Deixis prosodique multisensorielle : production et perception audiovisuelle de la focalisation contrastive en français

Marion Dohen

INSTITUT NATIONAL POLYTECHNIQUE DE GRENOBLE

N° attribué par la bibliothèque
|__|__|__|__|__|__|__|__|__|__|

T H E S E

pour obtenir le grade de

DOCTEUR DE L'INPG

Spécialité : « Sciences Cognitives »

préparée à l'Institut de la Communication Parlée
dans le cadre de **l'Ecole Doctorale** « ***Ingénierie pour le vivant : santé, cognition, environnement*** »

présentée et soutenue publiquement

par

Marion DOHEN

le 18 Novembre 2005

Titre :

DEIXIS PROSODIQUE MULTISENSORIELLE : PRODUCTION ET PERCEPTION AUDIOVISUELLE DE LA FOCALISATION CONTRASTIVE EN FRANÇAIS

_____

Directeurs de thèse :

Hélène LŒVENBRUCK & Jean-Luc SCHWARTZ

_____

JURY

M. Jeanny HÉRAULT	, Président
Mme Valérie HAZAN	, Rapporteure
Mme Mariapaola D'IMPERIO	, Rapporteure
M. Jean-Luc SCHWARTZ	, Directeur de thèse
Mme Hélène LŒVENBRUCK	, Co-encadrante
M. David HOUSE	, Examinateur

Remerciements

Je souhaite avant tout remercier Hélène Lœvenbruck sans qui cette belle lisse poire n'aurait pu être écrite. Son soutien, son attention, sa confiance et sa présence sans faille m'ont sans cesse propulsée. Dès notre rencontre, elle n'a jamais hésité à partager avec moi sa propre expérience. Nos discussions ont été, sont et seront encore longtemps j'espère, très stimulantes et enrichissantes. Elles ont toujours tenu du partage et du débat. Il y a trois ans, je savais à peine ce que *prosodie* signifiait et aujourd'hui j'ai écrit plus de 200 pages sur sa perception audiovisuelle. Je sais l'étendue qu'il demeure à explorer, mais avec l'aide d'Hélène, je suis devenue une exploratrice méthodique et zélée qui s'enrichit de tout ce qu'elle trouve sur son chemin et qui a appris que les choses qui ne brillent pas au premier regard méritent d'être polies.

Hélène, merci pour ta disponibilité, ton enthousiasme, tes encouragements, ta sagesse face à la science et à la vie ; merci de m'avoir éclairée de ton immense culture ; et merci aussi de m'avoir fait rencontrer Philippe le docteur africain, le petit Nicolas qui jardine chez son vieux copain, puis Jean le gastronome. Dans mon cœur, je resterai toujours ton padawan …

Je remercie Jean-Luc Schwartz, mon directeur de thèse, qui, en quelques instants, et avec pour seuls outils une craie et quelques transparents, m'a passionnée à vie pour la perception de la parole. Je le remercie tout particulièrement de m'avoir présenté Hélène et d'avoir ainsi orienté mon parcours. Merci pour son aide, sa confiance, ses explications si limpides et les nombreux conseils qu'il m'a donnés si généreusement et qui ont jalonné mon travail de thèse … Merci d'avoir lu, critiqué et commenté ces pages avec tant de finesse et surtout si rapidement.

Je suis aussi très reconnaissante envers Pierre Escudier, d'abord en tant que directeur du DEA de sciences cognitives en 2001-2002, pour avoir accepté ma candidature, puis en tant que directeur de l'ICP à mon arrivée, pour m'avoir accueillie à bras ouverts en m'accordant sa confiance. C'est lui qui a rendu possible ma présence ici et je l'en remercie infiniment.

Je remercie très chaleureusement mes deux rapporteures, Mariapaola D'Imperio et Valérie Hazan d'avoir accepté si promptement et avec tant d'enthousiasme de se plonger dans mon travail. Je suis très fière qu'elles aient toutes deux accepté.

Je remercie David House, l'examinateur de ce jury, d'avoir accepté cette responsabilité avec tant d'enthousiasme malgré la surcharge de travail que représente pour lui la lecture d'un manuscrit en français. Je le remercie pour ses encouragements et l'intérêt constant qu'il a porté à mon travail depuis que je l'ai rencontré sur les rives d'un lac un mois de septembre 2003.

Je remercie également Jeanny Hérault, président de mon jury, pour le temps qu'il a accepté d'accorder à la lecture et à la critique de mon travail.

Pour ses conseils et sa disponibilité, mais aussi pour ses connaissances qu'elle a si volontiers partagées, je remercie chaleureusement Marie-Agnès Cathiard. Merci d'avoir toujours répondu si rapidement à toutes mes interrogations.

Je remercie tous les locuteurs et tous les sujets qui ont participé aux expériences décrites dans ce mémoire. Je suis tout particulièrement reconnaissante à ceux qui m'ont aidée pour la mise en place technique et le bon déroulement de mes expériences : Alain Arnal, Nino Medves et Christophe Savariaux. Je souhaite remercier Guillaume Rolland pour avoir mis au point et enregistré le corpus AV1. Je remercie aussi tout particulièrement Coriandre Vilain de m'avoir aidée à pondérer les intensités …

Je tiens aussi à remercier Christian Abry pour tant de discussions surprenantes et passionnantes. Je le remercie même pour un certain après-midi pénible qui m'a appris à défendre mon travail de façon acharnée, méthodique et argumentée. Alberich n'est somme toute pas si brutal …

Je remercie très chaleureusement tous les membres de l'ICP gare et campus, permanents et moins permanents pour leur accueil, leur chaleur, leurs encouragements et parfois aussi leur amitié. Merci à Nino d'avoir si souvent volé à mon secours et notamment très récemment quand Word n'était pas très coopératif. Je n'oublie pas Monique et Christian si efficaces et agréables. Merci à l'équipe administrative de l'ICP : Nadine Bioud pour sa gentillesse, son efficacité, sa patience et sa bonne humeur et Mme Gaude pour son efficacité, son dynamisme, sa chaleur et sa disponibilité. Sans elles, mes humeurs professionnelles vagabondes auraient été contrariées.

Merci à Harold Hill de m'avoir accueillie aux laboratoires Advanced Telecommunications Research au Japon en 2004, puis de m'y avoir invitée en 2005. Je suis très honorée d'avoir la possibilité d'effectuer mon post-doctorat avec lui. Je le remercie tout particulièrement pour l'intérêt qu'il a tout de suite porté à mon travail et pour le temps qu'il m'a consacré en janvier 2005 pour l'enregistrement des données Optotrak. Merci à lui pour sa patience, son flegme (britannique ?) et sa connaissance de la si complexe culture japonaise qui m'a évité de commettre des impairs. Je tiens aussi à remercier tous ceux et celles qui m'ont aidé pour mes enregistrements Optotrak à ATR : Guillaume, Harold, Ishi-san, Kura, Marcia. Merci à Benoît, Etienne, Guillaume, Ishi-san, Kamachi-san, Kinoshita-san, Kura, Mathieu, Mitsudo-san et Valentin pour leur patience pendant les enregistrements (et même parfois pour le sacrifice de leur barbe !). Pour avoir rendu mes séjours au Japon si plaisants, je remercie Yves, Choi-Lin et Mireille, Nakamura-san et Mizumachi-san. Je remercie Eric Vatikiotis-Bateson de m'avoir mise en relation avec Harold.

Je remercie la société d'acoustique américaine (*Acoustical Society of America*) d'avoir récompensé mon travail en me décernant la bourse Stetson. Le fait que mes recherches s'ancrent si profondément dans la continuité de celles de Stetson n'a fait que renforcer l'honneur que j'ai ressenti quand j'ai reçu ce prix. Je remercie tout particulièrement M. Pickett pour ses mails de félicitations et d'encouragements suite à l'obtention de cette bourse. Je le remercie de m'avoir communiqué ce que Stetson aurait pensé de mes travaux et d'avoir montré tant d'intérêt pour en suivre l'évolution. J'espère que vous pourrez lire ces pages …

Je remercie Mamoun et Papoun d'avoir toujours protégé leur oisillon qui s'est maintenant envolé si loin du nid. Sans eux, mon petit monde serait moins ensoleillé. Leur attention constante pour que je sois heureuse n'a pas été vaine et leurs encouragements ont jalonné mon parcours. Merci pour votre amour.

Je tiens à remercier Nichette, ma petite artiste, d'apporter à sa Ionion tant d'affection et de lui insuffler tant de force. N'oublie jamais de croire en toi et en tes rêves.

Pour sa patience et sa compréhension, je remercie mamy Do. Je sais que mes choix sont parfois obscurs pour toi mais je te remercie de toujours les respecter. Je n'oublie pas le sourire de ma mamy Crapet qui me manque tant. Je remercie aussi tout particulièrement mon papy Milou, qui m'a plus que quiconque et dès ma plus tendre enfance, transmis cette curiosité et cette fascination pour la connaissance qui sont aujourd'hui les principaux moteurs de mon travail. Je sais combien tu serais fier et cette fierté me manque cruellement. Papy Noël, je ne t'oublie pas, ton amour me manque.

Je remercie aussi Brigitte et Jean-Louis qui m'ont si souvent encouragée et qui apprécient si bien ce que cette thèse représente pour moi.

Je tiens à remercier Zig, pour qui le mot *amie* est trop faible. Le fait d'être comprise avant même de parler est si rassurant pour moi. Merci d'être si attentive et drôle, de toujours m'encourager et de m'épauler quand les vents sont contraires et d'être aussi fofolle que moi. Et merci aussi d'avoir relu mes premières productions en étant toujours honnête et objective. Ton amitié est pour moi un trésor.

Je remercie également Pauline pour ses encouragements et son soutien pendant cette période difficile et pour avoir relu et éclairer de ses lumières le chapitre II.

Pour leur soutien, leurs encouragements et leur amitié, je remercie : Audinette, Mélie, Vivi, JP, Pauline, Tom, Claire & Jérèm, Fanny & Xavier.

Merci à Guillaume d'avoir toléré mes angoisses, mes colères, mes états d'âmes, mes coups de folies, mes posters et surtout mon « bazar » alors qu'il est si ordonné. Merci d'avoir partagé avec moi plus que ton bureau.

Merci à Amélie, Annemie, Antoine, Bertrand, Guillaume, Julie, Nico, Pauline et Virginie mes compagnons de pause café et plus, j'espère. Merci d'avoir rendu mon quotidien doux, drôle et agréable. Merci de m'avoir fait rire en ces derniers temps troublés. Et je n'oublie pas Rv et son frère GrG.

Je tiens à remercier tout particulièrement mes compagnons de voyage Lisbonnais : Amélie, Antoine, Julie, Mohammad et Pauline sans oublier João, pour avoir partager avec moi ces quelques jours de soleil de début septembre avant d'affronter les vents les plus violents de la tempête. Merci à eux de m'avoir aidée à gérer mon stress de dernière minute avec beaucoup d'attention et de gentillesse alors que j'étais loin de ceux qui habituellement m'aident à maintenir le cap.

Enfin, je remercie Bunit sans qui ces pages seraient restées désespérément blanches. Merci d'avoir cru en mes projets dès leurs balbutiements et d'avoir vécu avec moi toutes les étapes de ce parcours avec intérêt, admiration mais aussi esprit critique. Merci pour ton aide et ton immense implication et notamment tes talents de « Matlabeur » et de relecteur averti. Merci d'être toi.

Résumé

Le travail présenté dans ce mémoire est sous-tendu par trois observations majeures. D'abord, de nombreux travaux ont mis en évidence que la parole n'était pas uniquement de nature auditive mais aussi visuelle. D'autre part, la prosodie, domaine de l'intonation, du rythme et du phrasé joue un rôle crucial en parole. Enfin, la deixis ou monstration est un phénomène au cœur de la communication parlée et de son acquisition par les jeunes enfants. Or celle-ci peut, entre autre, s'exprimer uniquement avec la parole : il est possible de « montrer de la voix » par la focalisation prosodique par exemple. Ces observations et constatations permettent d'émettre l'hypothèse que la focalisation contrastive prosodique se manifesterait non seulement par la modalité auditive ,comme il a déjà été largement exploré, mais aussi par la modalité visuelle. C'est la piste que les travaux de ce mémoire visent à explorer pour le cas particulier du français. Plusieurs analyses en production de la parole ont ainsi permis, grâce aux enregistrements de six locuteurs avec deux systèmes de mesure différents et complémentaires, de mettre en évidence les stratégies de signalisation visuelle de la focalisation. Il semble que les locuteurs produisent des indices articulatoires visibles selon deux stratégies principales : la stratégie de signalisation absolue et la stratégie de signalisation différentielle. Les analyses ont également permis de montrer que d'autres gestes faciaux non articulatoires (mouvements des sourcils et de la tête) pourraient être liés à la production de la focalisation mais de façon très variable non seulement d'un locuteur à l'autre mais aussi pour un même locuteur. Par ailleurs, des analyses parallèles en perception, ont permis de montrer que les indices visuels mis en évidence en production, étaient effectivement utilisés en perception et qu'ils permettent d'extraire l'information de focalisation quand la modalité auditive est indisponible ou dégradée. Il a été mis en évidence que les indices visuels identifiés en production correspondent au moins en partie à ceux utilisés en perception audiovisuelle. Ces travaux montrent ainsi que la focalisation contrastive en français est « visible » et est « vue ». Ces résultats permettent d'esquisser un modèle cognitif de la production et de la perception audiovisuelles de la focalisation contrastive en français.

Mots clés : multisensorialité, focalisation, deixis, prosodie, production et perception de la parole, français, articulation, gestes faciaux, contrôle multisensoriel.

Abstract

The work described in this dissertation is grounded by three major findings. Firstly, numerous researchers have shown that speech is not only auditory but also visual. Secondly, prosody *i.e.* intonation, rhythm and phrasing, plays a key role in speech. Thirdly, deixis is a core phenomenon in spoken communication and its acquisition by infants. Deixis can be achieved using speech: it is indeed possible to "show with the voice" using prosodic focus for example. These observations enable us to assume that prosodic contrastive focus is rooted not only in audition, as has already been widely explored, but also in vision. The various works presented in this dissertation explore this hypothesis for French. Several production studies analyzing the recordings of six speakers using two different and complementary measurement techniques have shown that focus is signaled visually. Speakers use two different strategies regarding the visible articulatory movements: an absolute strategy and a differential one. The measurements have also shown that other non-articulatory facial gestures may be linked to the production of contrastive focus such as eyebrow and head movements. The link is however widely inter and intra speaker dependent. In parallel, perceptual experiments have enabled us to show that the visual correlates of focus are used for focus information extraction when the auditory modality is absent or degraded. It was also shown that the visual correlates identified in the production studies correspond at least in part to those used in audiovisual perception. These studies have thus shown that prosodic contrastive focus is "visible" and "seen". The findings allow us to sketch a cognitive model of the audiovisual production and perception of contrastive focus in French.

Key words: multisensoriality, focus, deixis, prosody, speech production and perception, French, articulation, facial gestures, multisensory control.

Table des matières

Liste des figures

Liste des tables

– Notes et indices de lecture –

Le but de cette section est de donner au lecteur quelques outils pour comprendre les conventions qui ont été adoptées dans ce mémoire. Ces notes lui permettront de trouver son chemin facilement et il pourra s'y reporter lorsqu'il trébuchera sur une notation ou une convention.

A. Abréviations

Les abréviations suivantes seront utilisées de façon récurrente dans ce manuscrit afin d'en rendre la lecture plus claire et synthétique.

AL	aire intéro-labiale
ANOVA	Analyse de la Variance
EL	étirement des lèvres
F0	fréquence fondamentale
FS	focalisation sur le sujet
FV	focalisation sur le verbe
FO	focalisation sur l'objet
ICP	Institut de la Communication Parlée
MM	mouvements de la mandibule
O	objet
OL	ouverture des lèvres
P	protrusion de la lèvre supérieure
S	sujet
SA	Syntagme Accentuel
SI	Syntagme Intonatif
V	verbe

B. Convention typographique

Dans les exemples qui émaillent ce mémoire, la focalisation sera signalée par des capitales *e.g.* dans l'exemple ci-dessous, c'est le constituant *Romain* qui est focalisé :

ROMAIN mange la pomme.

C. Conventions

C.1. Cas neutre

Dans un but de clarté, le terme *cas neutre* sera utilisé dans ce mémoire en référence au cas où aucun des constituants d'un énoncé n'est focalisé. Bien que conscients du fait que le terme exact à employer en linguistique est *focalisation large* (*vs. focalisation étroite* quand l'un des constituants est focalisé), nous avons opté pour cette convention afin de faciliter la lecture à tous, et surtout aux non linguistes.

C.2. Contrastes intra- et inter-énoncés

Dans la suite de ce mémoire, la terminologie suivante sera utilisée pour faire référence aux différents phénomènes analysés :

- Contraste intra-énoncé : différence syntagmatique entre le constituant focalisé et le reste de l'énoncé ;
- Contraste inter-énoncés : différence paradigmatique entre le constituant focalisé et ce même constituant dans le cas neutre.

C.3. Hyper-articulation

Le terme *hyper-articulation* sera utilisé dans ce mémoire pour référer à l'augmentation d'un paramètre articulatoire précis et non pour référer à une tendance globale impliquant tous les paramètres articulatoires.

C.4. Hypo-articulation

Le terme *hypo-articulation* sera utilisé dans ce mémoire pour référer à la diminution d'un paramètre articulatoire précis et non pour référer à une tendance globale impliquant tous les paramètres articulatoires.

C.5. Seuil de signification pour les tests statistiques

Dans la totalité de ce mémoire, les résultats des tests statistiques ont été considérés comme étant significatif à un niveau de 5%, c'est-à-dire pour $p<0,05$. Les niveaux de p obtenus seront néanmoins cités afin que le lecteur puisse se faire une idée du niveau de signification.

C.6. Vitesse

Dans l'intégralité de ce mémoire, lorsque le terme *vitesse* sera utilisé, il fera référence à la vitesse de variation d'un paramètre c'est-à-dire à sa dérivée première en fonction du temps.

D. Locuteurs enregistrés

Au total, 6 locuteurs différents ont été enregistrés pour les expériences en production. Certains ont participé à plusieurs des expériences décrites dans ce manuscrit. Ils seront ci-après nommés locuteurs A, B, C, D, E et F sachant que les locuteurs A et B ont été enregistrés plusieurs fois et que leurs productions ont été notamment utilisées pour fabriquer les stimuli de tous les tests perceptifs. Tous les locuteurs étaient de sexe masculin, âgés de 20 à 45 ans et francophones de langue maternelle ayant tous grandi et appris à parler en France métropolitaine.

– Introduction –

Language is a new machine built out of old parts […] emerging from a nexus of skills in attention, perception, imitation and symbolic processing that transcend the boundaries of 'language proper'

Bates & Dick

L'intonation de ses petites phrases coupées, le clignement de ses yeux achèvent de me révéler sa situation exacte dans la maison du capitaine Mauger.

Le journal d'une femme de chambre
Octave Mirbeau

Finalement les mots sortent par le nez, la bouche, les oreilles, les pores, entraînant avec eux tous les organes […] dans un envol puissant, majestueux, qui n'est autre que ce qu'on appelle, improprement, la voix, se modulant en chant ou se transformant en un terrible orage symphonique avec tout un cortège… des gerbes de fleurs des plus variées, d'artifices sonores : labiales, dentales, occlusives, palatales et autres, tantôt caressantes, tantôt amères ou violentes.

La leçon
Eugène Ionesco

A panda walks into a cafe. He orders a sandwich, eats it, and then draws a gun and fires two shots in the air.

"Why?" asks the confused waiter, as the panda makes towards the exit. The panda produces a badly punctuated wildlife manual and tosses it over his shoulder.

"I'm a panda," he says, at the door. "Look it up."

The waiter turns to the relevant entry and, sure enough, finds an explanation.

"**Panda**. Large black-and-white bear-like mammal, native to China. Eats, shoots and leaves."

d'après *Eats, shoots and leaves*
Lynne Truss

« ne va pas t'aviser au moins de suggérer une pareille idée à mon père » s'écria Octave du ton d'un homme effrayé.

Cinq cents millions de la Begum
Jules Verne

Monsieur B… rugit encore une fois, mais sur un ton interrogatif…

Monsieur B… rugit ainsi qu'il a été dit.

… Cela veut dire : « Veux-tu venir te promener avec moi ? » Or, si Madame B… répond sur un ton lassé, chromatique et descendant… (*Madame B… rugit lamentablement.*) … cela veut dire : « Non ! Je suis fatiguée, restons à la maison. »

Si, au contraire, elle répond avec entrain (*Madame B… fait entendre un « Iroum » joyeux*) … cela veut dire : « Oui, je veux bien. Sortons ! »

Si enfin elle rugit avec allégresse deux ou trois fois en sautillant… (*Madame B… fait comme il est dit.*) … cela veut dire : « Allons au cinéma ! »

Ce que parler veut dire
Jean Tardieu

Le flegmatique gentleman l'écoutait, en apparence au moins, avec la plus extrême froideur sans qu'une intonation, un geste décelât en lui la plus légère émotion.

Le tour du monde en 80 jours
Jules Verne

Dans le commerce des humains, bien souvent les mouvements du corps, les intonations de la voix et l'expression du visage en disent plus long que les paroles.

Un mot pour un autre
Jean Tardieu

Harry Caul est détective. Son travail consiste à enregistrer discrètement des conversations pour ses clients à qui il vend ensuite les cassettes. Il vit rongé par le remord d'avoir ainsi vendu une cassette puis d'avoir découvert que l'un des protagonistes de l'enregistrement avait été assassiné ainsi que sa famille.

Alors qu'il vient d'enregistrer une conversation entre un homme et une femme qui sont apparemment amants, il essaie d'éliminer le bruit pour clarifier certaines zones du dialogue. Ce faisant, il entend l'homme dire :

"He'd kill us if he'd got the chance."

Cette phrase le plonge dans une angoisse profonde. En enquêtant, il découvre que le commanditaire de l'enregistrement n'est autre que le mari de la jeune femme enregistrée. Il se persuade alors que la jeune femme va se faire assassiner par son mari jaloux. Se souvenant d'un rendez-vous que les amants s'étaient fixés dans un hôtel, il loue la chambre juste à côté le même jour. Alors qu'il attend en redoutant ce qui risque d'arriver, quelqu'un se fait assassiner.

Le remords envahit alors Harry Caul et, dans un accès de folie, il se rend là où le mari jaloux travaille pour découvrir que c'est en fait lui qui a été tué … L'amant avait en fait dit:

"He'd kill US if he'd got the chance."

résume du film *The Conversation*
film de Francis Ford Coppola

Puis elle ajouta un : « Monsieur je vous remercie » dont l'intonation équivalait à un congé.

Etudes de mœurs – 1er livre – Scènes de la vie privée – T. 1 – La paix du ménage
Honoré de Balzac

« Vous aviez donc une lentille, monsieur ? demanda Harbert à Cyrius Smith.
- Non, répondit celui-ci, mais j'en ai fait une. »
Et il montra l'appareil qui lui avait servi de lentille.

L'île mystérieuse
Jules Verne

Mais le marin n'en était pas à cinquante pas qu'il s'arrêtait, poussait de nouveau un hourra formidable, et, tendant la main vers l'angle de la falaise : « Harbert ! Nab ! Voyez ! » s'écriat-il.

L'île mystérieuse
Jules Verne

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

La grammaire est une chanson douce
Erik Orsenna

Il est vrai que certains de ces extraits sont amusants ou intrigants, mais vous vous demandez certainement ce qu'ils font ici. Ils ne semblent avoir ni de rapport entre eux ni de lien avec le titre de ce manuscrit et le sujet qui est censé y être traité. Pourtant, détrompez-vous ! Ces petites touches glanées ici ou là vont permettre d'aborder la problématique de façon exploratoire et intuitive. Elles évoquent ou analysent en effet toutes, des situations de communication naturelles qui font appel à divers mécanismes communicationnels. Les phénomènes auxquels elles font référence et leur analyse permettront de construire le corps de la problématique qui sera, par la suite, abordée de façon expérimentale.

A. La parole n'est pas qu'auditive …

… contrairement à l'idée répandue, elle est en effet aussi visuelle.

Pour l'apprécier de façon intuitive, pensons à une situation dans laquelle chacun de nous s'est un jour retrouvé : au téléphone. Il est en effet parfois difficile de bien se comprendre au téléphone même quand il n'y a pas de « friture ». Prenons un exemple typique : lorsqu'on épelle un mot, il est bien souvent nécessaire de dire « m, comme dans *maman* » sinon il y a des chances que l'auditeur à l'autre bout du fil fasse une erreur. Les sons /m/ et /n/ sont souvent confondus parce que très semblables au niveau acoustique. Pourtant, dès que l'on voit notre interlocuteur, cette distinction devient nettement plus facile. Pour le /m/ le locuteur va en effet fermer la bouche alors que pour le /n/ non.

Selon nous, la parole, qui sera l'objet principal de notre attention dans ce mémoire, doit donc être envisagée sous l'angle de la multisensorialité et surtout en tant qu'objet audiovisuel. Le but de cette section sera ainsi d'introduire la notion de bimodalité en parole et de proposer au lecteur les bases scientifiques d'une démarche expérimentale explorant cette notion. Il ne s'agit pas ici de faire une revue exhaustive de la littérature concernant la parole audiovisuelle mais simplement d'en introduire les fondements. Des revues détaillées de la littérature dans ce domaine pourront néanmoins être trouvées dans : Summerfield [1987], Massaro [1987, 1989], Campbell [1988], Cathiard [1988/1989, 1994], Vroomen [1992], Robert-Ribes [1995] et Schwartz [2004].

A.1. Se voir pour mieux se comprendre

Afin de bien appréhender le fait que la parole est aussi visuelle, intéressons-nous d'abord aux personnes sourdes et malentendantes. Chacun sait que nombre d'entre elles peuvent en partie lire sur les lèvres : c'est la lecture labiale. Les sourds et malentendants ne sont pourtant pas les seuls à en être capables. Bien que cette faculté soit moins développée chez les personnes bien entendantes, elle joue un rôle important et inconscient dans la communication. La seule lecture labiale permet en effet de discriminer 40 à 60% des phonèmes d'une langue et de 10 à 20% des mots (données reprises dans Schwartz [2004]). On notera cependant que l'intelligibilité visuelle des mots dépend aussi fortement du locuteur.

C'est lorsque le signal audio est dégradé que l'on se rend le mieux, et le plus facilement, compte de l'intérêt de se voir en se parlant. La vision nous aide, en effet, considérablement à communiquer en milieu bruité et de façon générale, quand l'information auditive est dégradée. Ceci a d'ailleurs été mis en évidence lors à de nombreuses études (Sumby & Pollack [1954], Miller & Nicely [1955], Neely [1956], Erber [1969, 1975], Binnie *et al.* [1974], Summerfield [1979], MacLeod & Summerfield [1987], Grant & Braida [1991], Benoît *et al.* [1994]).

Sumby & Pollack [1954] ont ainsi montré que le fait de fournir la modalité visuelle en plus de la modalité auditive pour percevoir la parole en milieu bruité était équivalent à améliorer le rapport signal sur bruit de 15 dB. MacLeod & Summerfield [1987] ont effectué la même mesure et ont mis en évidence une contribution de la lecture labiale équivalente à une amélioration du rapport signal sur bruit de 11 dB. Or Miller *et al.* [1951] trouvent qu'une amélioration de 1 dB du rapport signal sur bruit peut correspondre à une augmentation de 5 à 10% de l'intelligibilité. On peut ainsi, grâce à ces données chiffrées, mieux se représenter l'importance de l'apport de la modalité visuelle pour la perception de la parole en milieu bruité.

Ce qui est intéressant dans le fait que la modalité visuelle aide à mieux se comprendre en milieu bruité, ce n'est pas simplement que celle-ci fournisse des informations rendues indisponibles par les conditions sonores, c'est également la complémentarité qui existe entre les modalités auditive et visuelle qu'il convient d'apprécier. Il apparaît en effet que les composantes de la parole les moins robustes au bruit acoustique correspondent à des paramètres articulatoires bien visibles alors que les composantes difficilement visibles correspondent, quant à elles, souvent à des indices acoustiques robustes au bruit. La vision permet souvent d'extraire l'information de lieu d'articulation alors que l'acoustique permet d'obtenir l'information de mode d'articulation (*e.g.* voisé *vs* non voisé). Benoît *et al.* [1994] ont ainsi montré que l'intelligibilité en modalité audio seul de la voyelle [a] est plus importante que celle de la voyelle [i] et que celle de la voyelle [y] alors qu'en modalité visuelle seule, celle de [y] (voyelle pour laquelle les lèvres sont très arrondies et protruses) est plus importante que celle de [a] et que celle de [i]. Il apparaît ainsi clairement que la modalité visuelle n'est pas un simple substitut ou un auxiliaire pour la perception de la parole mais qu'elle est réellement complémentaire à la modalité auditive.

Bien que l'apport de la modalité visuelle soit le plus facile à mesurer en milieu bruité, d'autres études montrent que la vision est systématiquement utile pour la perception de la parole. Il apparaît ainsi, par exemple, que les indices auditifs sont renforcés perceptivement lorsque la modalité visuelle est disponible même en milieu non bruité. Lorsque l'on voit notre interlocuteur parler, on l'entend mieux et on a l'impression qu'il parle plus fort. Grant & Seitz [2000] ont montré que le seuil de détection auditive de la parole (et non de compréhension, attention !) pouvait être abaissé de 1 à 2 dB lorsque la modalité visuelle était disponible en plus de la modalité auditive. L'apport de la modalité visuelle pour la perception de la parole a également été démontré lorsque la modalité auditive était intacte. Reisberg *et al.* [1987] ont en effet montré que la modalité visuelle améliorait la perception de la parole lorsque les signaux de parole étaient acoustiquement très clairs. Ils ont en effet utilisé trois types de signaux pour lesquels les performances perceptives en modalité audio seul étaient très moyennes, mais pas à cause du bruit. Les signaux de parole utilisés correspondaient soit à une langue avec laquelle les sujets n'étaient pas familiers, soit à des productions d'un locuteur étranger parlant avec un accent, soit à un contenu sémantique très complexe (*La Critique de la raison* pure de Kant). Dans tous ces cas, les performances perceptives étaient meilleures quand les sujets pouvaient voir le locuteur parler en plus de l'entendre.

Il vous paraîtra maintenant évident que la modalité visuelle intervient dans les processus de perception et de compréhension de la parole. Mais peut-être pensez-vous qu'il s'agit là simplement d'une aide qui nous est bien utile lorsque le son ne suffit pas. La vision joue pourtant un rôle à part entière dans la perception de la parole. Elle ne nous aide pas toujours d'ailleurs. Elle nous induit même parfois en erreur comme c'est le cas pour la célèbre illusion de McGurk (McGurk & MacDonald [1976]). McGurk & MacDonald ont ainsi été les premiers à mettre en évidence que des stimuli auditif et visuel conflictuels (/ba/ audio et /ga/ vidéo) présentés simultanément donnent lieu à un percept différent des deux stimuli d'origine (/da/ dans ce cas). Il apparaît ainsi clairement que la vision est pleinement actrice du processus de perception de la parole. Les informations auditives et visuelles semblent en effet être combinées lors du processus de perception de la parole. Cette combinaison n'a d'ailleurs pas seulement été mise en évidence au niveau infra-lexical de la syllabe. Dodd [1977] a ainsi effectué une étude au niveau lexical. Il rapporte que, lorsque les informations auditive et visuelle sont conflictuelles pour un mot donné, il arrive que les sujets combinent les deux informations et perçoivent un troisième mot différent, à la fois du mot présenté acoustiquement et de celui présenté visuellement.

Vous comprendrez maintenant que lorsque l'on écoute quelqu'un parler, il n'y a pas que nos oreilles et notre cortex auditif qui travaillent mais aussi nos yeux et notre cortex visuel. De façon claire, la vision intervient dans la perception de la parole et il est donc absolument primordial d'envisager la parole sous l'angle audiovisuel.

A.2. Voir pour mieux apprendre à parler ?

A.2.1. Modalité visuelle et développement du langage

De nombreux chercheurs ont étudié le rôle que pourrait jouer la modalité visuelle dans l'acquisition du langage par les enfants. La multisensorialité se développe en effet de manière très précoce. Comme le souligne Bahrick [2003], les enfants montrent très tôt des capacités intermodales et des connections entre les sens. Le développement de ces capacités se fait très rapidement pendant la première année de la vie et la multimodalité permet même d'accélérer certaines phases du développement.

Kuhl & Meltzoff [1982, 1984] ont observé que des bébés de 18 à 20 semaines préfèrent déjà des stimuli de parole pour lesquels il y a correspondance entre acoustique et visuel, à des stimuli audiovisuels incohérents. Ils ont même tendance à essayer d'imiter les stimuli cohérents. Legerstee [1990] a montré que des bébés de trois à quatre mois imitaient les voyelles /a/ et /u/ prononcées par un adulte uniquement lorsque les stimuli audio et vidéo qui leur étaient présentés étaient concordants. Ces observations pourraient refléter une connaissance précoce des liens entre audition et articulation. La notion de multisensorialité serait donc associée très tôt à la parole. L'information multimodale pourrait même jouer un rôle dans l'acquisition des sons de la parole en favorisant leur apprentissage.

Suite à plusieurs études, Lewkowicz [1998] conclut que le développement des processus d'intégration des informations visuelle et auditive chez les très jeunes enfants dépend de nombreux facteurs tels que l'expérience personnelle et le stade de développement du langage ou encore le mode d'élocution (parole dirigée vers des adultes, vers des bébés ou chant).

Gogate & Bahrick [1998] ont observé chez des bébés de sept mois que la synchronie temporelle entre vocalisations et mouvement d'un objet facilitait l'apprentissage de relations parole-objet arbitraires. Il semblerait donc que les informations visuelles non forcément relatives à la parole (*i.e.* le fait de voir l'objet correspondant au mot bouger en même temps que le mot est prononcé par exemple) puissent faciliter l'apprentissage lexical.

Enfin, il a été observé que les enfants aveugles avaient plus de difficultés à acquérir certains contrastes phonémiques (Mulford [1988]). La distinction entre [m] et [n] par exemple est bien visible (occlusion labiale pour le premier et alvéolaire pour le second) mais très faiblement audible et les enfants aveugles mettent plus de temps à l'acquérir que les autres.

A la lumière de ces résultats, on pourra conclure que, bien qu'on ne sache pas encore précisément comment la modalité visuelle intervient dans l'acquisition du langage, il paraît évident qu'elle y joue un rôle qui est, dans certains cas, assez important.

A.2.2. Modalité visuelle et apprentissage d'une langue étrangère (L2)

Certaines études montrent que la perception d'une langue étrangère semble favorisée par la lecture labiale (Reisberg *et al.* [1987] et Kim & Davis [2003]). On pourrait donc supposer que la vision est utile lors de l'apprentissage d'une L2. Pourtant les résultats des études abordant ce sujet sont assez mitigés.

Hazan *et al.* [2002] ont étudié l'influence de l'information visuelle sur la perception de contrastes phonémiques de l'anglais (/b/-/v/ et /b/-/p/) par des espagnols apprenant l'anglais et pour qui ces contrastes n'étaient pas natifs. Les auteurs n'ont noté aucun avantage audiovisuel pour désambiguïser les contrastes. Une analyse détaillée des résultats en fonction des sujets et notamment de leur niveau d'apprentissage de l'anglais, a permis aux auteurs de conclure que la sensibilité aux indices visuels était apprise tout comme celle aux indices acoustiques. Les auteurs notent aussi une dépendance assez forte par rapport au sujet en ce qui concerne la sensibilité aux indices visuels.

Les mêmes chercheurs (Sennema *et al.* [2003]) ont par la suite étudié la perception audiovisuelle du contraste /l/-/r/ de l'anglais chez des sujets japonais apprenant l'anglais. Les performances perceptives en audiovisuel par rapport à l'audio seul étaient meilleures pour seulement 18,5% des sujets. Ces résultats sont en contradiction avec ceux de Hardison [2003] qui avait noté un avantage audiovisuel pour la perception du même contraste. Sennema *et al.* suggèrent que cette différence puisse être due au fait que les indices visuels correspondant au contraste /l/-/r/ sont plus forts pour l'anglais américain, qu'avait utilisé Hardison, que pour l'anglais britannique qu'ils avaient eux-mêmes utilisé. Sennema *et al.* ont noté que les sujets britanniques étaient sensibles à l'information visuelle puisque leur score d'identification du contraste /l/-/r/ en visuel seul est excellent (80%). Les sujets japonais ne sont pas sensibles à ces mêmes informations puisque leur score est de 55,2% (proche du hasard). Les auteurs concluent que, bien que l'information visuelle ne puisse peut-être pas aider directement pour l'apprentissage d'une L2, il est possible que le développement d'une sensibilité visuelle puisse être utile, par la suite, pour la communication en langue étrangère.

A.3. Que peut-on voir qui nous aide à mieux comprendre ?

> Finalement les mots sortent par le nez, la bouche, les oreilles, les pores, entraînant avec eux tous les organes.
>
> *La Leçon*
> Eugène Ionesco

Maintenant qu'il paraît sans doute plus clair que la modalité visuelle joue un rôle important dans la communication parlée, vous vous demandez peut être *comment* elle joue un rôle *i.e.* qu'est-ce qui est présent dans le flux visuel qui permet de rendre la communication plus efficace ?

Souvenons-nous de la lecture labiale évoquée précédemment, elle permet apparemment de recouvrer beaucoup d'informations quant au message lorsque le signal audio est dégradé ou non disponible. On peut penser, comme le suggère le terme qui a été emprunté pour désigner ce phénomène (*lecture labiale*), que ce sont les mouvements des lèvres qui fournissent des informations. Leur forme, par exemple, doit certainement être une information utile pour discriminer les phonèmes les uns des autres : arrondissement (comme pour /u/), étirement (comme pour /i/), ouverture (comme pour /a/). Au moins une partie des informations utiles et utilisées doit ainsi se trouver dans la zone de la bouche. Les formes des lèvres, de la partie visible des dents et de la langue permettent de déterminer les lieux d'articulation de nombreux phonèmes et ainsi de remonter à ce qui a été dit lorsque le bruit masque l'information acoustique (McGrath *et al.* [1984]). Pourtant, il a déjà été suggéré par le passé que d'autres indices présents dans le visage sont importants, surtout au niveau paralinguistique et notamment les mimiques et expressions faciales (Ekman [1999]).

Benoît *et al.* [1996b] ont mené une étude comparative de l'intelligibilité visuelle lorsque tout le visage du locuteur est visible ou seulement ses lèvres. Il apparaît que les scores d'intelligibilité sont plus élevés lorsque tout le visage du locuteur est disponible. Les auteurs concluent que: *« lips alone carry on average the two-thirds of the speech intelligibility carried out by the whole natural face »* (les lèvres seules véhiculent environ les deux tiers de l'intelligibilité de la parole véhiculée par le visage naturel tout entier). On peut donc penser que les informations visuelles utilisées ne correspondent pas toutes aux mouvements de la bouche et que le visage dans son intégralité doit être visible pour que la communication soit optimale.

Certains chercheurs ont avancé que l'aspect dynamique du mouvement perçu quand quelqu'un nous parle, est primordial. Campbell [1992] a en effet montré que, sans capacité à percevoir le mouvement, il est difficile de lire sur les lèvres de façon efficace. Rosenblum & Saldaña [1996] ont utilisé le dispositif du *Point Light Display* qui consiste à placer sur le visage d'un locuteur des points lumineux qui seront ensuite les seuls indices visuels disponibles pour la perception. Suite à cette étude, ils ont conclu que c'était la dynamique visuelle qui fournissait les informations pour la perception audiovisuelle de la parole. D'autres études montrent néanmoins que ce n'est pas forcément la partie dynamique qui fournit toutes les informations visuelles pour la perception visuelle de la parole. Cathiard [1994] a étudié de façon approfondie et détaillée les différences perceptives observées pour des tests menés avec des images (statiques) ou des films (dynamiques). Elle conclut finalement que : « L'apport de la dynamique n'apparaît pas ainsi donné comme un primat, mais comme un indice parmi d'autres pour récupérer l'information sur le trait dans les conditions variables de communication ». Le mouvement ne servirait en fait que quand l'information statique est

insuffisante, mais l'information récupérée serait bien une information sur la forme statique. Une réanalyse récente de données sur la perception visuelle chez les macaques, entreprise par Perret et collègues (cf. Jellema & Perret [2003] et Perret [2005]), va d'ailleurs dans ce sens. Les neurones sensibles à une orientation particulière des visages répondent de façon plus nette (avec une précision temporelle et une amplitude de réponse plus élevée) lorsque l'information visuelle est statique que lorsqu'elle est dynamique.

Plusieurs études (*e.g.* Grant & Seitz [2000], Schwartz *et al.* [2004]) ont exploré la possibilité que les indices fournis par l'information visuelle puissent aussi être temporels (et non seulement spatiaux). Schwartz *et al.* [2004] ont montré que ce n'est pas seulement l'information visuelle en elle-même qui est importante mais sa cohérence temporelle avec le son. Il apparaît ainsi que pour des gestes articulatoires visibles exactement identiques (même signal vidéo) combinés à des signaux auditifs différents, la discrimination phonémique est meilleure en audiovisuel qu'en audio seul. Il semblerait en fait que la modalité visuelle apporte aussi une information temporelle qui permet d'attirer l'attention du spectateur sur la partie du signal à écouter et donc de faciliter la compréhension. Schwartz *et al.* [2004] ont de plus montré que ces informations temporelles étaient spécifiques à la parole. Ils ont en effet renouvelé leur expérience en fournissant comme signal visuel aux spectateurs, une barre de volume qui bougeait exactement de la même façon que les lèvres dans leur première expérience. Or l'effet de l'avantage audiovisuel n'est plus mesuré dans cette situation.

On notera aussi que plusieurs études se sont penchées sur la visibilité des mouvements articulatoires « internes » et donc *a priori* non directement visibles tels que les mouvements de la langue. Intuitivement, on pourra affirmer que certains mouvements de la langue doivent en effet être visibles lorsque la bouche est simultanément ouverte, comme l'abaissement de l'apex de la langue lors de la production de la syllabe /la/. Yehia *et al.* [1998], Kuratate *et al.* [1999] et Jiang *et al.* [2000] ont ainsi mesuré, pour un même signal de parole, les mouvements faciaux et les mouvements de la langue. Ces études ont toutes montré qu'il existait une forte corrélation entre les mouvements faciaux et la forme de la langue. Elles ne fournissent cependant pas d'information sur ce qui pourrait être récupéré comme indices sur la forme de la langue à partir des données faciales. Bailly & Badin [2002] ont, quant à eux, montré que l'information visuelle n'était pas suffisante pour extraire des informations telles que la constriction linguale. Ces résultats ont été confirmés par l'étude de Engwall & Beskow [2003] qui a montré que les informations visuelles permettaient de bien reconstruire l'ouverture de la mâchoire et les mouvements de l'apex de la langue mais pas de prédire une constriction non alvéolaire du conduit vocal. Il ressort de l'ensemble de ces études que certaines informations sur les mouvements de la langue peuvent être récupérées à partir de l'observation des mouvements faciaux. L'information visuelle est néanmoins loin d'être suffisante pour remonter à la forme de la langue.

En considérant les études décrites ci-dessus, on pourra conclure que ce qui est essentiellement utile perceptivement pour la parole, ce sont les lèvres. Le visage tout entier fournit cependant plus d'informations que les lèvres seules et peut même permettre de recouvrer certaines informations sur le mouvement des articulateurs internes. Les informations temporelles et dynamiques jouent apparemment aussi un rôle. La modalité visuelle fournit donc un ensemble d'informations qui pourront être utilisées à divers niveaux des processus de perception et de compréhension de la parole. Ces informations forment un tout dont chaque partie peut être utile.

A.4. Et que regarde-t-on ?

Bien que, comme il vient d'être mis en évidence, plusieurs types d'informations soient disponibles visuellement, on pourra se demander lesquelles attirent le plus l'attention de l'auditeur/spectateur et donc quelles parties du visage il regarde. Pour répondre à cette question, quelques études ont été consacrées à la trajectoire de notre œil lorsque l'on regarde quelqu'un parler.

