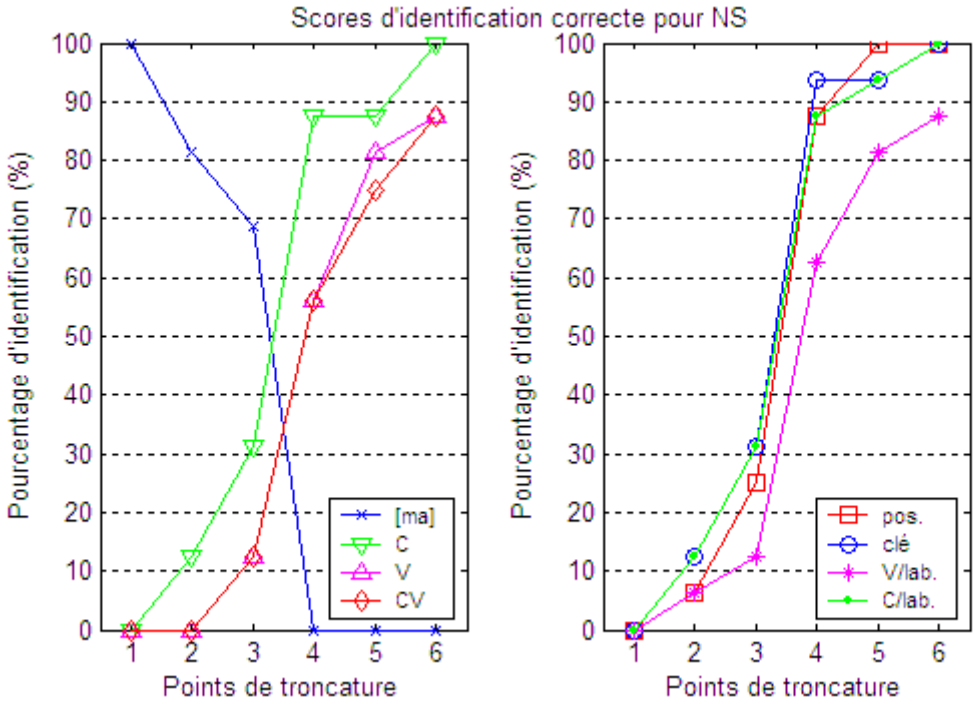
ND a 25 ans. Il a une surdité bilatérale très sévère. Sa surdité a été dépistée à l'âge de 18 mois. Il est appareillé des deux côtés depuis l'âge de 2 ans. Il utilise surtout la lecture labiale pour communiquer avec ses parents qui sont entendants. Il a eu un peu de LPC quand il était petit vers l'âge de 3-4 ans. Puis il a eu du code durant tout le collège par un professeur de soutien qui codait environ 2 heures par jour le soir. Depuis il n'a plus du tout de code. Sa faculté à lire sur les lèvres est bonne.



%	1	2	3	4	5	6
[ma]	100	93,75	62,5	0	6,25	0
consonne	0	6,25	31,25	68,75	93,75	93,75
voyelle	0	0	18,75	37,5	62,5	62,5
syllabe	0	0	18,75	31,25	62,5	56,25
clé	0	6,25	37,5	100	93,75	100
position	0	6,25	25	81,25	93,75	81,25
consonne/labialité	0	6,25	31,25	68,75	93,75	93,75
voyelle/labialité	0	0	18,75	37,50	62,50	68,75

Sujet NS

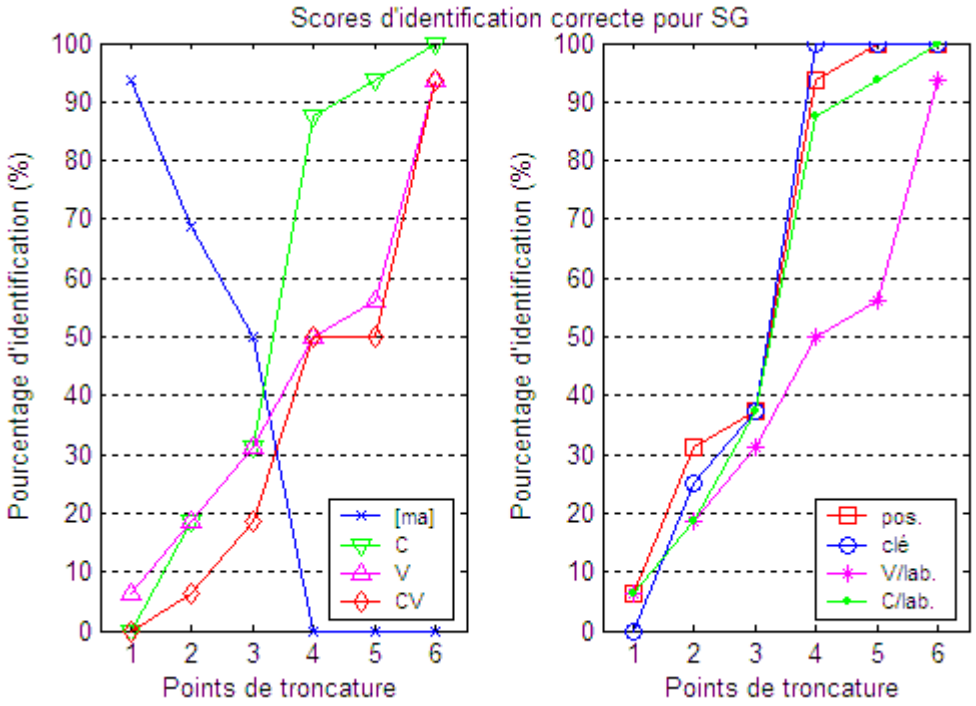
N.S. a 35 ans. Elle est sourde profonde, sa surdité ayant été dépistée à l'âge de 2 ans. Ses parents sont entendants et ne codaient pas et ne signaient pas non plus. Elle lisait sur les lèvres pour communiquer et a appris à oraliser. Pour des raisons professionnelles, elle a appris à décoder le LPC. Elle est donc amenée à décoder en moyenne une fois par mois. Elle a d'abord été appareillée avec contours, puis a été implantée à l'âge de 30 ans.



%	1	2	3	4	5	6
[ma]	100	81,25	68,75	0	0	0
consonne	0	12,50	31,25	87,50	87,50	100
voyelle	0	0	12,50	56,25	81,25	87,50
syllabe	0	0	12,50	56,25	75	87,50
clé	0	12,50	31,25	93,75	93,75	100
position	0	6,25	25	87,50	100	100
consonne/labialité	0	12,50	31,25	87,50	93,75	100
voyelle/labialité	0	6,25	12,50	62,50	81,25	87,50