Vatikiotis-Bateson *et al.* [1998] décrivent une expérience de suivi du regard lors de la perception audiovisuelle de la parole. Les auteurs ont utilisé un dispositif infrarouge pour enregistrer les mouvements oculaires. Cette étude concerne le japonais et l'anglais et le test a été mené pour différents niveaux de bruit environnant. Elle a permis de montrer que, en condition de lecture labiale, nos yeux se fixent principalement sur la bouche et les yeux du locuteur avec une nette préférence pour l'un des deux yeux. Pour des niveaux de bruits environnants faibles, les fixations se font 65% du temps sur l'un des deux yeux du locuteur. Quand le niveau de bruit environnant devient plus important, les sujets ont tendance à regarder plus la bouche bien qu'ils fixent l'un des yeux encore 45% du temps. Les saccades de l'œil vers la bouche sont aussi deux fois moins fréquentes quand il y a beaucoup de bruit. Les auteurs observent également que les comportements relevés sont très peu différents de ceux qui avaient été identifiés dans d'autres études portant sur les mouvements oculaires lors de l'observation d'un visage non parlant. Vatikiotis-Bateson *et al.* font l'hypothèse que *« phonetically relevant visual information occurs all over the face, not just the perioral regions defined by the lips »* (des informations phonétiquement pertinentes sont disponibles sur le visage tout entier et non uniquement dans la région péri-orale définie par les lèvres). Or, on se souviendra que Benoît *et al.* [1996b] avaient obtenu de meilleurs résultats perceptifs en fournissant à leurs sujets tout le visage du locuteur plutôt que ses lèvres seules. Il est en effet évident que lorsque la mâchoire et les lèvres bougent, il y a des répercussions sur tout le visage et il a même été observé qu'il était possible de recouvrer certaines informations labiales lorsque seules d'autres parties du visage (sans les lèvres) étaient visibles. Finalement, Vatikiotis-Bateson *et al.* pensent que *« perceivers may produce habituated eye movement patterns that serve both phonetic and higher level, sociolinguistic criteria »* (il est possible que les sujets produisent des mouvements oculaires systématiques conditionnés non seulement par des critères phonétiques mais aussi par des critères sociolinguistiques de plus haut niveau). Ceci expliquerait aussi que les fixations ne portent pas uniquement sur la bouche du locuteur. Les auteurs n'ont relevé aucune différence significative d'une langue à l'autre. On retiendra de cette étude qu'il apparaît que le système oculomoteur est très actif lors du processus de perception audiovisuelle de la parole et que, bien que la bouche soit beaucoup fixée, elle est loin d'être l'unique point d'intérêt lorsque l'on regarde quelqu'un parler.

Paré *et al.* [2003] ont étudié l'influence des lieux de fixation oculaire sur la perception de l'effet Mc Gurk décrit précédemment. Leur étude concernait l'anglais canadien. Le résultat principal confirme l'étude décrite précédemment : les fixations oculaires des sujets percevant de la parole audiovisuelle se répartissent principalement entre la bouche et les yeux. Les auteurs ont également observé qu'il n'y avait pas de lien entre l'endroit de fixation et l'efficacité de l'effet Mc Gurk. Même quand la direction du regard des sujets était attirée en dehors du visage du locuteur, ce n'est que pour une déviation de plus de 10° à 20° que l'influence sur la perception s'est faite sentir, cette influence n'étant de plus que très légère. Les auteurs concluent que *« subjects do not fixate and, more important need not fixate exclusively on the talker's mouth to perceive linguistic information »* (les sujets ne fixent pas et même, n'ont pas besoin de fixer, exclusivement la bouche du locuteur pour percevoir des informations linguistiques). Les auteurs demeurent néanmoins prudents dans leurs interprétations,

notamment sur les liens entre attention et direction du regard à cause du nombre d'inconnues et de biais expérimentaux qui pourraient exister. Ils soulignent l'importance de mener des études complémentaires.

Il apparaît ainsi que lorsque nous regardons quelqu'un parler, nous regardons non seulement l'évolution de sa bouche, manifestation la plus directement liée à la parole, mais aussi dans une large proportion ses yeux. Il semble de plus que nous n'ayons pas besoin de fixer exclusivement la bouche ni même de fixer un quelconque emplacement du visage pour percevoir les informations relatives à l'articulation. La vision périphérique suffit à extraire les informations visuelles importantes pour la perception de la parole. C'est donc, selon toute vraisemblance, le visage dans sa globalité qui joue un rôle dans la perception audiovisuelle de la parole et non uniquement la bouche et la mâchoire comme il aurait pu intuitivement être inféré.

A.5. Intégration des informations auditives et visuelles

La vision et l'audition interagissent donc dans la compréhension de la parole et ce de façon systématique et non nécessairement consciente. Il reste cependant à déterminer comment s'opère la fusion des informations auditives et visuelles pour donner lieu à une décision unique. C'est un problème auquel les chercheurs tant psychologues, ingénieurs que linguistes se sont intéressés depuis longtemps. Plusieurs revues de travaux sont ainsi disponibles sur le sujet de l'intégration audiovisuelle, c'est-à-dire du traitement conjoint et corrélé des informations auditives et visuelles. Le lecteur pourra se reporter à Summerfield [1987], Robert-Ribes [1995], Schwartz *et al.* [1998] et Schwartz [2004]. De nombreuses études comportementales ont également été menées en psychologie et en neurophysiologie.

A.5.1. Comment la fusion s'opère-t-elle ?

Robert-Ribes [1995] et Schwartz *et al.* [1998] mettent en évidence quatre architectures possibles pour la fusion audiovisuelle :

- **Modèle à « Identification Directe »** noté ID : dans ce modèle, les flux auditifs et visuels sont compilés, la classification se fait directement sans étape préliminaire de mise en forme commune des données ;
- **Modèle à « Identification Séparée »** noté IS : la classification phonétique se fait séparément sur les entrées auditives et visuelles et la fusion ne s'opère qu'après cette classification séparée : la fusion est tardive et il s'agit dans ce cas d'une fusion de décision ;
- **Modèle à « Recodage dans la modalité Dominante »** noté RD : dans ce type de modèle, la modalité auditive est considérée comme modalité dominante et l'entrée visuelle est recodée sous un format compatible avec celui des représentations auditives, il s'agit là d'une fusion précoce ;
- **Modèle à « Recodage commun des deux entrées sensorielles vers la modalité Motrice »** noté RM : dans ce type de modèle, les principales caractéristiques articulatoires

sont estimées à l'aide des informations auditives et visuelles et sont ensuite soumises à un processus de classification. Il s'agit d'un modèle de fusion précoce.

Schwartz [2004] indique que les modèles ID et IS sont ceux qui sont les plus fréquemment utilisés en reconnaissance de parole. Les données fournies par la psychologie expérimentale l'ont cependant conduit, lui et ses collègues, à privilégier le modèle RM (cf. Schwartz *et al.* [1998]). Ce modèle tire en partie son inspiration de la célèbre théorie motrice de la perception de la parole proposée par Liberman & Mattingly [1985]. Selon cette théorie *« phonetic information is perceived in […] a module specialized to detect the intended gestures of the speaker that are a basis for phonetic categories »* (l'information phonétique est perçue par un module spécialisé dans la détection des gestes planifiés par le locuteur qui sont la base des catégories phonétiques).

A.5.2. Où s'opère-t-elle ?

Au delà des interactions sous-corticales largement décrites par Stein & Meredith [1993] dans le colliculus supérieur pour la localisation chez le chat, les données récentes en neuroimagerie font ressortir deux séries d'observations qui s'ancrent bien sur les résultats expérimentaux décrits ci-dessus. D'une part, les interactions corticales sont précoces, incluant notamment des interactions dès le cortex auditif primaire (Calvert *et al.* [1997]) et largement dans le cortex auditif secondaire (Lebib *et al.* [2003], Besle *et al.* [2004], Ghazanfar *et al.* [2005], Ojanen *et al.* [2005] et van Wassenhove *et al.* [2005]). Ensuite un circuit temporo-pariéto-frontal a été décrit récemment dans le cadre de l'observation et de l'imitation de mouvements. Selon Iacoboni [2005], au sein de ce réseau, le cortex temporal supérieur fournit une description visuelle de l'action à imiter aux neurones miroirs du cortex pariétal postérieur. Ce dernier élabore une information somato-sensorielle sur l'action à imiter qui est envoyée aux neurones miroirs du cortex frontal inférieur. Le cortex frontal inférieur à son tour code le but de l'action à imiter. De plus, des copies des commandes motrices fournissant les conséquences sensorielles prédites de l'action planifiée sont renvoyées au cortex temporal. La description visuelle de l'action et les prédictions sont alors comparées pour optimiser le déroulement du plan moteur. Or ce réseau est clairement impliqué dans le processus d'interaction audiovisuelle (Callan *et al.* [2003], Calvert & Campbell [2003] et Watkins *et al.* [2003]) ce qui est en bon accord avec l'hypothèse d'un recodage moteur mentionnée précédemment.

Il est donc important de se voir pour bien se comprendre et ce pas uniquement lorsqu'il y a du bruit mais dans nombre d'autres situations. La modalité visuelle est vraisemblablement liée à la parole très tôt dans l'acquisition du langage par les bébés. Il semblerait que les lèvres ne soient pas les seules à fournir de l'information et que l'on perçoive en fait un « tout » visuel dont chacune des « parties » est importante. L'intégration des informations des deux modalités se fait assez tôt et apparemment dans un cadre de recodage vers la modalité motrice ou, en tout état de cause, dans un réseau cortical impliquant des relations entre aires sensorielles et motrices.

Dans le commerce des humains, bien souvent les mouvements du corps, les intonations de la voix et l'expression du visage en disent plus long que les paroles.

Un mot pour un autre
Jean Tardieu

B. Place et rôles de la prosodie dans la communication parlée

Analysons maintenant de façon plus approfondie la plaisanterie du panda citée au tout debut de l'introduction :

> A panda walks into a cafe. He orders a sandwich, eats it, and then draws a gun and fires two shots in the air.
>
> "Why?" asks the confused waiter, as the panda makes towards the exit. The panda produces a badly punctuated wildlife manual and tosses it over his shoulder.
>
> "I'm a panda," he says, at the door. "Look it up."
>
> The waiter turns to the relevant entry and, sure enough, finds an explanation.
>
> "**Panda**. Large black-and-white bear-like mammal, native to China. Eats, shoots and leaves."
>
> d'après *Eats, shoots and leaves*
> Lynne Truss

Intéressons-nous tout particulièrement au comportement étrange du panda : pourquoi donc tire-t-il ces coups de feu en l'air avant de partir ? Il s'explique en fournissant un livre à la ponctuation pour le moins bâclée. La définition du panda dans cet ouvrage tel qu'il est ponctué se traduit par : « grand mammifère noir et blanc ressemblant à un ours et originaire de Chine. Mange, tire et part. ». Les auteurs voulaient bien sûr dire : « Mange des pousses de bambou et des feuilles. » mais avec une telle ponctuation associée à la prononciation qui en découle, l'interprétation de leurs dires est toute autre.

Cet exemple amusant ramène à un phénomène primordial de la communication parlée : la prosodie. Il s'agit du domaine de l'intonation, du phrasé, de l'accentuation et du rythme. La prosodie est une composante qui est l'objet de nombreux travaux de recherche allant de la sémantique à la reconnaissance de la parole en passant par la synthèse de la parole. C'est un champ très ancien de la philologie et de la linguistique. Le but de cette section sera d'expliquer comment et à quel niveau la prosodie intervient dans la communication parlée.

B.1. Quelques aspects subjectifs de l'apport de la prosodie pour la communication

Pour mieux appréhender ce qu'est la prosodie et le rôle qu'elle peut jouer, commençons par les aspects qui paraissent les plus évidents à appréhender en tant que néophytes. Analysons l'extrait suivant :

> « ne va pas t'aviser au moins de suggérer une pareille idée à mon père » s'écria Octave du ton d'un homme effrayé.
>
> *Cinq cents millions de la Begum*
> Jules Verne

Dans cet extrait c'est le ton d'Octave qui permet apparemment de déceler chez lui la peur. La prosodie permet en effet, comme nous venons de le comprendre, de véhiculer les attitudes et émotions du locuteur. Ne vous est-il jamais arrivé de comprendre rien qu'à son intonation de voix qu'une personne était en colère par exemple ? Bolinger [1989] a ainsi mis en valeur le fait que dans une langue donnée, la prosodie peut véhiculer des attitudes telles que l'ironie, l'incrédulité ou encore le doute. L'expression des attitudes et émotions par la prosodie diffère d'ailleurs d'une langue à l'autre. L'utilisation d'un contour intonatif exprimant la colère dans sa langue maternelle peut parfois susciter une incompréhension totale si on le calque directement sur une seconde langue. Il s'agit là de ce qu'on qualifie communément de conventions intonatives. La prosodie permet ainsi de communiquer de nombreux éléments qui sont souvent qualifiés de paralinguistiques (Ladd [1996]). Elle permet en effet, outre l'état émotionnel du locuteur, de communiquer un certain nombre d'informations communicationnelles telles que l'agression, la condescendance, la solidarité ou même l'approbation ou la désapprobation. C'est ce qu'illustre l'extrait suivant :

> Puis elle ajouta un : « Monsieur je vous remercie » dont l'intonation équivalait à un congé.
>
> *Etudes de mœurs - 1er livre - Scènes de la vie privée - T. 1 - La paix du ménage*
> Honoré de Balzac

On pourra d'ailleurs souligner que l'on a conscience du fait que la prosodie véhicule ce type d'informations et qu'on s'attend à les trouver :

> Le flegmatique gentleman l'écoutait, en apparence au moins, avec la plus extrême froideur sans qu'une intonation, un geste décelât en lui la plus légère émotion.
>
> *Le tour du monde en 80 jours*
> Jules Verne

La prosodie permet également de manipuler les émotions d'autrui. Une simple intonation de la voix suffit parfois à rendre son (ou ses) interlocuteur(s) en colère ou triste(s) par exemple. Elle permet même parfois de déceler certains éléments sur le locuteur tel que sa situation sociale comme c'est le cas dans l'exemple ci-dessous.

> L'intonation de ses petites phrases coupées, le clignement de ses yeux achèvent de me révéler sa situation exacte dans la maison du capitaine Mauger.
>
> *Le journal d'une femme de chambre*
> Octave Mirbeau

B.2. Quelle place pour la prosodie dans la parole ?

Bien que les éléments décrits ci-dessus (état émotionnel du locuteur et informations communicationnelles) soient évidemment essentiels à la communication, la prosodie est ancrée de façon plus profonde dans la parole. Elle permet de découper, mettre en relation, organiser, structurer les unités de l'axe syntagmatique. Elle joue un rôle crucial dans la démarcation des unités, la hiérarchisation des énoncés, la topicalisation, la spécification du type d'acte de parole ou la prise de tour de parole.

Si ces notions vous paraissent quelque peu ambiguës, les quelques exemples ci-dessous, étayés à la fois par des travaux scientifiques et des exemples culturels, vous donneront un aperçu (non exhaustif) des rôles cruciaux joués par la prosodie.

La prosodie permet ainsi bien souvent de désambiguïser des énoncés sémantiquement ambigus. C'est d'ailleurs indirectement le propos du livre de Lynne Truss *Eats, shoots, and leaves* qui traite de l'importance de la ponctuation à l'écrit. La ponctuation est en effet à la communication écrite une partie de ce que la prosodie est à la communication orale. La ponctuation marque par exemple le rythme, les pauses mais aussi certaines inflexions intonatives telles que l'interrogation, l'affirmation ou encore l'exclamation. C'est ce qu'explique Lynne Truss : *« Punctuation directs you how to read, in the way musical notation directs a musician how to play »* (la ponctuation donne des indications sur la façon de lire tout comme la notation musicale dicte au musicien la façon de jouer), *« On the page, punctuation performs its grammatical function, but in the mind of the reader it does more than that. It tells the reader how to hum the tune. »* (Sur la page, la ponctuation joue son rôle grammatical mais dans l'esprit du lecteur, elle fait d'avantage. Elle indique au lecteur comment fredonner la mélodie). Nombre des illustrations cocasses et amusantes proposées par Lynne Truss peuvent relever de la prosodie si on les considère du point de vue de la communication orale. Dans l'exemple ci-dessous, la même phrase peut avoir plusieurs sens selon la ponctuation qui lui est attribuée et donc aussi selon la façon dont elle peut être lue, prononcée et « prosodiée ». A vous de voir quelle version vous préférez, mais attention, quelle que soit celle que vous choisirez prononcez-là correctement sinon, on pourrait se méprendre …

> A woman, without her man, is nothing.
>
> A woman: without her, man is nothing.[1]

La prosodie permet également de segmenter un énoncé en mots et en groupe de mots. Cette étape du traitement du flux acoustique est primordiale pour la compréhension. Il existe en effet nombre d'exemples pour lesquels il est possible de segmenter l'énoncé en mots de plusieurs façons

[1] *A woman, without her man, is nothing.* Une femme sans son homme, n'est rien.
A woman: without her, man is nothing. Une femme: sans elle, l'homme n'est rien.

et ainsi de mal interpréter ce que l'on nous dit. Marc Monnier nous le fait d'ailleurs remarquer de façon tres poétique grâce à ses vers holorimes :

Gall, amant de la reine, alla, tour magnanime,
Galamment de l'arène à la Tour Magne, à Nîmes.

Or la prosodie fournit des indices accentuels et rythmiques pour la segmentation. Les indices accentuels ont d'ailleurs été étudiés en détails par Pauline Welby [2003]. Elle a mis en évidence l'importance du contour intonatif au niveau de la frontière entre mot outil et mot de contenu dans la production et la perception de la segmentation pour le français.

La prosodie permet également d'identifier les intentions pragmatiques du locuteur. Elle permet notamment de savoir s'il nous pose une question ou s'il affirme simplement quelque chose. D'Imperio & House [1997] montrent ainsi qu'en fonction du décours temporel du contour intonatif d'un énoncé (courbe de fréquence fondamentale de même forme mais alignée de manière différente), les sujets perçoivent soit une affirmation soit une question. Ils mettent ainsi en avant l'importance des indices prosodiques pour cette distinction perceptive et notamment de leur alignement temporel.

Certains chercheurs ont tenté d'associer des contours intonatifs à divers éléments du discours ou à des intentions pragmatiques particulières. Delattre [1966], par exemple, a ainsi répertorié pour le français dix contours intonatifs types, censés véhiculer respectivement la continuation mineure, la continuation majeure, la finalité, l'interrogation, la question, l'écho, l'implication, la parenthèse, l'exclamation et le commandement.

Bien que, comme il vient d'être mis en avant, la prosodie joue un rôle privilégié dans la structuration morpho-syntaxique et sémantico-pragmatique des énoncés, nombreux sont ceux qui la considèrent comme un phénomène subordonné à la syntaxe et à la sémantique et sans structure propre. Pourtant, de plus en plus de chercheurs proposent que la prosodie possède une organisation phonologique propre. Ladd [1996] impute cet intérêt croissant pour une approche phonologique à deux avancées parallèles. La première est le développement de phonologies génératives non linéaires telles que la phonologie autosegmentale et la phonologie métrique (Liberman & Prince [1977]). Ces phonologies ont en effet offert un statut phonologique aux phénomènes suprasegmentaux. La seconde est l'expansion rapide des recherches en technologie de la parole, qui ont montré l'apport considérable de la prosodie pour le naturel de la synthèse vocale et l'efficacité de la reconnaissance automatique de la parole.

Beckman [1996] poursuit l'idée que la prosodie possède une structure phonologique propre et va plus loin en disant : *« prosody is not another word for "suprasegmentals" ; Rather it is a complex grammatical (phonological) structure that must be parsed in its own right »* (le mot prosodie n'est pas une autre appellation du domaine suprasegmental ; il s'agit plutôt d'une structure grammaticale (phonologique) qui doit être analysée pour elle-même). Elle suggère ainsi que la prosodie doit, elle aussi, être décomposée par l'interlocuteur pour percevoir la structure linguistique dans le flux de parole. Elle insiste sur le fait que, dans le cadre général de la phonologie générative non linéaire actuelle, la prosodie d'une énonciation est vue, non pas comme un ensemble de traits distinctifs, mais comme une structure organisée de façon hiérarchique (de la syllabe à l'énoncé, en passant par le mot et des unités telles que le syntagme accentuel, le syntagme intonatif, *etc.*). Il apparaît ainsi que la prosodie est primordiale dans la communication parlée et qu'elle doit être régie par des règles et un système qui lui sont propres.

Il est des cas où la prosodie supplée même aux mots, où c'est elle seule qui parvient à véhiculer le sens. Comme l'ont montré Dohalská & Mejvaldová [2000] pour le tchèque, dans une situation de communication professionnelle (gare de triage) en milieu bruité, quand les mots ne peuvent plus être identifiés, c'est la prosodie qui fournit l'information notamment sur l'état d'urgence de certaines annonces.

La prosodie permet aussi d'identifier une langue. Plusieurs études (Ohala & Gilbert [1981] et Maidment [1983]) ont permis de montrer que des sujets adultes non entraînés étaient capables de différencier plusieurs langues à partir de l'information prosodique seule (parole filtrée).

B.3. Prosodie et acquisition du langage

On notera la place toute particulière occupée par la prosodie dans l'acquisition de la langue par l'enfant. Les travaux de Fernald [1993] en prosodie articulatoire incitent d'ailleurs à prendre en compte l'intonation maternelle au niveau biologique dans une perspective adaptative de l'évolution. Des revues complètes de la littérature sur le développement et l'acquisition de la prosodie sont disponibles chez Vihman [1996] et Snow & Balog [2002]. On tentera simplement ici de souligner les principales conclusions qui ont pu être tirées d'un certain nombre d'études.

Il apparaît que les enfants ont très tôt une préférence pour la prosodie de leur langue maternelle par rapport à la prosodie d'une autre langue (Mehler *et al.* [1988]). Ils peuvent dès quatre jours distinguer leur langue maternelle d'une langue étrangère alors qu'ils ne parviennent pas à distinguer deux langues étrangères entre elles. Des tests complémentaires avec des signaux filtrés permettent aux auteurs de conclure que c'est grâce aux indices acoustiques que cette distinction est possible. On notera, en effet, que seule la prosodie leur parvient lorsqu'ils sont encore dans le ventre de leur mère. Dans les premiers instants de leur développement, c'est donc uniquement la « mélodie » de leur langue maternelle qu'ils perçoivent. Ramus [2002] fait un bilan sur le thème de la discrimination des langues chez les nouveau-nés et pose la question de savoir si cette discrimination fait intervenir des indices purement rythmiques ou intonatifs. Les expériences décrites suggèrent que ce sont les indices rythmiques qui jouent un rôle. Néanmoins, il apparaît aussi que, sans indices phonotactiques et avec les indices rythmiques seuls, la discrimination est plus difficile. Bien qu'il paraisse donc indéniable que les indices perçus par les nouveau-nés et leur permettant de discriminer des langues soient essentiellement prosodiques, la nature exacte des indices utilisés (rythmiques ou intonatifs) n'est pas encore clairement établie. Il est possible que la forte corrélation entre intonation et rythme joue un rôle important.

Lorsqu'ils parlent à de très jeunes enfants, les adultes accentuent leur prosodie pour diverses raisons (attirer l'attention, faire mieux ressortir certains contours mais aussi certains groupes syntaxiques importants … cf. Stern *et al.* [1982], Hirsh-Pasek *et al.* [1987], Fernald [1989]). Fernald [1991] décrit de façon précise les fonctions de la prosodie de la parole adressée aux enfants pendant leur première année de vie.

Boysson-Bardies [1996] propose que les enfants sont sensibles aux indices prosodiques dès deux mois. Elle cite plusieurs études sur le développement de l'utilisation de ces mêmes indices, qui montrent que, dès sept à onze mois, les enfants utilisent certains indices prosodiques spécifiques à leur langue que des enfants du même âge, mais d'une autre langue maternelle, n'utilisent pas (cf. Hallé *et al.* [1991] et Levitt & Wang [1991]).

Mandel *et al.* [1994] ont constaté que l'organisation prosodique de la parole en propositions aidait les bébés à mémoriser les propriétés phonétiques des mots dès l'âge de deux mois.

Par ailleurs, en ce qui concerne la production de la parole, l'étude de Boyssons-Bardies *et al.* [1984] constitue une autre preuve que la prosodie commence à être produite très tôt. Cette étude montre que les adultes français parviennent à différencier dans plus de 70% des cas, les babillages de bébés de huit mois français de ceux de bébés d'autres pays. C'est bien que ces bébés produisent, dès huit mois, des indices rythmiques et intonatifs non universels et spécifiques à la prosodie de leur langue maternelle, les séquences articulatoires produites lors du babillage étant similaires. Dès 1975, Halliday ([1975, 1979]) avait d'ailleurs proposé que des formes prosodiques étaient utilisées avant même l'apparition des premiers mots pour différencier certains proto-mots.

Les résultats de Galligan [1987] montrent que, bien que le développement de l'utilisation pragmatique de l'intonation dépende des individus et des interactions verbales auxquelles ils participent, le développement de l'utilisation grammaticale de l'intonation se fait à peu près au même moment chez les différents individus. Galligan note que l'utilisation accrue de l'intonation à des fins grammaticales coïncide à peu près avec la transition vers la syntaxe aux alentours de 14 à 15 mois.

Il est ainsi clair que la prosodie se développe tôt et joue un rôle dans l'acquisition du langage. Les enfants l'utilisent aussi pour avoir des retours sur leurs productions (cf. *e.g.* Galligan [1987]). Quand ils produisent leurs premiers mots par exemple, ils peuvent les accompagner d'un contour montant (*i.e.* interrogatif) pour avoir un retour sur la justesse de ce qu'ils ont produit pour désigner telle ou telle chose.

Une question principale concernant le développement de la prosodie a été l'objet de nombreuses études. Il s'agit de celle de la primauté de l'acquisition de l'intonation par rapport à celle des mots *i.e.* la prosodie est-elle acquise avant le lexique et la syntaxe ? Vihman [1996] fait le bilan de plusieurs études et conclut que l'utilisation linguistique de la fréquence fondamentale et de l'allongement de fin de syntagme est bien maîtrisée lorsque l'enfant entre dans la phase d'acquisition de la syntaxe. Snow & Balog [2002] font aussi une synthèse des études ayant été menées dans ce domaine. A l'issue de cette revue détaillée, les auteurs concluent que les enfants contrôlent certains aspects de l'intonation avant qu'ils ne produisent des combinaisons de deux mots mais pas avant qu'ils ne produisent leurs premiers mots. Cette conclusion confirme que la prosodie serait, au moins en partie, maîtrisée avant la syntaxe.

Tous les travaux évoqués ci-dessus, bien que parfois contradictoires sur certains détails, suggèrent néanmoins que la prosodie intervient très tôt dans le processus d'acquisition du langage et qu'elle y tient une place particulière. Il est difficile de savoir si elle précède le stade lexical mais il apparaît assez clairement qu'elle précède l'apparition des premières constructions syntaxiques.

B.4. Que pourrait apporter une meilleure connaissance de la prosodie ?

Les recherches sur la prosodie permettront de progresser dans plusieurs domaines. Il est en effet indéniable par exemple qu'une meilleure connaissance des phénomènes prosodiques et de leur fonctionnement permettra d'augmenter la qualité de la synthèse de la parole à partir du texte (cf. Romary & Pierrel [1989], Sorin [1991], Barbosa & Bailly [1994] et Rilliard & Aubergé [2001]).

Comme il a été exposé précédemment, la prosodie contribue, d'une part, à la segmentation et à l'organisation hiérarchique du message et permet, d'autre part, d'exprimer des attitudes et des émotions. Avec l'aide des résultats des recherches sur le contrôle des caractéristiques prosodiques d'une langue donnée, il sera possible d'élaborer des systèmes de synthèse adaptative de la parole plus naturels pour les auditeurs.

La synthèse de la parole n'est cependant pas le seul domaine qui pourrait bénéficier de travaux de recherche sur la prosodie. Les systèmes de reconnaissance exploitent en effet des modèles de production tels que celui de Deng *et al.* [1997] qui pourraient bénéficier, pour l'accès lexical par exemple, de connaissances sur les corrélats articulatoires de la structure prosodique, qui prédisent la variabilité acoustique pour une même séquence de phonèmes. Ces mêmes systèmes pourraient également obtenir des informations pour la segmentation de l'énoncé afin d'éviter des erreurs de « découpage » de l'énoncé en mots souvent commises par ces systèmes. La connaissance précise des corrélats prosodiques acoustiques permet aussi de recouvrer des informations précieuses telles que savoir si le locuteur pose une question ou s'il insiste sur un mot …

Il n'y a pas que la technologie qui bénéficiera d'avancées dans les recherches sur la prosodie. C'est en effet la prosodie d'une langue qui se révèle un des éléments les plus complexes à acquérir lorsque l'on apprend une langue étrangère. Elle n'est pourtant que très rarement et de façon assez rudimentaire prise en compte lors de l'enseignement des langues étrangères. Or la mauvaise utilisation de la prosodie d'une langue par le locuteur étranger est souvent la source de mauvaises interprétations par l'auditeur natif, ce qui nuit à l'efficacité de la communication (cf. *e.g.* Beckman [1996]).

La connaissance des gestes articulatoires qui sous-tendent les productions prosodiques des nouveau-nés ou des enfants permet de discerner des régularités rythmiques et spectrales qui, avec des connaissances sur l'influence du milieu linguistique, peuvent alimenter des théories sur l'ontogenèse de la parole.

La prosodie est donc une composante essentielle de la communication parlée. Elle permet de véhiculer de nombreuses informations souvent impossibles à transmettre d'une autre façon et joue un rôle important dans l'acquisition du langage. Une meilleure connaissance de son fonctionnement permettra de rendre la synthèse de la parole plus naturelle et la reconnaissance de la parole plus efficace. Elle permettra aussi peut être de compléter l'enseignement des langues étrangères.

Comment nommer les êtres et les choses, sinon par les gestes que nous faisons vers eux ou par les mouvements qui les conduisent à nous ?

Pages d'écriture
Jean Tardieu

C. Deixis, communication et prosodie

Souvenons-nous de deux des extraits de la page d'accueil, qui mettent en évidence un phénomène souvent utilisé en communication : le fait de montrer du doigt (bien souvent l'index) ce dont on est en train de parler :

> Mais le marin n'en était pas à cinquante pas qu'il s'arrêtait, poussait de nouveau un hourra formidable, et, tendant la main vers l'angle de la falaise :
> « Harbert ! Nab ! Voyez ! » s'écria-t-il.
>
> *L'île mystérieuse*
> Jules Verne

> « Vous aviez donc une lentille, monsieur ? demanda Harbert à Cyrius Smith.
> - Non, répondit celui-ci, mais j'en ai fait une. »
> Et il montra l'appareil qui lui avait servi de lentille.
>
> *L'île mystérieuse*
> Jules Verne

Ce geste de pointage du doigt permet de désigner « ce dont on parle », ce qui permet à notre (ou nos) interlocuteur(s) d'avoir une référence concernant « ce dont il est question ». Le procédé communicatif utilisé alors est ce qu'on appelle la deixis. Or la deixis occupe une place très importante dans la communication et notamment dans la communication parlée. On va voir qu'il existe un lien tout particulier entre deixis et prosodie, ce par la focalisation verbale. Cette section a donc pour but d'établir ces liens théoriquement afin de justifier l'évolution vers la problématique. Il n'est donc en aucun cas question d'effectuer une revue bibliographique complète du domaine de la deixis. De telles revues, partielles ou complètes, pourront être trouvées dans Ducey [2002] et Kita [2002]. Néanmoins quelques principes seront exposés ici, qui permettront par la suite de bien comprendre l'évolution logique vers la problématique de ce mémoire.

C.1. La deixis et son importance dans la communication parlée : pointage bracchio-manuel et acquisition du langage

Nous parlerons ici de deixis au sens littéral de désignation, monstration et pointage. Comme on vient de le mettre en évidence, la deixis est utile en communication en ce sens qu'elle permet bien souvent de désigner « ce dont il est question ». Mais les liens qui l'unissent à la communication sont

encore plus profonds et commencent à se manifester dès les premières phases de l'acquisition du langage puis au cours de son évolution (pour une revue de la littérature, le lecteur pourra se reporter à Bates & Dick [2002]). C'est en effet la deixis qui sous-tend la construction du lexique et l'émergence de la syntaxe. C'est de plus l'un des premiers dispositifs communicatifs acquis par les bébés.

Plusieurs « évènements », qui constituent des pierres angulaires pour l'acquisition du langage, ont été analysés et décrits dans de nombreux travaux récents : Volterra & Caselli [1986], Capirci *et al.* [1996], Goldin-Meadow [1999], Bates & Dick [2002], Butterworth [2003] et Goldin-Meadow & Butcher [2003]. Il a été observé que dès 8 mois, le bébé commence à porter alternativement son regard d'un objet qui l'intéresse vers les yeux d'un adulte et à produire ainsi un regard déictique afin d'attirer l'attention de l'adulte vers quelque chose de précis. Puis vers 9 à 11 mois, *i.e.* lors de la compréhension des premiers mots, le bébé produit ses premiers gestes communicatifs déictiques c'est-à-dire pointer vers un objet avec son index pour le désigner. Le geste de pointage permet la mise en place de l'attention conjointe entre l'enfant et l'adulte et semble jouer un rôle crucial dans la construction du lexique (Bruner [1983]). A l'âge moyen de 11 mois, on constate l'apparition du pointage canonique avec l'index comme chez les adultes. Au début du pointage canonique, les bébés pointent puis se tournent vers l'adulte comme s'ils vérifiaient que l'adulte comprenait. A 12 mois, 60% de tous les gestes produits par les enfants sont des gestes de pointage. A 16 mois, lors du pointage canonique, les bébés commencent par regarder l'adulte puis font un geste de pointage. Ce comportement témoigne du fait qu'ils ont compris que c'est important d'avoir l'attention de l'autre pour communiquer. Enfin vers 16 à 20 mois, le bébé construit des paires composées d'un geste déictique (comme pointer, montrer, requérir) et d'un mot, alors même qu'émerge la morphosyntaxe *i.e.* le passage du stade à un mot au stade à deux mots. Il semble ainsi qu'il existe un lien entre les productions de combinaisons geste-mot et mot-mot. Les énoncés à deux mots qui commencent à être produits pendant cette période sont d'ailleurs bien souvent déictiques. Le nombre de mots qui seront produits à 20 mois peut ainsi être prédit à partir du nombre de gestes et de combinaisons geste-mot produits à 16 mois (Morford & Goldin-Meadow [1992], Capirci *et al.* [1996], Goldin-Meadow & Butcher [2003]). Globalement, la façon dont les enfants utilisent le pointage laisse présager de leur développement langagier futur.

Il apparaît ainsi clairement que la deixis est à l'œuvre, non seulement dans la construction du lexique, mais également dans l'émergence de la morphosyntaxe.

C.2. Montrer en parole : la focalisation

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

La grammaire est une chanson douce
Erik Orsenna

Cette phrase d'Erik Orsenna est parfaite pour faire la transition entre geste physique de pointage (avec l'index) et parole. La fonction première d'un mot est en effet de désigner quelque chose *i.e.* de montrer cette chose grâce à un espace abstrait : l'espace lexical. Cette monstration est plus forte que le pointage bracchio-manuel puisqu'elle permet de désigner quelque chose qui n'est pas physiquement là (et qu'on ne peut donc pas montrer du doigt) ou même qui est inexistant physiquement comme un concept ou une idée (et qu'on ne pourra jamais montrer du doigt).

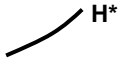
Chez l'adulte, notre définition de la deixis verbale, dans son acceptation étymologique, correspond à ce que les linguistes nomment focalisation. La focalisation consiste pour un locuteur à mettre en valeur la partie de l'énoncé qu'il veut faire passer comme étant la plus informative (cf. par exemple Halliday [1967], Gussenhoven [1983], Selkirk [1984], Nølke [1994], Birch & Clifton [1995] et Ladd [1996]). Il s'agit d'une mise en relief d'un constituant afin de le placer comme étant le plus informatif de l'énoncé. La focalisation permet de concentrer ou d'attirer l'attention de l'interlocuteur sur un constituant (Nølke [1994]). C'est un processus qui, comme nous allons le voir par la suite, est très souvent utilisé en communication parlée.

C.2.1. Différents types de focalisation

Il existe deux types de focalisation : la focalisation informative et la focalisation contrastive. La *focalisation informative* consiste en la mise en valeur d'un constituant apportant une information nouvelle. Dans la réponse à la question de l'exemple (1), le constituant *Sarah* est focalisé afin de signaler le fait que Sarah soit la personne qui ait mangé la pomme, est une information nouvelle. La courbe schématique de fréquence fondamentale correspondante est donnée en dessous du constituant focalisé. La *focalisation contrastive* consiste quant à elle à sélectionner le constituant focalisé dans la dimension paradigmatique (Selkirk [1984], Touati [1987], Pierrehumbert & Hirshberg [1990], Bartels & Kingston [1994], Dahan & Bernard [1996], Ladd [1996] et Di Cristo [2000]). Concrètement, elle permet de contraster une information par rapport à une autre, de même place dans l'énoncé. Dans l'exemple (2), le constituant *Sarah* est focalisé de façon contrastive afin de mettre en valeur le fait que ce soit Sarah et non Carol qui a accompli l'action de manger la pomme. Si le lecteur souhaite approfondir cette distinction entre focalisation informative et contrastive, il pourra se reporter par exemple à Touati [1987], Pierrehumbert & Hirshberg [1990], Bartels & Kingston [1994] ou Di Cristo [2000].

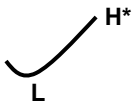
(1) *Who ate the apple ?* (Qui a mangé la pomme ?)

SARAH ate the apple. (SARAH a mangé la pomme.)



(2) *Did Carol eat the apple ?* (Est-ce que Carol a mangé la pomme?)

No, SARAH ate the apple. (Non, SARAH a mangé la pomme.)



On notera que bien que proches, ces deux types de focalisation sont légèrement différents. Une des principales différences qui les caractérisent est le contour intonatif associé, comme le montrent les courbes schématiques de fréquence fondamentale dans les exemples (1) et (2).

En français, comme pour beaucoup d'autres langues, la focalisation peut être réalisée de deux façons. La première consiste en un processus d'extraction syntaxique *i.e.* l'utilisation d'un présentatif « c'est X qui/que » comme dans la réponse B1 de l'exemple (3). La seconde façon de véhiculer la focalisation, est d'utiliser la prosodie en affectant un contour intonatif portant une focalisation sur l'élément à désigner (Touati [1989], Morel & Danon-Boileau [1998], Rossi [1999], Di Cristo [2000], Touratier [2000]) comme dans la réponse B2 de l'exemple (3). Dans ce cas, le constituant *Sarah* portera un contour intonatif montant, accompagné d'un allongement de la durée de ses syllabes.

(3) Carole a mangé la pomme ?

deixis syntaxique

B1 : Non, **c'est** Sarah **qui** a mangé la pomme.

deixis prosodique

B2 : Non, **SARAH** a mangé la pomme.

La focalisation peut donc être informative ou contrastive et s'exprimer soit par la syntaxe soit par la prosodie. Comme il sera rappelé dans la suite, c'est à la focalisation contrastive prosodique que nous nous intéresserons plus particulièrement dans ce mémoire.

C.2.2. Importance de la focalisation en parole

Comme nous venons de le voir, la focalisation est utile pour signaler à l'interlocuteur qu'une partie du discours est plus importante qu'une autre. Or ce type d'information est primordial pour la compréhension du message véhiculé par le signal de parole. Pour le comprendre, analysons le film *The Conversation* de Francis Ford Coppola dont un bref résumé a été donné en page d'accueil. Il apparaît que l'intrigue même du film tient au fait que Harry Caul, le personnage principal, a écouté un message dégradé et dont le contour prosodique avait été altéré. Après avoir modifié le signal afin d'en extraire le bruit et d'entendre ce que les locuteurs disent, il entend :

« He'd kill us if he'd got the chance. »

Il nous tuerait s'il en avait la possibilité.

alors que le locuteur avait en fait dit :

« He'd kill US if he'd got the chance. »

Il NOUS tuera s'il en a la possibilité.

Or, dans ce cas précis, la réplique prend un tout autre sens lorsque le mot *us* est focalisé prosodiquement. Lorsque ce mot n'est pas focalisé, on peut en effet penser que le locuteur est angoissé à l'idée que, si l'homme dont ils parlent en a l'opportunité, il n'hésitera pas à les tuer. Or si ce même mot est focalisé, on comprendra que le locuteur insinue que de toutes façons, s'ils ne le font pas les premiers, l'homme dont ils parlent les tuera. Or ces deux phrases ont un sens tout à fait différent, surtout dans la perspective psychologique de Harry Caul. Ce film permet ainsi d'apprécier le rôle prépondérant que peut jouer la focalisation prosodique dans la compréhension de la parole.

C.2.3. Place de la focalisation dans l'acquisition du langage

Les observations sur l'acquisition relativement précoce de la prosodie chez les enfants par rapport à la syntaxe (cf. ci-dessus), permettent de penser que la deixis prosodique serait maîtrisée avant la deixis syntaxique. Les bébés maîtrisent en effet les variations de fréquence fondamentale et de durée avant d'apprendre à réaliser des constructions syntaxiques. Plusieurs études sur l'acquisition de l'utilisation de constructions relatives (cf. *e.g.* Diessel & Tomasello [2000] et Tomasello [2003]) montrent en effet que celle-ci est progressive et ne commence qu'à plus d'un an alors que les bébés ont déjà acquis une certaine utilisation de la prosodie. Les premières constructions relatives utilisées sont de plus assez simples : une proposition principale et une relative avec un verbe intransitif. La construction syntaxique de la focalisation (« C'est X qui … »), aussi appelée relative présentationnelle, est semble-t-il acquise plus tard et de façon progressive. A notre connaissance, aucune donnée n'est disponible sur l'acquisition de la focalisation prosodique en particulier.

Cette capacité précoce à faire porter l'attention de l'allocutaire sur un élément précis semble bien être une des racines de la communication parlée. Elle apparaît comme étant petit à petit maîtrisée, probablement d'abord en pointant avec l'œil, puis le doigt, en désignant par l'intonation et enfin, de façon formalisée, par la syntaxe, mais semble bien être présente et jouer un rôle important à tous les stades de l'acquisition du langage.

La focalisation prosodique sera ainsi l'objet d'étude de ce mémoire car elle est l'une des formes que prend la deixis et est ainsi au cœur des problématiques sur l'acquisition du langage et les représentations linguistiques.

D. Venons-en à la problématique …

Les principales briques de construction ayant été posées, nous allons maintenant pouvoir les assembler et comprendre comment, ensemble, elles constituent le corps de la problématique qui sera abordée dans ce mémoire. Rappelons d'abord très brièvement ce que nous ont apporté ces différentes briques. D'une part, la parole est multimodale, non seulement entendue mais aussi vue et a depuis maintenant quelque temps été envisagée comme telle. La prosodie, élément crucial de la communication verbale, a quant à elle était envisagée principalement sous l'angle acoustique certainement parce qu'elle est souvent considérée uniquement comme étant la « mélodie » ou la « musique » de la parole. Pourtant, il y a des raisons de penser qu'elle est, tout comme la parole dont elle fait partie, elle aussi multimodale *i.e.* non seulement entendue mais aussi visible et vue. D'autre part, il a été souligné que la deixis et ainsi la désignation joue un rôle très important dans la communication. Or celle-ci peut être réalisée directement par la voix. Il est en effet possible de « montrer avec la voix » et ce en utilisant la focalisation linguistique.

L'objet sur lequel se concentrera notre étude sera la deixis ou focalisation prosodique ou encore la façon d'exprimer « la chose dont il est question » dans le discours. Soulignons que cette « chose » peut d'ailleurs être physique ou abstraite ou encore une référence spatiale ou temporelle. C'est cette notion de référence du discours qui est primordiale et qui marque le fort ancrage de la deixis dans la communication verbale : comment comprendre « ce que l'on nous dit » si on ne sait pas « ce dont il est question ». C'est sur la base de cet ancrage que va se construire la problématique.

Quelle que soit la langue considérée, la focalisation prosodique et sa perception, lorsqu'elles sont étudiées, sont bien souvent envisagées comme essentiellement auditives (cf. Dahan & Bernard [1996] pour le français, Baum *et al.* [1982], Bryan [1989], Gussenhoven [1983], Weintraub *et al.* [1981] pour l'anglais, D'Imperio [2001] pour l'italien, Krahmer & Swerts [à paraître] pour le néerlandais et l'italien, Brådvik *et al.* [1991] pour le suédois). Or, à la lumière de ce qui a été exposé précédemment, il semble primordial d'envisager cet objet sous un angle nouveau, celui de la multimodalité. La focalisation prosodique est avant tout *prosodique* et si la prosodie peut être visuelle, il en va de même pour la focalisation prosodique. Ensuite de façon plus théorique, il paraît contre nature de n'envisager la production et la perception de la focalisation que du point de vue auditif. Rappelons en effet que les références déictiques sont d'abord spatiales et se communiquent visuellement (le bébé pointe vers un objet). Puis, peu à peu, la deixis devient plus abstraite jusqu'à ce qu'on puisse la manipuler avec un outil qui n'a, *a priori*, plus aucun lien avec une représentation spatiale visuelle : la syntaxe. C'est donc d'abord sur la base visuelle et non auditive que la deixis se construit. Peu à peu elle se développe et, avec l'apparition du langage, elle se verbalise pour devenir plus puissante : il n'est en effet pas toujours possible de « montrer » avec l'index. Pourtant bien qu'elle soit essentiellement verbale à l'âge adulte, elle s'est développée sur une base visuelle : le pointage bracchio-manuel. Les mécanismes sous-jacents à sa production et à sa perception sont donc sûrement, au moins en partie, reliés à des mécanismes faisant intervenir la modalité visuelle.

Il y a ainsi toutes les raisons de penser que d'une manière générale la prosodie, et plus particulièrement la focalisation, doivent être considérées comme audiovisuelles et donc analysées comme telles. La focalisation prosodique doit ainsi certainement se rapporter doublement à la modalité visuelle et l'exploration de cette possibilité constituera le principal objectif de ce mémoire.

La focalisation prosodique est ainsi typiquement un de ces éléments du langage qui font certainement appel à des mécanismes qui transcendent les frontières de ce qui est propre au langage ainsi que l'évoquent par Bates & Dicks[2].

[2] Voir la citation de la page d'introduction.

– Chapitre Premier –

De la « visibilité » de la focalisation prosodique : des indices dans la littérature

Le devoir social serait le mot, mais l'instinct serait le geste.

Jean Tardieu

Ce mémoire va donc porter sur la « visibilité » de la deixis ou focalisation prosodique. Avant de détailler les expériences qui ont permis d'analyser de façon approfondie ce sujet, il convient de relever et de décrire les informations fournies par la littérature. Le premier point d'intérêt était ainsi de chercher des indices dans la littérature qui valideraient la pertinence d'une telle étude. Nous nous sommes ainsi demandés si certains chercheurs avaient déjà constaté que la focalisation prosodique était visible.

A. A-t-on déjà vu la focalisation prosodique ?

Bien que nous ne disposions que de très peu d'informations à ce sujet, certains résultats repérés dans différents domaines de recherche nous permettent de justifier du fait que la focalisation soit certainement visible. Plusieurs études montrent en effet, pour d'autres langues que le français, qu'il est possible de percevoir la focalisation prosodique à partir de la modalité visuelle seule. On remarquera que les études qui seront citées ci-dessous sont principalement extraites de la littérature dans le domaine de l'exploration des possibilités d'aides transmettant des informations prosodiques pour améliorer les performances en lecture labiale chez les sourds et les malentendants. C'est-à-dire que les études qui seront décrites ici s'insèrent, le plus souvent, dans le cadre d'études perceptives mettant en évidence l'apport de ces aides pour la perception des informations prosodiques dans le discours. Néanmoins, les tests perceptifs mis en œuvre, comportent souvent un aspect de perception en visuel seul et c'est cet aspect que nous décrirons et discuterons ici. On remarquera aussi que ces études sont, pour la plupart, assez anciennes. La prise en compte de la visibilité (ou de la non visibilité) des informations prosodiques n'est donc pas une problématique nouvelle. Bien que pendant un temps elle semble avoir été oubliée, elle a déjà en partie été étudiée. On retiendra de plus que c'est une problématique qui a été bien souvent abordée dans le cadre d'étude sur les aides possibles à la lecture labiale chez les sourds et les malentendants.

La première étude qui sera citée ici est très ancienne. Elle a été menée en 1934 par Dorothy Thompson (Thompson [1934]). Cette étude portait déjà sur l'apport d'aides tactiles pour la perception visuelle de l'*« emphasis »*[3] (emphase) en anglais. Elle a montré que pour des phrases comportant quatre ou cinq mots pouvant être sous emphase, les quatre sujets (bien entendants) parvenaient à identifier le lieu de l'emphase correctement en moyenne dans 63,7% des cas lorsque seule la modalité visuelle était disponible (*i.e.* sans aide tactile). Ces résultats sont nettement supérieurs au niveau de hasard qui est dans ce cas de 20 à 25%. L'écart-type mesuré est certes assez important, mais globalement, on pourra conclure que bien que cette transmission ne soit pas parfaite, les informations prosodiques de l'emphase peuvent être transmises visuellement. Suite à une étude approfondie des résultats, il est apparu que la détection était moins bonne lorsque les mots sous emphase se trouvaient en fin de phrase. L'emphase s'est révélée plus facile à détecter sur des mots comportant des voyelles « arrière » que sur des mots comportant des voyelles « avant ». Il est apparu que les consonnes dentales favorisaient également la détection visuelle de l'emphase. La « visibilité » partielle de l'emphase a donc été mise en évidence très tôt.

[3] Équivalent au terme *focalisation* selon la terminologie utilisée dans ce mémoire.

Arne Risberg et ses collègues ont plus tard mené quelques études sur le suédois qui vont ici attirer notre attention. Ils ont analysé (Risberg & Agelfors [1978]) les performances perceptives de sujets malentendants pour ce qu'ils nomment également *« emphasis »*[4] sous trois conditions (visuel seul, audio seul et audiovisuel). Les sujets, au nombre de 16, avaient pour tâche de dire quel mot ils avaient perçu comme étant sous emphase dans une phrase composée de trois mots. Les résultats montrent que les sujets sont parvenus à identifier correctement le lieu de l'emphase en visuel seul dans 50,6% des cas pour un niveau de hasard de 33%. Les auteurs qualifient cette performance de mauvaise. Ils la placent en effet dans une perspective de perception unimodale et veulent justifier de la nécessité d'aides pour la perception d'éléments prosodiques chez les malentendants. Une perception juste dans seulement 50,6% peut en effet être améliorée. Nous noterons néanmoins que l'identification rapportée du lieu de l'emphase en visuel seul est nettement supérieure au hasard. Cette observation suggère une fois de plus que des informations visuelles sont disponibles pour détecter la focalisation, même si celles-ci ne semblent pas suffire.

Dans une autre étude, Risberg & Lubker [1978] ont étudié les apports d'aides tactiles transmettant des informations prosodiques pour la perception de la prosodie. Dans le cadre de cette étude, ils ont analysé la perception visuelle de ce qu'ils nomment *« emphasis »* chez des enfants bien entendants. Deux phrases porteuses, dans lesquelles étaient insérés quatre mots clés pouvant être focalisés, ont été utilisées pour le test perceptif. Le pourcentage de réponses correctes d'identification du lieu de l'emphase en visuel seul était de 42,8% pour un niveau de hasard de 25%. Bien que l'écart-type correspondant soit assez important, il apparaît que les sujets peuvent percevoir la focalisation à partir de la modalité visuelle seule. Il est cependant difficile d'analyser ces résultats en détail, étant donné le fait qu'on ne dispose que de très peu d'informations sur les locuteurs, leurs productions et la tâche perceptive utilisée.

Bernstein *et al.* [1989] ont aussi mené une étude préliminaire qui nous intéressera ici au plus haut point et ce, dans le cadre d'une étude générale sur les apports d'aides tactiles pour la perception visuelle de la prosodie chez les sourds. Cette étude préliminaire consistait à demander à des personnes anglophones bien entendantes de localiser la focalisation à partir de la modalité visuelle seule, dans une phrase prononcée en anglais. Or elle a montré que le pourcentage de réponses correctes (76%) était nettement supérieur au niveau de hasard (33,3%).

Ces études montrent que, bien que la perception visuelle de la focalisation ne soit pas parfaite, elle est possible au moins pour l'anglais et le suédois. L'information de focalisation est donc aussi en partie visible.

B. Ce qui pourrait être « visible »

La focalisation prosodique pourrait ainsi être « vue ». Le deuxième point qu'il conviendra d'analyser est donc l'identification des informations liées à la focalisation contrastive qui pourraient être « visibles ». On notera que le thème des informations visuelles utilisées en perception de la parole a déjà été abordé en introduction. Il s'agit donc ici simplement d'identifier les informations visuelles qui pourraient être corrélées à la focalisation prosodique.

[4] Équivalent au terme *focalisation* dans la terminologie utilisée dans ce mémoire.

Dans le cadre de l'étude de la bimodalité de la parole, il a été mis en évidence que les événements « visibles » de la parole étaient avant tout les mouvements des lèvres, de la mâchoire et tout ce qui peut se passer dans la zone de la bouche. Ce qui est « visible » dans la parole c'est donc avant tout l'articulation, qui est par essence directement liée à la parole (cf. introduction). On notera cependant aussi l'existence d'un certain nombre d'études qui mettent en avant la perception visuelle d'indices temporels qui favoriseraient la perception acoustique. La perception n'est dans cette perspective pas directement visuelle mais les indices visuels permettent littéralement d'attirer l'oreille et donc de favoriser la perception auditive (cf. introduction). Il se pourrait donc que les indices « visibles » de la focalisation prosodique soient au moins en partie articulatoires et temporels.

Il a été argumenté en introduction que, bien que ce sujet soit controversé, certains articulateurs « internes » pourraient être « visibles » de l'extérieur. Or Lœvenbruck [1999] avait mis en évidence des corrélats articulatoires prosodiques au niveau de la langue et il se pourrait donc que la focalisation prosodique s'accompagne de mouvements visibles de la langue.

Il existe aussi bien sûr d'autres indices potentiellement visibles de la focalisation. On pourra penser à toutes les mimiques faciales tels les mouvements des sourcils, les grimaces … Il est également possible que les mouvements de la tête et même d'autres gestes tels que les mouvements de bras ou autres jouent un rôle. Il a déjà été montré en introduction que le rôle de la deixis bracchio-manuelle était crucial dans l'acquisition du langage et la communication parlée adulte.

Nous allons ainsi maintenant explorer la littérature, à la recherche d'informations sur les corrélats potentiellement visibles de la focalisation prosodique.

C. Une étude prometteuse

Keating *et al.* [2003] ont mené une étude combinée en production et en perception sur le *« lexical stress »* (accent de mot inexistant en français) et le *« phrasal stress »* (équivalent de *focalisation* dans la terminologie employée dans ce mémoire). Ils ont analysé les productions de trois locuteurs de l'anglais américain. Les intelligibilités segmentales des trois locuteurs avaient été jugées au préalable par cinq adultes sourds comme étant respectivement forte, moyenne et mauvaise. Nous nous attacherons tout particulièrement ici à décrire les mesures et résultats concernant le *« phrasal stress » i.e.* la focalisation. Six prénoms bisyllabiques commençant par une consonne labiale ou alvéolaire et comprenant quatre voyelles différentes (Mimi, Pammy, Bobby, Timmy, Debby, Tommy) ont été insérés dans une phrase porteuse dans trois positions différentes. Ces prénoms étaient tour à tour focalisés. Parfois aucun ne l'était et la phrase était donc neutre. L'enregistrement de trois locuteurs a été effectué avec le système d'analyse des gestes faciaux Qualysis intégré dans un dispositif d'enregistrement (cf. Bernstein *et al.* [2000] pour une description du dispositif). Des pastilles réflectives avaient au préalable été collées sur le visage des locuteurs de la façon indiquée par la figure I.1. Le système utilisé a permis d'extraire les coordonnées tridimensionnelles de ces pastilles. Aucune mesure de durée n'a été effectuée, les auteurs ayant observé lors d'une étude préliminaire que l'analyse des pics de vitesse fournissait des informations équivalentes.

Les données articulatoires (déplacement et pics de vitesse) des lèvres et du menton ont été relevées pour les gestes d'ouverture et de fermeture de la bouche. La distance entre lèvres supérieure et inférieure a également été analysée ainsi que les mouvements du sourcil gauche et ceux de la tête. On notera que les mesures n'ont été analysées que pour la voyelle de la syllabe

accentuée (au sens de l'accent de mot) *i.e.* toujours la première syllabe dans le cas présent. De façon générale, aucune différence articulatoire n'a été notée entre les cas focalisés et non focalisés pour les gestes de fermeture. Il est néanmoins apparu que toutes les mesures correspondant au geste d'ouverture étaient influencées de façon significative par la focalisation. La première voyelle du mot focalisé était ainsi accompagnée de *« larger and faster mouth opening movements, more open mouth positions, and head movements »* (des mouvements d'ouverture de la bouche plus grands et plus rapides, plus de positions « bouche ouverte », et plus de mouvements de la tête). Les auteurs ont également bien souvent relevé un mouvement de haussement du sourcil gauche lors de la focalisation : *« talkers raised an eyebrow on almost all stressed words »* (les locuteurs haussaient un sourcil sur presque tous les mots focalisés). Ils ont également observé que les différences articulatoires étaient les plus marquées chez le locuteur moyennement intelligible au niveau segmental. En ce qui concerne les mouvements de la tête, c'est chez le locuteur le mieux intelligible au niveau segmental, que les différences étaient les plus nettes.

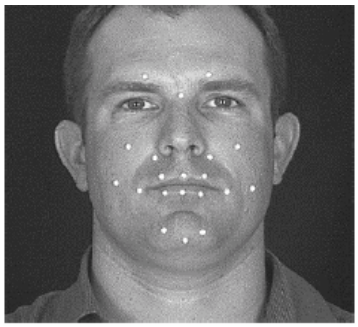


FIGURE I.1 – *Photo extraite de Keating et al. [2003] montrant l'emplacement des 20 pastilles réflectives collées sur le visage du locuteur.*

Un test de perception visuelle a ensuite été mené avec les données vidéo enregistrées simultanément à la capture des mouvements. La tâche était de déterminer quel prénom (trois possibilités dans une même phrase) avait été focalisé, les sujets pouvaient répondre *aucun*. Il s'est avéré que la perception de la focalisation en visuel seul était bien supérieure au niveau du hasard (54% de réponses correctes pour un niveau de hasard de 25%). Les performances perceptives correspondant au locuteur dont les productions étaient les plus marquées sur le plan articulatoire et à celui dont les productions étaient les plus marquées pour les mouvements de la tête étaient à peu près identiques. Ces résultats suggèrent que les deux types de mouvements auraient une importance perceptive équivalente. Les auteurs concluent pourtant que *« the mouth area is most important, but head movement can help make up for lack of information in the mouth area »* (la zone correspondant à la bouche est la plus importante, mais les mouvements de la tête peuvent permettre de compenser un manque d'information dans la zone de la bouche). Peut-être est-ce à la lumière d'autres analyses non décrites dans leur article qu'ils ont pu tirer de telles conclusions. Les données permettent également de conclure à la possibilité de percevoir des différences assez subtiles bien qu'il soit évidemment également possible que les mesures effectuées ne soient pas exhaustives.

Les auteurs soulignent enfin qu'il leur apparaît que *« It is better to vary several articulatory parameters than just one »* (il est préférable de faire varier plusieurs paramètres articulatoires plutôt qu'un seul), et qu'il existe une différence entre intelligibilité visuelle segmentale (reconnaissance de mots) et intelligibilité visuelle prosodique.

Cette étude préliminaire pertinente, effectuée simultanément à nos propres premières études[5], montre qu'il existe sûrement, au moins pour l'anglais, des corrélats articulatoires et faciaux à la focalisation et que ceux-ci permettent d'extraire l'information de focalisation à partir de la modalité visuelle seule. On pourra ainsi s'attendre à obtenir des résultats au moins en partie similaire pour le français.

On pourra néanmoins regretter le fait qu'aucun paradigme expérimental ne soit mis en place pour induire la focalisation chez les locuteurs : il leur est simplement demandé de focaliser. On peut donc craindre que ces productions soient peu naturelles voire exagérées. Il est de plus difficile de tirer des conclusions globales à partir de mesures effectuées sur la première voyelle de six prénoms bisyllabiques. On s'attendrait en effet intuitivement à ce que le mot soit affecté intégralement et non seulement sa première voyelle. Il pourrait également exister des différences en fonction du nombre de syllabes des mots, du nombre de mots du constituant focalisé (il peut en effet y en avoir plusieurs *e.g.* un adjectif et un nom) ou encore des voyelles étudiées (seulement quatre ici). Cette étude est donc loin d'être exhaustive quant aux possibilités de variabilité qu'offre la parole. Il est également dommage que les auteurs ne se soient intéressés qu'au constituant focalisé et non également à ce qui l'entoure. Certaines études qui seront décrites ci-après (*e.g.* Erickson [1998]) ont en effet montré que l'articulation des constituants voisins au constituant focalisé est elle aussi affectée, renforçant ou réduisant ainsi le contraste visible.

Cette étude offre donc une base très intéressante pour l'étude de la focalisation du point de vue visuel. Plusieurs aspects méritent cependant d'être approfondis. Notons de plus qu'elle concerne uniquement l'anglais et qu'il est regrettable qu'elle n'ait pas été approfondie.

D. Les effets de la focalisation prosodique sur l'articulation

Plusieurs études ont été menées dans diverses langues quant aux conséquences articulatoires de la focalisation et ce aussi bien pour les articulateurs potentiellement « visibles » tels que la mâchoire et les lèvres, que pour ceux qui sont plutôt non « visibles » tels que la langue. Bien que nous soyons conscients de l'existence de conséquences de l'accentuation sur les mouvements de la langue ou du larynx (cf. Kent & Netsell [1971], Ostry *et al.* [1983], Macchi [1985], de Jong *et al.* [1993], de Jong [1995], et Lœvenbruck [2000]), nous nous limiterons ici à la description des corrélats les plus directement « visibles » par les interlocuteurs, c'est-à-dire les mouvements des lèvres et de la mandibule. Il a en effet déjà été évoqué en introduction que la visibilité des mouvements des articulateurs « internes » était pour l'instant un sujet assez controversé. On notera que les études ayant été menées sur les relations entre l'articulation et la focalisation concernent principalement l'anglais.

[5] Les premières publications sur certains travaux décrits dans ce mémoire datent en effet également de 2003 : Dohen *et al.* [2003a, 2003b].

D.1. Quelques pistes

Stone [1981] a étudié les déplacements de la mâchoire et leur vitesse pour des voyelles insérées dans des phrases sans sens composées de huit à neuf syllabes CV *i.e.* consonne-voyelle (C : /d/ et V : /ə,ɑ,e/). Ces mesures ont été effectuées pour plusieurs niveaux de *« stress »* (le niveau de *« stress »* le plus important étant à peu près équivalent à ce qui est appelé focalisation dans ce mémoire). Elle a trouvé que la vitesse d'ouverture de la mâchoire était le meilleur corrélat pour signaler le niveau de *« stress »*.

La thèse de Macchi [1985] citée par de Jong *et al.* [1993] et de Jong [1995] indique qu'il existe, pour les voyelles [u], [i], [ɛ], un effet important de l'accentuation sur la hauteur de la mandibule.

Westbury & Fujimura [1989] ont étudié les productions articulatoires de huit locuteurs de l'anglais américain. Ils ont conclu que *« maximum displacements and peak velocities for the tongue blade and lower lip […] increase during a syllable under the influence of contrastive emphasis »* (les déplacements et pics de vitesse maximum correspondant à la langue et la lèvre inférieure augmentent lorsque la syllabe considérée est sous emphase contrastive).

Goffman & Smith [1996] ont analysé les mouvements de la lèvre inférieure pour des phrases avec ou sans *« contrastive stress »*[6] chez huit locuteurs anglophones. Ils ont pu conclure que : *« increased duration and amplitude occurred in stressed contexts »* (il se produit une augmentation de la durée et de l'amplitude en contexte focalisé).

D.2. Les travaux de Kelso et collègues

Kelso *et al.* [1985] ont mené une étude portant sur les mouvements de la mandibule et de la lèvre inférieure pour différents niveaux de *« stress »* (accentuation) et différents débits de parole. On notera que ce que les auteurs entendent par *« stress »* dans leur étude est loin d'être clair et il semblerait qu'il s'agisse d'accents lexicaux plutôt que de focalisation. Nous en détaillerons tout de même ici les résultats compte tenu de la relative « ressemblance » de ces phénomènes au niveau prosodique. Cette étude concernait les productions de deux locuteurs de langue maternelle anglaise. Les déplacements verticaux de la mâchoire et de la lèvre inférieure ont été mesurés à l'aide d'un système de suivi optoélectronique du type *Selspot* utilisant des diodes infrarouges placées sur les articulateurs. Les données articulatoires ont été échantillonnées à 200Hz et analysées après élimination des mouvements verticaux de la tête et filtrage.

Les locuteurs ont été enregistrés en train de produire des séquences extraites du *« rainbow passage »* en parole délexicalisée (*i.e.* toutes les syllabes étaient remplacées par les syllabes /ba/ ou /ma/ mais la structure prosodique de la phrase était maintenue). Les auteurs ont analysé les paramètres cinématiques (déplacements et vitesse) des paramètres articulatoires pour les gestes d'ouverture et de fermeture. Ils ont ainsi constaté que les gestes accentués présentaient des amplitudes et des durées plus importantes que les gestes non-accentués, l'effet étant plus fort pour les gestes d'ouverture que pour les gestes de fermeture. Ils ont aussi remarqué que la relation entre les mesures spatiales (amplitude des gestes) et temporelles (durées des gestes) n'était pas la même

[6] Équivalent au terme *focalisation contrastive* selon la terminologie utilisée dans ce mémoire.

pour les deux locuteurs. Pour le premier, la relation est en effet apparue comme étant linéaire mais pour le second les deux mesures étaient indépendantes. Les pics de vitesse mesurés étaient plus importants pour les gestes d'ouverture (par rapport aux gestes de fermeture) et pour les gestes accentués (par rapport aux gestes non-accentués).

Cette étude a été reprise et élargie à d'autres langues (français et japonais) dans Vatikiotis-Bateson & Kelso [1993]. Pour cette étude, les locuteurs (cinq pour l'anglais, quatre pour le français et cinq pour le japonais) furent enregistrés en parole délexicalisée comme pour l'étude de Kelso *et al.* [1985]. Pour le français, les auteurs ont analysé l'accent naturel (assez faible) de fins de mots : il ne s'agit donc pas d'une étude portant sur la focalisation. On notera de plus que les phrases de base pour la production de la parole délexicalisée en français étaient extraites des *Maximes* de LaRochefoucauld et ne correspondaient donc pas à une prosodie naturelle parlée. Les résultats sont ainsi à prendre avec précaution. Les auteurs ont observé qu'en français et en anglais, les gestes accentués correspondaient à des amplitudes, des durées et des pics de vitesse plus importants que les gestes non-accentués. En anglais, l'accentuation se reflétait à la fois sur le plan temporel et sur le plan spatial mais en français il semble qu'amplitude et durée étaient moins liées. Les auteurs notent que les gestes correspondant aux amplitudes les plus importantes correspondaient aussi aux durées les plus longues et aux pics de vitesse les plus élevés. Une corrélation forte a été observée entre pics de vitesse et amplitude, ainsi qu'entre durée du mouvement et amplitude mais de façon moins nette. Le lecteur pourra se référer à la section A.2.1.1.c du chapitre III pour une discussion sur l'intérêt et la signification de la mesure des pics de vitesse.

D.3. L'étude de Summers

Summers [1987] a étudié les corrélats articulatoires (mouvements de la mâchoire) et formantiques du *« contrastive stress »*[7] (accentuation contrastive). Les cibles étudiées, insérées dans une phrase porteuse, étaient des logatomes CVC *i.e.* consonne-voyelle-consonne (C_1VC_2 tel que V : /a/ ou /æ/ ; C1 : /b/ et C2 : /b/, /p/, /v/ ou /f/), la consonne finale pouvant être voisée ou non. La focalisation pouvait avoir lieu soit sur le CVC lui-même, soit sur le mot le précédant directement. Seuls les paramètres correspondant à la voyelle centrale ont été analysés. Les mouvements de la mandibule et de la lèvre inférieure de trois locuteurs de langue maternelle anglaise ont été mesurés et échantillonnés à 200Hz.

Le premier résultat confirme des études menées précédemment, c'est-à-dire que la durée de la voyelle augmente lorsque le CVC auquel elle appartient est focalisé. L'effet est encore plus important quand la consonne finale est voisée.

L'auteur a ensuite observé que le geste d'abaissement de la mâchoire, de la première bilabiale à la voyelle, était plus long et plus rapide lorsque la séquence était sous accentuation contrastive. Selon lui, cette augmentation de la vitesse et de la durée permettrait des positions articulatoires plus extrêmes (plus basses). Il apparaît aussi que la position stable après le geste d'abaissement et avant le début du geste d'élévation, est maintenue plus longtemps lorsqu'il y a accentuation contrastive.

Le geste d'élévation de la mandibule est plus rapide avec l'accentuation contrastive permettant d'atteindre des positions plus élevées. Il semble par contre n'y avoir aucun effet sur la durée du mouvement d'élévation de la mandibule.

[7] Équivalent au terme *focalisation contrastive* selon la terminologie utilisée dans ce mémoire.

Il apparaît aussi que si le CVC suit un mot focalisé, la durée de sa voyelle diminue ainsi que la vitesse, la durée et l'amplitude des mouvements articulatoires.

Suite à une analyse formantique détaillée, l'auteur propose que l'augmentation d'amplitude de la mandibule sous accentuation contrastive serait reliée à l'augmentation d'amplitude du premier formant. Il précise enfin que ses résultats concernant l'effet de l'accentuation contrastive sur les mouvements articulatoires vont dans le sens de ceux obtenus par d'autres chercheurs (Kozhevnikov & Chistovich [1965], Kent & Netsell [1971], Stone [1981] et Kelso *et al.* [1985]) *i.e.* lorsqu'il y a accentuation contrastive, les positions articulatoires sont plus extrêmes.

D.4. Les travaux de de Jong et collègues

De Jong [1995] (voir aussi de Jong *et al.* [1993]) a aussi étudié les conséquences articulatoires de la *« prominence »*[8] (proéminence). Son étude, qui concerne trois locuteurs de l'anglais américain, comprend à la fois des mesures des mouvements de la mandibule, de la langue et des lèvres. Les locuteurs prononçaient une phrase dans laquelle trois mots étaient successivement focalisés. On notera que l'auteur a utilisé une technique intéressante pour induire la proéminence. Celle-ci a permis d'obtenir des productions relativement naturelles : il ne demandait pas simplement aux locuteurs de produire la proéminence sur tel ou tel mot mais incitait le locuteur, grâce à un stimulus auditif, à effectuer une tâche de correction. Nous nous attacherons ici à décrire les résultats obtenus concernant les mouvements de la mandibule et des lèvres pour lesquels les mesures, obtenues par radiographie *« microbeam »*[9], ont été échantillonnées respectivement à 50Hz et à 100Hz. Les mouvements de la mandibule sont caractérisés par deux pastilles placées à la base d'une incisive et sur le haut d'une molaire. Trois pastilles placées sur les bords des lèvres supérieures et inférieures permettent de décrire les mouvements des lèvres. Deux pastilles placées sur le nez permettent de soustraire les mouvements de la tête.

Les observations permettent de penser que, quand il y a proéminence, l'amplitude et le pic de vitesse des mouvements augmentent simultanément. Aucun effet systématique de la durée n'est mesuré. Il apparaît ainsi que la proéminence correspond le plus souvent à des positions plus basses de la mandibule pour les voyelles et des positions plus hautes pour les consonnes antérieures. Il semble donc que la principale stratégie mise en place pour la proéminence soit d'atteindre une cible articulatoire plus extrême ou au moins de se rapprocher au mieux de la cible articulatoire. De Jong note, dans plusieurs cas, une augmentation de l'amplitude des mouvements sans allongement de la durée. Il en déduit que les augmentations d'amplitude des mouvements articulatoires liées à la proéminence ne peuvent pas simplement être imputées à une modification du contrôle temporel.

De Jong conclut que la proéminence correspond le plus souvent à une modification des cibles articulatoires. Bien que le contrôle temporel soit aussi parfois modifié, cette modification n'est pas systématique : les locuteurs ont le choix entre ces stratégies et peuvent bien sûr combiner les deux. Selon lui, la production de la proéminence met en œuvre des effets phonétiques correspondant à une hyper-articulation locale.

[8] Équivalent au terme *focalisation* selon la terminologie utilisée dans ce mémoire.
[9] http://www.medsch.wisc.edu/udeam/

D.5. Le travail de Harrington et collègues

Dans le cadre d'une étude portant sur la comparaison de modèles de coproduction articulatoire, Harrington *et al.* [1995] ont étudié les productions du mot /barb/ avec et sans *« accentuation »*[10] (accentuation). Le mot était inclus dans une phrase porteuse. Quatre locuteurs de l'anglais australien ont été enregistrés mais seules les données de trois d'entre eux ont pu être analysées. Les mouvements de la mandibule ont été mesurés à l'aide du système de mesure électromagnétométrique « *Movetrack »* permettant de suivre les mouvements horizontaux et verticaux des lèvres supérieure et inférieure et de la mandibule. Les données articulatoires ont été échantillonnées à 200Hz.

Les auteurs ont constaté que les durées des voyelles augmentaient de façon significative avec l'accentuation qu'il s'agisse d'un geste d'ouverture ou de fermeture. Chez deux locuteurs, il est apparu que l'amplitude des mouvements d'ouverture et de fermeture de la mâchoire était plus importante pour les voyelles accentuées que non accentuées. Pour le troisième locuteur, la différence d'amplitude due à l'accentuation n'était significative que pour le geste d'ouverture. En ce qui concerne la vitesse des mouvements mandibulaires, les auteurs ont noté, pour deux sujets, une augmentation du pic de vitesse en condition accentuée pour le geste d'ouverture, mais les différences observées n'étaient pas significatives pour le geste de fermeture. Pour le troisième sujet, le pic de vitesse augmentait également pour les voyelles accentuées en ouverture, mais pour le geste de fermeture ce même pic est étrangement apparu comme étant plus élevé en condition non-accentuée.

D.6. Les travaux de Erickson et collègues

Les travaux les plus exhaustifs et les plus détaillés sur l'influence de la focalisation prosodique sur l'articulation ont été effectués par Donna Erickson et ses collègues. Ils ont ainsi analysé l'influence de la *« contrastive emphasis »* (emphase contrastive) selon leur terminologie qui est l'équivalent du terme *focalisation contrastive* utilisé dans ce mémoire, sur l'ouverture de la mandibule en anglais américain (Erickson [1998, 2004], Erickson *et al.* [1994], Erickson & Fujimura [1996], Erickson & Honda [1996], Erickson *et al.* [1998] et Erickson *et al.* [2000]).

Leurs premiers travaux (Erickson *et al.* [1994], Erickson & Fujimura [1996], Erickson & Honda [1996]) suggèrent que l'emphase contrastive sur un chiffre (corpus constitué de chiffres focalisés ou non) est accompagnée d'une augmentation de l'amplitude de l'ouverture de la mandibule. L'étude de Erickson *et al.* [1994] suggère que c'est la portion VC (voyelle-consonne) de la syllabe CVC (consonne-voyelle-consonne) qui est la plus sensible à l'emphase contrastive pour les mouvements de la mandibule.

L'étude de Erickson [1998] est celle qui est la mieux détaillée. Cette étude porte sur l'anglais américain et permet d'avoir une idée assez précise des conséquences de l'emphase contrastive sur les mouvements de la mandibule. Trois locuteurs de l'anglais américain ont lu un texte pour lequel ils devaient focaliser les mots écrits en lettres majuscules. Ces mots correspondaient à des chiffres placés soit au début, soit au milieu, soit à la fin d'une séquence de trois chiffres incluse dans une phrase porteuse comme dans l'exemple : *« I work at 959 Pine Street »* (Je travaille au 959 Pine

[10] Équivalent au terme *focalisation* selon la terminologie utilisée dans ce mémoire.

Street.). Pendant cette lecture, des images obtenues par rayons X ont été enregistrées alors que des bobines en or avaient été placées sur les articulateurs du locuteur, ainsi que le montre la figure I.2. Après validation acoustique du corpus, l'ouverture de la mandibule a été calculée comme étant la coordonnée verticale correspondant au déplacement maximal (*i.e.* à la valeur maximale correspondant à l'ouverture maximale) à partir du plan occlusal.

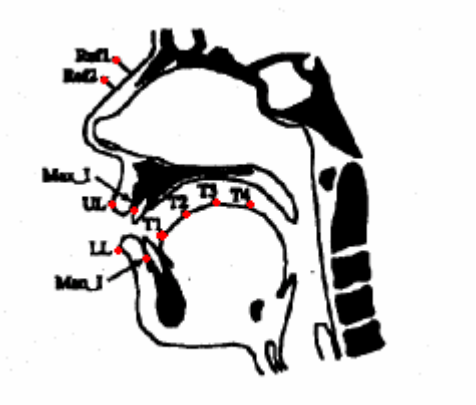


FIGURE I.2 – *D'après Erickson et al. [1998] : emplacement des bobines en or sur la langue, les lèvres, la mâchoire et deux points de référence.*

Les mesures effectuées ont montré que l'emphase contrastive correspond systématiquement à une ouverture significativement plus grande de la mandibule par rapport à une référence pour laquelle le même chiffre n'était pas sous emphase contrastive. Cette observation a été faite chez tous les locuteurs sauf un pour lequel les résultats n'étaient pas significatifs.

L'auteur a également mesuré un effet significatif de la position du chiffre sous emphase contrastive dans la séquence, sur l'amplitude du geste articulatoire. Il apparaît en effet que l'augmentation de l'ouverture de la mandibule par rapport au cas neutre est plus importante en position centrale, puis en position initiale et enfin en position finale.

Erickson a donc observé des effets de l'emphase contrastive sur l'articulation des mouvements de la mandibule du mot sous emphase contrastive. Néanmoins, elle a également remarqué qu'il y a des effets sur le mot qui suit directement le chiffre sous emphase contrastive et de façon plus générale sur toute la séquence qui suit ce même élément. En ce qui concerne le mot qui suit directement le mot sous emphase contrastive, l'auteur observe une réduction significative de l'amplitude des mouvements de la mandibule. Il apparaît que la tendance à la réduction articulatoire post-emphase dépend de l'emplacement du chiffre sous emphase contrastive dans la séquence de trois chiffres. Si l'emphase contrastive a lieu en position initiale, la réduction de l'amplitude des mouvements de la mandibule sur le mot qui suit le chiffre sous emphase contrastive est moins importante que si l'emphase a lieu en position médiane puis en position finale.

L'auteur note également un fait très intéressant : le locuteur qui n'augmentait pas l'amplitude des mouvements articulatoires correspondant au chiffre sous emphase contrastive est aussi celui qui la diminue le plus sur le mot qui suit (comparer les graphiques des figures 8 et 9 dans Erickson [1998]). Quel que soit le locuteur, il existe donc une différence significative (diminution) entre l'amplitude des mouvements de la mandibule correspondant au chiffre sous emphase contrastive et le mot qui le suit. Néanmoins, cette diminution est relative à l'augmentation mesurée pour le digit sous emphase contrastive et peut en réalité ne donner lieu à aucune diminution par rapport à l'articulation du même

mot dans la version neutre de l'énoncé. Il apparaît en effet que, selon les locuteurs, cette diminution forte peut être due soit à une augmentation forte sur le chiffre sous emphase contrastive suivie d'un retour à une articulation d'amplitude « normale » pour l'élément suivant, soit à une diminution forte de l'articulation sur le mot qui suit le chiffre sous emphase contrastive (par rapport au même mot dans la version neutre de la phrase), soit à une combinaison des deux phénomènes. En ce qui concerne la totalité de la séquence qui suit le mot sous emphase contrastive, l'auteur observe une tendance à la diminution globale des amplitudes des mouvements de la mandibule. Celle-ci commencerait sur le mot qui suit directement le chiffre sous emphase contrastive et se poursuivrait jusqu'à la fin de l'énoncé. L'auteur ne peut cependant pas conclure de façon certaine, les énoncés analysés dans cette étude étant trop courts.

La conclusion générale de cette étude est que, non seulement l'emphase contrastive a un effet localisé sur le mot contrasté (augmentation de l'amplitude des mouvements de la mandibule), comme ce à quoi on aurait pu s'attendre, mais aussi un effet global sur le mot suivant le chiffre sous emphase contrastive et de façon plus générale certainement sur toute la séquence qui suit ce même chiffre (réduction soudaine de l'amplitude des mouvements de la mandibule par rapport au chiffre sous emphase contrastive puis graduelle jusqu'à la fin de l'énoncé). Or une étude précédente (Erickson & Lehiste [1995]) sur les variations de durée liées à l'emphase contrastive en anglais, avait montré que la durée du mot sous emphase contrastive augmentait par rapport au cas neutre alors même que celle des autres mots de l'énoncé diminuait. Il apparaît ainsi que le patron d'évolution lié à l'emphase contrastive est le même pour la durée et pour les mouvements de la mandibule. Erickson propose ainsi que l'emphase pourrait être appliquée au sommet de la hiérarchie prosodique et affecterait donc l'énoncé dans son intégralité.

Les résultats d'une étude qui a suivi (Erickson *et al.* [1998]) ont permis de confirmer les observations décrites ci-dessus pour les mêmes chiffres et les mêmes phrases mais cette fois en situation de dialogue (tâche de correction) et non plus pour de la parole lue. La situation était donc plus naturelle et on peut penser qu'elle a donné lieu à des mises sous emphase également plus naturelles. Cette étude a montré pour quatre locuteurs de l'anglais américain, que l'augmentation de l'ouverture de la mandibule était un corrélat qui véhiculait de façon efficace l'emphase contrastive. Le croisement des données articulatoires mesurées avec des tests de perception auditive a en effet montré que plus l'augmentation de l'ouverture de la mandibule était importante plus la perception acoustique de l'emphase était bonne.

On pourra simplement regretter le fait que toutes ces études aient été effectuées pour le même jeu de phrases prononcées certes par des locuteurs différents mais dans lesquelles seuls deux mots différents (chiffres *five* et *nine i.e.* cinq et neuf) ont été analysés. Les analyses n'ont donc porté que sur deux mots différents aux propriétés syntaxiques identiques et variant simplement de position dans un groupe nominal décrivant un nombre à trois chiffres. Ces études n'ont donc pas couvert un grand domaine des possibilités de variabilité de l'anglais.

L'étude Erickson *et al.* [2000] remédie en partie à ce problème puisqu'elle confirme les résultats obtenus mais pour un nouveau corpus. Le but de cette étude était de comparer deux voyelles (avant ou arrière) sous emphase contrastive : /i/ et /æ/. Cette étude est d'autant plus intéressante qu'elle a été menée pour un grand nombre de locuteurs de l'anglais américain (45). Deux analyses ont été réalisées en parallèle : une acoustique et une articulatoire. Il en ressort que les pics de F0 et l'ouverture de la mandibule sont les seuls paramètres pour lesquels il y a une différence significative entre voyelles dans les cas avec et sans emphase contrastive. Il apparaît que l'emphase est toujours accompagnée d'une augmentation de l'amplitude de l'ouverture de la mandibule sur la voyelle quelle

que soit sa nature (avant ou arrière). Enfin, les auteurs mettent en avant deux types de stratégies utilisées par les locuteurs. Certains locuteurs utilisent à la fois F0 et l'ouverture de la mandibule pour signaler l'emphase contrastive et d'autres n'utilisent que F0. Les pics de F0 obtenus pour ces derniers sont plus importants que pour les locuteurs utilisant aussi l'articulation pour signaler la focalisation.

D.7. L'étude de Cho

Cho [2005] a mené une étude sur les réalisations articulatoires des voyelles /ɑ, i/ sous *« accentuation »* [11] (accentuation) chez six locuteurs de l'anglais américain. Des paires de syllabes CV comportant les voyelles citées et séparées par une frontière lexicale étaient insérées dans une phrase porteuse. Soit l'une de ces deux syllabes était accentuée, soit les deux, soit aucune. Les mesures des mouvements de la langue, de la mâchoire et des lèvres ont été effectuées par électromagnétométrie (EMA, Carstens). L'auteur montre que les effets articulatoires de l'accentuation et de la frontière prosodique sont différenciés : sous accentuation, l'ouverture des lèvres et de la mâchoire était plus grande alors qu'en fin de mot, seule l'ouverture des lèvres était plus importante. Ceci suggère une différenciation des deux phénomènes prosodiques au niveau articulatoire. Ceci implique en outre qu'il est important de prendre en compte le niveau dans la hiérarchie prosodique auquel la focalisation s'applique comme le suggérait également Erickson [1998].

D.8. Bilan : existe-t-il des indices articulatoires « visibles » à la focalisation prosodique ?

Les différentes études présentées ci-dessus qui, rappelons-le, concernent toutes la langue anglaise, permettent de conclure que, de façon générale, lorsqu'une syllabe ou un mot est focalisé (ces études ne portent jamais sur la focalisation d'un groupe de mots), les gestes articulatoires lui correspondant sont amplifiés. Il apparaît ainsi que la focalisation affecte non seulement les paramètres acoustiques (F0, intensité et durée) mais aussi les paramètres articulatoires. Il est en effet souvent observé que la focalisation est associée à une ouverture de la bouche et de la mandibule plus importante et plus rapide. Erickson [1998] note de plus une activité réduite de la mandibule sur les mots qui suivent l'élément focalisé.

L'augmentation en amplitude et en vitesse des mouvements des lèvres et de la mandibule, souvent nommée hyper-articulation, est associée par les auteurs aux phénomènes acoustiques sous-jacents. Jamais à notre connaissance, n'est-elle envisagée sous l'angle de la multimodalité. Or si la bouche est plus ouverte, on a de bonnes raisons de penser que cela va se voir … et donc que l'hyper-articulation pourra être un indice « visible » de la focalisation prosodique. C'est dans cette perspective que les corrélats articulatoires de la focalisation seront envisagés dans ce mémoire.

Comme il a été précisé, aucune des études présentées ci-dessus ne concerne le français. Elles permettent donc seulement d'établir des bases pour notre recherche. On notera aussi qu'il existe quelques petits inconvénients à ces études. De façon générale, seule une très petite portion du mot focalisé a été étudiée (bien souvent uniquement la voyelle accentuée). Ces voyelles étaient de plus

[11] Équivalent au terme *focalisation* selon la terminologie utilisée dans ce mémoire.

bien souvent intégrées dans des non-mots ou alors constamment dans le même mot pour toute l'étude. On pourra ainsi déplorer le manque de variabilité des cibles étudiées (phrases sans sens composées de huit à neuf syllabes parmi trois possibles dans l'étude de Stone [1981], huit voyelles insérées dans des CVC dans celle de Summers [1987], un seul mot étudié dans celles de de Jong [1995] et de Harrington *et al.* [1995] et trois mots différents seulement chez Erickson [1998]). Certaines des études présentées, comme celle de Kelso *et al.* [1985], concernent la parole délexicalisée et demanderaient d'être complétées par des études sur la parole réelle. Les études concernant la parole réelle portent souvent sur de la parole lue pour laquelle on a explicitement demandé au(x) locuteur(s) de focaliser un élément. Il ne s'agit donc pas de parole spontanée (sauf dans l'étude Erickson *et al.* [2000]). Enfin, il est important de remarquer que dans la plupart des études présentées ici (sauf Summers [1987] et Erickson [1998]) les auteurs se sont intéressés uniquement au constituant focalisé et ni d'une part à ce qui le précède ni d'autre part à ce qui le suit. Summers et Erickson ont pourtant tous deux observé une réduction de l'activité articulatoire après la focalisation.

Il sera donc non seulement intéressant d'étudier les corrélats articulatoires « visibles » de la focalisation prosodique pour le français, qui n'ont jamais été mis en évidence, mais aussi d'étudier de la parole plus spontanée et de tenter de couvrir un plus large spectre de la variabilité offerte en parole. On notera de plus l'importance d'étudier l'énoncé focalisé dans son intégralité et non uniquement le constituant focalisé.

E. D'autres indices visibles ?

Il apparaît donc qu'il existe certainement plusieurs corrélats articulatoires potentiellement visibles de la focalisation contrastive prosodique. Les gestes articulatoires ne sont néanmoins certainement pas les seuls indices potentiellement visibles. Il paraît en effet tout à fait intuitif que de nombreux gestes faciaux et même corporels accompagnent souvent le signal de parole. Quand on parle, on bouge tous les muscles de la face même ceux qui ne sont pas intrinsèquement liés à l'articulation comme les sourcils, les muscles des joues et aussi la tête. Ekman [1976] suggère que tous les gestes faciaux travaillent en collaboration avec le contenu linguistique afin de contribuer à véhiculer un sens ou un contenu. Slama-Cazacu [1976] propose que l'utilisation des gestes faciaux permettrait au locuteur de véhiculer des idées ou des concepts difficiles à exprimer uniquement avec les mots. Nous nous intéresserons ici plus particulièrement aux gestes faciaux (visage et tête) concomitants à des événements prosodiques afin de déterminer si certains travaux décrits dans la littérature ont pu mettre en évidence un lien entre ces gestes et la prosodie.

E.1. Des gestes et des voix : Explorons les possibles …

On se rappellera que Keating *et al.* [2003] (étude décrite à la section C du présent chapitre) ont observé que les mots focalisés sont souvent accompagnés d'un haussement de sourcil et d'un mouvement de la tête. Bien que les auteurs donnent très peu d'informations sur la nature exacte de ces mouvements, leur durée, leur amplitude et surtout leur rôle au niveau perceptif, ils précisent

néanmoins qu'ils existent. Ce résultat suggère qu'il pourrait y avoir un lien entre les mouvements des sourcils, de la tête et la focalisation.

D'autres études ont exploré les liens possibles entre les gestes faciaux de façon générale et les variations de fréquence fondamentale. Tel que le suggère Bull [2001] dans sa revue d'idées sur la communication non-verbale, les locuteurs produisent toujours des gestes lorsqu'ils parlent et ce bien souvent sans même en être conscients. D'après Guaïtella [1994], dès 1957, Heese [1957] suggérait, dans un cadre encore préliminaire, qu'il existait des liens entre gestes et intonation. Bolinger [1985] a quant à lui mis en évidence des relations entre geste et intonation complexes, non systématiques et nuancées. Les études concernant les gestes faciaux liés à la prosodie seront décrites ci-dessous, puis on s'intéressera plus spécifiquement à celles qui concernent les mouvements des sourcils et de la tête qui sont les plus fréquentes et les mieux documentées.