Sujet SG

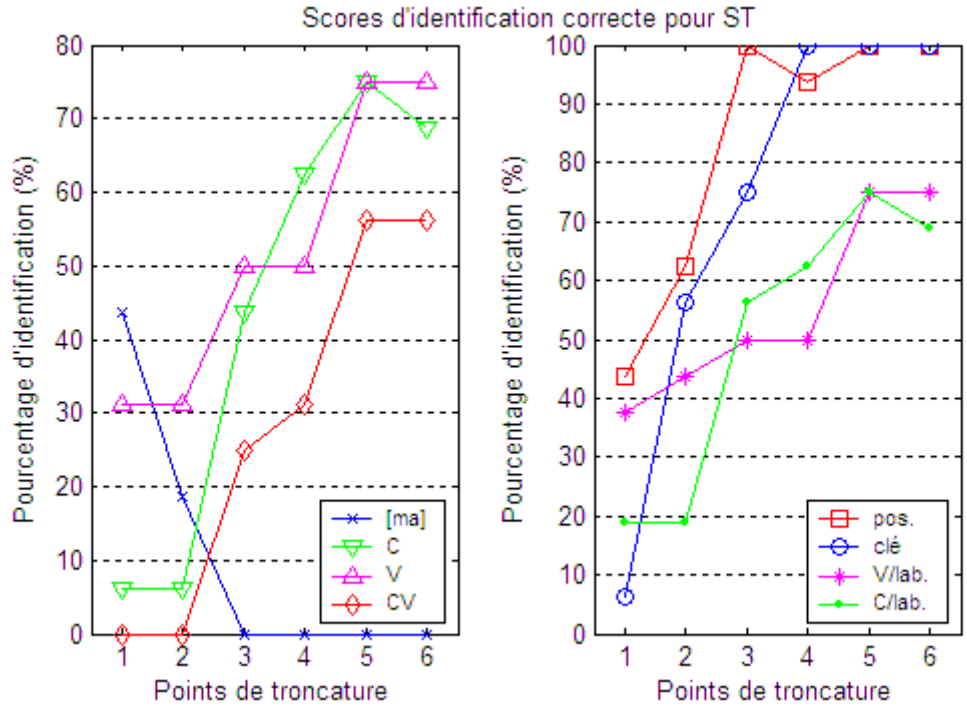
SG a 17 ans. Elle a une surdité profonde bilatérale suite a un virus lors de la grossesse. Elle a porté des appareils et a très vite arrêté car elle ne récupérait absolument rien. Ses parents sont entendants et utilisent principalement le LPC et la lecture labiale pour communiquer avec elle. Elle a aussi eu un peu de LSF pour certains mots comme « dormir », « manger »… Elle est exposée au LPC depuis l'âge de 8 mois. Jusqu'à cette année elle avait 7 heures de code par jour en moyenne mais cette année elle n'en a plus que 2 heures par semaine suite à un changement d'établissement. Sa capacité à lire sur les lèvres est moyenne.



%	1	2	3	4	5	6
[ma]	93,75	68,75	50	0	0	0
consonne	0	18,75	31,25	87,5	93,75	100
voyelle	6,25	18,75	31,25	50	56,25	93,75
syllabe	0	6,25	18,75	50	50	93,75
clé	0	25	37,5	100	100	100
position	6,25	31,25	37,5	93,75	100	100
consonne/labialité	6,25	18,75	37,50	87,50	93,75	100
voyelle/labialité	6,25	18,75	31,25	50	56,25	93,75

Sujet ST

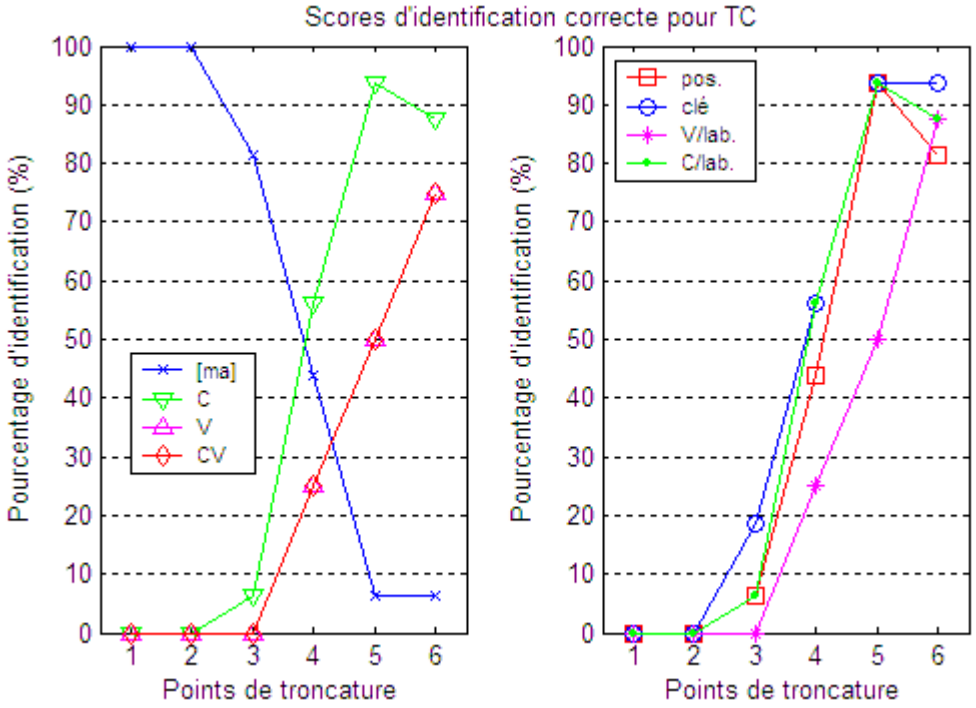
ST a 13 ans, et est sourde profonde de naissance. Ses parents codent en LPC. Elle ne s'est intéressée sérieusement au LPC qu'à partir de 8 ans. Elle bénéficie de code en moyenne 3 heures par jour à l'école et le soir avec ses parents.