Guaïtella [1991, 1994] suggère que les variations intonatives seraient liées à plusieurs types de gestes (mouvements de la tête, des yeux, des sourcils, de la main droite et de la main gauche) produits en accord avec ces variations. Gestes et intonation seraient associés selon une « combinaison sémiotique » et non simplement comme « les éléments des diverses modalités pris indépendamment ».

Chovil [1991/1992] a étudié les *« facial displays »* dans la conversation chez 28 sujets de langue anglaise. Elle s'est intéressée à tous les gestes faciaux sauf les clignements des yeux, la déglutition, l'inhalation, le rire et les mouvements articulatoires liés à l'action de parler. Elle a aussi omis les sourires trop nombreux et donc difficiles à lier à un quelconque événement linguistique. Elle a réparti son analyse entre les gestes liés à des événements syntaxiques, sémantiques (redondants ou non redondants aux mots), discursifs (marques d'attention ou demandes de commentaires par exemple) ou encore ce qu'elle nomme *« adaptators »* (catégorie des messages non linguistiques tels que la communication de la nervosité par exemple). Nous nous limiterons ici à la partie de son étude qui analyse les gestes faciaux liés à la focalisation. Il apparaît ainsi que les gestes faciaux correspondant à ce que Chovil qualifie de *« syntactic displays »* étaient liés, dans 50% des cas, à une focalisation. Les gestes les plus fréquents dans ce cas étaient de loin les mouvements de sourcils (haussement ou abaissement) mais l'auteur note aussi des élargissements ou rétrécissement des yeux. On notera que l'auteur n'est pas très claire sur le type de focalisation qu'elle considère. Il apparaît en effet qu'elle mélange dans la même catégorie (*« emphasizers »*) la focalisation syntaxique et prosodique. Son analyse rejoint cependant les observations précédemment décrites que des mouvements de sourcils seraient potentiellement liés à la focalisation.

Boyer *et al.* [2001] ont étudié les liens entre voix, gestes et focalisation. Ils ont analysé pour des dialogues spontanés à la fois les variations de fréquence fondamentale et les gestes des locuteurs. Les gestes considérés étaient les mouvements des bras, des mains, le regard, et les mouvements de la tête et des sourcils. Les auteurs ont relevé « les points de départ des gestes ainsi que leur points culminants et leur localisation par rapport aux événements lexicaux ». Leur analyse montre que sur toutes les séquences de focalisation répertoriées, 58% correspondent à une association synchronisée du geste et de la F0 et 16% à cette même association mais désynchronisée. Les auteurs donnent des exemples de gestes observés avec la focalisation : mouvement de sourcil, point culminant d'un geste de pointage, clignement des yeux, changement d'orientation du regard. Lorsque geste et F0 sont synchronisés, les auteurs proposent « qu'une plus grande quantité d'indices (gestes des différentes parties du corps, F0) à un endroit précis du continuum va augmenter le « poids » informatif de la focalisation à l'endroit auquel elle se produit ». Lorsqu'un (des) geste(s) est (sont) produit(s) de façon désynchronisée par rapport à F0, le geste anticipe bien souvent les indices sonores. Or des travaux

antérieurs de Boyer ont montré que cette anticipation était mieux acceptée perceptivement par rapport au cas où les indices sonores précèdent le geste. Les auteurs proposent deux interprétations possibles. La première consiste à dire que « la synchronisation entre les divers éléments n'a pas besoin d'une grande précision, l'important est que la localisation des divers éléments se situe sur le lexème. ». Il s'agirait d'une manière d'éviter l'accumulation des indices. La seconde propose que « le décalage temporel entre la voix et le geste focalise l'une après l'autre les deux syllabes du lexème. ». Les auteurs concluent que la synchronisation et la désynchronisation pourraient correspondre à des stratégies communicatives différentes.

Argyle & Cook [1976] avaient déjà observé des effets de coordination du regard entre interlocuteurs vers un objet d'intérêt mutuel et des liens éventuels entre production de la parole et direction du regard. On notera aussi que Massaro [2002] déclare que certaines études non publiées ont montré que la focalisation prosodique était souvent accompagnée d'un élargissement oculaire.

E.1.1. Sourcils et F0

Dans cette section, nous tenterons de voir ce qu'il en est des études ayant été menées sur le lien entre mouvements de sourcils et F0 et ce que nous pouvons en tirer comme conclusions.

On notera qu'à ce jour, relativement peu d'études se sont penchées sur ce problème. D'après les informations fournies par Cavé *et al.* [1996], dès les années 70, Bridwhistel [1970] et Condon [1976] suggéraient que les mouvements rapides des sourcils pourraient jouer un rôle dans les processus intonatifs. Morgan [1953] ou Bolinger [1985] ont aussi suggéré que lorsque la fréquence fondamentale monte ou descend, les sourcils suivent une évolution dynamique du même type *i.e.* ils montent ou descendent. D'autres études plus récentes (*e.g.* Cosnier [1991] et Pentland & Darell [1994]) ont établi des liens entre certaines structures intonatives (*e.g.* l'interrogation) et les mouvements des sourcils. Bernstein *et al.* [1989] rapportent une observation non publiée de Ken Grant. Grant aurait en effet remarqué, dans le cadre d'une étude sur les possibilités offertes par les aides tactiles pour la perception d'informations prosodiques chez les sourds et les malentendants, qu'une des locutrices enregistrées avait tendance à hausser les sourcils lorsqu'elle produisait un contour final de F0 montant. Lorsqu'elle produisait un tel mouvement, les performances perceptives en visuel seul étaient très bonnes, alors que si elle ne produisait pas le geste, les performances ne dépassaient pas le niveau de hasard. Bien que très peu des études citées ci-dessus ne comportent de mesures précises pour plusieurs locuteurs des caractéristiques des mouvements des sourcils (amplitude, durée ou alignement temporel avec la parole), elles suggèrent toutes qu'il existerait un lien entre les variations de F0 et certains mouvements de sourcils (principalement des mouvements rapides de haussement).

En ce qui concerne le français, plusieurs études ont été effectuées (Cavé *et al.* [1993], Guaïtella *et al.* [1993] et Cavé *et al.* [1996]). Les premières études menées (Cavé *et al.* [1993] et Guaïtella *et al.* [1993]) ont montré que des mouvements de montée puis de descente des sourcils étaient associés à des courbes de F0 ayant la même forme. L'étude de Cavé *et al.* [1996] a permis d'effectuer des mesures plus précises de ces mouvements dans un cadre conversationnel (dialogues). Un système d'analyse automatique des mouvements composé de deux caméras (*Elite*) a été utilisé pour faire les mesures. Les mouvements des sourcils (durée de la montée et amplitude) de trois locuteurs ont été étudiés en correspondance avec les variations de fréquence fondamentale. Il est apparu qu'il existait une grande variabilité d'un locuteur à l'autre quant au nombre de mouvements produits et à leur amplitude. La durée des mouvements s'est avérée quant à elle être étonnamment très constante (376

ms en moyenne). Il est apparu que 38% des mouvements des sourcils correspondaient à du silence (*i.e.* le locuteur ne parlait pas quand il les a produits). Les auteurs constatent ainsi que *« The fundamental frequency pattern was not related to the duration of the eyebrow movements, nor to the magnitude of the movements of the right eyebrow, but was significantly linked to the movements of the left eyebrow. »* (le patron de fréquence fondamentale n'était pas relié à la durée des mouvements des sourcils, ni à l'amplitude des mouvements du sourcil droit, mais était significativement lié aux mouvements du sourcil gauche). Les auteurs proposent que des différences fonctionnelles entre hémisphères cérébraux lors des processus mis en œuvre pendant la communication, pourraient être à l'origine du déséquilibre entre les amplitudes des mouvements des deux sourcils. Le principal résultat mis en valeur par rapport à l'étude précédente (Cavé *et al.* [1993]) est l'existence de mouvements de haussement des sourcils alors que la F0 est plate ou seulement légèrement montante. Il n'existe donc pas de lien systématique entre la F0 et les mouvements des sourcils et les auteurs pensent que ces mouvements seraient plutôt une conséquence des choix linguistiques et communicationnels faits par le locuteur.

D'autres études montrent qu'il existe des liens non systématiques entre les mouvements des sourcils et l'intonation interrogative ou la prise de tour de parole (Purson *et al.* [1999]) ou les signaux de *« backchannel »* (marqueurs de la régulation du dialogue) (Bertrand *et al.* [1995]).

On retiendra donc qu'il peut exister des liens entre variations de F0 et mouvements de sourcils. Ces liens ne sont néanmoins pas systématiques, ils dépendent du locuteur mais aussi certainement de nombreux autres facteurs. Peu d'études ont pour l'instant été menées et il convient donc de rester prudent sur les conclusions tirées. Les liens potentiels entre variations de F0 et mouvements de sourcils paraissent donc être potentiellement intéressants mais leur analyse devra être approfondie pour pouvoir tirer des conclusions nettes.

E.1.2. Tête et F0

Plusieurs chercheurs se sont penchés sur le problème des corrélations possibles entre les mouvements de la tête, le flux de parole en général et la prosodie en particulier.

Hadar *et al.* [1983] ont mené une étude sur la corrélation entre les mouvements de la tête et le *« stress »*[12]. Cette étude s'est basée sur l'opposition entre, d'une part, ceux qui avaient trouvé un lien entre la F0 et des mouvements faciaux de diverses natures (Birdwhistell [1970], Kendon [1978] et Raffler Engel [1980]) et, d'autre part, ceux qui mettaient en doute la notion même de kinésie suprasegmentale (Dittman & Llewelyn [1969] et Dittman [1974]). Hadar *et al.* ont enregistré les mouvements de tête de quatre locuteurs à l'aide d'un goniomètre à lumière polarisée (*Crane Electronics*). Les locuteurs étaient en situation de communication avec l'expérimentateur pendant 5 à 10 minutes. Après analyse des mesures effectuées, les auteurs ont trouvé que le *« stress »* était relativement souvent accompagné d'un mouvement rapide de la tête. Aucune correspondance systématique n'a cependant pu être établie entre l'intensité du mouvement et son amplitude et le niveau de *« stress »* considéré. Les auteurs suggèrent avec une grande réserve que les mouvements de la tête pourraient être associés à des phénomènes prosodiques forts tels que le *« stress »* pour suppléer au fait que l'organe vocal ne puisse pas produire une *« energy »* (énergie) suffisante. Les mouvements de la tête seraient alors utilisés pour renforcer la signalisation acoustique insuffisante. Ils

[12] Équivalent au terme *focalisation* selon la terminologie utilisée dans ce mémoire.

demeurent tout de même très critiques sur cette hypothèse qui demanderait d'autres analyses. On la gardera néanmoins en mémoire.

Cerrato & Skhiri [2003] ont mesuré et analysé les mouvements de la tête chez quatre sujets suédois dans une situation de dialogue. Les mouvements de la tête les plus souvent mesurés étaient des hochements de têtes verticaux. Les auteurs concluent qu'il n'y a pas de correspondance spécifique entre un geste et une expression verbale ou une intonation. Il existe en effet une forte variabilité intra- et inter-locuteurs.

Munhall *et al.* [2004] ont trouvé que les six composantes du mouvement de la tête (3 rotations et 3 translations) expliquaient 63% de la variance de la fréquence fondamentale d'un locuteur japonais pendant son flux de parole. Un test perceptif avec une tête parlante leur a ensuite permis de montrer que l'intelligibilité audiovisuelle de la parole dans le bruit était plus grande quand les mouvements de la tête naturels (ceux de la tête parlante étaient calqués sur ceux de la tête du locuteur mesurés dans la première partie de leur étude) étaient présents.

Ces diverses études suggèrent l'existence d'un lien entre les variations de F0 et les mouvements de la tête. La nature exacte de ces liens et leurs caractérisations quantitatives sont encore très peu connues et méritent encore d'être approfondies. Il ressortira néanmoins des études décrites ci-dessus que le lien potentiel entre intonation et mouvements de la tête n'est pas systématique. Il apparaît en effet qu'il existe de fortes variations inter- et intra-locuteurs.

E.1.3. Tête, sourcils et prosodie: l'étude de Graf *et al.* [2002]

Graf *et al.* [2002] se sont intéressés à la prosodie visuelle et aux mouvements faciaux accompagnant la parole de façon générale. Les auteurs ont ainsi principalement étudié les indices potentiellement « visibles » qui ne sont pas directement liés à la parole c'est-à-dire non articulatoires. Leur analyse a porté sur les liens entre la structure prosodique d'un texte et les mouvements de la tête et autres gestes faciaux de locuteurs en train de lire ce texte.

Les auteurs ont analysé les enregistrements vidéo à l'aide d'un dispositif évaluant les positions de plusieurs points du visage de façon précise et sans avoir besoin de marqueurs. Ce système (cf. Graf *et al.* [2000] pour une description) permet d'accéder à l'information de posture de la tête, de mouvement des sourcils, de forme des yeux et de direction du regard (voir figure I.4). La figure I.3 donne les conventions de repère et d'orientation choisies par les auteurs. Plusieurs types de mouvements de tête ont ainsi été observés. D'abord, des mouvements de basse fréquence (de 0 à 2 Hz), c'est-à-dire sur plusieurs syllabes voire plusieurs mots. Les auteurs précisent que ce type de mouvements correspondait le plus souvent à un changement de posture et qu'ils n'avaient pas de lien apparent avec la parole. Les auteurs ont aussi noté la présence de mouvements de plus haute fréquence (de 2 à 15 Hz). Ces mouvements peuvent correspondre à plusieurs types d'événements dans le signal de parole. Certains correspondent ainsi à des mots accentués qui sont d'ailleurs souvent marqués par plusieurs hochements de tête. Les auteurs remarquent également qu'après une pause, il est assez fréquent que le locuteur baisse légèrement la tête puis la relève lorsqu'il reprend la parole. Il apparaît que les hochements de tête (rotations autour de l'axe nommé x, cf. figure I.3) sont de loin les mouvements de tête les plus fréquemment détectés. Il arrive aussi parfois que les auteurs observent un mouvement de gauche à droite (ou de droite à gauche) de la tête (rotations autour de l'axe nommé y, cf. figure I.3) comme quand on dit « non », mais jamais que le locuteur tourne la tête autour du troisième axe (nommé z, cf. figure I.3). Enfin les auteurs observent aussi beaucoup de

mouvements en diagonale c'est-à-dire des combinaisons de rotations autour des axes x et y. Les amplitudes des mouvements varient énormément mais il apparaît que l'alignement temporel avec F0 varie peu d'un mouvement à l'autre. Globalement, les auteurs concluent que les corrélations entre mouvements de la tête et F0 ne sont pas systématiques et dépendent beaucoup du locuteur mais elles existent. Les mouvements sont en effet bien « visibles » et devraient pouvoir être utiles en perception. Les auteurs observent aussi que *« rises of eyebrows are often placed at prosodic events, sometimes with head nods, at other times without »* (des haussements de sourcils sont souvent présents lors des événements prosodiques, parfois accompagnés d'un hochement de tête et parfois non). Les auteurs ne précisent malheureusement pas à quels « événements » ils font référence. Enfin, ils soulignent l'importance de valider leur étude par une étude qui concernerait de la parole non lue et donc plus naturelle.

FIGURE I.3 – *D'après Graf et al. [2002] : conventions de repère et d'orientation choisies par Graf et al..*

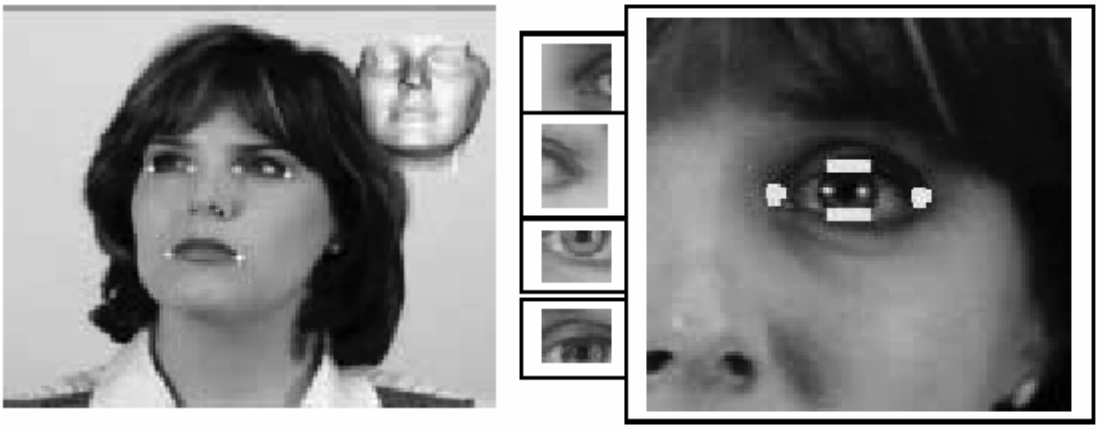


FIGURE I.4 – *D'après Graf et al. [2002] : (gauche) points du visage identifiés automatiquement ; la posture de la tête est calculée à partir des coins des yeux et des narines ; (droite) localisation de certaines parties de l'œil : les coins et les parties supérieures et inférieures.*

E.2. Sourcils, mouvements de la tête et perception audiovisuelle de la focalisation

Bien que peu de résultats soient pour l'instant disponibles à ce sujet, plusieurs chercheurs ont néanmoins tenté d'explorer ce domaine. Comme nous le verrons dans la suite, les tests perceptifs qui

ont été menés dans cette perspective utilisent souvent des systèmes de synthèse multimodale de type « tête parlante ». C'est en effet souvent dans l'optique d'améliorer le réalisme et l'intelligibilité de leur tête parlante que les chercheurs se penchent sur les problèmes de l'influence des mouvements de la tête et des sourcils sur la perception de la parole. Or comme nous venons de le voir, il se pourrait qu'il existe un lien entre les mouvements de la tête et des sourcils et les variations de F0 de façon générale et donc la focalisation prosodique en particulier. Etant donné que la focalisation prosodique est une composante importante de la communication et permet souvent d'en améliorer l'efficacité, il semble logique que les chercheurs tentent d'intégrer à leurs systèmes de synthèse multimodale tous les indices « visibles » possibles pour mieux la communiquer. On notera que pour l'instant aucune étude n'a été menée sur le français. Les principales études ont été menées par Emiel Krahmer et Marc Swerts de l'université de Tilburg aux Pays-Bas sur le néerlandais et par Björn Granström et David House du KTH à Stockholm en Suède sur le suédois. Quelques autres études isolées ont également été menées.

E.2.1. Les travaux de Krahmer et Swerts

Krahmer et collègues (Krahmer *et al.* [2002a, 2002b] et Krahmer & Swerts [à paraître]) ont étudié l'intervention des sourcils dans la perception de la *« prominence »*[13] (proéminence) en néerlandais à l'aide d'une tête parlante. La première étude (Krahmer *et al.* [2002a] reprise dans Krahmer & Swerts [à paraître]) avait pour but de déterminer les contributions perceptives respectives de F0 d'une part et des mouvements rapides des sourcils pouvant accompagner la proéminence d'autre part. Une tâche de perception de la proéminence a été mise en place utilisant le paradigme de reconstruction de l'historique d'un dialogue (*« reconstruct dialogue history »*, cf. Swerts *et al.* [2002]). La tête parlante prononçait l'expression *blauw vierkant* (carré bleu) et les sujets devaient en déduire ce que le précédent énoncé du dialogue aurait pu décrire (soit un carré rouge auquel cas *blauw* était focalisé soit un triangle bleu auquel cas *vierkant* était focalisé soit un triangle rouge auquel cas *blauw* et *vierkant* étaient focalisés). La proéminence était signalée soit par un accent (pic de F0), soit par un mouvement rapide des sourcils, soit les deux, sauf dans le cas où *blauw* et *vierkant* étaient tous deux focalisés. Dans ce cas, seul l'un des deux mots était accompagné d'un mouvement de sourcils. La figure I.5 donne l'exemple de deux images extraites d'une part d'une séquence pendant laquelle la tête parlante hausse les sourcils (gauche) et d'autre part d'une autre séquence sans aucun mouvement des sourcils (droite). On notera aussi l'utilisation de stimuli conflictuels c'est-à-dire pour lesquels F0 indique que l'un des deux mots est focalisé alors que le mouvement des sourcils est sur l'autre mot. Les auteurs distinguent aussi les cas où la voix de la tête parlante était synthétique et les cas où elle était naturelle.

Les résultats montrent que l'information acoustique (F0) aussi bien que l'information visuelle (mouvements des sourcils) ont un effet significatif sur la perception de la proéminence. Néanmoins, les effets diffèrent dans leur amplitude, l'effet acoustique (F0) étant nettement plus important que l'effet visuel (sourcils). Il apparaît que les indices visuels ont une contribution plus importante à la perception lorsque les indices acoustiques ne permettent pas de conclure de façon satisfaisante. Les sourcils n'ont par contre aucun effet lorsque les indices acoustiques sont clairs même dans les cas conflictuels. Enfin, il apparaît que les résultats sont les mêmes que la voix utilisée soit synthétique ou naturelle.

[13] Équivalent au terme *focalisation* selon la terminologie utilisée dans ce mémoire.

Ces résultats ont surpris les auteurs dans la mesure où, de façon globale, les jugements de proéminence sont moins bons que dans une étude précédente (Swerts *et al.* [2002]) pour laquelle seule l'information acoustique était disponible. Ils avancent qu'il serait possible qu'en ajoutant une autre modalité, la charge cognitive devienne trop importante et que la tâche perceptive soit rendue trop complexe.

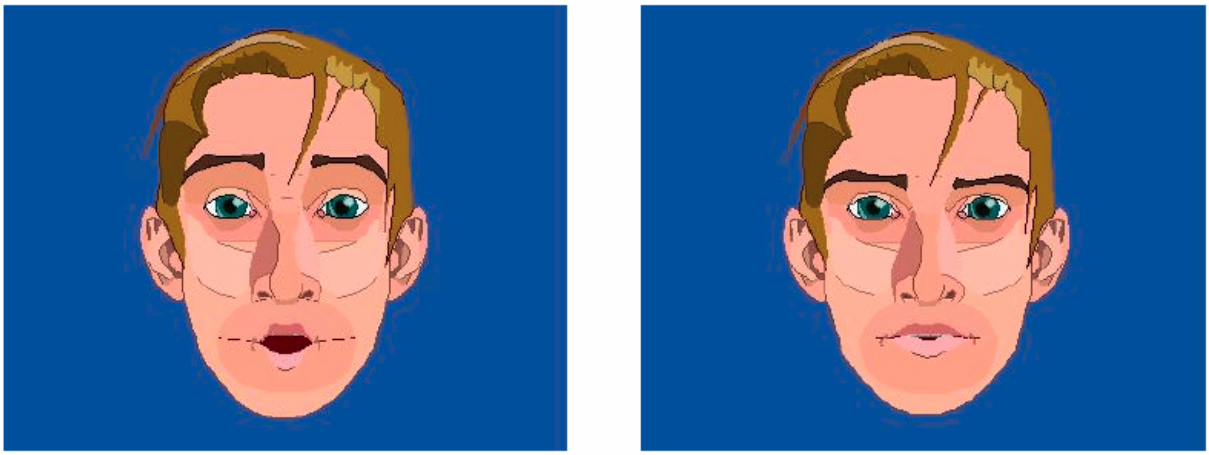


FIGURE I.5 – *D'après Krahmer et al. [2002a] : deux images de la tête parlante utilisée pour le test perceptif en train de prononcer « blauw vierkant » (carré bleu) avec les sourcils haussés (gauche) et pas de mouvements des sourcils (droite).*

Dans la lignée de ces observations, les auteurs ont donc mené une nouvelle étude (Krahmer *et al.* [2002b]) au cours de laquelle deux nouvelles expériences de perception ont été mises en place avec la même tête parlante et toujours pour le néerlandais. La première expérience avait pour but de déterminer si les sujets avaient une préférence pour un stimulus parmi deux. Pour l'un des deux stimuli, l'accent de F0 et le mouvement des sourcils étaient sur le même mot et pour l'autre sur deux mots différents. La tâche était ainsi de déterminer quel stimulus était le plus naturel dans le sens de la synchronisation entre le son et l'image. Il est apparu que les sujets préféraient nettement les stimuli pour lesquels l'accent de F0 et le mouvement des sourcils étaient sur le même mot. Et les auteurs de conclure *« rapid eyebrow movements may indeed serve the same purpose as pitch accents, i.e., to render prominence to a word »* (il se peut en effet que les mouvements rapides des sourcils aient la même utilité que les accents de fréquence fondamentale, *i.e.*, de souligner la proéminence d'un mot). La seconde expérience de cette étude traite la question de savoir si la perception de la focalisation est affectée ou non par la présence ou l'absence de mouvements des sourcils. Les sujets avaient ainsi pour tâche de dire dans quel cas parmi deux, le mot considéré (communiqué au préalable par les expérimentateurs) avait été focalisé. Les deux énoncés différaient de par le fait que le mot focalisé (c'est-à-dire auquel avait été affecté un accent de F0) était accompagné d'un mouvement des sourcils ou non. Cette expérience a montré que la présence d'un mouvement des sourcils a un effet clair sur la perception de la proéminence et ce de deux façons. D'abord, les mouvements des sourcils renforcent l'effet de l'accent de F0, les effets de F0 et des mouvements des sourcils apparaissent ainsi comme étant additifs. De plus, les mouvements de sourcils semblent réduire les perceptions de la focalisation sur des mots environnants *i.e.* réduire la fréquence des confusions. Ils renforceraient ainsi la signalisation de la proéminence.

A la lumière de ces nouveaux résultats, les auteurs s'étonnent encore plus d'avoir trouvé si peu d'effet des mouvements de sourcils dans leur première étude (Krahmer *et al.* [2002a]). Ils émettent

l'hypothèse que les mouvements des sourcils servent peut-être plutôt d'autres fonctions discursives que la proéminence et que la perception de la proéminence serait surtout basée sur l'acoustique. Ils évoquent ainsi la nécessité de faire des études de la production naturelle des mouvements des sourcils pendant le discours et de leur corrélation avec la proéminence.

Dans l'article Krahmer & Swerts [à paraître], les auteurs mettent en avant une observation intéressante. Ils ont en effet effectué des tests perceptifs auditifs conjoints de la proéminence prosodique en néerlandais et en italien. Ces tests leur ont permis de constater que la perception auditive de la proéminence était beaucoup moins bonne voire inexistante en italien alors qu'elle était très bonne en néerlandais. Les auteurs émettent ainsi l'hypothèse que les indices visuels, et en particulier les mouvements des sourcils, auraient un effet beaucoup plus grand sur la perception audiovisuelle de la proéminence en italien par rapport au néerlandais. Ce test perceptif comparatif n'a apparemment pas encore été mené mais il sera très intéressant d'en connaître les résultats.

Plusieurs critiques peuvent être formulées à ces études (Krahmer *et al.* [2002a, 2002b], Krahmer & Swerts [à paraître]). La principale critique, que formulent d'ailleurs d'eux-mêmes les auteurs à la toute fin de la seconde étude, concerne le manque d'informations préalables sur ce qui se passe réellement au niveau des sourcils lors de la focalisation dans la communication parlée naturelle. Les auteurs se basent sur plusieurs études, dont une seule porte réellement sur des données expérimentales (Cavé *et al.* [1996]) lesquelles concernent de plus le français et non le néerlandais. Aucune information n'est donc disponible sur l'organisation temporelle précise des mouvements des sourcils et de l'accentuation (F0) et encore moins sur l'amplitude de tels mouvements. On notera d'ailleurs que cette amplitude n'est évoquée dans aucun des articles. On n'a donc aucune information sur la façon dont elle a été choisie. Les tests perceptifs ont de plus été menés avec une tête parlante ce qui crée un environnement de communication très peu naturel. Il se peut en effet que la perception de la parole soit au moins en partie différente lorsque les sujets ont affaire à une tête parlante ou à un humain. On pourrait d'ailleurs répondre à l'étonnement des auteurs devant leurs résultats pour la première étude en comparaison aux résultats en audio seul, que les différences viennent peut-être tout simplement du fait qu'ils utilisent une tête parlante. C'est peut-être cela qui gêne quelque peu les sujets dans leur perception et non les mouvements des sourcils eux-mêmes. Enfin, lors de la première expérience décrite dans l'étude de Krahmer *et al.* [2002b] il aurait été très intéressant d'inclure une condition audiovisuelle pour laquelle l'information n'aurait été qu'acoustique en plus des conditions pour lesquelles les mouvements de sourcil et l'accentuation étaient synchronisés ou non. Les sujets auraient peut-être en effet jugé que cette condition était la plus naturelle.

E.2.2. Les travaux du KTH (Stockholm, Suède)

Plusieurs études sur l'intervention des mouvements des sourcils et des mouvements de la tête dans la perception audiovisuelle de la *« prominence »*[14] (proéminence) ont été menées au KTH de Stockholm par Björn Granström, David House et collègues (Granström *et al.* [1999], House *et al.* [2001a, 2001b] et Granström & House [2004, 2005]). Ces études s'inscrivent dans le cadre d'un projet de grande envergure dont le but est d'améliorer la synthèse audiovisuelle de la parole afin de la rendre plus naturelle et surtout plus efficace. C'est ainsi que Björn Granström, David House et collègues se sont intéressés aux gestes faciaux non articulatoires potentiellement liés à la parole et surtout à la prosodie et à l'expressivité *i.e.* les mouvements de la tête, la forme des sourcils, les mouvements des sourcils

[14] Équivalent au terme *focalisation* selon la terminologie utilisée dans ce mémoire.

(haussement), les mouvements et clignements des yeux. Les articles de Granström & House [2004, 2005] constituent des synthèses des études précédentes.

Nous nous attacherons ici à décrire uniquement les parties de ces travaux qui concernent directement la proéminence. Dans cette perspective, les études menées au KTH visaient essentiellement à répondre à trois questions majeures : (1) dans quelle mesure les mouvements de sourcils peuvent-ils être des indices pour la perception de la proéminence ?; (2) les mouvements de la tête constituent-ils un indice plus fort pour la perception de la proéminence que les mouvements des sourcils ?; (3) quelle est la sensibilité perceptive à l'alignement temporel des mouvements de la tête et des sourcils avec la syllabe accentuée ? Afin de répondre à ces questions, plusieurs tests perceptifs ont été mis en place pour le suédois en utilisant le système de synthèse multimodale du KTH (cf. Beskow [1997]).

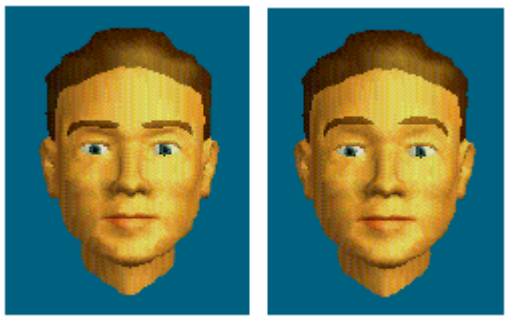


FIGURE I.6 – *D'après Granström et al. [1999] : images de la tête parlante (Alf) utilisée pour les tests perceptifs décrits dans Granström et al. [1999] sans mouvement des sourcils (à droite) et en train de hausser les sourcils (à gauche).*

Dans l'étude de Granström *et al.* [1999] (reprise ensuite dans Granström & House [2004, 2005]), les auteurs ont mené un test perceptif pour lequel, à paramètres acoustiques et articulatoires identiques, étaient associés divers patrons de mouvements des sourcils (haussement). Ces mouvements ont été ajoutés manuellement à la synthèse automatique. Peu d'indications sont fournies sur l'amplitude des mouvements synthétisés ou sur la façon dont les auteurs l'ont choisie. La seule information fournie est que *« The degree of eyebrow movement was chosen to create a subtle movement that was distinctive although not too obvious. »* (le degré de mouvement des sourcils a été choisi de façon à créer un mouvement subtil distinctif mais pas trop évident) et que la durée des mouvements était de 500ms. La figure I.6 illustre un cas de mouvements des sourcils par rapport à un cas sans mouvement de sourcils. Le corpus utilisé pour le test était constitué d'une seule phrase qui comportait cinq mots de contenu lesquels étaient associés tour à tour à un mouvement de sourcils. Le signal acoustique ne comportait quant à lui aucun indice pouvant signaler la proéminence. Le test a été passé par 21 sujets, dont six pour qui le suédois n'était pas la langue maternelle bien qu'ils le parlent très bien. La tâche des sujets était de détecter le mot qui leur avait paru le plus focalisé dans la phrase (*« most stressed/most prominent »*). Il est ressorti de ce test que le haussement de sourcil est un bon indice pour signaler la proéminence sur un mot dans une phrase. Il apparaît que c'est un indice indépendant des indices acoustiques et articulatoires. Les auteurs observent de plus que les sujets dont le suédois n'était pas la langue maternelle étaient encore plus sensibles que les autres aux mouvements des sourcils pour la détection de la proéminence. Ils proposent ainsi que les mouvements de sourcils pourraient être des indices plus universels pour signaler la proéminence. Les indices acoustiques signalant la focalisation varient en effet beaucoup d'une langue à l'autre en quantité et en nature. On pourra regretter que cette étude ne comporte pas d'analyse comparative

avec les indices acoustiques mais les auteurs précisent qu'ils prévoient de le faire dans une prochaine étude.

Dans la lignée de ces résultats, une autre étude a été menée (House *et al.* [2001a, 2001b] reprise ensuite dans Granström & House [2004, 2005]). Elle a combiné l'étude de l'influence des mouvements des sourcils (*« raising » :* haussement) et hochement de la tête (*« nodding »*) pour la perception de la proéminence. Une même phrase a été utilisée pendant tout le test. Deux mots de contenu de cette phrase, séparés par un mot de fonction, étaient marqués d'un accent focal (*« focal accent »*) en acoustique. Des mouvements de tête et de sourcils ont ensuite été ajoutés de diverses façons. La figure I.7 donne des exemples d'images extraites de séquences pour lesquelles la tête parlante ne bouge ni les sourcils ni la tête (gauche) et pour lesquelles la tête parlante bouge les deux (droite). Encore une fois, peu d'indications sont données sur l'amplitude des mouvements des sourcils et de la tête. On sait cependant que les mouvements de la tête sont un *« slight vertical lowering »* (léger abaissement vertical) dont l'amplitude ne dépasse pas 3% de la dynamique totale possible. Les mouvements des sourcils sont apparemment d'amplitude comparable à ceux qui avaient été synthétisés pour l'étude de Granström *et al.* [1999] c'est-à-dire qu'ils sont assez subtils (pas plus de 4% par rapport à la dynamique totale possible). On voit, sur la figure I.7, que les mouvements sont assez subtils et paraissent difficiles à percevoir en statique mais les auteurs précisent qu'en dynamique ces mouvements sont nettement visibles. Les durées des mouvements de la tête et des sourcils sont de 300ms. Deux jeux de stimuli ont été confectionnés. Pour le premier, les mouvements de la tête et des sourcils étaient synchronisés et six alignements différents ont été testés allant d'un alignement parfait avec la voyelle accentuée du premier mot à un alignement parfait avec la voyelle accentuée du second mot accentué. Pour le deuxième jeu de stimuli, les mouvements de la tête et des sourcils n'étaient plus synchronisés. Dans trois cas, les mouvements de la tête étaient systématiquement alignés avec la voyelle accentuée du second mot accentué et les mouvements des sourcils variaient de position, du début de la voyelle accentuée du premier mot accentué vers le second mot accentué. Pour les trois autres cas, les mouvements des sourcils étaient systématiquement alignés avec la voyelle du second mot accentué et les mouvements de la tête variaient de position, du début de la voyelle du premier mot accentué vers le second mot accentué. Un total de 33 sujets a été testé, leur tâche étant de dire quel était le mot (parmi les deux mots considérés) qui était le plus accentué (*« most prominently accented »*). Pour le premier jeu de stimuli, il a été constaté de façon nette que l'alignement des mouvements avec l'audio influençait la perception. Pour le second jeu, les résultats étaient moins clairs. Les auteurs ont conclu que les mouvements des sourcils aussi bien que les mouvements de la tête étaient des indices puissants pour signaler la proéminence lorsqu'ils étaient alignés avec la voyelle accentuée d'un mot potentiellement *« prominent »* (*i.e.* accentué acoustiquement). Il apparaît que la sensibilité temporelle à l'alignement avec l'audio est d'environ 100ms. Cependant, quand les mouvements ne sont pas parfaitement alignés avec la voyelle accentuée du mot potentiellement *« prominent »*, les sujets ont tendance à les intégrer au mot potentiellement *« prominent »* le plus proche. Cette étude ne permet pas de discriminer les mouvements de la tête des mouvements des sourcils en terme de poids d'influence. Les auteurs observent tout de même un léger avantage pour les mouvements de la tête. Enfin, les auteurs concluent que *« synchronization with the stressed syllable is important, but perhaps not absolutely critical as a large degree of visual integration seems to occur within 100ms of synchronization with the syllable. »* (la synchronisation avec la syllabe accentuée est importante, mais peut-être pas critique puisqu'une large part de l'intégration visuelle semble avoir lieu avec une synchronisation avec la syllabe ne dépassant pas les 100 ms). On pourra souligner le fait que pendant le test perceptif, les sujets pouvaient visionner les stimuli autant de fois qu'ils le désiraient avant de

répondre. Il se peut ainsi que la perception ne soit plus tout à fait naturelle. Dans le doute, les sujets auront pu visionner plusieurs fois et détecter des indices non détectés spontanément et donc non perçus « instinctivement ».

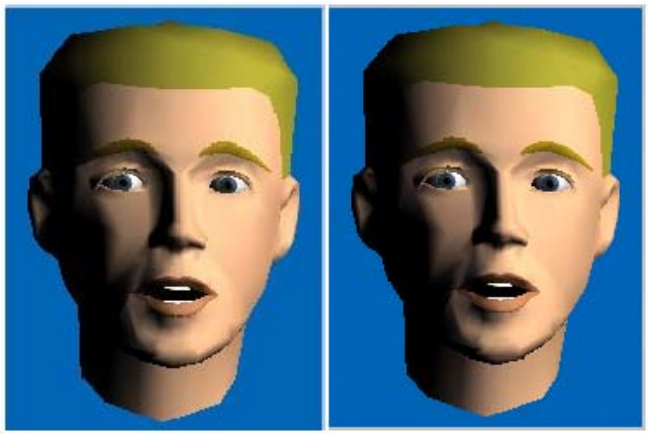


FIGURE I.7 – *D'après House et al. [2001a] : images de la tête parlante utilisée pour les tests perceptifs décrits dans House et al. [2001a] sans mouvement des sourcils (gauche) et avec haussement des sourcils et abaissement de la tête (droite).*

Ces études permettent donc de penser que les mouvements des sourcils et de la tête peuvent être de bons indices pour la perception de la focalisation. Il apparaît aussi que la synchronisation temporelle de ces mouvements avec le signal acoustique peut jouer un rôle. Néanmoins, ces études ne permettent pas de mieux comprendre comment ces mouvements pourraient être contrôlés en liaison avec le signal de parole. Il semble donc difficile pour le moment de les intégrer à un système de synthèse. Rappelons en effet que pour les tests décrits ci-dessus, les mouvements ont été programmés à la main.

E.2.3. D'autres pistes

Massaro [2002] rapporte une étude menée sur l'anglais en collaboration avec Jonas Beskow du KTH à Stockholm sur l'influence de F0, de l'intensité, de l'élargissement oculaire et des mouvements des sourcils sur la perception du *« stress »*[15]. Un test perceptif a été mis en place avec une tête parlante et une voix de synthèse. Un total de 20 phrases différentes de type nom1-verbe-nom2 a été testé. Les quatre variables testées signalaient le *« stress »* sur l'un des deux noms, F0 quant à elle était parfois maintenue à son niveau neutre pour les deux noms. La tâche perceptive était d'indiquer pour chaque nom à quel degré de *« stress »* il correspondait (*« indicate the degree to which a given word in a sentence was stressed »* indiquer le point auquel un mot donné de la phrase était focalisé). Les auteurs ont conclu que toutes les variables testées avaient une influence sur la perception du *« stress »* mais qu'il semblait que c'était l'intensité qui en avait le plus. En réalité, d'après les courbes communiquées dans l'article de Massaro [2002], il est assez difficile de conclure et de séparer les effets de chacune des variables. On remarque clairement que si trois des quatre indices portent sur le nom 1 (respectivement le nom 2) alors que le dernier, quel qu'il soit, porte sur l'autre, c'est le nom 1 (respectivement le nom 2) qui est détecté comme étant focalisé. Les autres résultats sont à notre sens

[15] Équivalent au terme *focalisation* selon la terminologie utilisée dans ce mémoire.

difficiles à interpréter notamment dans la mesure où c'est l'effet du mouvement des sourcils isolément qui nous intéressait ici tout particulièrement.

E.2.4. Discussion générale sur les liens éventuels entre mouvements des sourcils et de la tête et focalisation prosodique

On pourra conclure qu'aussi bien au niveau de la production que de la perception, aucune des études décrites ci-dessus ne permet de tirer de conclusion claire sur le lien entre les mouvements des sourcils et la focalisation. Bien qu'elles fournissent de bonnes pistes, d'autres études seront nécessaires pour mieux comprendre ces liens complexes. Les relations entre parole en général et prosodie en particulier et mouvements des sourcils ou de la tête sont en effet très complexes puisque ces mouvements ne sont pas directement liés à la parole : on n'est pas obligé de hausser les sourcils pour augmenter F0 …. Peu d'études ont été menées sur la production de tels mouvements en liaison avec la focalisation. Pourtant ces études sont nécessaires pour explorer et tenter de comprendre le fonctionnement du contrôle des mouvements des sourcils ou de la tête en liaison avec le processus de parole. House *et al.* [2001a, 2001b] et Granström & House [2004, 2005] soulignent d'ailleurs l'importance d'effectuer de telles études. Même au niveau perceptif, la plupart des tests décrits ci-dessus utilisent des têtes parlantes. Or il a déjà était montré (cf. Benoît *et al.* [1996b]) que leur intelligibilité était moins bonne que celle de visages naturels (voir courbes comparatives des pages 324 et 326 de Benoît *et al.* [1996b]). A notre connaissance, le seul test portant sur la perception de visages naturels, et qui sera décrit ci-après, a été mis en œuvre en demandant explicitement aux locuteurs de produire un mouvement de sourcil (Swerts & Krahmer [2005]). Avec une telle contrainte, on ne peut pas non plus tirer de conclusion quant à la validité des résultats pour le lien entre la focalisation et les mouvements des sourcils pour la parole naturelle. Il faudra donc, pour obtenir un compte rendu précis des relations entre focalisation prosodique et mouvements des sourcils, mener des mesures précises sur de la parole plus naturelle. De plus, peu d'études ont été menées pour le français (cf. Cavé *et al.* 1996 en production) or il semblerait qu'il puisse exister des différences inter-linguistiques au niveau de l'influence des mouvements des sourcils sur la perception de la focalisation (cf. Krahmer & Swerts [à paraître]).

En conclusion, on soulignera ici la prudence qui doit être de mise quant à la tendance générale à accorder une importance toute particulière aux mouvements des sourcils ou de la tête pour la prosodie en générale et la focalisation en particulier. Les premières études menées sur le sujet suggèrent que ces mouvements pourraient avoir une influence mais il reste à définir comment et surtout dans quelle mesure.

Cette prudence est renforcée par un certain nombre d'études psycholinguistiques ayant montré qu'il était parfois possible de percevoir des événements ne se produisant jamais. Tout d'abord, les études citées ci-dessus, utilisent l'intelligibilité de la focalisation comme test de la qualité de leur système de synthèse audiovisuelle. Or il a été montré (Benoît *et al.* [1996a]) qu'à cause de la redondance d'informations véhiculées dans un message de parole, les signaux synthétiques même les moins réussis sont très intelligibles. Par conséquent d'autres méthodes sont préconisées par Benoît *et al.* pour évaluer les systèmes de synthèse notamment l'utilisation de phrases sémantiquement imprévisibles (*« Semantically Unpredictable Sentences »*). Il semble donc être délicat d'étendre les résultats perceptifs obtenus avec une tête parlante à la perception naturelle de la parole audiovisuelle.

F. Perception audiovisuelle de la focalisation prosodique : interactions entre modalités auditive et visuelle

Swerts & Krahmer [2004] ont mené une étude sur la perception audiovisuelle de la *« prominence »*[16]. Cette étude concernait le néerlandais et visait à déterminer les effets respectifs de l'information auditive et visuelle pour percevoir la *« prominence »*. Pour cela, les auteurs ont conçu deux expériences. La première avait pour but de déterminer si les sujets pouvaient détecter une syllabe accentuée dans une séquence de trois syllabes sans sens dans trois conditions différentes : audio seul, visuel seul et audiovisuel. La seconde expérience avait pour but d'analyser les comportements perceptifs dans le cas où les informations auditives et visuelles sont conflictuelles. Les auteurs ont filmé les productions de 20 locuteurs de face. Seules les productions de cinq d'entre eux ont été retenues pour le test perceptif. La figure I.8 donne des exemples d'images enregistrées en version neutre *vs.* accentuée. Les locuteurs devaient lire une séquence de trois syllabes CV sans sens (soit /ma ma ma/ soit /ga ga ga/) en rendant l'une d'entre elle *« more prominent than the other two »* (plus proéminente que les deux autres). Si les auteurs ont choisi d'utiliser une consonne labiale et une consonne vélaire, c'était dans le but d'étudier si *« frontal sounds would have clearer visual correlates of prominence than sounds produced in the back »* (les sons frontaux correspondraient à des corrélats visuels plus marqués que les sons produits à l'arrière). Toutes les productions ont été enregistrées pour deux modes d'élocution : un mode « normal » (*« natural speaking mode »*) et un mode « exagéré » (*« exaggerated » speaking mode*) pour lequel les locuteurs devaient imaginer qu'ils parlaient à quelqu'un qui était plus loin (« *someone standing at a larger distance »)*. Un total de 45 sujets a été testé soit en audiovisuel (AV) soit en audio seul (A) soit en visuel seul (V). Leur tâche était de dire quelle syllabe ils percevaient comme ayant été produite avec *« the strongest accent »* (l'accent le plus fort). Chaque stimulus a été présenté deux fois. Les auteurs ont trouvé des effets significatifs de la position de l'accent (il semblerait que l'accent soit mieux détecté s'il porte sur la première ou la dernière syllabe), de la modalité (A~AV>V), du mode d'élocution (résultats non communiqués) et du locuteur (résultats non communiqués). Des interactions significatives ont été mesurées entre l'accent et la modalité, entre la modalité et le locuteur et entre le locuteur et la syllabe considérée (*i.e.* /ma/ ou /ga/). On remarquera qu'aucun effet principal de la consonne n'a été mesuré. Les pourcentages de réponses correctes sont les suivant : en audiovisuel, 97,11%, en audio seul, 97,33% et en visuel seul 92,89% (notons que ces scores élevés surtout en visuel seul sont liés à la tâche : focalisation sur une seule syllabe très ouverte). On remarque donc clairement grâce aux performances en visuel seul, qu'il existe des indices « visibles » de la proéminence puisque les sujets parviennent très bien, à partir de la modalité visuelle seule, à détecter et localiser la proéminence. Etant donné l'importance des scores mesurés, les auteurs concluent à l'existence d'un effet plafond et pensent donc qu'il est difficile, d'après ces résultats de déterminer l'importance relative des indices acoustiques et visuels sur la perception.

Les auteurs ont donc mis en place une seconde expérience pour laquelle ils ont manipulé les stimuli enregistrés pour les deux locuteurs les mieux perçus de la première expérience afin que les informations acoustiques et visuelles deviennent conflictuelles. Un troisième locuteur a été enregistré.

[16] Équivalent au terme *focalisation* selon la terminologie utilisée dans ce mémoire.

Il avait pour tâche d'exagérer les expressions faciales produites. Cet enregistrement a été effectué dans le but de pouvoir tester s'il existait un gradient des effets des indices visuels sur la perception. Un total de 55 sujets a passé le test. La tâche perceptive était exactement identique à celle de la première expérience mais cette fois-ci le test n'avait lieu que pour la condition audiovisuelle. Chaque stimulus leur était également présenté deux fois. Les auteurs ont mesuré des effets significatifs de l'accent acoustique (*« auditory accent »*), de l'accent visuel (*« visual accent »*) et du locuteur et une interaction significative entre le locuteur et l'accent visuel. Globalement, les réponses des sujets tendent pour une large majorité à être en faveur de la syllabe ayant reçu l'accent acoustique. Cependant, lorsque cela n'est pas le cas, les réponses s'orientent le plus souvent vers la syllabe ayant reçu l'accent visuel. Les auteurs concluent que *« the auditory cues are stronger than the visual cues though the latter cannot be ignored »* (les indices acoustiques sont plus forts que les indices visuels bien que ces derniers ne puissent être ignorés). Ils notent aussi un effet de la position de la syllabe accentuée. L'effet des indices acoustiques est en effet plus fort pour la syllabe initiale que pour la syllabe finale. Les auteurs suggèrent que ceci serait dû au phénomène de déclinaison de F0 qui rendrait ainsi un pic de F0 sur la dernière syllabe moins fort qu'un pic de F0 sur la première syllabe. On pourra noter que cette hypothèse n'est peut-être pas la bonne puisqu'il a été montré que les auditeurs savent compenser le phénomène de déclinaison (Liberman & Pierrehumbert [1984]). Les auteurs constatent que les indices acoustiques ont moins d'effet chez le locuteur ayant exagéré les expressions faciales. Les auteurs précisent de plus que les résultats perceptifs bien que s'orientant vers l'accent acoustique sont beaucoup moins bons que les résultats obtenus pour la première expérience. Les sujets ont de plus apparemment trouvé le test difficile et les stimuli parfois étranges ce qui n'avait pas été rapporté lors de la première expérience. Les sujets ont donc clairement été perturbés par ces informations conflictuelles.

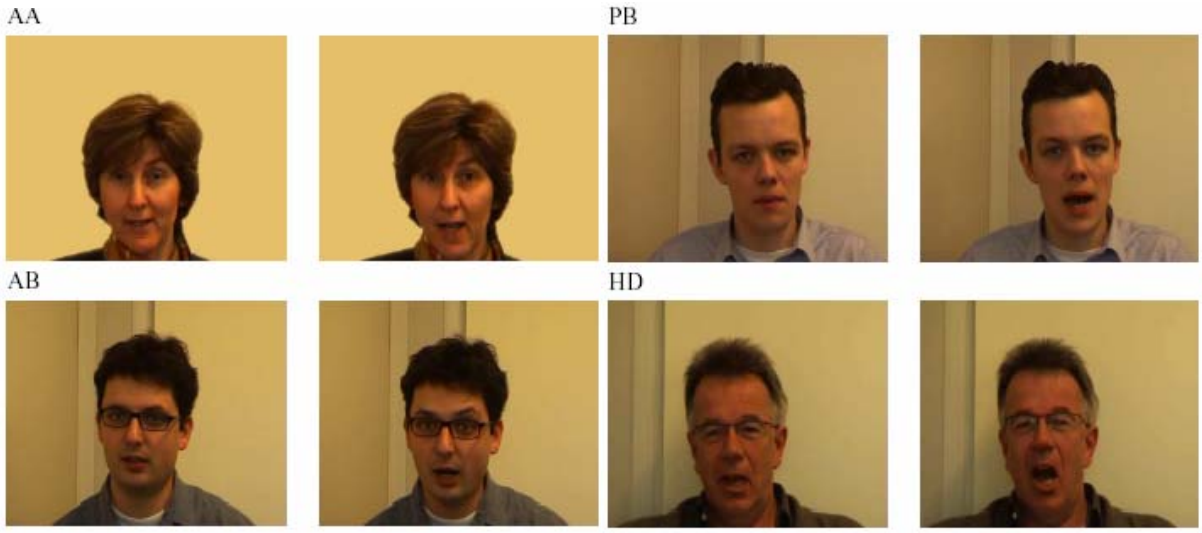


FIGURE I.8 – *D'après Swerts & Krahmer [2004] : huit images extraites des enregistrements de quatre des locuteurs pour une syllabe inaccentuée (gauche) et accentuée (droite).*

Il existe donc des indices visuels à la *« prominence »* et ceux-ci sont importants compte tenu des résultats obtenus en visuel seul pour la première expérience. La deuxième expérience montre que l'acoustique prévaut sur le visuel mais les auteurs disent tout de même que *« visual cues can interfere with auditory information, in attracting some of the perceived accents »* (les indices visuels peuvent interférer avec l'information acoustique, en attirant vers eux certains des accents perçus). Pour la

suite, les auteurs pensent développer une expérience de perception pour laquelle seule une partie du visage du locuteur serait montrée aux sujets et ce afin d'évaluer les différences inter-locuteurs en ce qui concerne les indices « visibles » produits. Ils prévoient également d'étudier les seuils de combinaison des indices acoustiques et visuels à l'aide de continua de F0 et d'informations visuelles. Ils voudraient de plus évaluer la charge cognitive imposée aux sujets lorsqu'ils doivent évaluer des stimuli pour lesquelles les informations auditives et visuelles sont conflictuelles. Enfin ils reconnaissent que les résultats obtenus ici ne sont que préliminaires puisque l'étude ne portait pas sur de « vrais » mots ou de « vrais » énoncés et que la situation de communication n'était pas naturelle.

Swerts & Krahmer [2005] décrivent une autre étude conduite dans le but d'étudier deux points cruciaux. Le premier est l'importance des indices faciaux en comparaison aux indices acoustiques et le second est l'exploration des zones faciales qui seraient les plus importantes pour la perception des mots proéminents. Cette étude concerne aussi le néerlandais. Huit locuteurs ont été filmés de face en train de prononcer une phrase *« each time with emphasis on one of the [three] words »* (avec à chaque fois, focalisation sur l'un des [trois] mots). Deux tests perceptifs ont eu lieu. Le premier était en fait une expérience de mesure du temps de réaction. Les enregistrements décrits précédemment ont été manipulés afin que, pour certains, les informations auditives et visuelles soient congruentes et pour d'autres non. La tâche des sujets était d'indiquer *« which word they perceived as the most prominent one »* (quel mot ils percevaient comme étant le plus proéminent) et ce aussi vite que possible. Il apparaît que le temps de réaction est plus lent lorsque les informations acoustiques et visuelles ne sont pas congruentes. L'analyse n'a été menée que pour les stimuli conflictuels pour lesquels la réponse a porté sur le mot marqué par l'information acoustique. Les auteurs concluent que *« subjects are sensitive to visual information to prominence, even in cases where they do not use this information in their actual choice »* (les sujets sont sensibles à l'information visuelle même dans les cas pour lesquels ils n'utilisent pas cette information pour effectuer leur choix final).

Pour le second test perceptif, les enregistrements ont été manipulés afin qu'ils correspondent tous à une F0 monotone (*i.e.* pas d'information acoustique) avec un accent visuel sur l'un des trois noms. Les sujets voyaient soit la partie haute, soit la partie basse du visage et soit la partie gauche, soit la partie droite du visage. Ils étaient placés à une distance de soit 50cm, soit 250 cm, soit 380 cm de l'écran. La tâche était exactement la même que pour le premier test. Les résultats montrent que la détection de la proéminence est de plus en plus difficile à mesure que la distance à l'écran augmente. Il apparaît que la partie haute du visage donne plus d'information aux sujets que la partie basse. Les résultats perceptifs sont aussi meilleurs pour la partie gauche que pour la partie droite du visage. Notons que le fait que la partie haute du visage apparaisse comme étant la plus utile est peut être dû au fait que les mouvements de sourcils produits semblaient d'amplitude assez forte. Rappelons que, bien que les enregistrements utilisés correspondaient à des productions « humaines », les locuteurs savaient qu'ils devaient produire une focalisation (pas de tâche naturelle pour la production) et qu'il est ainsi possible qu'ils aient exagéré certains de leurs mouvements. Dans le futur les auteurs voudraient distinguer « l'effet locuteur » de « l'effet observateur » notamment en menant exactement le même test avec les images miroirs de celles utilisées ici.

Ces deux études permettent donc de penser qu'il existe bien des informations visuelles à la focalisation prosodique (au moins pour le néerlandais). Il apparaît de plus que le processus de perception combine les deux modalités (auditives et visuelles) pour prendre une décision perceptive unique, puisque quand ces informations sont conflictuelles, non seulement les performances perceptives sont moins bonnes mais les temps de réaction sont plus longs. Les deux modalités jouent donc apparemment un rôle conjoint. Il apparaît aussi que les parties haute et gauche du visage

fourniraient le plus d'informations visuelles utiles. On notera qu'aucune tâche naturelle n'avait été mise au point pour la production de la focalisation et qu'il est ainsi possible que les productions analysées ne soient pas tout à fait naturelles.

G. Bilan : la « visibilité » de la focalisation dans la littérature

Cet état de l'art a permis de montrer que la littérature pouvait fournir des données de départ pour approfondir la problématique que nous nous étions fixée en fin d'introduction. Tout d'abord, on a vu que plusieurs études avaient constaté qu'il était en partie possible de voir la focalisation prosodique. Puis nous nous sommes intéressés aux paramètres visuels qui pourraient jouer un rôle dans la production de la focalisation prosodique. Cette investigation a permis de comprendre que plusieurs gestes faciaux intervenaient certainement dans cette perception visuelle. La focalisation prosodique possède apparemment un certain nombre de corrélats articulatoires potentiellement visibles. Lorsqu'un locuteur focalise, les cibles articulatoires de la bouche et de la mandibule sont plus extrêmes. La vitesse et la durée des gestes effectués par ces mêmes articulateurs semblent aussi être affectées. Il semble de plus que l'énoncé dans son intégralité soit affecté par la focalisation et pas seulement l'élément focalisé. Ceci doit sans doute créer un contraste visible au sein de l'énoncé entre ce qui est focalisé et ce qui ne l'est pas. Il est également possible que la focalisation prosodique soit accompagnée d'autres gestes faciaux tels que des mouvements des sourcils ou de la tête. La façon dont parole et gestes faciaux sont corrélés est encore très mal comprise. Il est pour l'instant difficile d'apprécier les apports relatifs des informations articulatoires et des informations fournies par d'autres gestes faciaux.

La littérature fournit ainsi des résultats importants pour l'étude qui va nous intéresser. On soulignera cependant l'intérêt particulier que pourront avoir de nouvelles études. Notons déjà que la plupart des études décrites ci-dessus ne concernait pas le français. Ces études permettent ainsi d'avoir une idée de ce à quoi il conviendra de s'intéresser mais il reste encore un long chemin à parcourir. Ces mêmes études se sont aussi bien souvent concentrées sur des aspects particuliers et précis n'analysant qu'une infime partie de l'étendue des phénomènes à explorer. La parole offre en effet une variabilité considérable qui a certainement une influence sur le sujet qui nous intéresse ici. Les productions étaient bien souvent peu naturelles et on peut penser que les résultats ont pu en être affectés.

On soulignera de plus le manque frappant qui existe en ce qui concerne l'étude de la visibilité des gestes articulatoires liés à la focalisation. Le champ d'exploration de la bimodalité de la production et de la perception de la focalisation contrastive prosodique en français demeure donc immense et nous allons maintenant nous y aventurer.

– Chapitre II –

Analyse préliminaire de l'acoustique de la deixis prosodique en français

Avant de nous lancer dans l'étude détaillée de la problématique, nous commencerons par préciser le cadre général de la prosodie acoustique dans lequel nous nous positionnerons. Bien que l'intérêt principal de ce mémoire soit l'analyse de la focalisation prosodique du point de vue visuel, il n'en demeure pas moins que la prosodie s'ancre de façon très forte dans l'acoustique. Or l'étude de l'acoustique de la prosodie du français est assez ancienne et la littérature offre une multitude de modèles et de cadres théoriques. Il convient donc de faire le point et de se positionner clairement dans ce cadre afin d'obtenir une base solide pour l'étude qui nous concernera ensuite.

A. Le modèle prosodique de Jun & Fougeron

De nombreux modèles phonologiques de la structure prosodique du français ont été échafaudés (Rossi [1985], Hirst & Di Cristo [1993], Mertens [1993], Vaissière [1997], Di Cristo [1998, 2000], Post [2000] et Jun & Fougeron [2000, 2002] et *inter alia*). Ces modèles considèrent, pour la plupart, que l'intonation française est constituée de suites de mouvements mélodiques montants et descendants et que l'accent est post-lexical. De plus, l'énoncé y est souvent organisé en niveaux prosodiques hiérarchisés. Ces modèles divergent du point de vue du nombre de niveaux hiérarchiques, du type de hiérarchie, des représentations tonales et intonatives sous-jacentes et, de façon plus générale, au niveau de la notion d'accent en français.

Dans le cadre des travaux décrits dans ce mémoire, le modèle phonologique auto-segmental de Jun & Fougeron (Jun & Fougeron [2000, 2002]) a été choisi. Ce modèle, qui s'inscrit dans le cadre de la phonologie prosodique développée particulièrement par Pierrehumbert [1980], Selkirk [1984], Pierrehumbert & Beckman [1988] et Beckman [1996], est en effet en accord avec la plupart des descriptions de l'intonation du français et utilise un système de transcription qui a montré son utilité dans la description prosodique de nombreuses langues : le système ToBI et ses diverses versions (Beckman *et al.* [2005]).

Le modèle de Jun & Fougeron comporte principalement deux unités prosodiques de niveaux hiérarchiques croissants : la *« Accentual Phrase »* souvent traduit par Syntagme Accentuel (SA) et la *« Intonational Phrase »* souvent traduit par le Syntagme Intonatif (SI). L'unité la plus basse, Syntagme Accentuel (SA), correspond environ au « groupe de force » de Passy, au « Mot/Syntagme Prosodique » de Vaissière, à l'« Arc Accentuel » de Fonagy, au « Mot Phonologique » de Milner & Regnaud, à l'« Intonème Mineur » de Rossi, à l'*« Intonation Group »* de Mertens, au « mot rythmique » de Pasdeloup et au « groupe rythmique » de Delais-Roussarie (cf. Jun & Fougeron [2002]). Il est au-dessus de l' « Unité Tonale » de Di Cristo & Hirst et correspond parfois à leur « Unité Rythmique ». Le SA contient un (ou plusieurs) mot(s) de contenu (au moins un) et éventuellement un (ou plusieurs) mot(s) de fonction. En moyenne, un SA contient ainsi 2,3-2,6 mots (ou 1,2 mots de contenu) et 3,5-3,9 syllabes (selon Jun & Fougeron [2000]). Il est marqué à droite par un accent mélodique (pitch accent) démarcatif : l'accent primaire bitonal LH* (*Low-High**, montée de F0 constituée d'un ton bas suivi d'un ton haut). Le ton H* est associé à la dernière syllabe pleine (sans e-muet final) du SA. On note que l'étoile (*) est utilisée de façon conventionnelle pour indiquer une association fixe à une syllabe en particulier. Le SA est de plus marqué par un allongement final. Il est parfois aussi marqué par une montée initiale de F0 d'emplacement variable (phrase accent) : l'accent secondaire LHi (*Low-High initial*). Contrairement au ton H*, sommet de l'accent primaire, le ton Hi,

sommet de l'accent secondaire, n'est pas toujours réalisé sur une syllabe spécifique et n'entraîne pas un allongement de la syllabe qui le porte (Pasdeloup [1990], Mertens *et al.* [2001], Jun & Fougeron [2000] et Welby [2003]). Le ton Hi est le plus souvent porté par la première, la deuxième ou plus rarement la troisième syllabe du premier mot de contenu du SA. Cette montée initiale est considérée comme étant un « accent de syntagme » *(« phrase accent »*) par Jun & Fougeron [2002] plutôt qu'un « accent de hauteur » (*« pitch accent »*) comme le LH*. La séquence LHi n'est en effet pas associée à des syllabes précises, Jun & Fougeron ont observé que le L peut s'étendre sur plusieurs syllabes clitiques précédant le premier mot de contenu du SA. LHi est ainsi simplement associé à la frontière gauche du SA. Bien que le L de LHi ne soit pas associé à une syllabe précise et qu'il puisse s'étendre à plusieurs syllabes, Jun & Fougeron [2002] notent qu'il est toujours réalisé à la frontière gauche du SA et pas plus tard. Un SA commence ainsi généralement par un L. Selon Welby [2002, 2003], LHi aurait également une autre fonction que celle d'accent de syntagme. Jun & Fougeron précisent d'ailleurs que le L peut s'étendre sur plusieurs mots de fonction précédant le premier mot de contenu mais les associations tonales de leur modèle ne le précisent pas. Or Welby montre, notamment avec les données exposées dans Welby [2002], que le point d'inflexion du L vers le Hi, nommé « coude » (*« elbow »*), est le plus souvent réalisé soit très tard dans le dernier mot de fonction, soit très tôt dans la première syllabe du premier mot de contenu du SA. Welby conclut, grâce à d'autres études notamment en perception, que ce coude aurait pour cible la frontière mot de fonction/mot de contenu et serait ainsi un indice fort à la segmentation pour identifier les frontières entre les mots ou le début du premier mot de contenu. Elle propose ainsi (Welby [2003]) l'hypothèse d'une double association du ton bas (L) de l'accent de syntagme (montée initiale) avec la frontière gauche du premier mot de contenu du SA et de manière facultative avec la frontière gauche du SA ou avec le début d'une autre syllabe. Cette double association rejoint celles décrites par Pierrehumbert & Beckman [1988] et Grice *et al.* [2000] pour d'autres langues. Un schéma exposant les associations tonales proposées par Welby [2002] est donné figure II.1.

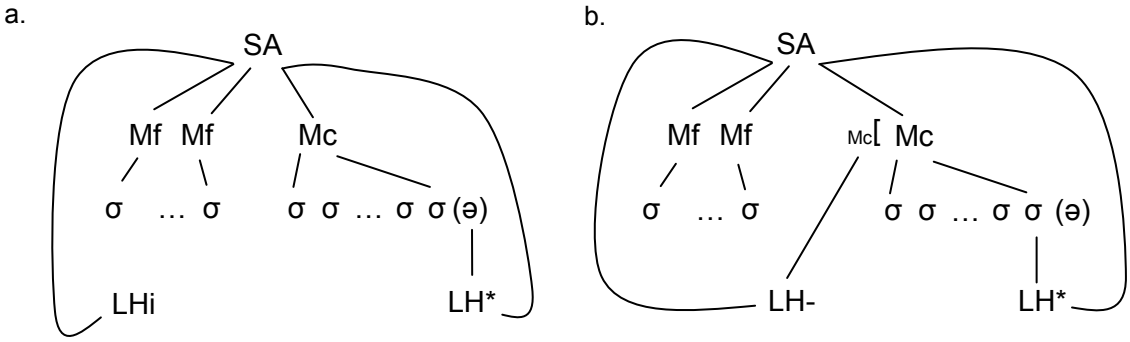


FIGURE II.1 – *Associations tonales pour le SA : a. (gauche) modèle de Jun & Fougeron [2002] ; b. (droite) révision proposée par Welby [2002].*
(Mf : mot de fonction ; Mc : mot de contenu ; σ : syllabe).

Pour Jun & Fougeron, le patron tonal par défaut d'un SA est donc /LHiLH*/. Une illustration en est donnée pour le deuxième SA de la figure II.2 (*i.e.* pour le verbe *ranima*).

Dans le modèle de Jun & Fougeron, l'unité intonative de niveau le plus élevé est le Syntagme Intonatif (SI). Le SI correspond approximativement à l'« Intonème Majeur » de Rossi, à l'« Unité Intonative » de Di Cristo & Hirst et au « Groupe de Souffle » de Vaissière. Il est composé d'en moyenne 2-2,7 SA et de 7,3-10,1 syllabes (selon Jun & Fougeron [2000]). Le SI est marqué à sa

droite par un ton de frontière : L% ou H%. L% (*Low%*) correspond à une chute finale majeure et souvent à un énoncé déclaratif. H% (*High%*) correspond à une continuation finale majeure et plutôt à une interrogation ou une continuation (en effet un énoncé peut être composé de deux SI et dans ce cas le premier SI peut être terminé par une montée de continuation H%). Le SI est aussi marqué par un allongement final parfois suivi d'une pause. L'allongement final est plus important en fin de SI qu'en fin de SA (cf. *e.g.* Di Cristo [1985, 1998] et Jun & Fougeron [2000]). Un SI peut de plus moduler le SA qu'il contient. En effet, dans le modèle de Jun & Fougeron, lorsqu'un SA est en fin de SI, son accent primaire (LH*) est remplacé par le ton de frontière du SI : L% ou H% comme c'est le cas pour le dernier SA de la figure II.2 (*i.e.* l'objet *la jolie maman*).

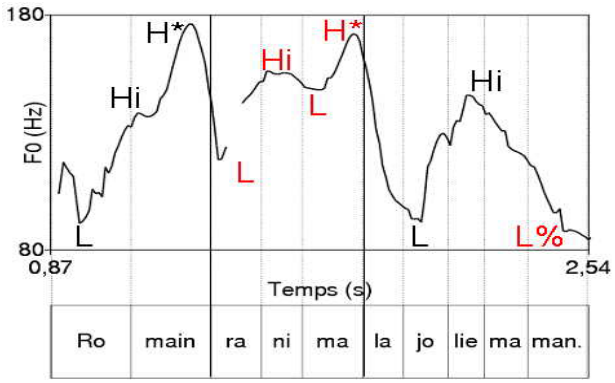


FIGURE II.2 - *Suivi de F0 pour un SI comprenant 3 SA.*
L'énoncé est {[Romain]$_{SA}$[ranima]$_{SA}$[la jolie maman.]$_{SA}$}$_{SI}$ réalisé {[LHiH][LHiLH*][LHiL%]}.*

Aux représentations sous-jacentes ou par défaut ainsi définies pour les SA (LHiLH*) et SI (L% ou H%) peuvent correspondre diverses réalisations phonétiques effectives. Au niveau du SA, les réalisations phonétiques dépendent de la longueur du SA (notamment du nombre de syllabes qu'il contient) et de facteurs rythmiques ou pragmatiques. L'accent primaire est le seul quasi-obligatoire. La réalisation /LHiLH*/ n'est ainsi observée en pratique que pour des SA de plus de trois syllabes de contenu (Jun & Fougeron [2000]). Jun & Fougeron [2000] rapportent l'observation de cinq réalisations tonales simplifiées souvent par manque de temps parce que le syntagme est trop court pour réaliser les quatre cibles tonales. Ces représentations sont résumées dans le tableau II.1.

La troisième colonne du tableau II.1 (ajoutée par moi-même) donne la longueur en syllabes des SA pour lesquels on observe le plus souvent les patrons décrits dans les deux colonnes précédentes. Dans la première version de leur modèle, Jun & Fougeron [2000] précisaient qu'en principe, chacun des quatre tons L, Hi, L et H* était associé à une syllabe, rendant le patron /LHiLH*/ impossible sur les syntagmes de moins de quatre syllabes. Dans une version ultérieure (Jun & Fougeron [2002]), elles sont revenues sur cette association forte, et seul H* est associé précisément à une syllabe, ce qui rend compte de l'observation de syntagmes de trois syllabes ayant un schéma /LHiLH*/. Cette deuxième version du modèle est plus conforme à mes propres données.

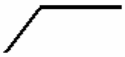
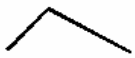
	/L Hi L H*/	plus de 3 syllabes de contenu
	a. [L (Hi L] H*]	1, 2 ou 3 syllabes de contenu
	b. [L (Hi) L H*]	3 syllabes de contenu
	c. [L Hi (L) H*]	3 syllabes de contenu (cas le moins fréquent)
	d. [(L) Hi L H*]	3 syllabes de contenu et plus
	e. [L Hi (L) L*]	3 syllabes de contenu et plus quand le SA est en fin de SI ou quand le SA suivant commence avec un Hi.

TABLE II.1 – *Les cinq types de réalisations sous-jacentes possibles d'un SA (/LHiLH*/) dans le cas où l'un ou plusieurs des quatre tons n'est pas réalisé. Le (ou les) ton(s) entre parenthèses ne sont pas réalisés. D'après Jun & Fougeron [2000] (p. 216).*

Le tableau II.1 montre qu'un SA d'une syllabe sera réalisé [LH*] comme c'est le cas pour le SA central de la figure II.3 (*i.e.* le verbe *vit*). Les SA de 3 syllabes par exemple peuvent être réalisés de diverses façons. Un facteur important qui détermine la réalisation prosodique qui sera utilisée est le contenu lexical du SA. Le premier L ne sera par exemple souvent pas réalisé si le SA commence par un mot de contenu et non par un mot outil tel un article (Jun & Fougeron [2000]). Welby [2003] montre ainsi que la réalisation est de toute évidence conditionnée par la segmentation du flux de parole en mots. Le coude du L vers le Hi en début de SA marque en général la passage d'un mot outil vers un mot de contenu. En l'absence de mot outil avant le premier mot de contenu, un coude peut tout de même être réalisé en tout début de ce mot.

En outre, d'autres patrons, non décrits par Jun & Fougeron semblent exister tel LHi comme l'expose Welby [2003] (p. 89).

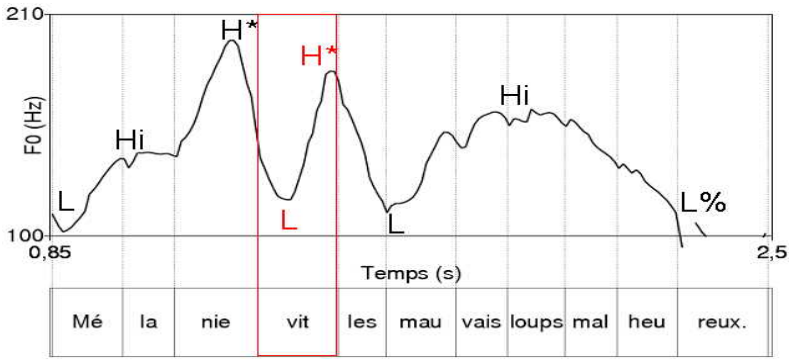


FIGURE II.3 – *Suivi de F0 pour un SI comprenant un SA central de 1 syllabe avec la réalisation tonale [LH*]. L'énoncé est {[Mélanie]$_{SA}$[vit]$_{SA}$[les mauvais loups malheureux.]$_{SA}$}$_{SI}$.*

On constate que la réalisation [HiLH*] est observable non seulement pour des SA de trois syllabes mais aussi pour des SA de plus de trois syllabes. Le ton initial L peut en effet ne pas être réalisé lorsque le SA commence par un mot lexical, un ton Hi étant alors réalisé sur sa première syllabe. La réalisation [LHiL*] est la réalisation la moins courante de toutes dans le corpus étudié par Jun & Fougeron [2002]. Les réalisations [LHiLL*] et [LHiL*] sont observées lorsqu'un SA est suivi d'un autre SA commençant par Hi et ceci pour éviter une succession de trois tons hauts comme ce serait le cas pour la séquence : [L**HiH***]$_{SA1}$[**Hi**LH*]$_{SA2}$ qui serait ainsi plutôt réalisée [LHi**L***]$_{SA1}$[HiLH*]$_{SA2}$.

On soulignera enfin qu'en français, comme dans la plupart des langues du monde, le niveau global de F0 baisse de façon constante du début à la fin d'un énoncé (anglais américain : Pike [1945], Maeda [1976] ; anglais britannique : De Pijper [1980] ; néerlandais : 't Hart *et al.* [1990] ; japonais : Fujisaki & Sudo [1971] ; danois : Thorsen [1980] ; suédois : Bruce [1977] ; français : Vaissière [1971], Delgutte [1978] et voir aussi le cas particulier du Hausa : Meyers [1976]). C'est ce que les chercheurs nomment phénomène de déclinaison.

B. État de l'art sur les corrélats acoustiques de la focalisation contrastive prosodique en français

Les corrélats intonatifs acoustiques de la deixis prosodique ou focalisation contrastive prosodique en français ont été étudiés à de nombreuses reprises (*e.g.* Touati [1987], Di Cristo [1998], Clech-Darbon *et al.* [1999], Di Cristo & Jankowski [1999], Rossi [1999], Jun & Fougeron [2000], Astesano [2001] et Delais-Roussarie *et al.* [2002]). Il est en effet à la fois plus naturel et plus facile expérimentalement d'étudier les corrélats acoustiques de la focalisation contrastive que les corrélats articulatoires. Ceci explique que tant d'études se soient penchées sur les corrélats acoustiques en français comme d'ailleurs dans d'autres langues. Ces études ont abordé aussi bien l'analyse de l'élément focalisé lui-même que de l'élément post-focal. L'élément pré-focal n'a quant à lui été que très peu étudié.

B.1. Constituant focalisé

Les nombreuses études menées sur le sujet s'accordent pour dire que le constituant focalisé est marqué par une augmentation forte et soudaine de la fréquence fondamentale (F0) et/ou de l'intensité sur le syntagme focalisé (Dahan & Bernard [1996], Di Cristo [1998], Rossi [1999] et Jun & Fougeron [2000]).

Certaines études mettent en valeur une augmentation de la durée des syllabes focales (*e.g.* Dahan & Bernard [1996]) avec allongement encore plus important de la syllabe portant la focalisation et plus spécifiquement de la consonne d'attaque du constituant focalisé. Néanmoins, d'autres auteurs trouvent qu'il y a invariance de la durée des syllabes focales (Rossi [1999]).

B.2. Séquence post-focale

De façon générale, les auteurs ayant étudié la séquence post-focale concluent à une compression globale de la plage de variation de F0 et de l'intensité de la séquence post-focale. De nombreuses études rapportent la réalisation d'un plateau de F0 bas jusqu'à la fin de l'énoncé. Ce phénomène est appelé désaccentuation dans la littérature : tous les accents ou tons sont supprimés.

Di Cristo & Jankowski [1999] ont ainsi observé une réduction drastique de l'amplitude des variations de F0 mais *sans* élimination des contrastes tonals. Les auteurs soulignent cependant que cette observation a surtout été faite pour des énoncés complexes ou des séquences post-focales relativement longues. En effet, quand les énoncés sont simples, ils observent également un plateau bas ou légèrement descendant sur la séquence post-focale.

D'après Delais-Roussarie *et al.* [2002] la « désaccentuation » post-focale ne correspond pas forcément à une séquence uniformément plate. En effet, leur étude met en évidence trois réalisations prosodiques possibles pour la séquence post-focale en français selon deux dimensions : la longueur de la séquence post-focale et le statut informationnel de cette séquence. Lorsque la séquence post-focale est relativement courte, syntaxiquement simple et non informative, la séquence post-focale est réalisée par un plateau bas. Si la séquence post-focale est complexe et/ou qu'elle apporte un élément informatif nouveau, les auteurs observent une compression par rapport au patron tonal qu'on observerait pour la même séquence dans le cas neutre[17]. Enfin, lorsque la séquence post-focale est courte mais qu'elle apporte une information nouvelle, il y a décroissance constante de F0 jusqu'à la fin de l'énoncé (Delais-Roussarie *et al.* [2002]). Dans la même lignée, Astésano *et al.* [2004b] constatent que « en position post-focale, les patrons de F0 seraient aplatis ».

Néanmoins, selon la plupart des auteurs ayant étudié la question, cette désaccentuation n'est pas forcément synonyme de réorganisation prosodique de la séquence en question : l'information de marquage prosodique est toujours véhiculée par les indices temporels comme par exemple l'allongement de fin d'unité (Di Cristo [1998], Di Cristo & Jankowski [1999], Jun & Fougeron [2000] et Delais-Roussarie *et al.* [2002]). Di Cristo & Jankowski [1999] précisent même qu'il y aurait une tendance à contre-balancer la désaccentuation par un allongement final aux frontières syntaxiques exagéré par rapport au cas neutre.

De façon générale, on observe donc une désaccentuation de la séquence post-focale mais les informations de marquages prosodiques sont conservées grâce aux informations de durée (allongements finaux). Cette caractéristique est différente de ce que l'on peut observer en anglais, par exemple, où tous les types d'information de marquage prosodique (F0 et durée) sont effacés (Erickson & Lehiste [1995]).

[17] Rappelons ici au lecteur que le terme *cas neutre* fait référence à ce que les linguistes nomment *focalisation large* (cf. la section *Notes et indices de lecture* au début de ce mémoire pour plus de détail sur ce choix de notation).

B.3. Autres corrélats

Dans la lignée des théories des linguistes du Cercle de Prague, Touati [1987] distingue, pour le français, *focalisation* et *contraste* comme deux cas de figure de la rhématisation[18]. Selon lui il existe des différences intonatives subtiles entre ces deux phénomènes pragmatiques. Dans les deux cas, il a analysé l'intonation de l'élément pré-focal (ou pré-contraste). Dans le cas de la focalisation, il observe une « montée de F0 d'ampleur réduite » pour l'élément directement pré-focal. Par contre, pour le contraste, il note que les montées tonales de l'élément pré-contraste « suivent la hiérarchie accentuelle mise en évidence dans l'énoncé neutre ». Astésano *et al.* [2004b] notent quant à eux une diminution de la durée de l'élément pré-focal. Jun & Fougeron [2000] observent de plus une réduction du nombre de frontières syntagmatiques.

Dans les langues en général, la focalisation contrastive prosodique semble donc mettre en valeur le constituant focalisé en rehaussant les paramètres prosodiques sur le constituant focalisé mais aussi en les diminuant sur les autres constituants voisins, stratégie notée par Lehiste [1970] et Erickson [1998] dans d'autres langues. Erickson [1998] (p. 166) écrit ainsi pour l'anglais:

> *It is as if emphasis were applied to the top of the prosodic hierarchy to affect everything in the sentence ; […] In acoustic studies, we see an increase of duration (among other things) on the unit receiving emphasis, and a decrease on the other ones.* (« C'est comme si l'emphase [équivalent à notre focalisation contrastive] était appliquée au sommet de la hiérarchie prosodique pour tout affecter dans la phrase ; […] Les études acoustiques montrent une augmentation de la durée (parmi d'autres paramètres) de l'unité recevant l'emphase et une diminution pour les autres unités. »)

B.4. La focalisation constrastive dans le cadre du modèle de Jun & Fougeron

Dans le cadre du modèle de Jun & Fougeron (Jun & Fougeron [2000]), la focalisation est marquée par un accent haut fort noté Hf (*pour* focalisation) suivi le plus souvent d'un plateau bas sur les syllabes de la séquence post-focale (jusqu'à la fin du SI). Hf remplace le plus souvent Hi comme dans le cas de la figure II.4 : dans le cas neutre (figure II.4.a) le SA central (*i.e.* le verbe *ranima*) est réalisé [LHiLH*] mais lorsque ce même SA est focalisé (figure II.4.b) il est réalisé [LHf] et on voit clairement que le pic Hf est porté par la syllabe [ni] comme le pic Hi du cas neutre. Néanmoins, selon Jun & Fougeron [2000], Hf peut également remplacer à la fois Hi et H* (*i.e.* la montée de F0 est portée par toutes les syllabes du syntagme et culmine sur la dernière syllabe). Après le ton Hf, la F0 diminue vers une cible tonale L le plus souvent atteinte sur la première syllabe du mot de contenu suivant. On voit ainsi assez clairement sur la figure II.4.b que la cible tonale L est atteinte sur la première syllabe du mot de contenu qui suit l'élément focalisé : le [o] de *jolie*. La durée de la chute vers L est variable et dépend principalement de la position de Hf : elle est plus longue si Hf remplace Hi que s'il remplace H*.

[18] La rhématisation est le processus par lequel un constituant de l'énoncé devient le rhème, autrement dit le focus ou commentaire ou encore prédicat psychologique.

Jun & Fougeron observent parfois une pause après la focalisation et la formation d'un nouveau SI aux variations de F0 identiques à celles observées dans le cas neutre. Il y a également parfois une réorganisation prosodique avec moins de frontières que dans le cas neutre. Enfin, Jun & Fougeron [2000] notent une conservation des informations de marquage prosodique de durée (allongements en fin de SA et de SI).

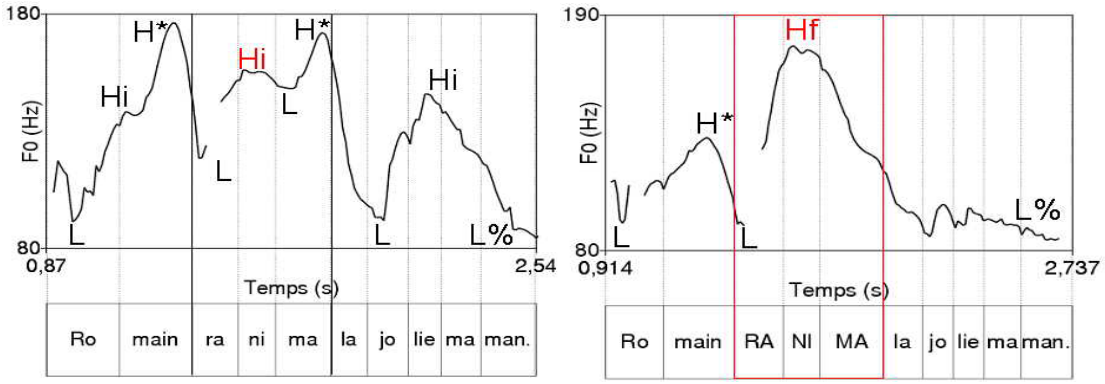


FIGURE II.4 - Suivi de F0 pour un SI comprenant 3 SA. a. (gauche) cas neutre. b. (droite) focalisation sur le SA verbal. On observe le remplacement de Hi par Hf, comme décrit par Jun & Fougeron. L'énoncé est {[Romain]$_{SA}$[ranima]$_{SA}$[la jolie maman]$_{SA}$}$_{SI}$.

C. Analyse des corrélats acoustiques de la focalisation contrastive en français : expérience

Bien que, comme il a été décrit précédemment, il existe de nombreuses études portant sur les corrélats acoustiques de la focalisation contrastive prosodique en français, j'ai tout de même souhaité, et ce pour plusieurs raisons, effectuer moi-même une étude sur le sujet. Peu des études présentées précédemment s'appuyaient en effet sur un modèle prosodique précis. Or les analyses qui seront présentées ci-après s'appuieront sur le modèle de Jun & Fougeron décrit précédemment se plaçant ainsi sous un angle différent par rapport aux autres études. De plus, très peu d'études ont analysé les aspects pré-focaux. A ma connaissance, il n'en existe en effet que deux pour le français : celle de Touati [1987] et celle de Astésano *et al.* [2004b]. L'élément post-focal a aussi été assez peu étudié bien que, comme il a été décrit précédemment, l'on dispose à son sujet de plus d'informations. Or Lehiste [1970] propose que la saillance puisse se faire à la fois en rehaussant le constituant focalisé mais aussi en rabaissant ses voisins.

C.1. Protocole expérimental

C.1.1. Corpus[19]

Le corpus utilisé pour cet enregistrement est composé de huit phrases de structure sujet-verbe-objet (SVO). Tous les items sont constitués de syllabes consonne-voyelle (CV). Chaque phrase est susceptible d'avoir une réalisation prosodique du type {[LHiLH*]$_S$ [LHiLH*]$_V$ [LHiLL%]$_O$}. Dans la mesure du possible, nous avons utilisé des consonnes sonores afin de faciliter le suivi de F0. Les huit phrases du corpus sont [20] :

(1) [Jean]$_{S1}$ [veut ménager]$_{V4}$ [nos jolis nouveaux navets]$_{O7}$

(2) [Romain]$_{S2}$ [ranima]$_{V3}$ [la jolie maman]$_{O5}$

(3) [Mélanie]$_{S3}$ [vit]$_{V1}$ [les mauvais loups malheureux]$_{O7}$

(4) [Véroniqua]$_{S4}$ [mangeait]$_{V2}$ [les mauvais melons]$_{O5}$

(5) [Les mauvais loups]$_{S4}$ [mangeront]$_{V3}$ [Jean]$_{O1}$

(6) [Mon mari]$_{S3}$ [veut ranimer]$_{V4}$ [Romain]$_{O2}$

(7) [Les loups]$_{S2}$ [suivaient]$_{V2}$ [Marilou]$_{O3}$

(8) [Le beau marin]$_{S4}$ [vit]$_{V1}$ [Véroniqua]$_{O4}$

C.1.2. Enregistrement

Le corpus a été enregistré pour le locuteur A[21]. L'enregistrement a été réalisé en chambre sourde (à l'ICP) avec un microphone sur pied placé à une distance constante du sujet. Chaque phrase a été enregistrée pour quatre conditions : focalisation contrastive sur le sujet, le verbe ou l'objet et version neutre. La version neutre correspond à une énonciation de la phrase pour laquelle aucun des constituants n'est mis en valeur par rapport aux autres[22]. Deux répétitions ont été enregistrées pour chaque type d'énoncé. Au total, 64 énoncés ont donc été enregistrés.

C.1.3. Méthode d'obtention de la focalisation contrastive

De façon à rendre la production de la focalisation contrastive la plus naturelle possible, une technique d'obtention de motifs intonatifs a été mise en place. Le locuteur entendait en effet un prompt audio[23] dans lequel la phrase à prononcer était légèrement modifiée. Sa tâche était ensuite de répéter la phrase en corrigeant l'élément incorrect. Le locuteur produisait ainsi de la focalisation contrastive sur l'élément qui avait était modifié dans le prompt (S, V ou O). L'exemple (II.1) permet de comprendre la

[19] Le corpus décrit ici sera à nouveau utilisé dans la suite pour d'autres travaux et sera ci-après nommé corpus AV1.

[20] Les [ ] indiquent le découpage en composants sujet, verbe et objet. L'indice en bas à droite de chaque composant donne son rôle au sein de la phrase (S, V ou O) et le nombre de syllabes qui le compose.

[21] Voir la section D du chapitre *Notes et indices de lecture* pour un récapitulatif des locuteurs.

[22] Ce cas correspond à ce que les linguistiques nomment *focalisation large*.

[23] Prompt audio enregistré à l'avance par une locutrice de langue maternelle française et diffusé par des hauts-parleurs dans le but de déclencher une production spécifique chez l'interlocuteur.

façon dont s'est déroulé l'enregistrement (les majuscules signalent la focalisation qui porte ici sur l'objet).