%	1	2	3	4	5	6
[ma]	43,75	18,75	0	0	0	0
consonne	6,25	6,25	43,75	62,50	75	68,75
voyelle	31,25	31,25	50	50	75	75
syllabe	0	0	25	31,25	56,25	56,25
clé	6,25	56,25	75	100	100	100
position	43,75	62,50	100	93,75	100	100
consonne/labialité	18,75	18,75	56,25	62,50	75	68,75
voyelle/labialité	37,50	43,75	50	50	75	75

Sujet TC

T.C. a 15 ans, il est sourd profond de naissance. Ses parents ont appris à coder mais le font rarement. Il bénéficie de code depuis l'âge de 5 ans, mais a commencé à décoder surtout à partir de 11 ans (son entrée au collège). Il pratique également la LSF et lit beaucoup sur les lèvres. Fréquence de décodage par jour : 3 heures au collège.



%	1	2	3	4	5	6
[ma]	100	100	81,25	43,75	6,25	6,25
consonne	0	0	6,25	56,25	93,75	87,50
voyelle	0	0	0	25	50	75
syllabe	0	0	0	25	50	75
clé	0	0	18,75	56,25	93,75	93,75
position	0	0	6,25	43,75	93,75	81,25
consonne/labialité	0	0	6,25	56,25	93,75	87,50
voyelle/labialité	0	0	0	25	50	87,50

Résumé

La *LPC* ou *Cued Speech* est un augment manuel qui permet au sourd de désambiguïser l'information phonologique visible sur le visage. L'efficacité de ce système pour l'acquisition de la phonologie de la langue est bien établie. Mais la production du code LPC n'avait jamais été étudiée, et nous l'avons fait par une technique de suivi des mouvements labiaux et manuels de quatre codeuses professionnelles. Notre résultat comportemental majeur est que le geste de la main – contre toute attente – *précède* le geste des lèvres. Cette anticipation donne un rôle inattendu à la parole visible : celui de venir désambiguïser le geste manuel, conçu au départ pour désambiguïser la parole… Notre hypothèse est que le système de Cornett a été recodé en termes neuralement compatibles pour le contrôle des gestes des voyelles et des consonnes dans la LPC et la parole. Ainsi le contrôle des *contacts* vocaliques manuels va se trouver en phase avec celui des contacts consonantiques visibles. Ce phasage est assez précis pour que, quelles que soient les variations de la durée de la production de la syllabe CV, l'aboutissement de la détente (*stroke*) du système main-bras se produise dans la phase de tenue de l'attaque consonantique. L'incorporation de la main et de la face dans un espace de contrôle neural commun peut être ainsi pleinement réalisée dans la LPC.

Mots-clés : code LPC, Cued Speech, surdité, production de parole, coordination gestes manuel et articulatoire, perception/intégration

Abstract

Cued Speech is a manual method that allows deaf people to disambiguate the phonological information through the visual channel. Its efficiency for phonological speech acquisition is well established, but a study of Cued Speech production is lacking. Therefore the aim of this work is to investigate the temporal organization of French Cued Speech production on four experienced cuers. Our main result is that the hand gesture anticipates the lip gesture. It is hypothesized that the found pattern of coordination results from the neural compatibility between movement control of consonants and vowels in Cued Speech and visible speech. Thus the vocalic manual contact control is in-phase with the consonantal contact control of visible speech. This phasing is maintained regardless the variability of the syllable duration. This way, in Cued Speech, hand and face are completely incorporated in a common neural control space.

Key-words: Cued Speech, deafness, speech production, manual and articulatory gestures coordination, speech perception/integration