(II.1) *Le locuteur lit :* Les loups suivaient Marilou.
Le locuteur entend : Les loups suivaient Aurélie.
Le locuteur dit : Les loups suivaient MARILOU.

Aucune indication n'a été donnée au locuteur sur la manière de produire la focalisation (*e.g.* quelle(s) syllabe(s) devai(en)t être accentuée(s)). Pour obtenir la version neutre de la phrase, le prompt était constitué de l'énoncé sans erreur et le locuteur effectuait alors une tâche de répétition.

C.1.4. Données et mesures

Les signaux audio enregistrés ont ensuite été numérisés et échantillonnés à 16kHz[24]. Tous les signaux ont été segmentés en syllabes selon la méthode décrite ci-dessous. Les durées des syllabes et le paramètre F0 ont été mesurés à l'aide du logiciel PRAAT (Boersma & Weenink [2005]). Une méthode d'auto-corrélation a été utilisée pour la détection de F0.

Afin de vérifier qu'il y avait bien eu production de focalisation contrastive sur le constituant désiré, un test de perception informel a été mené. Ce test a permis de vérifier que la focalisation contrastive était bien perçue sur le constituant désiré.

C.1.4.1. Technique de segmentation acoustique des données

Avant traitement, tous les énoncés enregistrés ont été segmentés en syllabes. Les frontières gauche et droite de toutes les syllabes ont ainsi été étiquetées à la main à l'aide du logiciel PRAAT (Boersma & Weenink [2005]). Des règles de segmentation ont été adoptées afin d'établir des critères précis pour que la segmentation soit la plus homogène possible à travers tout le corpus. Toutes les étiquettes correspondent ainsi à des passages par zéro du signal. Le début et la fin de l'énoncé ont été étiquetés de la façon suivante : le début de l'énoncé correspondait au premier front montant après que l'amplitude du signal ait dépassé un seuil défini au préalable et la fin de l'énoncé, au dernier front descendant avant que l'amplitude du signal devienne inférieure à ce même seuil. Les frontières de fin de chaque syllabe ont d'abord été établies approximativement à l'aide du spectrogramme et des formants. Ces frontières ont ensuite été affinées afin de correspondre au front descendant marquant la fin acoustique de la voyelle de la syllabe CV *i.e.* quand il ne restait plus aucune « trace » de la voyelle. Une vérification auditive a ensuite été réalisée. La segmentation a donc été effectuée essentiellement sur des critères acoustiques.

[24] Avec le logiciel CoolEdit : www.cooledit.com.

C.2. Analyse de la structure prosodique des énoncés neutres

Comme il a déjà été expliqué dans la section C.1.1 du présent chapitre et en se basant sur ce qui a été expliqué dans la section A du présent chapitre, le patron tonal par défaut auquel on s'attend pour les énoncés neutres est {[LHiLH*]$_S$ [LHiLH*]$_V$ [LHiL%]$_O$}. Les sommets des accents secondaires (Hi) du sujet et du verbe sont optionnels ainsi que tous les tons bas. Une analyse de la structure prosodique des énoncés neutres enregistrés a été menée afin de vérifier que tel était le cas.

Cette analyse a montré que le patron {[LHiLH*]$_S$ [LHiLH*]$_V$ [LHiL%]$_O$} pouvait être observé mais seulement quand le nombre de syllabes de l'objet était supérieur ou égal à cinq. Quand le nombre de syllabes de l'objet était inférieur à cinq, le verbe et l'objet étaient le plus souvent regroupés pour ne former qu'un seul SA et ceci quel que soit le nombre de syllabes du verbe. Dans ce cas, le patron tonal observé devient donc : {[LHiLH*]$_S$ [LHiL%]$_{V+O}$}. Un exemple de ce groupement du verbe et de l'objet est donné figure II.5. sur laquelle on voit clairement la réalisation [LHiL%] sur V+O.

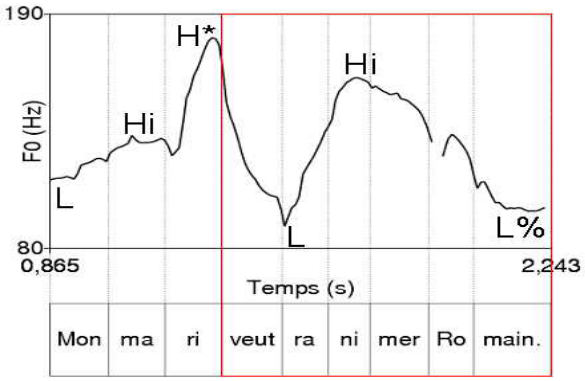


FIGURE II.5 - *Suivi de F0 pour un SI comprenant 2 SA dans le cas neutre (groupement prosodique de V et O). L'énoncé était {[Mon mari]$_{AP}$ [veut ranimer Romain]$_{AP}$}$_{IP}$.*

La syllabe portant le ton de frontière de SA, H* (respectivement le ton de frontière de SI, L% ou H%) est la dernière syllabe du SA (respectivement du SI). Comme il est expliqué par Di Cristo [1985] puis par Astésano [2001] et Jun & Fougeron [2000], ces syllabes de frontière sont allongées. Pour ce locuteur, nous avons en effet mesuré un allongement de la syllabe finale des SA de 20,3%. Pour effectuer cette mesure, nous avons calculé la moyenne des durées de toutes les syllabes des versions neutres des énoncés. Puis nous avons divisé les valeurs des durées des syllabes de fin de SA non final d'un SI par cette valeur moyenne. Si la valeur obtenue était supérieure à 1, la syllabe était considérée comme allongée par rapport à la moyenne, sinon elle n'était pas considérée comme allongée. Nous avons ensuite effectué un test t (comparaison à 1) qui a permis de montrer que cet allongement est significatif : t=3,826 (p=0,001). De plus, cet allongement devrait être perceptivement pertinent puisqu'il est supérieur au seuil de différence perceptible auditif *(« Just Noticeable Difference »*) de 20% évoqué par Astésano [2001] (p. 57). Nous avons également mesuré un allongement de la syllabe finale des SI de 31,7% également significatif : t=8,501 (p<0,001). On notera que l'allongement en fin de SI est significativement différent (supérieur) à l'allongement en fin de SA : t=1,424 (p=0.161), ce qui a déjà été noté dans la littérature (cf. section A du présent chapitre).

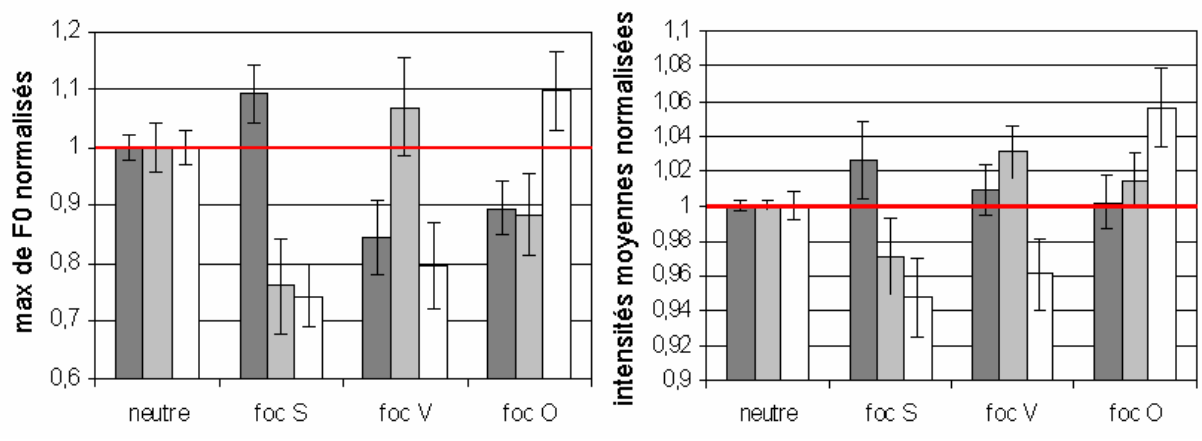
C.3. Analyse de la structure prosodique des énoncés focalisés

Les graphiques de la figure II.6 regroupent les mesures effectuées. Nous avons mesuré :

- le maximum de F0 sur chaque syntagme et pour chaque condition (graphique a. de la figure II.6) ;
- l'intensité moyenne sur chaque syntagme et pour chaque condition (graphique b. de la figure II.6) ;
- la moyenne des durées des syllabes sur chaque syntagme et pour chaque condition (graphique c. de la figure II.6).

Toutes ces données ont été normalisées par division par les valeurs de F0, de l'intensité et de la durée des mêmes syllabes dans les versions neutres des énoncés. Ainsi, après normalisation, une valeur supérieure à 1 indique-t-elle une augmentation par rapport au cas neutre et une valeur inférieure à 1, une diminution par rapport au cas neutre[25]. Les résultats de ces mesures sont décrits en détail dans les paragraphes suivants.

Les analyses des différents paramètres acoustiques (F0, intensité, durée) seront à chaque fois présentées à la fois pour les syntagmes pré-focal, focal et post-focal. L'analyse des contrastes intra-énoncés, c'est-à-dire entre l'élément focalisé et les autres éléments de l'énoncé, sera aussi exposée pour tous les paramètres. Il semblerait en effet que, bien que la majorité des études se soient intéressées à l'élément focalisé lui-même, il existe également des corrélats acoustiques de la focalisation sur le reste de l'énoncé (séquences pré- et post-focales) renforçant par là-même le contraste entre ce qui est focalisé et ce qui ne l'est pas au sein de l'énoncé. C'est pourquoi j'ai choisi d'étendre mon étude des corrélats acoustiques de la focalisation contrastive aux autres éléments de l'énoncé.



[25] La valeur de référence pour la condition neutre est en fait la moyenne des deux énoncés neutres enregistrés (toutes les phrases ont été enregistrées deux fois dans chaque condition : FS, FV, FO et neutre). Les énoncés neutres sont donc eux-mêmes normalisés par cette moyenne et leurs valeurs normalisées peuvent être légèrement différentes de 1 d'où les écarts types qui pourront être observés sur les graphiques qui seront présentés ultérieurement.

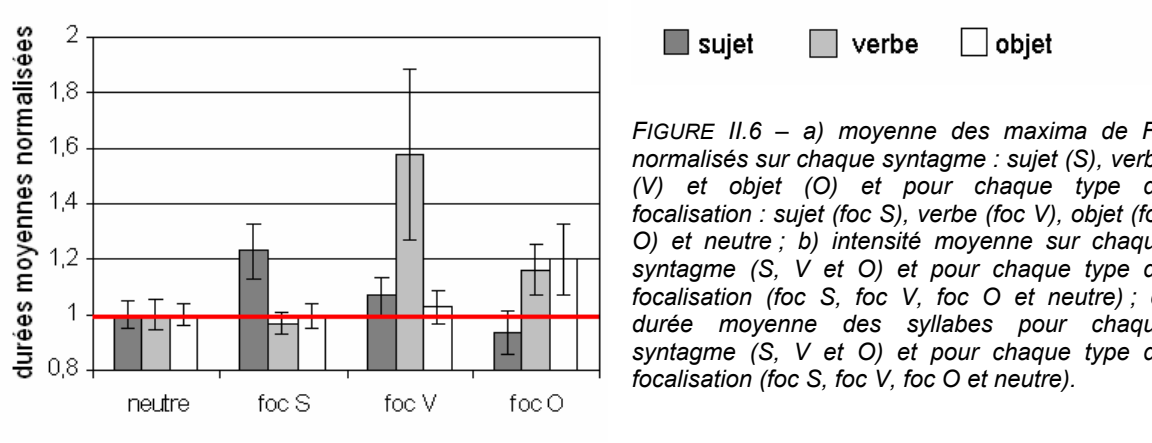


FIGURE II.6 – a) moyenne des maxima de F0 normalisés sur chaque syntagme : sujet (S), verbe (V) et objet (O) et pour chaque type de focalisation : sujet (foc S), verbe (foc V), objet (foc O) et neutre ; b) intensité moyenne sur chaque syntagme (S, V et O) et pour chaque type de focalisation (foc S, foc V, foc O et neutre) ; c) durée moyenne des syllabes pour chaque syntagme (S, V et O) et pour chaque type de focalisation (foc S, foc V, foc O et neutre).

C.3.1. Fréquence fondamentale

C.3.1.1. Position du ton Hf

Comme il a été mentionné plus tôt, Jun & Fougeron [2000] observent que le constituant focalisé est marqué par un ton noté Hf qui remplace le plus souvent le ton (Hi) et que tous les tons post-focaux sont supprimés (dont H*). En ce qui concerne le locuteur A, nous avons identifié quatre cas de figure différents :

- Hf remplace Hi ;
- Hf remplace H* ;
- Hf remplace à la fois Hi et H* et on observe un plateau haut de F0 sur l'élément focalisé ;
- Hf remplace Hi mais on observe aussi un résidu de H*, *i.e.* un ton H* abaissé par rapport au cas neutre, sur la dernière syllabe du SA.

Le tableau II.2 donne les fréquences de réalisation de chacune de ces possibilités. Nous constatons ainsi que Hf remplace le plus souvent :

- Lors d'une focalisation sur le sujet : à la fois Hi et H* ;
- Lors d'une focalisation sur le verbe : Hi avec une forte probabilité d'avoir un résidu de H* ;
- Lors d'une focalisation sur l'objet : Hi (on se souviendra ici du fait qu'il n'y a pas de H* sur le dernier SA d'un SI).

	FS	FV	FO
Hi	13,3	6,7	100
H*	26,7	33,3	
Hi&H*	40	13,3	
Hi + H* résidu	20	53,3	

TABLE II.2 – *Fréquences respectives des diverses localisations possibles de F0 (en pourcent par rapport au nombre total de Hf) pour les cas de focalisation sur le sujet (FS), sur le verbe (FV) et sur l'objet (FO).*

Ces résultats sont en accord avec ce qui avait été décrit précédemment par Jun & Fougeron [2000].

C.3.1.2. Contrastes intra-énoncés

Le graphique a de la figure II.6 montre que lorsqu'un énoncé est focalisé, le maximum de F0 du syntagme focalisé est plus important que ceux des autres syntagmes du même énoncé. Le maximum de F0 du syntagme focalisé est ainsi en moyenne 33% plus élevé que celui des autres syntagmes du même énoncé (45,2% pour FS, 30,2% pour FV et 23,6% pour FO). Une ANOVA à deux facteurs intra-sujets (congruence[26] et type de focalisation[27]) permet ainsi de mettre en valeur un effet congruence significatif : $F(1,15)=643,746$ ($p<0,001$). L'effet type de focalisation est également significatif : $F(1,385,30)=12,187$ ($p=0,001$) et ceci car le maximum de F0 d'un syntagme est plus élevé dans le cas de focalisation sur l'objet quel que soit le syntagme considéré. L'interaction congruence×type de focalisation est aussi significative : $F(2,30)=8,326$ ($p=0,001$). Lorsque la focalisation porte sur le sujet, le contraste au sein de l'énoncé entre la F0 du syntagme focalisé (S) et celle des autres syntagmes (V+O) est en effet significativement plus marqué. On verra dans la suite que la séquence post-focale est désaccentuée (contour de F0 plat) et donc dans le cas où la focalisation porte sur le sujet, tout le reste de l'énoncé est désaccentué. Il est donc logique de noter dans ce cas un contraste plus important.

C.3.1.3. Constituant focal

La figure II.6.a montre que le maximum de F0 d'un élément focalisé est plus élevé que le maximum de F0 du même élément dans le cas neutre (les barres correspondant à S&FS, V&FV et O&FO sont toutes au-dessus de 1). Cette augmentation est en moyenne de 8,7% (9,3% pour S&SF, 7% pour V&VF et 10% pour O&OF) et elle est significative ($t=8,817$ $p<0,001$).

Ce résultat est en accord avec les études menées précédemment par d'autres qui avaient aussi trouvé une augmentation importante de la F0 du constituant focalisé (Dahan & Bernard [1996], Di Cristo [1998], Rossi [1999] et Jun & Fougeron [2000] notamment).

C.3.1.4. Séquence pré-focale

Abaissement des accents

De façon générale, on constate un abaissement des maxima de F0 et donc des tons hauts (*downstepping*) sur le constituant pré-focal. Cet abaissement est en moyenne de 7,1% et est significatif ($t=-13,525$ $p<0,001$).

Pour 59% des stimuli ayant un sujet pré-focal (énoncés à focalisation sur le verbe ou sur l'objet), le patron tonal de ce sujet pré-focal était semblable à celui de ce même sujet dans le cas neutre. Cependant, de façon générale bien que la forme du contour intonatif soit la même, les accents sont abaissés de 17,3% pour les Hi et de 17,3% pour les H*.

Pour 28% des stimuli ayant un sujet pré-focal, l'accent Hi a été supprimé et le H* a été abaissé de 18,3% comme dans l'exemple donné en figure II.7 pour lequel la séquence tonale [LHiH*] du sujet dans le cas neutre est remplacée par la séquence [LH*] pour le cas de focalisation sur le verbe (sujet pré-focal).

[26] Facteur congruence à deux niveaux : cas congruents (S&FS, V&FV et O&FO) et cas incongruents (S&FV, S&FO, V&FS, V&FO, O&FS et O&FV).
[27] Facteur type de focalisation à trois niveaux : FS, FV et FO.

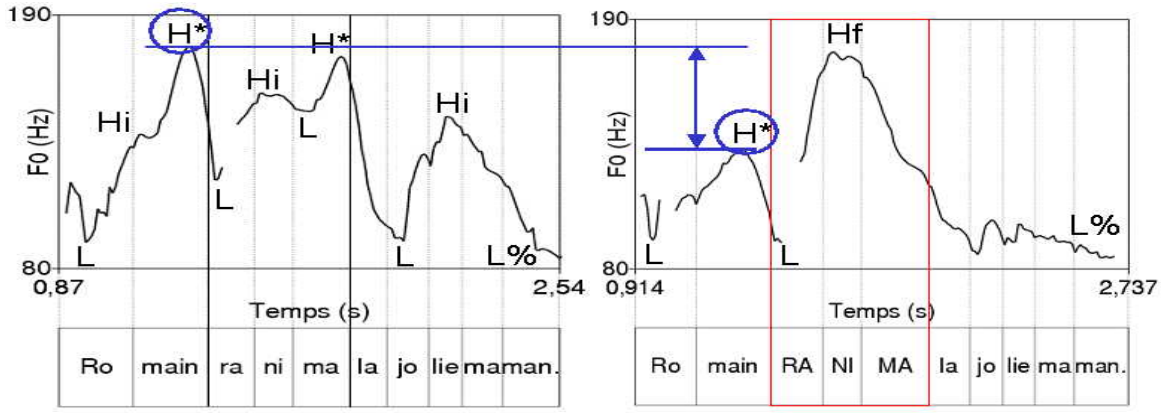


FIGURE II.7 – *Suivi de F0 pour un SI contenant 3 SA. a. (gauche) cas neutre. b. (droite) focalisation sur le verbe. L'énoncé était {[Romain]$_{AP}$[ranima]$_{AP}$[la jolie maman]$_{AP}$}$_{IP}$. On note l'abaissement pré-focal du ton H* sur le sujet.*

Ce résultat est très intéressant puisqu'il montre qu'il existe aussi des conséquences de la focalisation sur le constituant pré-focal. Or ce constituant a jusqu'ici été très peu étudié. On notera que l'étude de Touati [1987] avait cependant conduit à des observations similaires dans le cas de ce qu'il nomme *focalisation*.

Réorganisation prosodique

Il a été expliqué précédemment que dans le cas neutre, lorsque le nombre de syllabes de l'objet est inférieur à cinq, le verbe et l'objet sont regroupés en un seul SA. La question est maintenant de savoir ce qu'il se passe quand l'objet ou le verbe de ce SA regroupé est lui-même focalisé. Dans 62,5% des cas[28], le verbe et l'objet sont séparés en deux SA. En effet, lorsque l'objet est focalisé, un accent H* apparaît sur le verbe et la durée de la syllabe finale du verbe augmente de 30,6% par rapport à la durée de cette même syllabe dans le cas neutre : c'est l'allongement final de fin de SA (corrélat de l'accent primaire H*). Ce phénomène est illustré figure II.8.

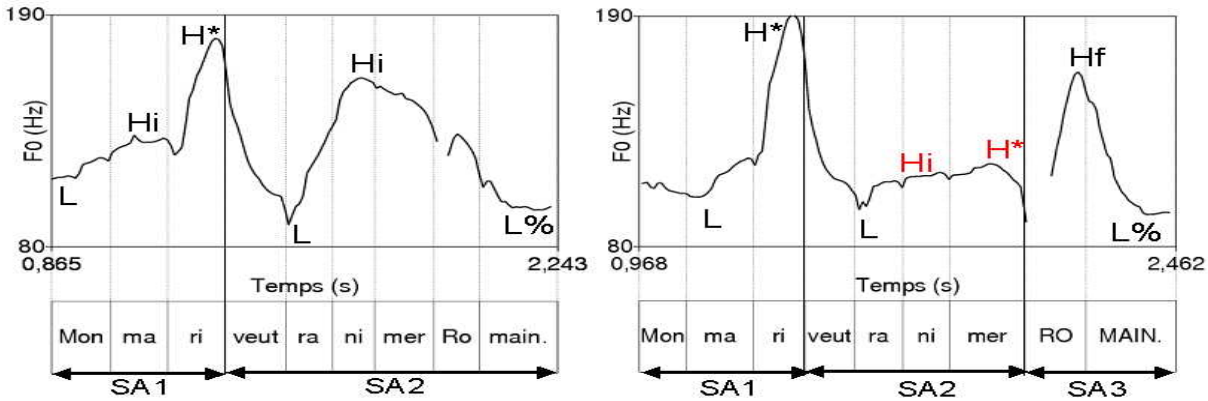


FIGURE II.8 – *Suivi de F0 pour un SI contenant : a. (gauche) 2 SA dans le cas neutre et b. (droite) 3 SA dans le cas de focalisation sur l'objet. On note la réorganisation prosodique liée à la focalisation sur l'objet. L'énoncé était {[Mon mari]$_{AP}$[veut ranimer]$_{AP}$[Romain]$_{AP}$}$_{IP}$.*

[28] Cas de focalisation sur l'objet

De plus, on sait que lorsque le nombre de syllabes du verbe est supérieur ou égal à cinq, le verbe et l'objet sont séparés en deux SA dans le cas neutre. Il se trouve que lorsque ces mêmes énoncés sont focalisés sur l'objet, le verbe et le sujet sont alors regroupés en un SI composé de 2 SA et séparé d'un autre SI composé de l'objet. En effet, nous avons observé que l'accent H* de frontière du SA verbal était nettement relevé dans le cas de focalisation sur l'objet par rapport au cas neutre. De plus lors de la focalisation sur l'objet, le pic de F0 correspondant à l'accent primaire H* du verbe est 13,7% plus important que celui correspondant au sujet. Ceci n'est pas du tout en accord avec la déclinaison de F0 habituellement observée au cours d'un syntagme intonatif (cf. section A du présent chapitre pour plus de détails sur ce phénomène). La dernière syllabe du verbe était également allongée de 26,3% par rapport au cas neutre. Cet allongement ainsi que l'augmentation de F0 nous permettent de conclure que cet accent est en fait un ton de frontière d'un SI de continuation (un H% donc). Un exemple d'une telle réorganisation est donné figure II.9.

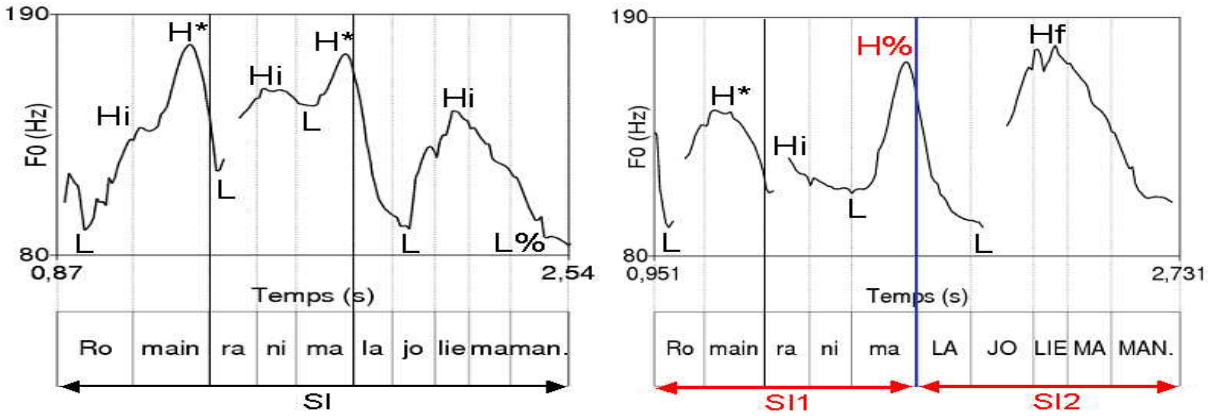


FIGURE II.9 – *Suivi de F0 pour un énoncé comprenant : a.(gauche) un SI dans le cas neutre et b.(droite) deux SI dans le cas focalisation sur l'objet. L'énoncé prononcé était {[Romain]$_{AP}$[ranima]$_{AP}$[la jolie maman.]$_{AP}$}$_{IP}$.*

C.3.1.5. Séquence post-focale

Pour ce corpus et pour ce locuteur, on observe que l'élément focalisé est suivi d'un plateau bas de F0 comme celui qu'on peut observer sur l'objet après le verbe focalisé de la figure II.7.b). La cible tonale basse (L) est généralement atteinte sur la première ou la deuxième syllabe post-focale. La séquence post-focale est donc désaccentuée (on n'observe plus aucun accent H). La cible tonale basse est en moyenne de 87,9 Hz ce qui n'est pas significativement différent de la cible tonale L% de fin de SI des énoncés neutres.

Ce résultat de désaccentuation de la séquence post-focale est en accord avec les travaux précédents de la littérature (Di Cristo [1998], Di Cristo & Jankowski [1999], Delais-Roussarie *et al.* [2002] et Astésano *et al.* [2004b]).

C.3.2. Intensité

C.3.2.1. Contrastes intra-énoncés

Le graphique b. de la figure II.6 montre que l'intensité moyenne des syntagmes focalisés est plus importante que l'intensité moyenne des syntagmes non focalisés du même énoncé. Il existe donc un contraste à l'intérieur de l'énoncé entre ce qui est focalisé et ce qui ne l'est pas du point de vue de l'intensité. Ce contraste intra-énoncé moyen est de 5,5% (7,1% pour FS, 4,6% pour FV et 4,8% pour FO). Un test ANOVA à deux facteurs intra-sujets (congruence[29] et type de focalisation[30]) a été mené et a montré que l'effet congruence était significatif : $F(1,15)=253,93$ ($p<0,001$). De plus l'effet de type de focalisation était également significatif : $F(2,30)=40,241$ ($p<0,001$). L'intensité moyenne dans le cas de la focalisation sur l'objet était en effet supérieure aux intensités moyennes dans les autres cas de focalisation et ceci quel que soit le syntagme considéré. L'effet d'interaction congruence$\times$type de focalisation était tout juste significatif : $F(2,30)=3,427$ ($p=0,046$). Globalement, on note en effet un contraste intra-énoncé plus important pour la focalisation sur le sujet.

C.3.2.2. Constituant focal

Le graphique b. de la figure II.6 permet de constater que l'intensité moyenne d'un syntagme est augmentée lorsqu'il y a focalisation contrastive par rapport au cas neutre (les barres S&FS, V&FV et O&FO sont toutes au-dessus de 1). Cette augmentation est en moyenne de 3,8% (2,7% pour FS, 3,1% pour FV et 5,6% pour FO), elle est de plus significative[31] : $t=9,195$ ($p<0,001$).

C.3.2.3. Séquence post-focale

On constate grâce au graphique b. de la figure II.6 que l'intensité moyenne des séquences post-focales est nettement réduite par rapport au cas neutre (les barres V&FS, O&FS et O&FV sont toutes en-dessous de 1). Cette baisse est en moyenne de 4% (4,1% après FS et 3,9% après FV) et est significative[32] : $t=11,225$ ($p<0,001$).

C.3.2.4. Séquence pré-focale

Au niveau du constituant directement pré-focal, le graphique b. de la figure II.6 permet de constater une très faible augmentation de l'intensité moyenne par rapport au cas neutre : 1,2% en moyenne (0,9% avant FV et 1,4% avant FO). Néanmoins, cette augmentation est significative[33] : $t=3,618$ ($p=0,001$).

[29] Facteur congruence à deux niveaux : cas congruents (S&FS, V&FV et O&FO) et cas incongruents (S&FV, S&FO, V&FS, V&FO, O&FS et O&FV).
[30] Facteur type de focalisation à trois niveaux : FS, FV et FO.
[31] Test t de comparaison à 1.
[32] Test t de comparaison à 1.
[33] Test t de comparaison à 1.

C.3.3. Durées

C.3.3.1. Contrastes intra-énoncés

Le graphique c. de la figure II.6 montre que la durée moyenne des syntagmes focalisés est nettement supérieure à la durée moyenne des autres syntagmes du même énoncé. Un test ANOVA a été mené avec deux facteurs intra-sujets (congruence[34] et le type de focalisation[35]). Ce test a montré que l'effet congruence était significatif : $F(1,15)=149$ ($p<0,001$), ce qui signifie que lorsqu'un syntagme est focalisé la durée moyenne de ses syllabes est supérieure à la durée moyenne des autres syllabes du même énoncé. Ce contraste intra-énoncé est en moyenne ici de 29,9% (25% pour S&FS, 50,4% pour V&FV et 14,3% pour O&FO). De plus, on constate un effet significatif du type de focalisation : $F(1,59,30)=20,22$ ($p<0,001$), ceci est dû au fait que globalement, lorsqu'il y a focalisation sur le verbe, la durée moyenne de toutes les syllabes de l'énoncé est plus importante. Enfin l'effet d'interaction congruence×type de focalisation est aussi significatif : $F(1,458,30)=15,685$ ($p<0,001$), ceci est dû au fait que le contraste entre ce qui est focalisé et ce qui ne l'est pas au sein d'un même énoncé est plus important dans le cas de la focalisation sur le verbe.

C.3.3.2. Constituant focal

Syllabes focales

Le graphique c. de la figure II.6 montre que la durée moyenne des syllabes focalisées est supérieure à celle de ces mêmes syllabes dans le cas neutre (les barres sont au-dessus de 1). Cette augmentation par rapport au cas neutre est de 33,6% en moyenne (23% pour S&FS, 57,7% pour V&FV et 20,1% pour O&FO). Un test t de comparaison à 1 permet ainsi de montrer que la durée moyenne des syllabes focalisées est significativement supérieure à la durée moyenne de ces mêmes syllabes dans l'énoncé neutre : $t=8,877$ ($p<0,001$). Cet allongement est perceptible puisqu'il est supérieur au seuil de perception qui est de 20% (cf. section C.2 du présent chapitre).

Ce résultat est en accord avec les observations faites dans les études citées précédemment.

Premier phonème

La durée du premier phonème, *i.e.* de l'attaque de la première syllabe (qui était toujours du type CV dans ce corpus), du syntagme focalisé a de plus été mesurée et comparée à celle du même phonème dans le cas neutre. Ceci a montré que le premier segment était allongé de 53% par rapport au cas neutre. Un test t de comparaison à 1 montre que cet allongement est significatif : $t=8,786$ ($p<0,001$). On note que cet allongement est nettement supérieur à l'allongement commun à toutes les syllabes focalisées (33,6%, cf. ci-dessus), le premier phonème est donc beaucoup plus allongé que les autres parties de l'élément focalisé.

[34] Facteur congruence à deux niveaux : cas congruents (S&FS, V&FV et O&FO) et cas incongruents (S&FV, S&FO, V&FS, V&FO, O&FS et O&FV).
[35] Facteur type de focalisation à trois niveaux : FS, FV et FO.

C.3.3.3. Séquence pré-focale

On constate que la syllabe précédant directement l'élément focalisé est allongée de 19,6% par rapport au cas neutre (10,9% pour FV et 28,4% pour FO). Cet allongement doit être perceptible auditivement puisqu'il correspond tout juste au seuil de perception (cf. section C.2 du présent chapitre). De plus, il est significatif[36] : t=6,782 (p<0,001).

Une mesure de l'allongement de toutes les syllabes pré-focales a été effectuée et a montré qu'il y avait un allongement significatif global des syllabes pré-focales (t=3,229 p=0,002). Cet allongement moyen n'est cependant que de 5,7% et donc non pertinent au niveau perceptif. C'est donc essentiellement la syllabe directement pré-focale qui est allongée de façon pertinente. Il s'agit sûrement là d'une stratégie d'anticipation. Cette hypothèse sera reprise et étayée plus tard (cf. section A.2.2.2.e.i du chapitre III).

Cet allongement significatif de la syllabe directement pré-focale est un résultat d'autant plus intéressant qu'il n'avait jamais été décrit avant. Comme on l'a vu précédemment, très peu d'études se sont en effet intéressées aux constituants pré-focaux et aucune aux indices temporels spécifiques de durée.

C.3.3.4. Séquence post-focale

On note que la durée moyenne des syllabes post-focales varie peu (voir graphique c. de la figure II.6 : barres V&FS, O&FS et O&FV proches de 1). On note d'ailleurs que l'on ne peut rejeter l'hypothèse nulle d'égalité de la moyenne à 1 : t=0,183 (p=0,856) et donc qu'il n'y aucune variation statistiquement significative. Du point de vue de la durée, il semble donc que la séquence post-focale soit identique à la même séquence dans le cas neutre.

Ce résultat est en accord avec ceux exposés par d'autres auteurs (Di Cristo [1998], Di Cristo & Jankowski [1999], Jun & Fougeron [2000] et Delais-Roussarie *et al.* [2002]).

C.3.4. Discussion sur la désaccentuation de la séquence post-focale

On a vu qu'il y avait désaccentuation de la séquence post-focale. Les indices de F0 de marquage prosodique sont donc effacés (il n'y a plus d'accents). On peut alors se poser la question de savoir s'il y a désorganisation prosodique globale de la séquence post-focale, c'est-à-dire un effacement total des indices de marquage prosodique. On sait en effet que les frontières de SA et de SI sont signalées non seulement par des accents mais aussi par des corrélats temporels. Or nous avons vu que les durées des syllabes de la séquence post-focale ne variaient pas de façon significative par rapport au cas neutre. En complément de cette constatation générale, nous avons mesuré les durées spécifiques de chaque syllabe de fin de SI appartenant à une séquence post-focale et nous avons comparé ces valeurs à celles mesurées dans les cas neutres. En effet, les syllabes de fin de SI étant marquées par un allongement significatif, si les corrélats de durée sont, comme les corrélats de F0, eux aussi effacés, on devrait trouver que la durée moyenne des syllabes de fin de SI appartenant à des

[36] Test t de comparaison à 1.

séquences post-focales est significativement inférieure à 1. Or il apparaît que les durées moyennes des syllabes de fin de SI post-focal ne sont pas significativement différentes de leurs valeurs dans le cas neutre[37] : t=-0,688 (p=0,497). On peut donc conclure que la séquence post-focale est certes désaccentuée mais que l'information de phrasé prosodique est conservée grâce aux corrélats de durée et notamment à l'allongement en fin de SI. Ce résultat est en accord avec les résultats présentés par Di Cristo & Jankowski [1999] ou encore Jun & Fougeron [2000] et Delais-Roussarie *et al.* [2002].

C.4. Conclusion

Cette étude avait pour but de décrire de façon précise et synthétique les corrélats acoustiques de la focalisation contrastive en français. Bien qu'elle n'ait été menée que pour un locuteur, le but de ma thèse n'étant pas fondamentalement de décrire ces corrélats de façon exhaustive, elle vient corroborer, synthétiser et compléter de nombreuses autres études menées auparavant par de nombreux chercheurs. En conclusion, nous pouvons donc résumer de la façon suivante les corrélats acoustiques de la focalisation contrastive prosodique chez le locuteur analysé :

- **constituant focal** : globalement, les maxima de fréquence fondamentale de l'élément focalisé sont supérieurs à la fois à ceux du reste de l'énoncé dans lequel il se trouve (+33%) et à leurs valeurs dans le cas neutre (+8,7%). Il en va de même pour l'intensité moyenne (respectivement +5,5% et +3,8%). On observe de plus un allongement des syllabes focalisées d'en moyenne 33,6% par rapport au cas neutre, avec un allongement encore plus important du premier phonème du constituant focalisé (+53%). La focalisation est essentiellement marquée par un ton Hf qui remplace le plus souvent le ton initial Hi ou à la fois les tons initial Hi et final H* ;
- **séquence pré-focale** : on note que les maxima de F0 de la séquence pré-focale sont globalement abaissés d'environ 7,1% ce qui crée un contraste encore plus important au sein de l'énoncé entre ce qui est focalisé et ce qui ne l'est pas. On observe de plus une très faible augmentation de l'intensité ainsi qu'un allongement de la syllabe directement pré-focale (+19,6%). On discutera par la suite de la possibilité de mise en place d'une stratégie d'anticipation de la focalisation chez ce locuteur. On note aussi une réorganisation prosodique de cette séquence lorsque l'objet est focalisé : si celui-ci a plus de cinq syllabes, les SA sujet et verbe sont regroupés en un SI séparé du SI de l'objet focalisé ; si l'objet a moins de cinq syllabes, les verbe et objet qui étaient regroupés en un seul SA dans le cas neutre sont séparés en deux SA ;
- **séquence post-focale** : cette séquence est désaccentuée mais n'est pas dépourvue d'informations de marquage prosodique. Celles-ci sont en effet fournies par les corrélats de durée. On observe de plus une faible baisse de son intensité moyenne.

En résumé, lorsqu'un constituant est focalisé, sa F0 et son intensité augmentent alors que celles des éléments voisins (pré- et post-focaux) diminuent ce qui crée un fort contraste au sein de l'énoncé. On constate de plus un allongement des syllabes focales et de la syllabe pré-focale. Enfin, on note une réorganisation prosodique de la séquence pré-focale (focalisation sur l'objet) et une désaccentuation de la séquence post-focale. On peut ainsi conclure que la focalisation n'affecte pas

[37] Test t de comparaison à 1.

seulement le SA sur lequel elle porte directement mais le SI en entier comme l'avaient déjà suggéré Lehiste [1970] et Erickson [1998]. Les résultats obtenus dans cette étude sont en accord avec les études de la littérature détaillées dans la section B du présent chapitre. Cette étude permet cependant de mettre en avant l'aspect global des effets de la focalisation sur l'énoncé en entier. Les études de la littérature portaient en effet le plus souvent soit uniquement sur le constituant focalisé soit uniquement sur la séquence post-focale et rarement sur l'énoncé dans sa globalité. De plus, les résultats permettent de décrire une observation qui n'avait jamais été faite auparavant *i.e.* l'existence d'un allongement pré-focal qui s'inscrit dans le cadre d'une stratégie d'anticipation de la focalisation chez ce locuteur.

– Chapitre III –

Analyse de la production visuelle de la deixis prosodique

Speech is rather a set of movements made audible than a set of sounds produces by movements.

Raymond H. Stetson

Le but général de ce chapitre est de déterminer s'il existe des indices visibles de la focalisation contrastive en français et d'identifier leur nature. Si de tels indices existent, la perception de la focalisation contrastive prosodique devrait ainsi être facilitée lorsque l'on peut voir, en plus d'entendre, notre interlocuteur. La communication de l'information prosodique de focalisation serait ainsi plus efficace surtout quand le signal audio est en partie ou même totalement dégradé. Comme il a été décrit dans la section A.3 de l'introduction, les indices visibles qui accompagnent le signal de parole peuvent être de natures diverses. Ils peuvent être directement liés au signal de parole comme les gestes articulatoires par exemple, qui sont en partie visibles, ou alors y être seulement indirectement liés comme les diverses expressions faciales qui, bien qu'accompagnant souvent la parole, en sont largement indépendantes. C'est ainsi qu'on pourra séparer deux types de corrélats visibles : les corrélats purement articulatoires et les autres gestes faciaux tels que les mouvements des sourcils, de la tête ou autres mimiques faciales. Ces deux types de corrélats ont été analysés et les résultats de ces analyses seront détaillés dans le présent chapitre.

A. Les gestes articulatoires comme indices visuels

A.1. Quels paramètres et pourquoi ?

Le but de cette section est de déterminer s'il existe des corrélats articulatoires visibles de la focalisation contrastive en français. Les gestes articulatoires sont en effet, au moins en partie, visibles. Les mouvements des lèvres et de la mâchoire sont directement visibles lorsque l'on parle à quelqu'un. Ce sont donc ces mouvements qui seront analysés dans la suite de cette section. Cependant, je suis consciente du fait que ce ne sont pas forcément les seuls mouvements articulatoires visibles. Il a été argumenté que certains mouvements de la langue pourraient en effet être visibles (cf. section A.3 de l'introduction). Mais les mesures effectuées dans ce sens ne permettent pas toujours de conclure de façon claire. De plus, les systèmes de mesure disponibles ne permettent pas d'effectuer des mesures simultanées de tous les mouvements articulatoires ce qui explique pourquoi il a fallu faire un choix des paramètres à analyser. Il paraissait donc tout à fait logique de commencer par ce qui semble être le plus directement visible c'est-à-dire les mouvements des lèvres et de la mandibule. Un autre argument important est la nature de la technique de mesure. Il est en effet possible de mesurer les mouvements de la langue pendant l'acte de parole mais le système de mesure le permettant (l'articulographe électromagnétométrique) est plus invasif, puisqu'il nécessite de coller des pastilles sur la langue du locuteur lesquelles sont chacune reliées par un fil à l'appareil de mesure et perturbe ainsi de façon assez importante l'acte de parole. La mesure des mouvements labiaux et mandibulaires est beaucoup moins invasive et c'est aussi une des raisons qui m'ont poussée à choisir ces paramètres. Les mouvements des lèvres et de la mandibule ont donc été étudiés d'abord avec un système de suivi automatique du contour des lèvres à partir de données vidéo. Une autre étude a ensuite été menée avec le système de suivi tri-dimensionnel de marqueurs placés sur le visage du locuteur (Optotrak). Ce système permet en effet de mesurer de façon plus précise certains paramètres (la protrusion de la lèvre supérieure par exemple) bien que la mesure soit moins précise pour d'autres paramètres (l'aire

intéro-labiale ou l'ouverture des lèvres notamment). C'est donc la complémentarité de ces deux systèmes qui sera exploitée ici.

A.2. Analyse de données vidéo

A.2.1. Protocole expérimental général

A.2.1.1. Plate-forme d'acquisition et de traitement de données vidéos de l'Institut de la Communication Parlée[38]

L'Institut de la Communication Parlée (ICP) dispose d'un système d'acquisition et de traitement de données fondé sur une détection des contours labiaux internes et externes à partir des images vidéos de face et de profil d'un locuteur (Lallouache [1991] et Audouy [2000]). Le but est de pouvoir suivre les mouvements articulatoires de la bouche au cours du flux de parole.

A.2.1.1.a. Acquisition des données

Le principe consiste à maquiller les lèvres du locuteur en bleu ainsi qu'à lui faire porter des lunettes opaques sur lesquelles sont collées des pastilles bleues de référence. Sur la branche des lunettes est aussi accrochée une réglette bleue de référence (graduée en centimètre) pour le traitement des images de profil. Le locuteur est ensuite filmé à l'aide de deux caméras 3CCD[39,40], l'une d'elle est placée en face de lui (images de face) et l'autre sur son côté droit (images de profil). La résolution des caméras est de 638x582 pixels et leur rapport signal sur bruit est de l'ordre de 55dB. Un obturateur de 1/250^{e} est utilisé sur les caméras ce qui nécessite un grand apport de lumière en compensation. Afin d'optimiser les enregistrements pour la détection des contours labiaux, le locuteur est donc éclairé de face et de profil grâce à des projecteurs néons fluorescents très puissants (5 500 Watts). Ces projecteurs étant d'une puissance assez importante, les lunettes que porte le locuteur sont opaques afin de préserver ses yeux. La tête du locuteur est fixée grâce à un casque (lui-même fixé au mur) ceci afin d'éviter que le locuteur ne bouge trop la tête pendant les enregistrements. Afin de détecter avec la meilleure précision possible les contours des lèvres, il faut en effet que le champ des caméras soit assez étroit. Par conséquent si la tête du locuteur bouge, ses lèvres risquent de sortir du champ des caméras. Les sorties des deux caméras sont mixées à l'aide d'une table de mixage de façon à obtenir sur une même bande les images de face et de profil. Un exemple des images obtenues après ce mixage est donné figure III.1. Le stockage des images est ensuite réalisé via des magnétoscopes Betacam SP[41].

[38] Des informations complémentaires sur les équipements et logiciels décrits dans cette section sont consultables en ligne sur la page web de Christophe Savariaux, ingénieur de recherche à l'ICP : http://www.icp.inpg.fr/~savario.
[39] Les CCD (Charge Coupled Device) sont des dispositifs d'analyse photosensibles de type matriciel dont la densité des photo-éléments détermine la résolution. Les CCD ont l'avantage de posséder une grande dynamique et une excellente précision géométrique (pas de déformation de l'image).
[40] Les caméras sont de marque JVC, référence JVC KY 15E.
[41] Les magnétoscopes sont de la marque Sony, références UVW 1400, UVW 1600 et UVW 1800.

Les lunettes portées par le locuteur ainsi que la réglette qui leur est attachée permettent d'obtenir des références à peu près immobiles. Ces références seront ensuite importantes pour le traitement des données.

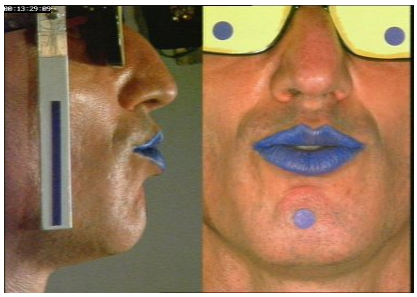


FIGURE III.1 – *Exemple d'image acquise grâce au banc d'acquisition d'images vidéo de l'ICP après mixage des images de face et de profil.*

A.2.1.1.b. Traitement des données vidéos

Après enregistrement, les énoncés cibles à étudier sont repérés grâce à leurs *time codes*[42] de début et de fin sur la bande vidéo. Les images correspondant à ces séquences cibles sont ensuite numérisées grâce à une carte d'acquisition vidéo et ce à la fréquence de 50 trames par seconde, une image vidéo correspondant à deux trames entrelacées. Ces images numérisées sont transmises au logiciel de détection des contours des lèvres : TACLE (Traitement Automatique du Contour des LEvres). Ce logiciel a été initialement mis au point à l'ICP par Lallouache [1990,1991] puis développé et maintenu par Christophe Savariaux avec l'aide de Marc Audouy [2000]. Cette application permet d'extraire des paramètres descripteurs des lèvres à partir de la séquence d'images numérisées. Le traitement consiste à appliquer un *chroma-key*[43] sur l'image, de manière à colorer en noir les zones teintes en bleu. Ces images sont filtrées par un filtre médian et les contours des zones noires sont détectés. Divers paramètres sont ensuite estimés à partir de ces contours et ceci pour chaque trame. L'application génère enfin un fichier dont chaque ligne correspond aux valeurs de tous les paramètres mesurés pour une trame. Les paramètres estimés sont détaillés ci-dessous (voir schéma de la figure III.2 pour une illustration) :

- paramètres de face : étirement labial (A), aperture labiale ou ouverture des lèvres (B), aire intéro-labiale (S) et position (coordonnées bi-dimensionnelles) d'un marqueur (Mf) placé sur le menton du locuteur ;
- paramètres de profil : protrusion des lèvres supérieure (P1) et inférieure (P2), position de la commissure (C), position (coordonnées bi-dimensionnelles) du centre de gravité de la règle de référence.

Ce système d'acquisition/traitement est très robuste et précis puisqu'il permet de détecter des aires intéro-labiales très faibles (jusqu'à 0,5 mm²).

[42] Time code : code temporel attribué à chacune des images enregistrées et permettant de référencer chaque image par rapport à la bande vidéo.
[43] Le chroma-key ou « incrustation en chrominance » est un code de chromacité qui sert à modifier les couleurs. C'est ce système qui permet de faire de l'incrustation vidéo au cinéma (on tourne sur un fond vert qui est ensuite remplacé par une image de fond).

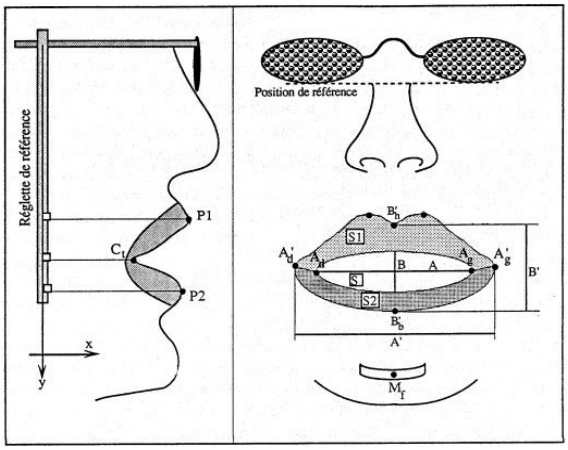


FIGURE III.2 – *Description des paramètres mesurés grâce à l'application TACLE.*

A.2.1.1.c. Paramètres articulatoires mesurés

Il est naturel de s'intéresser aux mouvements visibles des articulateurs de la parole tels que la mandibule et les lèvres. Mais les variations de l'amplitude de ces mouvements ne sont pas les seuls paramètres intéressants. La vitesse, et plus particulièrement les pics de vitesse, sont aussi des paramètres à prendre en compte. Il a en effet été montré que les mouvements articulatoires simples (unimodaux) peuvent se décomposer en deux phases : une phase d'accélération, lors de laquelle la vitesse augmente, et une phase de décélération, lors de laquelle la vitesse diminue. Ces deux phases sont délimitées par un pic de vitesse, le profil de la courbe de vitesse correspondant est donc en forme de cloche (*bell-shaped*). L'analyse des mouvements des membres humains, de la mâchoire ou de la langue a permis d'observer un fait remarquable : la relation entre amplitude du mouvement et pic de vitesse est stable et linéaire (voir *e.g.* Cooke [1980], Nelson [1983], Ostry *et al.* [1983], Kelso *et al.* [1985], et Perrier *et al.* [1989]). De façon particulièrement intéressante pour nous, il a été montré que pour les mouvements de parole, les variations prosodiques (telles le changement de rythme par exemple), influencent la pente de cette relation linéaire.

Certains chercheurs ont tenté de mettre en relation le pic de vitesse avec des entités physiques et de modéliser la relation linéaire entre pic de vitesse et amplitude à l'aide de systèmes mécaniques du second ordre du type masse-ressort, impliquant des paramètres tels que la masse, la raideur et la viscosité. Pour un modèle masse-ressort du second ordre non-amorti, le rapport pic de vitesse sur déplacement (noté Vmax/d ci-après) est proportionnel à la racine carrée de la raideur normalisée (*i.e.* raideur divisée par la masse). Une augmentation de la pente de la relation linéaire entre pic de vitesse et déplacement peut ainsi être modélisée par une augmentation de la raideur. Selon Cooke [1980] *« This observation accords with the common experience of tensing or co-contracting in the expectation of performing a very rapid movement. »* (Cette observation est en accord avec l'expérience commune de tension ou co-contraction produite en prévision de la production d'un mouvement très rapide.).

L'exactitude de la relation d'équivalence entre Vmax/d et la raideur est controversée, car elle n'est valable théoriquement que pour les mouvements non-amortis et non tronqués (voir *e.g.* Lœvenbruck [1996] et Byrd & Saltzman [1998]). Il a néanmoins été montré qu'il existait bien une relation entre pic de vitesse et *effort*. Nelson [1983] décrit un ensemble de coûts physiques associés à la réalisation de mouvements appris simples. Le coût d'impulsion (*impulse cost*) est défini comme proportionnel à

l'impulsion totale (l'intégrale temporelle de l'amplitude de la force) sur tout le mouvement. Nelson a montré que ce coût "equals the peak velocity, V, for all movements where the frictional forces are negligibly small compared to the applied forces, and where the velocity patterns are unimodal (have a single velocity peak). For all other cases, the peak velocities are always somewhat greater than this effort-cost." (est égal au pic de vitesse, V, pour tous les mouvements pour lesquels les forces de friction sont négligeables par rapport aux forces appliqués, et pour lesquels les profils de vitesse sont unimodaux (n'ont qu'un seul pic de vitesse). Pour tous les autres cas, les pics de vitesse sont toujours légèrement plus grands que ce coût lié à l'effort). Par conséquent, selon Nelson, le pic de vitesse est relié à la quantité d'effort fourni pendant le mouvement. Pour des résultats plus récents sur la notion d'effort en parole, le lecteur pourra se reporter aussi à Perkell *et al.* [2002].

Certains chercheurs ont tenté de mettre en relation les effets articulatoires des variations prosodiques (telles que l'accentuation ou les effets de frontières) avec les variations des paramètres des modèles articulatoires du second ordre. Dans ce cadre, les mouvements articulatoires liés à la parole accentuée ont parfois été associés à une diminution de la raideur, définie par le rapport Vmax/d (Ostry *et al.* [1983]). Ce résultat non intuitif, n'a pas toujours été observé par la suite (cf. Lœvenbruck [1996]). Il a notamment été montré que la relation entre l'accentuation et le rapport Vmax/d est loin d'être aussi simple et que de nombreux paramètres, parmi lesquels le séquencement temporel, semblent interagir dans la production de l'accentuation (Lœvenbruck [1996] et Cho [à paraître]). De plus, certains chercheurs ont suggéré que l'inverse de l'écart temporel entre le début du mouvement et son pic de vitesse est un meilleur indicateur de la raideur que ne l'est Vmax/d (Byrd & Saltzman [1998], Byrd *et al.* [2000]). Dans cette lignée, Keller [1987] et Gracco [1988], entre autres, ont décrit des données intéressantes sur le séquencement temporel des pics de vitesse d'un certain nombre d'articulateurs de la parole.

Par conséquent, dans la mesure du possible, nous prendrons en compte à la fois les amplitudes des gestes articulatoires et leurs pics de vitesse.

A.2.1.2. Méthode générale d'analyse des données

Ce paragraphe a pour but de présenter la méthode générale d'analyse des données qui a été utilisée pour toutes les analyses de données audiovisuelles. Les spécificités propres à chaque étude seront ensuite décrites dans les sections correspondantes. L'objectif général de ces analyses est de déterminer les variations articulatoires « visibles » lorsque le locuteur focalise un constituant d'un énoncé.

A.2.1.2.a. Segmentation acoustique

La segmentation acoustique consiste à identifier les frontières temporelles des syllabes afin d'isoler celles-ci dans le signal acoustique et de pouvoir les étudier séparément. Cette segmentation est réalisée à l'aide du logiciel Praat (Boersma & Weenink [2005]) et selon la méthode décrite à la section C.1.4.1 du chapitre II.

A.2.1.2.b. Validation acoustique

Une première étape consiste à vérifier que les productions acoustiques sont conformes aux attentes afin de ne pas étudier les corrélats articulatoires de données acoustiquement inadéquates. Le but est

ainsi de s'assurer que le locuteur a bien produit de la focalisation contrastive sur l'élément désiré. Deux types de validations sont menées pour chaque sous-ensemble de données. La première vise à étudier la **structure prosodique acoustique des énoncés produits**.

Elle consiste tout d'abord à mesurer les maxima de F0 et d'intensité de chaque énoncé puis à vérifier que ces maxima sont bien situés sur l'élément focalisé, ainsi que l'ont montré maintes études antérieures de la focalisation en français (cf. section B et C du chapitre II). Quand ce n'est pas le cas, il faut également vérifier que la baisse de F0 sur l'élément censé être focalisé n'est pas simplement due au phénomène de déclinaison évoqué dans la section A du chapitre II. La déclinaison peut ainsi rendre les pics de F0 d'un élément focalisé-objet (donc placé en fin d'énoncé) de mêmes amplitudes voire d'amplitudes plus faibles que ceux du sujet du même énoncé (en début d'énoncé). Néanmoins, il a été montré que les auditeurs compensent ce phénomène de déclinaison (Liberman & Pierrehumbert [1984]) et on peut donc penser que la perception de la focalisation sera correcte malgré tout.

Les F0 et intensité moyennes de chaque constituant de chaque énoncé sont également mesurées afin de faire des comparaisons entre éléments focalisés et non focalisés. En effet, les F0 et intensité moyennes de l'élément focalisé sont en général plus importantes que celles des autres éléments du même énoncé. Néanmoins, le phénomène de déclinaison peut également intervenir à ce niveau. Pour les énoncés ne vérifiant pas les critères suivants : les maxima de F0 et d'intensité sont sur l'élément focalisé et les F0 et intensité moyennes sont plus importantes sur ce même élément que sur le reste de l'énoncé, il faudra faire une analyse plus détaillée de chaque signal en particulier pour voir s'il s'agit d'une conséquence de la déclinaison ou s'il s'agit d'une production erronée.

La deuxième étape de la validation acoustique des données consiste en une **validation perceptive**. Un test de perception audio informel est ainsi mené, afin de vérifier que d'un point de vue acoustique, la focalisation contrastive est bien perçue. Ce test vise notamment à vérifier que lorsque le phénomène de déclinaison intervient, la perception audio de la focalisation est toujours correcte. Les résultats des deux types de validation sont ainsi mis en commun et les énoncés trop ambigus sont identifiés et éliminés pour l'étude articulatoire.

A.2.1.2.c. Visualisation et mesures : TRAP

Après avoir extrait les paramètres décrits à la section A.2.1.1.b du présent chapitre grâce à l'application TACLE, les signaux acoustiques et articulatoires peuvent être visualisés conjointement grâce au logiciel TRAP[44] (TRAitement de la Parole). Il s'agit d'une application intégrée pour le traitement des signaux de parole et de tous leurs dérivés (signaux articulatoires de divers types …). TRAP a été développé à l'Institut de la Communication Parlée (Lœvenbruck, Savariaux & Lefebvre) sous environnement Matlab et prend en compte divers types de signaux acoustiques (aux formats .wav, Matlab, articulographique EMA …) et de signaux articulatoires (aux formats électropalatographique EPG, articulographique EMA, vidéo TACLE …). Cette application permet aussi d'effectuer, soit de façon automatique, soit de façon manuelle, divers types de mesures : détection des extrema (locaux ou absolus), de valeurs moyennes, etc. et divers types de calculs : dérivées première et seconde, etc. Elle permet également d'incorporer les propres programmes de l'utilisateur en fonction des besoins. C'est cette application complétée par des programmes personnels qui sera utilisée pour toutes les études présentées ci-après.

[44] Disponible à l'adresse http://www.icp.inpg.fr/~savario/_activite/zip/trap_v4.zip.

A.2.1.2.d. Analyse statistique

Le même protocole de test statistique a été utilisé sur toutes les données qui seront présentées ci-après et ce afin d'une part de rendre aussi claire que possible la procédure de test et d'autre part de pouvoir faire des comparaisons de manière aisée et rigoureuse entre analyses. Les données seront donc soit utilisées telles quelles (pour la parole délexicalisée notamment), soit ramenées au cas neutre par simple calcul de variation par rapport au cas neutre (différence relative) ou enfin par une technique de normalisation qui sera exposée plus tard. Le cas neutre sera alors systématiquement éliminé des analyses statistiques puisqu'il est déjà pris en compte dans les autres données lors des calculs de variations. Il restera donc une valeur par syntagme (S, V et O) pour tous les énoncés. Une ANOVA à deux facteurs intra-sujets sera ensuite menée sur les données. Les deux facteurs seront : la congruence et le type de focalisation. Le facteur de congruence sera un facteur à deux niveaux : cas congruents et cas incongruents. Les cas congruents correspondront aux cas : sujet et focalisation sur le sujet (S&FS), verbe et focalisation sur le verbe (V&FV) et objet et focalisation sur l'objet (O&FO). Les cas incongruents correspondront aux cas : verbe et objet et focalisation sur le sujet (V&FS et O&FS), sujet et objet et focalisation sur le verbe (S&FV et O&FV) et sujet et verbe et focalisation sur l'objet (S&FO et V&FO). Le facteur type de focalisation aura trois niveaux : FS, FV et FO. Cette ANOVA permettra de tester la signification des contrastes, à l'intérieur d'un énoncé, définis ainsi : *le contraste entre le constituant focalisé et le reste de l'énoncé est-il significatif ?* Dans la suite de ce mémoire, ces contrastes seront appelés contrastes intra-énoncé.

Des tests t de comparaison au cas neutre seront ensuite effectués afin de voir s'il y a variation par rapport au cas neutre et si cette variation est significative. Trois types de test seront ainsi menés : un premier avec les valeurs sur les syntagmes focalisés comparées aux valeurs pour ces mêmes syntagmes dans le cas neutre, un deuxième avec les valeurs pour les syntagmes pré-focaux et un troisième avec les valeurs pour les syntagmes post-focaux toujours comparées aux valeurs pour ces mêmes syntagmes dans le cas neutre. Les résultats seront présentés sous forme de tableaux et analysés dans le texte.

A.2.2. Étude des productions du locuteur A

Comme il a été expliqué précédemment, le but est ici de déterminer s'il existe des corrélats articulatoires visibles de la focalisation contrastive en français. Si ces corrélats existent, le second but sera de déterminer quelle est leur nature et de les décrire de façon précise et aussi exhaustive que possible. Etant donné le fait que très peu d'études ont été menées à ce sujet -- et aucune pour le français (cf. chapitre I) --, ces travaux sont partis d'hypothèses vagues et intuitives sur les phénomènes potentiellement observables. Ces hypothèses sont fondées sur le fait que pour certaines langues autres que le français (l'anglais notamment, cf. section D du chapitre I) des corrélats articulatoires visibles tels qu'une plus grande ouverture de la mâchoire ont pu être mis en évidence lors de la focalisation (Summers [1987], De Jong [1995], Harrington *et al.* [1995], Erickson [1998]). On s'attend donc, *a priori*, à trouver une augmentation de l'amplitude de certains gestes articulatoires que nous nommerons ci-après hyper-articulation qui mettrait en valeur de façon visible le syntagme focalisé. Il est également possible que l'on observe une augmentation des pics de vitesse qui sera par la suite reliée à un effort articulatoire accru (cf. ci-dessous). Une étude pilote a été conduite pour un locuteur de façon à analyser de façon très précise les résultats et à savoir vers où diriger les études

suivantes s'il y avait lieu d'en mener. Cette section a ainsi pour but de décrire l'étude pilote qui a été mise en place sur les productions du locuteur A. Une étude préliminaire simple a d'abord été menée sur la parole délexicalisée (toutes les syllabes étant remplacées par la syllabe unique /ma/). Sur la base des résultats obtenus après cette étude, une étude de la parole réelle de ce même locuteur a été entreprise.

A.2.2.1. Mise en œuvre expérimentale

A.2.2.1.a. Corpus

Le corpus que nous avons utilisé pour ces enregistrements est le corpus AV1 qui a été décrit à la section C.1.1 du chapitre II.

A.2.2.1.b. Enregistrement

Le corpus a été enregistré pour le locuteur A[45] à l'aide de la plate-forme expérimentale décrite ci-dessus (section A.2.1.1 du présent chapitre). L'enregistrement a été réalisé dans la chambre sourde de l'ICP. Le signal audio a été enregistré avec un microphone sur pied placé à une distance constante du sujet. Chaque phrase a été enregistrée pour quatre conditions : focalisation sur le sujet, sur le verbe, sur l'objet et cas neutre. Tous les énoncés ont de plus été enregistrés pour quatre types d'élocution : en parole normale, en parole chuchotée, en parole délexicalisée et en parole délexicalisée chuchotée. La parole délexicalisée consiste à remplacer toutes les syllabes de l'énoncé par une syllabe unique. Ici c'est la syllabe /ma/ qui a été choisie pour ses caractéristiques acoustiques et articulatoires intéressantes (/a/ donne lieu à une nette et grande ouverture de la bouche après le /m/ correspondant à une fermeture des lèvres). Ce mode d'élocution, qui peut sembler artificiel de prime abord, a pour but d'effectuer facilement et clairement des comparaisons acoustiques et articulatoires d'une syllabe à l'autre. En effet, si la syllabe prononcée est toujours la même, ses caractéristiques acoustiques et articulatoires diffèrent uniquement à cause des variations prosodiques et non de façon inhérente par le fait qu'elle est constituée de phonèmes différents. Pour la parole chuchotée, la tâche était de se faire comprendre par un interlocuteur placé à une certaine distance et non de chuchoter à l'oreille de quelqu'un. Le but était ainsi que le locuteur se fasse comprendre sans utiliser les indices acoustiques et donc certainement en intensifiant les indices visibles. Deux répétitions ont été enregistrées pour chaque type d'énoncé (un type d'énoncé correspond à un type de focalisation et un type d'élocution). Au total 256 énoncés ont donc été enregistrés (8 phrases, 4 conditions de focalisation, 4 types d'élocution et 2 répétitions pour chaque cas).

La méthode d'obtention de la focalisation utilisée ici était la même que celle décrite dans la section C.1.3 du chapitre II. On notera que pour la parole délexicalisée, le prompt audio[46] était le même que pour la parole normale (soit en parole normale et non délexicalisée). Le locuteur était donc conditionné par un signal de nature linguistique. Le but était que le locuteur fasse simplement un changement de syllabe vers /ma/ juste avant la production et non qu'il accomplisse une tâche de programmation mélodique ou rythmique directement à partir des /ma/s. Cela permettait de s'assurer que la tâche restait le plus possible une tâche linguistique et non musicale.

[45] Voir la section D du chapitre *Notes et indices de lecture* pour un récapitulatif des locuteurs.
[46] Prompt audio : stimulus auditif enregistré à l'avance visant à déclencher une production spécifique chez le locuteur.

A.2.2.2. Étude préliminaire de la parole délexicalisée

A.2.2.2.a. Problématique : de l'intérêt et de la validité de la parole délexicalisée

L'intérêt d'utiliser ici la parole délexicalisée était de s'affranchir de la variabilité segmentale liée à la diversité phonologique du corpus, le but étant, en effet, d'identifier les corrélats articulatoires liés uniquement à la focalisation. Or les syllabes différentes ont des caractéristiques articulatoires différentes (par exemple /ma/ est articulé avec une plus grande ouverture mandibulaire que /mi/) et il semblait donc difficile, en utilisant de la parole réelle, de distinguer l'effet articulatoire lié à la différence segmentale de l'effet lié à la différence supra-segmentale (focalisation notamment). En utilisant la parole délexicalisée, toutes les syllabes étant identiques, on sait que les différences articulatoires mesurées seront liées à la prosodie et notamment à la focalisation. Ceci permet donc de justifier de l'intérêt d'utiliser la parole délexicalisée.

Cependant on peut maintenant se poser la question de la validité d'un tel type de parole. Il est en effet souvent objecté à l'usage de la parole délexicalisée, que le locuteur, et l'interlocuteur qui perçoit ce type de parole, seraient en train d'accomplir des tâches de nature plus musicale que langagière. Pourtant on notera que, très tôt, les chercheurs ont utilisé la parole délexicalisée pour étudier certains phénomènes de la communication parlée (*e.g.* Stetson [1905] ou Lindblom & Rapp [1973]). En fait, certains chercheurs pensent qu'il serait totalement inutile d'utiliser la parole délexicalisée en recherche et d'autant plus de faire des tests perceptifs l'utilisant. On ne testerait ainsi pas la perception de la parole mais un processus plus complexe. La production de la parole délexicalisée mettrait en effet en œuvre un transcodage de la parole naturelle vers la parole avec syllabes répétées et sa perception nécessiterait un transcodage inverse.

Liberman & Streeter [1978] se sont interrogés sur la validité de l'utilisation de la parole délexicalisée (*« nonsense-syllable mimicry »*) pour effectuer des analyses de durée. Or leur étude, sur la parole réitérée en /ma/, a permis de montrer que l'utilisation de ce type de parole permettait d'éliminer une grande partie de la variabilité segmentale sur la durée. Même si les auteurs nuancent leurs conclusions en soulignant l'intérêt d'une vérification préalable pour chaque locuteur, ils préconisent l'emploi de la parole délexicalisée pour la description des influences prosodiques, à la fois en production et en perception.

Larkey [1983] a également analysé la validité de l'utilisation de la parole délexicalisée *(« reiterant speech »*) aux niveaux acoustique et perceptif. Il conclut que la parole délexicalisée est un outil puissant et efficace pour la recherche en prosodie. Cependant il précise aussi que *« the evaluation of individual speakers' reiterant speech showed that many speakers do not produce good reiterant speech. »* (l'évaluation des productions de parole délexicalisée de chaque locuteur en particulier montre que de nombreux locuteurs ne produisent pas de la bonne parole délexicalisée). Il convient donc, avant toute chose, d'évaluer la capacité des locuteurs à produire ce type de parole.

Avant de mener des études en utilisant ce type de parole, nous avons ainsi tenu à effectuer quelques vérifications de base pour valider notre étude. Le but était de montrer qu'en ce qui concerne notre corpus, le locuteur avait bien effectué une tâche langagière lors de la production des phrases délexicalisées. Précisons que le locuteur A est habitué à produire ce type de parole et est « réputé » pour le produire correctement. Il a été décidé de comparer les profils prosodiques (F0) des phrases délexicalisées à ceux des phrases en parole lexicalisée. Le but était d'établir une correspondance entre les profils : forme générale (patron tonal), positions des accents à conditions de focalisation identiques ...

Or il est apparu clairement que les profils prosodiques des phrases délexicalisées étaient très proches de ceux des phrases lexicalisées. Le locuteur produisait donc bien des patrons rythmiques et intonatifs typiques du langage. Cette constatation permet donc de penser qu'il ne s'agit pas là simplement d'une tâche musicale pour laquelle on n'aurait pas observé des patrons prosodiques si proches de ceux observés en parole lexicalisée. On ne peut cependant pas conclure qu'il s'agit là d'une tâche purement langagière (certainement pas d'ailleurs). Néanmoins cette tâche fait sans aucun doute intervenir des processus similaires, au moins au niveau prosodique. C'est pourquoi, malgré les objections prononcées, il a été décidé d'utiliser la parole délexicalisée. On gardera cependant toujours en mémoire qu'il ne s'agit là que d'études préliminaires, à confirmer ensuite pour la parole lexicalisée et on se gardera à tout moment de tirer des conclusions générales, sachant que l'objet étudié est si spécifique et complexe.

A.2.2.2.b. Validation acoustique des données

Un total de 64 énoncés a été enregistré en parole délexicalisée. Parmi ces énoncés, 14 ont dû être éliminés parce que le locuteur s'était trompé dans le nombre de /ma/s (non identique au nombre de syllabes de la phrase de référence). Après élimination, il restait donc 50 énoncés dont 37 focalisés. Avant de commencer l'analyse articulatoire des données, les productions acoustiques de ces 50 énoncés ont été analysées attentivement selon la méthode décrite à la section A.2.1.2.b du présent chapitre.

Après mise en commun des résultats des deux validations parallèles (acoustique et perceptive), deux énoncés ont dû être rejetés parce que la focalisation pouvait vraiment être analysée et perçue de façon ambiguë. Après cette étape de validation acoustique, il restait donc 48 énoncés à analyser.

A.2.2.2.c. Mesures de durée

Comme il a déjà été expliqué précédemment, la durée peut à la fois être considérée comme purement auditive ou audiovisuelle. Elle sera ici étudiée en considérant qu'elle est visible par l'auditeur/le spectateur au même titre que les gestes faciaux qui seront étudiés après. Les durées de toutes les syllabes du corpus ont ainsi été mesurées sur la base de la segmentation acoustique effectuée à l'aide du logiciel Praat (Boersma & Weenink [2005]) et dont la méthode précise a été détaillée à la section C.1.4.1. du chapitre II. Une moyenne des durées syllabiques a ensuite été calculée pour chaque type de syntagme sous un type de focalisation. Les résultats de ces mesures sont consignés dans le graphique a. de la figure III.3. Un calcul du pourcentage d'augmentation de ces durées moyennes par rapport au cas neutre a également été effectué et les résultats de ce calcul sont exposés dans le graphique b. de la figure III.3.

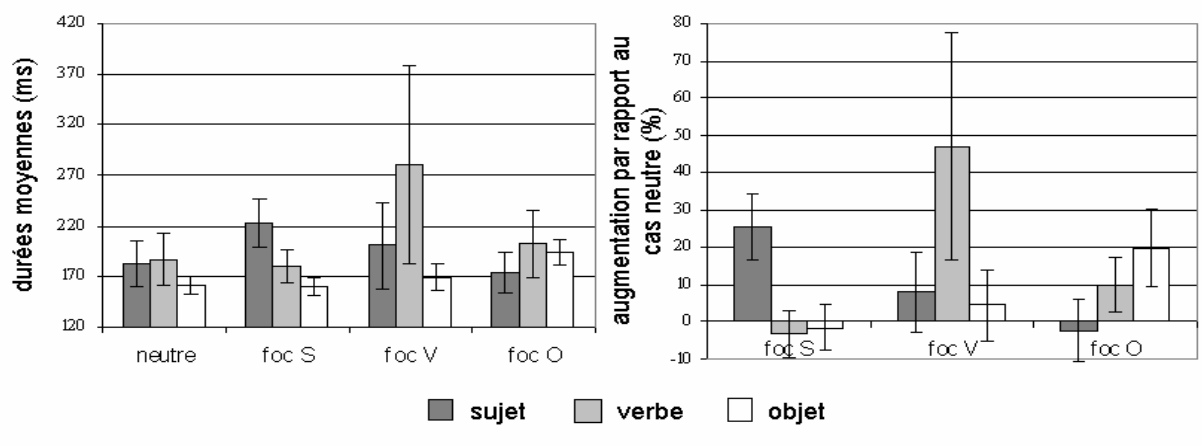


FIGURE III.3 – *a. (gauche) durées moyennes des syllabes de chaque type de syntagme et pour chaque type de focalisation (en ms) ; b. (droite) pourcentage d'augmentation de cette durée par rapport au cas neutre (en %).*

mesure	effet congruence	effet type de focalisation	interaction	test t congruence	test t post-foc	test t pré-foc
durées	F(1,12)=42,902 p<0,001	F(1,161,24)=5,520 p=0,005	F(1,158,24)=5,877 p=0,026	t=1,37 p=0,175	t=-0,222 p=0,825	t=1,256 p=0,213

TABLE III.1 – *Résultats des tests statistiques menés sur les données de durée de la fermeture initiale (début de syntagme) selon la méthode décrite à la section A.2.1.2.d du chapitre III.*

On constate ainsi sur le graphique a. que, de façon générale, la durée moyenne des syllabes est plus importante quand le syntagme auquel elles appartiennent est focalisé que quand elles appartiennent aux autres syntagmes du même énoncé. Ces contrastes intra-énoncés sont surtout visibles dans les cas de focalisation sur le sujet ou sur le verbe : 31,6% de contraste entre S&FS et la moyenne de V&FS et O&FS ci-après noté contraste FS et 51,9% de contraste entre V&FV et la moyenne de S&FV et O&FV ci-après noté contraste FV. Lorsque la focalisation porte sur l'objet, le contraste entre O&FO et la moyenne de S&FO et V&FO ci-après noté contraste FO, n'est que de 3,1% et ce surtout à cause du fait qu'on note un allongement assez net de la durée moyenne des syllabes du verbe pré-focal. Une analyse de la variance montre que les contrastes intra-énoncés sont significatifs (facteur congruence[47] de l'ANOVA[48] : F(1,12)=42,902 p<0,001). Le facteur type de focalisation[49] a également un effet significatif (F(1,161,24)=5,520 p=0,005) et ce car la durée syllabique moyenne est significativement plus importante lorsque c'est le verbe qui est focalisé et ce quel que soit le constituant considéré (contraste de l'ANOVA significatif à p=0,006). L'effet d'interaction[50] est lui aussi significatif (F(1,158,24)=5,877 p=0,026). Les contrastes intra-énoncés correspondant aux cas de focalisation sur le sujet ou sur le verbe sont en effet significativement plus importants que ceux correspondant au cas de focalisation sur l'objet (contraste de l'ANOVA significatif à p<0,001).

[47] Congruence : facteur intra-sujet à deux niveaux : cas congruents (S&FS, V&FV et O&FO) et cas incongruents (V&FS et O&FS, S&FV et O&FV et S&FO et V&FO).
[48] ANOVA à deux facteurs intra-sujets : congruence (deux niveaux) et type de focalisation (trois niveaux) ; pour une explication complète et détaillée du test voir la section 1.2.1.2.4 du chapitre II.
[49] Type de focalisation : facteur intra-sujet à trois niveaux : FS, FV et FO.
[50] Interaction congruence×type de focalisation

On notera que c'est lorsque la focalisation porte sur le verbe que le contraste intra-énoncé est le plus important. Ceci est peut-être dû au fait que les verbes du corpus comportent en moyenne moins de syllabes que les autres constituants. Or l'allongement dû à la focalisation est peut-être réparti entre toutes les syllabes du constituant comme il sera exposé dans la suite (cf. section A.2.4.3 du présent chapitre). Le verbe en comportant en moyenne moins verra donc l'allongement de ses syllabes « moins réparti et donc en moyenne plus important ». Ceci pourrait également expliquer le fait que l'allongement focal soit relativement faible lorsque la focalisation porte sur l'objet. Les objets de ce corpus sont en effet en moyenne les constituants les plus longs (en nombre de syllabes). La conjugaison du nombre de syllabes élevé pour les objets et faible pour les verbes explique peut-être aussi pourquoi dans le cas de la focalisation sur l'objet, la barre V&FO est légèrement plus haute que la barre O&FO.

Le graphique b. de la figure III.3 permet de plus de constater que lorsqu'un syntagme est focalisé la durée moyenne de ses syllabes est plus importante que dans le cas neutre. Cette augmentation est d'en moyenne 30,8% (25,5% pour S&FS, 47% pour V&FV et 19,8% pour O&FO). Néanmoins, elle n'est pas significative : $t=1,37$ $p=0,175$. Ceci est sans doute dû au fait que lorsque c'est le verbe qui est focalisé, on note des écarts types importants (cf. figure III.3). L'analyse des allongements en fonction du nombre de syllabes du constituant, qui sera détaillée à la section A.2.2.3.d.i du présent chapitre, montre que ce grand écart-type est principalement dû au fait que lorsque le verbe focalisé ne comporte qu'une seule syllabe, celle-ci est beaucoup plus allongée que lorsqu'il en comporte plusieurs. On se souviendra que chez ce locuteur la focalisation entraîne souvent une réorganisation prosodique. Dans le cas neutre, il arrive bien souvent que verbe et objet soient regroupés en un seul SA. Or quand le verbe est focalisé, celui-ci constitue la plupart du temps un SA à lui tout seul. Ceci implique que sa dernière syllabe porte l'allongement final de SA qui n'existait pas dans la version neutre. Or quand le verbe ne comporte qu'une syllabe, celle-ci porte à elle toute seule à la fois l'intégralité de l'allongement focal et un allongement dû à sa position en fin de SA d'où ce si grand allongement.

On notera que les allongements décrits ci-dessus sont tous supérieurs au seuil de perception de la variation de durée : la *« Just Noticeable Difference* » qui est de 20% (Astesano [2001], cf. section C.2 du chapitre II).

De façon générale, on n'observe aucune variation significative par rapport au cas neutre aussi bien pour les éléments pré- que post-focaux (cf. table III.1). Le graphique b. de la figure III.3 montre néanmoins que la durée moyenne des syllabes de l'élément directement pré-focal augmente par rapport au cas neutre (+8% pour S&FV, +10% pour V&FO). De façon à affiner l'étude, et compte tenu d'observations préliminaires sur les données, la durée de la syllabe directement pré-focale (dernière syllabe de l'élément pré-focal) a été étudiée de façon plus précise. Les durées de toutes les dernières syllabes des sujets et des verbes ont ainsi été relevées et les résultats sont consignés dans le graphique de la figure III.4. On constate que lorsque l'élément suivant est focalisé, la dernière syllabe d'un syntagme est allongée (les barres blanches sont plus grandes que les barres grises) et ce d'en moyenne 14,5% (9,4% pour le sujet pé-focal et 19,7% pour le verbe pré-focal). Cet allongement est significatif ($t=3,891$ $p<0,001$) mais inférieur au seuil de perception auditif des variations de durée (cf. section C.2 du chapitre II). Il est cependant intéressant à noter puisque l'on ne dispose, à notre connaissance, d'aucune information sur le seuil de perception visuelle de la variation de durée. On notera également qu'il s'agit peut-être d'une conséquence de l'emploi de la parole délexicalisée

puisque pour la parole réelle l'allongement de la syllabe pré-focale est supérieur au seuil de perception auditive de 20% (cf. section C.2 du chapitre II).

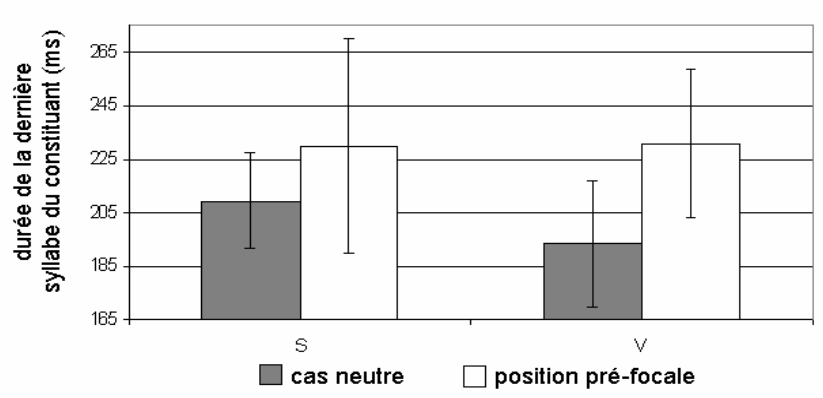


FIGURE III.4 – *Durée de la dernière syllabe d'un syntagme dans le cas neutre et dans le cas où le syntagme suivant est focalisé (parole délexicalisée, locuteur A).*

A.2.2.2.d. Mesures articulatoires

Les paramètres décrivant la forme et la protrusion des lèvres ainsi que la position du point sur le menton ont été extraits grâce à la méthode exposée dans la section A.2.1.1.b du présent chapitre. Des problèmes de traitement sont survenus pour deux énoncés pour lesquels il s'est révélé impossible d'extraire les paramètres souhaités. Un total de 46 énoncés a donc été traité correctement. La figure III.5 donne une illustration des paramètres mesurés, lesquels vont être détaillés dans la suite.

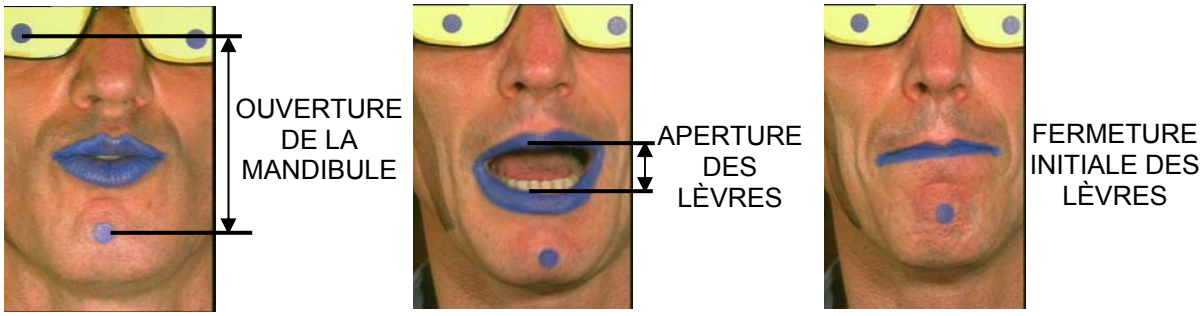


FIGURE III.5 – *Mesures articulatoires effectuées : (gauche) ouverture de la mandibule ; (milieu) aperture des lèvres ; (droite) fermeture initiale des lèvres.*

A.2.2.2.d.i. Mesure du geste d'ouverture de la mandibule

Le geste d'ouverture de la bouche a été analysé par l'intermédiaire de l'ouverture et de la vitesse de mouvement de la mâchoire inférieure ou mandibule. L'ouverture de la mandibule a été estimée en faisant la différence des ordonnées du point situé sur le menton du locuteur et du point sur les lunettes (référence) : voir figure III.5. Les mouvements obtenus sont ainsi les mouvements effectifs du point sur le menton par rapport au reste du visage. Nous sommes tout à fait conscients du fait que prendre le point sur le menton pour étudier les mouvements de la mandibule n'est pas rigoureusement correct.

En effet, il a été montré que les mouvements du menton ne reflètent pas toujours adéquatement les mouvements de la mandibule. Dans la séquence /iba/ par exemple, la mandibule commence à s'abaisser pendant que la lèvre inférieure monte pour former le /b/, entraînant ainsi le menton vers le haut (Badin, communication personnelle). Cependant, aucune méthode plus efficace n'a pour l'instant été mise au point dans le contexte du système labio-métrique utilisé ici. C'est pourquoi, bien qu'émettant cette réserve, nous utiliserons ensuite le terme de mouvements de la mandibule pour décrire les mouvements du menton. Un exemple de signal d'ouverture de la mandibule ainsi obtenu est donné figure III.6 (plus la valeur est importante, plus l'ouverture de la mandibule est importante).

Notre prédiction est que l'amplitude des mouvements de la mandibule sera plus importante pour les syllabes focalisées. Cependant, il est possible que les syllabes focales ne correspondent pas toujours à une amplitude plus importante du mouvement. En effet, lorsque le rythme d'élocution augmente, indépendamment du degré de focalisation, la durée allouée à l'ouverture de la mandibule décroît et il se peut que l'amplitude du mouvement soit ainsi réduite. En fait, comme il a été expliqué plus haut, il apparaît souvent que les pics de vitesse sont aussi des corrélats articulatoires importants. Stone [1981] qu'ils soient de meilleurs corrélats que l'amplitude des gestes.

Nous avons donc calculé la dérivée première (approximation par calcul du développement limité d'ordre 2) du signal d'ouverture de la mandibule au cours du temps. Le quatrième graphique de la figure III.6 donne un exemple de tracé de la vitesse du mouvement de la mandibule (une grande valeur absolue correspond aux ouvertures et fermetures les plus rapides). Dans cette étude, nous nous sommes particulièrement intéressés aux maxima d'ouverture (extrema positifs) du [m] vers le [a]. Nous ferons dans la suite référence à ces maxima par : maxima de vitesse d'ouverture de la mandibule. Notre prédiction est que les syllabes focalisées seront sans doute marquées par des maxima de vitesse d'ouverture de la mandibule plus importants.

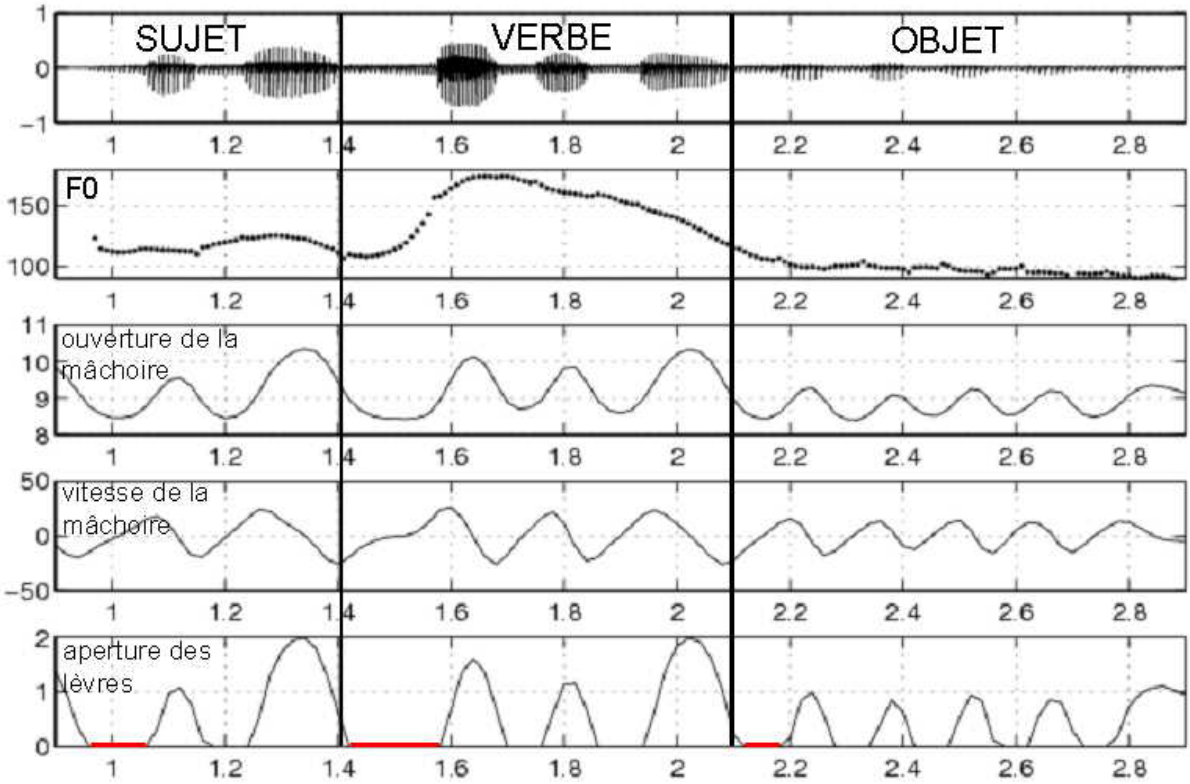


FIGURE III.6 – *a. (premier graphique en partant du haut) : signal acoustique ; b. (deuxième) suivi de F0 en Hz ; c. (troisième) ouverture de la mâchoire inférieure (cm) ; d. (quatrième) vitesse de mouvement de la mâchoire inférieure (cm/s) ; e. (cinquième) aperture des lèvres (cm). L'énoncé prononcé était {[mama] [MAMAMA] [ma mama mama.]}.*

A.2.2.2.d.ii. Mesure du geste de fermeture des lèvres

Le signal vidéo enregistré a été examiné attentivement avant de commencer les mesures, ce qui a permis de constater que la focalisation était souvent marquée en son début par une fermeture extrêmement longue des lèvres. C'est pourquoi nous avons également décidé de mesurer ce paramètre. Le geste d'ouverture de la bouche a été analysé grâce au paramètre d'aperture labiale (distance entre les lèvres supérieure et inférieure, voir figure III.5). Le dernier graphique de la figure III.6 donne un exemple de signal d'aperture des lèvres. Nous avons étudié la durée des plateaux de fermeture correspondant au premier [m] de chaque syntagme et représentée par des traits rouges en gras sur le dernier graphique de la figure III.6. Une fermeture correspond en effet à une valeur nulle de l'aperture des lèvres.

A.2.2.2.e. Résultats de l'analyse articulatoire

A.2.2.2.e.i. Ouverture de la mandibule

Les pics (maxima) d'ouverture et de vitesse d'ouverture de la mandibule ont été relevés pour chaque syllabe (/ma/) du corpus. Une moyenne de ces pics a ensuite été calculée pour chaque syntagme de chaque énoncé. Enfin une moyenne a été calculée sur tous les syntagmes de natures identiques (par exemple tous les S) pour un type de focalisation (par exemple FS). Les résultats de ces calculs sont exposés figure III.7 (graphique a. pour l'ouverture et graphique b. pour la vitesse).

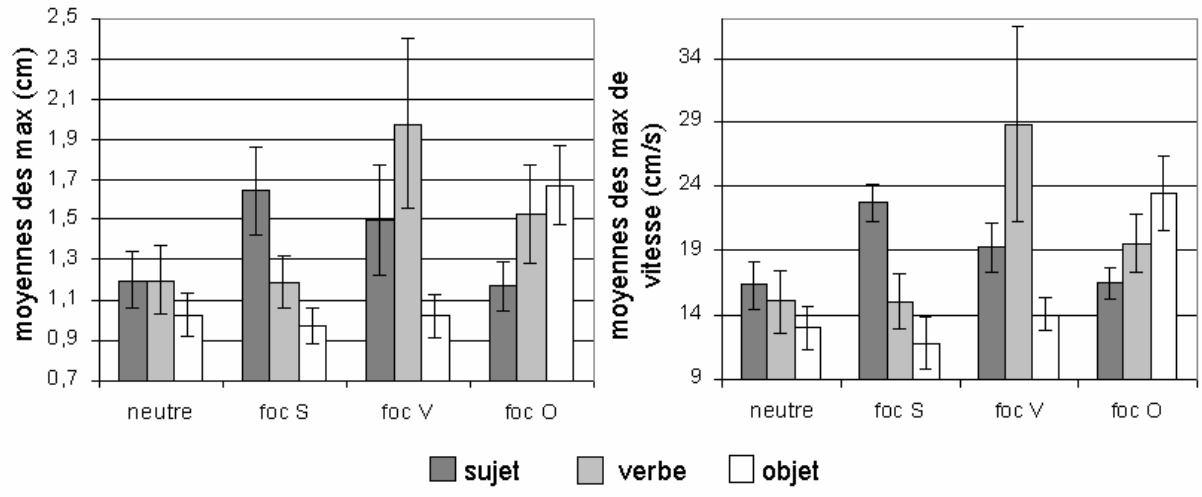


FIGURE III.7 – a. (gauche) moyennes des pics d'amplitude de chaque syllabe sur chaque syntagme puis pour chaque type de focalisation (en cm) ; b. (droite) moyennes des pics de vitesse de chaque syllabe puis pour chaque type de focalisation (en cm/s).

mesure	effet congruence	effet type de focalisation	interaction	test t congruence	test t post-foc	test t pré-foc

ouverture de la mandibule	F(1,12)=149,781 p<0,001	F(2,24)=8,068 p=0,002	F(2,24)=1,417 p=0,262	t=11,304 p<0,001	t=-0,743 p=0,46	t=4,278 p<0,001
pics de vitesse de la mandibule	F(1,11)=251,068 p<0,001	F(2,22)=13,778 p<0,001	F(1,423,22)=4 ,977 p=0,03	t=11,382 p<0,001	t=-0,177 p=0,86	t=5,212 p<0,001

TABLE III.2 – *Résultats des tests statistiques menés sur les données d'ouverture et de vitesse de la mandibule selon la méthode décrite section A.2.1.2.d du chapitre III.*

On constate donc que le locuteur A ouvre plus la mandibule sur l'élément focalisé que sur le reste de l'énoncé et ceci d'en moyenne 44,2% (51,9% pour FS, 57,1% pour FV et 23,7% pour FO). Les contrastes intra-énoncés sont significatifs (facteur congruence [51] de l'ANOVA [52], cf.table III.2 : F(1,12)=149,781 p<0,001). On constate de plus un effet significatif du type de focalisation[53] (cf. table III.2) : F(2,24)=8,068 (p=0,002). L'amplitude de l'ouverture de la mandibule est en effet plus importante lorsque la focalisation porte sur le verbe ou sur l'objet et ce pour tous les syntagmes (contraste de l'ANOVA entre FV&FO et FS : p=0,001). Bien que l'interaction[54] ne soit pas significative (table III.2 : F(2,24)=1,417 p=0,262), on constate que c'est lorsque la focalisation porte sur le verbe que le contraste entre ce qui est focalisé et ce qui ne l'est pas dans un énoncé est le plus important.

L'ouverture de la mandibule augmente significativement par rapport au cas neutre quand un syntagme est focalisé : t=11,304 (p<0,001). Ce test statistique compare les barres S&FS, V&FV et O&FO avec les barres S&neutre, V&neutre et O&neutre. Cette augmentation est en moyenne de 54,8%, la plus forte augmentation étant notée lorsque c'est le verbe qui est focalisé (36,9% pour FS, 64,6% pour FV et 63% pour FO). On n'observe aucune variation significative de l'ouverture de la mandibule sur les syntagmes post-focaux : t=-0,743 (p=0,46). Par contre, il y a augmentation significative de l'ouverture de la mandibule sur les syntagmes pré-focaux par rapport au cas neutre : t=4,278 (p<0,001). Cette augmentation est d'en moyenne de 16,6% (24,7% pour S&FV, -2,2% pour S&FO et 27,4% pour V&FO). En ce qui concerne cette augmentation pré-focale, on constate que c'est surtout l'amplitude des mouvements de la mandibule de l'élément directement pré-focal (S&FV et V&FO) qui augmente. Nous avons donc fait des tests t séparés pour les données directement pré-focales (S&FV et V&FO) et les données non directement pré-focales (S&FO). Ces tests ont montré que l'augmentation de l'ouverture de la mandibule n'était significative que pour les éléments directement pré-focaux (pour S&FV et V&FO : t=5,835 p<0,001 et pour S&FO : t=-0,549 p=0,588).

Comme il a été suggéré précédemment, il est possible que les pics de vitesse du mouvement de la mandibule soit un meilleur corrélat que les mouvements de la mandibule eux-mêmes. Ainsi constate-t-on grâce au graphique b. de la figure III.7 que globalement, le locuteur A ouvre ou ferme plus vite la mandibule sur l'élément focalisé par rapport au reste de l'énoncé et ce en moyenne de 57,5% (69,1% pour FS, 73,3% pour FV et 30,3% pour FO). Ce contraste intra-énoncé est significatif (facteur congruence[55] de l'ANOVA[56], voir table III.2 : F(1,11)=251,068 p<0,001). On constate également un

[51] congruence : facteur intra-sujet à deux niveaux : cas congruents (S&FS, V&FV et O&FO) et cas incongruents (V&FS et O&FS, S&FV et O&FV et S&FO et V&FO).
[52] ANOVA à deux facteurs intra-sujets : congruence (deux niveaux) et type de focalisation (trois niveaux) ; pour une description complète et détaillée du test voir la section A.2.1.2.d du chapitre III.
[53] Type de focalisation : facteur intra-sujet à trois niveaux : FS, FV et FO.
[54] Interaction congruence×type de focalisation
[55] congruence : facteur intra-sujet à deux niveaux : cas congruents (S&FS, V&FV et O&FO) et cas incongruents (V&FS et O&FS, S&FV et O&FV et S&FO et V&FO).

effet significatif du type de focalisation[57] : F(2,22)=13,778 (p<0,001). Ceci est dû au fait que, les pics de vitesse sont plus élevés lorsque la focalisation porte sur le verbe ou sur l'objet (contraste significatif à p<0,001). Enfin on trouve que l'interaction[58] a un effet faiblement significatif : F(1,423,22)=4,977 (p=0,03). Ceci est dû au fait que le contraste intra-énoncé est significativement plus important lorsque c'est le verbe qui est focalisé (contraste significatif à p=0,01).

On note aussi une augmentation significative des pics de vitesse par rapport au cas neutre lorsqu'il y a focalisation (t=11,382 p<0,001). Cette augmentation est d'en moyenne 71,7%, la plus forte augmentation étant notée lorsque la focalisation porte sur le verbe (40% pour FS, 90,2% pour FV et 85% pour FO). Il semble y avoir une faible tendance à la baisse de l'amplitude des pics de vitesse sur les éléments post-focaux (-0,2%) mais celle-ci n'est pas significative (t=-0,177 p=0,86). Enfin, on constate une augmentation significative des pics de vitesse des éléments pré-focaux (t=5,212 p<0,001). Celle-ci est d'en moyenne 17,1% (17,3% pour S&FV, 2,5% pour S&FO et 31,5% pour V&FO). Nous avons mené des tests t séparés pour les données directement pré-focales (S&FV et V&FO) et les données non directement pré-focales (S&FO). Ces tests ont montré que l'augmentation de l'ouverture de la mandibule n'était significative que pour les éléments directement pré-focaux (S&FV et V&FO : t=6,622 p<0,001 et pour S&FO : t=-0,051 p=0,96).

On note donc une augmentation de l'amplitude et des pics de vitesse des mouvements de la mandibule lorsqu'il y a focalisation et ce à la fois par rapport au reste de l'énoncé focalisé et par rapport au cas neutre. On constate néanmoins que les écarts types pour le syntagme verbal focalisé sont plus importants que les autres, un résultat qui avait été observé et commenté pour les mesures de durée.

Aussi bien en ce qui concerne l'amplitude des mouvements de la mandibule que leurs pics de vitesse, on observe également une augmentation significative sur l'élément directement pré-focal. Une analyse détaillée des signaux a même permis de constater que pour certains énoncés, l'amplitude des mouvements de la mandibule était plus importante sur la syllabe qui précède le syntagme focalisé que sur les syllabes du syntagme focalisé lui-même. La figure III.8 illustre ce phénomène. On y observe en effet les signaux acoustique et articulatoires correspondant à un énoncé dont l'objet est focalisé pour lequel la syllabe pré-focale (la dernière syllabe du verbe) est celle pour laquelle l'ouverture de la mandibule est la plus importante de tout l'énoncé. Deux explications peuvent être suggérées à ce phénomène.

La première est que la syllabe précédant directement l'élément focalisé est aussi la dernière syllabe d'un SA et de ce fait doit probablement arborer les corrélats articulatoires de l'accent primaire (H*). Il a en effet été suggéré par Lœvenbruck [1999] que ces corrélats incluent des mouvements articulatoires plus amples (au niveau de la langue et de la mandibule), un grand pic de vitesse et une durée plus longue. Il pourrait donc s'agir tout simplement d'une *marque de l'accent primaire*.

A la suite de l'analyse des mesures de durées effectuées pour la syllabe pré-focale et présentées à la section C.3.3.3 du chapitre II, on avait conclu à un allongement de la syllabe pré-focale. C'est de cette constatation que la seconde explication découle. Celle-ci est que le locuteur ralentit tout simplement, juste avant la focalisation. Ce faisant, il fournit plus de temps à la mandibule et à la

[56] ANOVA à deux facteurs intra-sujets : congruence (2 niveaux) et type de focalisation (3 niveaux) ; pour une description complète et détaillée du test voir la section A.2.1.2.d du chapitre III.
[57] Type de focalisation : facteur intra-sujet à 3 niveaux : FS, FV et FO.
[58] Interaction congruence×type de focalisation

bouche pour s'ouvrir et leur permet donc de s'ouvrir plus. Il s'agirait alors d'une *stratégie d'anticipation de la focalisation*. Cette interprétation est soutenue par le fait que lorsque le pic d'amplitude d'ouverture de la mandibule est plus important sur la syllabe pré-focale que sur les syllabes focales, il n'en va pas forcément de même pour le pic de vitesse. Or les corrélats articulatoires de l'accent primaire comprennent une augmentation du pic de vitesse aussi bien que de l'amplitude du geste (Lœvenbruck [1999]). La stratégie d'anticipation serait ainsi principalement fondée sur la durée et la plus grande ouverture de la mandibule serait simplement une conséquence de l'allongement.

L'hypothèse de la *marque de l'accent primaire* est de plus infirmée par le fait que l'augmentation de l'amplitude du geste mandibulaire est plus importante sur les syllabes pré-focales que sur les autres syllabes finales de SA qui pourtant portent de la même façon l'accent primaire. La figure III.8 montre ainsi que quand l'objet est focalisé, on observe une grande ouverture de la mandibule à la fois sur la dernière syllabe du sujet et sur la dernière syllabe du verbe. Néanmoins, l'augmentation d'amplitude est beaucoup plus importante sur la dernière syllabe du verbe (syllabe pré-focale). Si la première explication était la bonne, on devrait observer à peu près la même augmentation d'amplitude pour les dernières syllabes du sujet et du verbe.

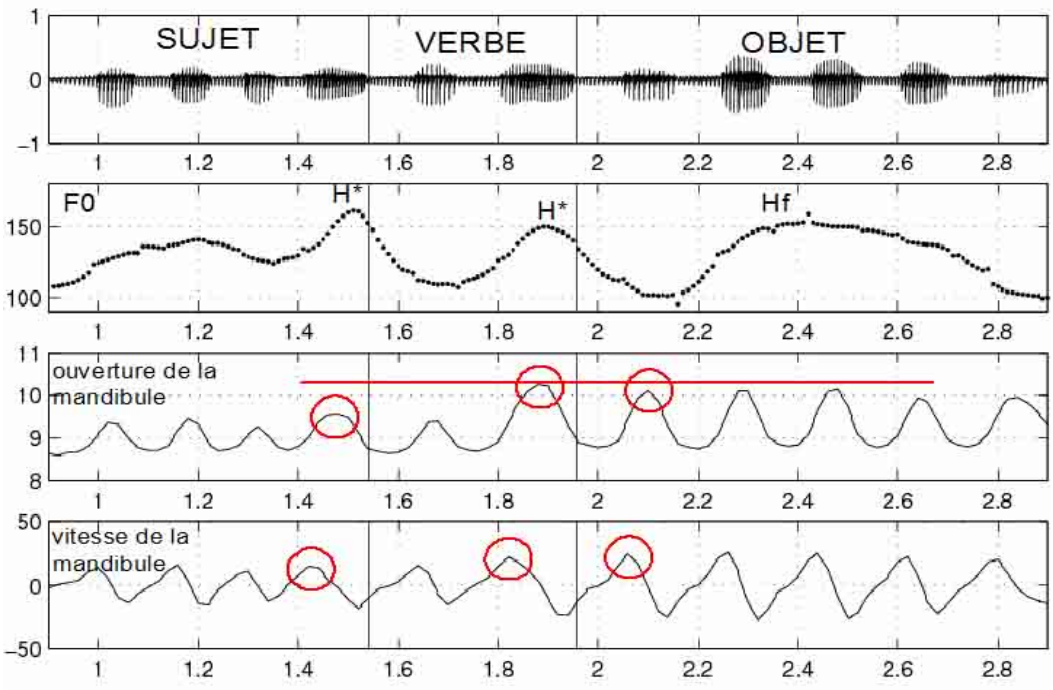


FIGURE III.8 – *a. (haut) signal acoustique, b. (deuxième en partant du haut) suivi de F0 en Hz, c. (troisième) ouverture (cm) et d. (quatrième) vitesse (cm/s) de la mandibule en fonction du temps (s) : illustration du phénomène d'anticipation pré-focale pour un énoncé où c'est l'objet qui est focalisé. L'énoncé prononcé était [mamamama][mama][MA MAMA MAMA.].*

A.2.2.2.e.ii. Durée de fermeture des lèvres

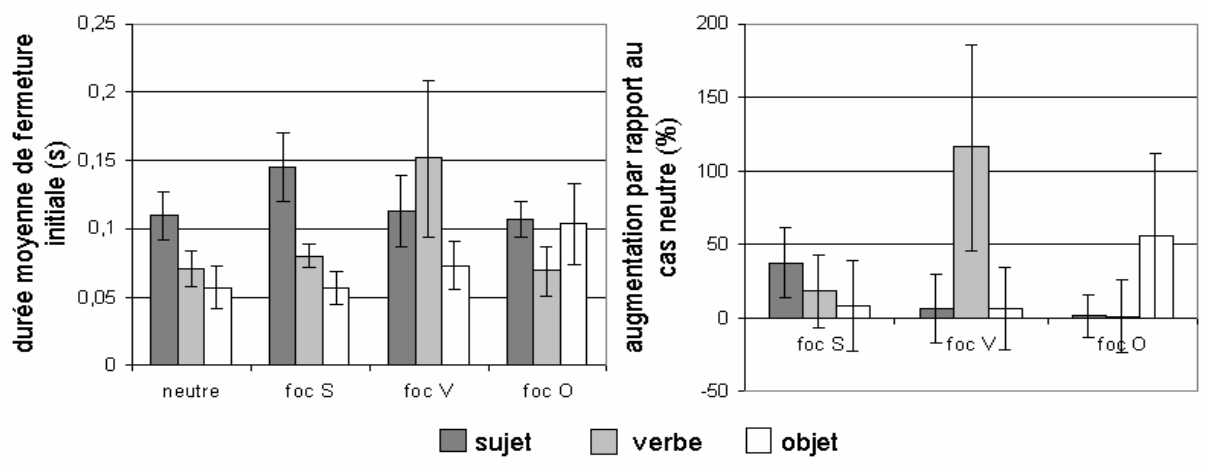


FIGURE III.9 – *a. (gauche) durées moyennes des fermetures initiales pour chaque type de syntagme sous chaque type de focalisation (en s) ; b. (droite) pourcentage d'augmentation de la durée moyenne de fermeture initiale par rapport au cas neutre (en %).*

mesure	effet congruence	effet type de focalisation	interaction	test t congruence	test t post-foc	test t pré-foc
durées des fermetures initiales	F(1,12)=86,688 p<0,001	F(2,24)=5,52 p=0,011	F(2,24)=6,107 p=0,007	t=6,729 p<0,001	t=-0,743 p=0,979	t=2,339 p=0,022

TABLE III.3 – *Résultats des tests statistiques menés sur les données de durée de la fermeture initiale (début de syntagme) selon la méthode décrite section A.2.1.2.d du chapitre III.*

Le graphique a. de la figure III.9 montre que lorsqu'un syntagme est focalisé, la durée de fermeture initiale du syntagme (/m/ du premier /ma/) en question est allongée par rapport aux durées de ces mêmes fermetures initiales pour les autres syntagmes de l'énoncé. Ce contraste est clair lorsque c'est le sujet ou le verbe qui est focalisé (112,2% pour FS et 62,5% pour FV) mais moins lorsqu'il s'agit de l'objet (17,5%). On constate sur la figure III.9 qu'en fait si le contraste est plus faible pour l'objet, c'est surtout parce que les durées initiales sont assez importantes pour le sujet non focalisé. Or on constate que, dans le cas neutre déjà, les durées de fermetures initiales sont plus importantes pour le sujet. Ceci est probablement dû au fait qu'au début de l'énoncé, il doit exister un allongement de la fermeture initiale inhérent qui marque ainsi le début de l'énoncé.

Les contrastes intra-énoncés sont néanmoins globalement significatifs (facteur congruence[59] de l'ANOVA[60], cf. table III.3 : F(1,12)=86,688 p<0,001). Le type de focalisation[61] a lui aussi un effet significatif : F(2,24)=5,520 (p=0,011) puisque l'allongement de la durée initiale de fermeture des lèvres est globalement plus important lorsque la focalisation porte sur le verbe (contraste de l'ANOVA significatif à p=0,016). Enfin, on constate une interaction[62] significative : F(2,24)=6,107 (p=0,007) et ce car le contraste intra-énoncé est significativement plus important lorsque c'est le verbe qui est focalisé (contraste de l'ANOVA significatif à p=0,007).

[59] congruence : facteur intra-sujet à deux niveaux : cas congruents (S&FS, V&FV et O&FO) et cas incongruents (V&FS et O&FS, S&FV et O&FV et S&FO et V&FO).
[60] ANOVA à deux facteurs intra-sujets : congruence (deux niveaux) et type de focalisation (trois niveaux) ; pour une description complète et détaillée du test voir la section A.2.1.2.d du chapitre III.
[61] Type de focalisation : facteur intra-sujet à trois niveaux : FS, FV et FO.
[62] Interaction congruence×type de focalisation

Le graphique b. de la figure III.9 permet de constater qu'il y a une nette augmentation de la durée de fermeture initiale par rapport au cas neutre lorsqu'il y a focalisation. Cette augmentation est en moyenne de 69,8% (37,5% pour S&FS, 116,1% pour V&FV et 55,8% pour O&FO) et elle est significative (t=6,729 p<0,001). En ce qui concerne les durées de fermetures initiales des éléments pré-focaux, on ne constate aucune variation significative (t=-0,743 p=0,979). On observe cependant un allongement de 9,9% de la durée de fermeture initiale des éléments post-focaux et cet allongement est significatif (t=2,339 p=0,022). Cet allongement est faible par rapport à celui du constituant focalisé. Il est peut-être dû à une articulation plus soignée de l'ensemble de l'énoncé en présence de focalisation.

A.2.2.2.f. Premiers pas vers un modèle de la production visuelle de la focalisation contrastive en français

Le but de cette étude sur la parole délexicalisée était, d'une part, de vérifier l'existence de corrélats articulatoires potentiellement visibles de la focalisation constrastive en français et, d'autre part, de déterminer la nature de ces corrélats. Les résultats de cette étude ont donc montré qu'il existe des corrélats articulatoires qui pourraient être visibles par l'auditeur/spectateur. D'après ces résultats, on peut déjà esquisser un modèle préliminaire des corrélats visibles de la production de la focalisation contrastive en français, pour ce locuteur.

La focalisation contrastive serait ainsi soulignée visiblement par un ***allongement des syllabes focales*** à la fois par rapport au reste de l'énoncé focalisé (en moyenne +28,9%) et par rapport au cas neutre (en moyenne +30,8%). C'est lorsque c'est le verbe qui est focalisé que cet allongement est le plus important, à la fois par rapport au reste de l'énoncé focalisé (contraste intra-énoncé) et par rapport au cas neutre. Néanmoins, comme il a été expliqué précédemment, ceci est certainement dû au fait que notre corpus contenait deux phrases pour lesquelles les verbes sont monosyllabiques. Il est donc probable que le locuteur ne disposant que d'une syllabe pour signaler la focalisation, les indices soient renforcés. C'est lorsque la focalisation porte sur l'objet que l'allongement est le moins important, à la fois par rapport au reste de l'énoncé et par rapport au cas neutre. Ceci est peut-être dû au fait qu'en moyenne, dans ce corpus, les constituants objets sont les constituants les plus longs (4,25 syllabes contre 2,9 pour les sujets et 2,6 pour les verbes). On peut en effet penser que plus il y a de syllabes dans le constituant à focaliser moins l'allongement sera prononcé : l'effort d'allongement est peut-être réparti entre toutes les syllabes et donc « plus » réparti lors des focalisations sur l'objet. Ceci sera à confirmer dans les analyses suivantes.

La focalisation contrastive s'accompagne de plus d'une ***hyper-articulation focale*** *i.e.* de mouvements de la mandibule plus amples et plus rapides et ce pour toutes les syllabes focales. Cette hyper-articulation se fait à la fois par rapport au reste de l'énoncé focalisé (en moyenne +44,2% pour l'ouverture elle-même et +57,5% pour la vitesse d'ouverture) et par rapport au cas neutre (en moyenne +54,8% pour l'ouverture et +71,7% pour la vitesse). Les contrastes et augmentations par rapport au cas neutre sont en général les plus importants lorsque la focalisation porte sur le verbe et ce sûrement pour la raison évoquée ci-dessus. Par contre, on note que l'augmentation articulatoire par rapport au cas neutre est en général plus importante lorsque la focalisation porte sur l'objet que lorsqu'elle porte sur le sujet. Cependant, les contrastes intra-énoncés sont plus importants lorsque c'est le sujet qui est focalisé. Ceci est dû au fait que lorsque c'est le sujet qui est focalisé il semble qu'il y ait une hypo-articulation de l'objet post-focal (les barres O&FS des graphiques de la figure III.7

sont en dessous de 1). Donc, bien que l'augmentation par rapport au cas neutre soit plus importante lorsque c'est l'objet qui est focalisé, le contraste est renforcé dans le cas de la focalisation sur le sujet par cette hypo-articulation en fin d'énoncé. Il existe de plus une hyper-articulation pré-focale sur le verbe lorsque la focalisation porte sur l'objet (les barres V&FO des graphiques de la figure III.7 sont au-dessus de 1) ce qui aura tendance à réduire le contraste intra-énoncé dans ce cas-là.

On note également que la focalisation est marquée par une ***fermeture initiale des lèvres allongée en durée*** au début du syntagme focalisé. Cet allongement est le plus net lorsque la focalisation porte sur le verbe, toujours pour la raison évoquée plus haut. C'est lorsque la focalisation porte sur le sujet que le contraste intra-énoncé est le plus important. Ce contraste est particulièrement faible lorsque l'objet est focalisé. En ce qui concerne l'augmentation par rapport au cas neutre, c'est l'inverse (l'objet est le plus focalisé) et cette inversion est due à la même raison que pour l'ouverture et la vitesse de la mandibule (voir paragraphe ci-dessus).

Il semblerait enfin qu'il y ait aussi un allongement pré-focal occasionnant une plus grande ouverture de la mandibule sur la syllabe directement pré-focale (dernière syllabe du syntagme précédant le syntagme focalisé). Nous expliquons ce phénomène par la mise en place d'une ***stratégie d'anticipation de la focalisation***. Cette anticipation est plus nette lorsque c'est l'objet qui est focalisé que lorsque c'est le verbe. Il est possible que lorsque la focalisation porte sur la fin de la phrase, le locuteur dispose de plus de temps pour mettre en place cette anticipation.

A.2.2.3. Étude de la parole lexicalisée

A.2.2.3.a. Problématique de l'étude

L'étude préliminaire décrite ci-dessus a permis de montrer qu'il existait apparemment des corrélats articulatoires visibles de la focalisation contrastive en français. Cependant cette étude a été menée sur la parole délexicalisée avec, comme il a été expliqué, tous les avantages mais aussi les inconvénients qu'elle suppose. Elle a permis de détecter facilement l'existence de ces paramètres et d'explorer leur nature. Cependant il paraît maintenant primordial, au vu des informations recueillies, de faire une étude complémentaire sur de la parole « réelle ». Or, comme il a été exposé à la section A.2.2.1.b du présent chapitre lors de la description de la procédure d'enregistrement, le corpus avait été enregistré également en parole « normale ». C'est donc à présent cette partie du corpus que nous allons étudier.

A.2.2.3.a.i. Le défi de la parole « réelle »

Jusqu'ici on a vu que les corrélats visibles pouvaient être de deux types : mouvements articulatoires et variations de durée. Lors de l'étude préliminaire de la parole délexicalisée, on a vu que les paramètres les plus indicatifs de la présence de focalisation sur un élément donné était l'ouverture ou aperture des lèvres et les mouvements de la mandibule. Ceci était évidemment lié au fait que la seule syllabe étudiée était la syllabe [ma]. Pour l'étude de la parole lexicalisée, ces paramètres devront donc être adaptés puisque l'on sera en présence de syllabes toutes différentes et a fortiori ayant des propriétés articulatoires très différentes. Le défi ici sera donc d'identifier des corrélats qui varient significativement pour toutes les configurations articulatoires présentes dans le corpus. Les paramètres articulatoires varieront en effet non seulement à cause des variations prosodiques mais aussi à cause des différences inter-syllabiques. Le but est d'étudier la parole « réelle » et non de se

limiter à l'étude d'un seul mot, voire « non-mot » comme il est souvent pratiqué dans ce genre d'étude afin de s'affranchir de la variation inter-syllabique inhérente.

A.2.2.3.a.ii. Corrélats articulatoires potentiels

Il s'agit donc à présent de déterminer les paramètres articulatoires dont les variations seront les plus significatives quelle que soit la syllabe en question. L'étude préliminaire a montré que la focalisation contrastive induisait une hyper-articulation. Or l'hyper-articulation peut être réalisée de manières différentes : pour un /a/ par exemple, ce seront les ouvertures verticales des lèvres et de la mandibule qui seront plus importantes mais pour un /i/, il s'agira plutôt de l'étirement labial horizontal (voir schéma de la figure III.2) et pour un /u/, plutôt de la protrusion labiale. Une variété importante de paramètres articulatoires pourrait donc être analysée et il semblerait qu'il faille considérer différents paramètres pour chaque syllabe. On peut d'ores et déjà éliminer le paramètre de protrusion car le corpus ne comprend que très peu de cas où elle serait affectée par une hyper-articulation (peu de voyelles protruses). Restent donc, l'ouverture (verticale) des lèvres (OL), l'étirement labial horizontal (EL) et les mouvements (verticaux) de la mandibule (MM). Comme il a déjà été exposé dans la section A.2.1.1.b du présent chapitre, nous disposons également d'un autre paramètre pouvant être étudié : l'aire intéro-labiale (AL). Or ce paramètre tient compte à la fois des variations de OL et de EL et serait ainsi plus robuste vis-à-vis des variations articulatoires inter-syllabiques. C'est donc les variations de ce paramètre qui seront étudiées ici ainsi que celles de MM.

A.2.2.3.a.iii. Corrélats de durée potentiels

Les durées de syllabes de natures différentes sont elles aussi intrinsèquement variables, tout comme les paramètres spécifiquement articulatoires, et une procédure de normalisation devra être envisagée afin de pallier ce phénomène et de pouvoir en toute rigueur mener des comparaisons d'une syllabe à l'autre (sur les différences de durée intrinsèque des voyelles du français, voir par exemple Benguerel [1971] ou O'Shaughnessy [1981]). En ce qui concerne les mesures effectuées, nous tenterons de voir si les allongements focal et pré-focal se confirment. Dans l'étude préliminaire, nous avions aussi étudié l'allongement significatif de la fermeture initiale des lèvres pour le premier segment du syntagme focalisé. Or, ce corrélat n'était peut-être qu'un artefact de la syllabe choisie (/ma/). En effet, dans le cadre de la parole lexicalisée, on n'observera pas forcément une fermeture initiale des lèvres pour chaque syntagme (comme pour le /m/ de /ma/). Afin d'étudier tout de même un paramètre se rapprochant, ce phénomène ayant peut-être un équivalent en parole « réelle », il a été décidé de mesurer la durée du premier phonème (consonne d'attaque) du syntagme focalisé qu'on nommera ci-après « durée du premier segment ».

A.2.2.3.b. Validation acoustique

Un total de 64 énoncés a été enregistré en parole lexicalisée. L'un de ces énoncés (phrase (3) avec focalisation sur le sujet en version 1) a dû être éliminé après l'enregistrement à cause d'une hésitation de la part du locuteur qui rendait l'énoncé inexploitable. Après élimination, il restait donc 63 énoncés dont 47 focalisés. Avant de commencer l'analyse articulatoire des données, les productions acoustiques de ces 63 énoncés ont été analysées attentivement selon la méthode décrite dans la section A.2.1.2.b du présent chapitre.

Après mise en commun des résultats des deux validations parallèles (acoustique et perceptive), aucun autre énoncé n'a dû être éliminé. Après validation acoustique, il restait donc toujours 63 énoncés à analyser.

A.2.2.3.c. Mesures et traitement

A.2.2.3.c.i. Durées

Les durées de toutes les syllabes du corpus ont été mesurées grâce à la segmentation acoustique du corpus en syllabes selon la méthode décrite à la section C.1.4.1 du chapitre II.

A.2.2.3.c.ii. Données articulatoires

On ne peut malheureusement pas ici employer la même technique que celle utilisée pour l'étude en parole délexicalisée, c'est-à-dire la détection automatique pour chaque paramètre articulatoire, des maxima sur chaque syllabe. On n'observe en effet pas systématiquement un pic d'aire intéro-labiale (resp. de vitesse de variation du paramètre en question) par syllabe comme c'était le cas pour l'étude préliminaire (un pic d'ouverture de la bouche et de vitesse du geste par syllabe /ma/). Les mouvements articulatoires en parole lexicalisée sont en effet co-articulés et on observe ainsi parfois un seul maximum d'aire aux lèvres (resp. de vitesse de variation du paramètre en question) sur deux voire trois syllabes (cf. par exemple sur les effets de la coarticulation Ostry *et al.* [1996]). C'est ici finalement l'amplitude moyenne des gestes articulatoires qui est intéressante : y a-t-il plus de mouvement des lèvres quand il y a focalisation ? Pour obtenir une mesure représentant au mieux cette quantité, nous avons estimé l'aire sous la courbe d'évolution temporelle de l'aire intéro-labiale. Cette aire a été estimée grâce à la méthode des trapèzes. Elle a été calculée sur chacun des syntagmes (S, V et O) comme il est illustré dans la figure III.10 et ce pour chacun des énoncés. Les valeurs d'aires (une par syntagme) ont ensuite été normalisées par la durée (division de la valeur d'aire obtenue par la durée du syntagme pour lequel elle a été mesurée). Cette normalisation a été effectuée pour ne récupérer que l'amplitude moyenne des mouvements. En effet, lorsqu'un syntagme est focalisé, on a vu que sa durée augmentait dans la grande majorité des cas. Or l'augmentation de cette durée provoque inévitablement une augmentation de l'aire sous la courbe considérée même si l'amplitude du signal en question ne varie pas. C'est justement cette variation d'amplitude qui nous intéresse et non celle de la durée, déjà mesurée par ailleurs. On obtient ainsi une valeur d'aire normalisée pour chaque syntagme de chaque énoncé (trois valeurs par énoncé). Les pics de vitesse de variation de l'aire intéro-labiale ont quant à eux été vérifiés à la main sur chaque syllabe afin de ne considérer que des maxima absolus réels (plusieurs maxima relatifs existent parfois sur une seule syllabe et une procédure de détection automatique ne fonctionne pas idéalement).

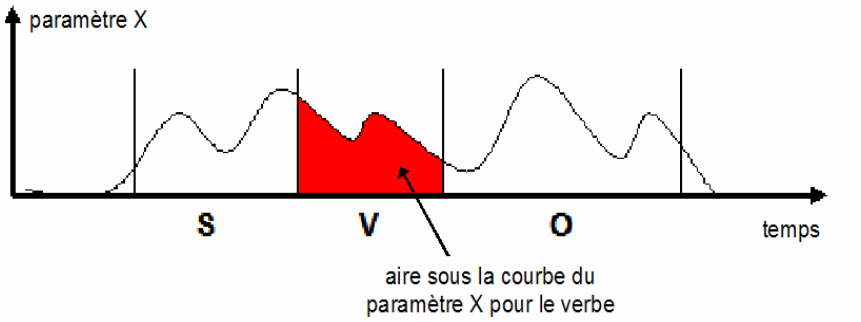


FIGURE III.10 – *Schéma illustrant la notion d'aire sous la courbe pour un paramétre X quelconque.*

A.2.2.3.c.iii. Problématique de comparaison : normalisation des données

Les comparaisons entre amplitudes des gestes articulatoires d'une syllabe à l'autre ne pouvaient pas être faites directement après mesure de l'aire sous la courbe correspondant à ces gestes. Cette aire est en effet fortement conditionnée par la ou les syllabes en question (l'amplitude moyenne de l'ouverture des lèvres pour /my/ par exemple est intrinsèquement bien plus faible que pour /ma/). Il en va de même pour la vitesse de variation (dérivée première). Une normalisation a par conséquent été effectuée. Celle-ci consistait à calculer tout d'abord, pour les différents types de données articulatoires, la moyenne des deux cas neutres pour chaque syntagme de chaque phrase du corpus, afin d'obtenir une valeur de référence par syntagme. Ensuite, pour chaque syntagme, dans chaque condition de focalisation possible, toutes les valeurs d'aires sous la courbe mesurées ont été divisées par la valeur de référence. Ceci revient en quelque sorte à calculer une variation par rapport au même geste dans le cas neutre.

En ce qui concerne les données sur la vitesse, une fois cette normalisation effectuée, une moyenne des pics de vitesse a été calculée pour chaque syntagme (il pouvait en effet y avoir plusieurs maxima par syntagme comme expliqué plus haut, or on veut obtenir une valeur par syntagme).

A la suite de ces traitements, il restait donc aussi bien pour l'aire intéro-labiale que pour sa vitesse, une seule valeur par syntagme de chaque énoncé soit trois valeurs par énoncé.

A.2.2.3.d. Analyse de la durée

A.2.2.3.d.i. Syllabes focales

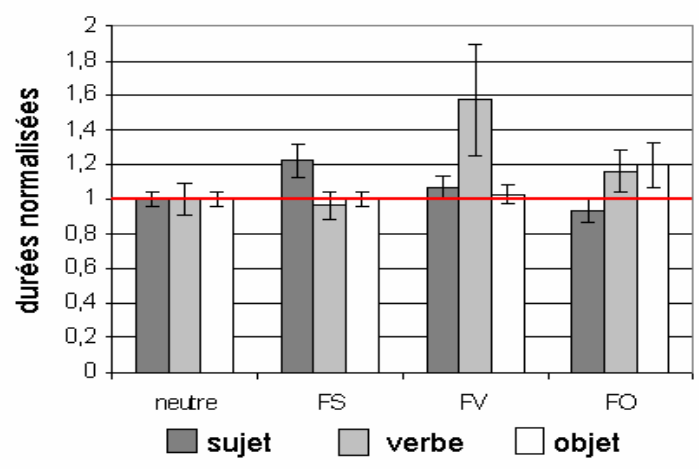


FIGURE III.11 - *Moyennes des durées normalisées des syllabes de chaque type de syntagme et pour chaque type de focalisation (le cas neutre correspond à la valeur 1).*

mesure	effet congruence	effet type de focalisation	interaction	test t congruence	test t post-foc	test t pré-foc	
durées	F(1,15)=158,922 p<0,001	F(1,496,30)=19,214 p<0,001	F(1,447,30)=13,603 p<0,001	t=8,583 p<0,001	t=-0,346 p=0,731	t=3,027 p=0,004	t=-3,205 p=0,006
							t=6,123 p<0,001

TABLE III.4 – *Résultats des tests statistiques menés sur les données de durées (parole lexicalisée) selon la méthode décrite à la section A.2.1.2.d du chapitre III.*

Les durées de toutes les syllabes du corpus ont été mesurées grâce à la segmentation acoustique effectuée selon la méthode décrite à la section C.1.4.1 du chapitre II. Le graphique de la figure III.11 présente les résultats de ces mesures[63].

De façon générale, la figure III.11 montre que la durée moyenne d'une syllabe est plus importante si celle-ci appartient au syntagme focalisé que si elle appartient à un autre syntagme du même énoncé. Le contraste intra-énoncé est en moyenne de 29,8% (24,6% pour FS, 50,3% pour FV et 14,4% pour FO) et il est significatif (facteur congruence [64] de l'ANOVA [65], cf. table III.4) : $F(1,15)=158,922$ ($p<0,001$). Le type de focalisation [66] a lui aussi un effet significatif : $F(1,496,30)=19,214$ ($p<0,001$) car les durées des syllabes de tout l'énoncé sont plus importantes lorsque la focalisation porte sur le verbe (contraste de l'ANOVA significatif à $p=0,001$). L'interaction[67] est aussi significative : $F(1,447,30)=13,603$ ($p<0,001$) et ce parce qu'on note des contrastes intra-énoncés significativement plus importants lorsque c'est le verbe qui est focalisé (contraste de l'ANOVA significatif à $p=0,001$).

Notre corpus comportant des constituants monosyllabiques, une explication possible de l'effet de la focalisation verbale sur la durée pourrait être liée à l'influence du nombre de syllabes comme il a été exposé plus haut. Pour examiner cette hypothèse plus en détail, j'ai cherché à évaluer l'influence du nombre de syllabes sur la durée. Ainsi, le graphique de la figure III.12 donne-t-il les valeurs des durées normalisées des constituants focaux (en ordonnée) en fonction à la fois du type de focalisation (courbes différentes) et du nombre de syllabes (en abscisse). On constate grâce à cette figure que lorsque le verbe est monosyllabique et qu'il est focalisé sa durée augmente beaucoup plus que quand il y a plus de syllabes ou quand il s'agit d'un autre type de focalisation. Comme suggéré plus haut, cette différence est probablement lié à une réorganisation prosodique de l'énoncé et donc à l'influence combinée de l'allongement focal et de l'allongement final de SA.

[63] Même principe de mesure, de calcul et de représentation que dans la section A.2.2.2.d du chapitre III.
[64] congruence : facteur intra-sujet à deux niveaux : cas congruents (S&FS, V&FV et O&FO) et cas incongruents (S&FV et S&FO, V&FS et V&FO et O&FS et O&FV).
[65] ANOVA à deux facteurs intra-sujets : congruence (deux niveaux) et type de focalisation (trois niveaux) ; pour une explication complète et détaillée de l'ANOVA voir la section A.2.1.2.d du chapitre III.
[66] type de focalisation : facteur intra-sujet à trois niveaux : FS, FV et FO.
[67] Interaction congruence×type de focalisation

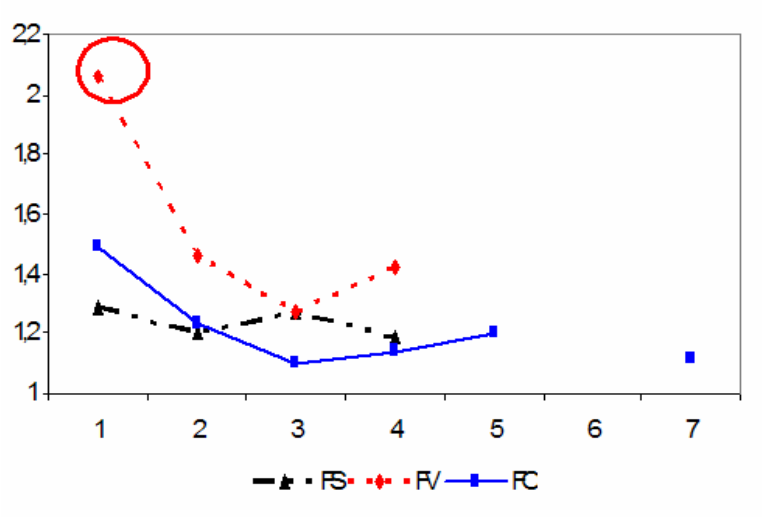


FIGURE III.12 – *Moyennes des durées normalisées des syntagmes focalisés pour chaque type de focalisation (courbes différentes) et en fonction du nombre de syllabes du constituant considéré (en abscisse).*

Le graphique de la figure III.11 permet également de constater que lorsqu'un syntagme est focalisé la durée de ses syllabes augmente par rapport au cas neutre (les barres S&FS, V&FV et O&FO sont toutes au-dessus de 1) et ce d'en moyenne 33,3% (22,2% pour FS, 57,7% pour FV et 20,1% pour FO). C'est le cas pour 100% des syntagmes focalisés. Cet allongement est significatif[68] (t=8,583 p<0,001) et est nettement supérieur au seuil de perception auditif de l'allongement qui est de 20% (cf. section C.2 du chapitre II).

Lorsqu'un syntagme est focalisé, il semblerait que les durées des syllabes de la séquence qui le suit ne soient pas modifiées (les barres V&FS, O&FS et O&FV sont très proches de 1). On ne constate d'ailleurs pas de différence significative par rapport à 1 (t=-0,346 p=0,731).

La figure III.11 permet enfin de constater une augmentation de la durée des syllabes du constituant pré-focal (les barres S&FV et V&FO sont au-dessus de 1) d'en moyenne 11,6% (7% pour S&FV et 16,3% pour V&FO). Cette augmentation est vérifiée uniquement pour le constituant pré-focal puisqu'elle n'est pas observée pour S&FO pour lequel on observe même une diminution de la durée d'en moyenne 6,3%. L'augmentation est significative pour les données concernant le constituant directement pré-focal (test t de comparaison à 1 pour les données directement pré-focales : t=6,123 p<0,001) mais la diminution observée pour S&FO est elle aussi significative[69] : t=-3,205 p=0,006. On observe donc un allongement limité au constituant pré-focal. Suite aux résultats obtenus en parole délexicalisée (cf. section A.2.2.2.c du présent chapitre), il était naturel de raffiner l'étude à la syllabe directement pré-focale.

A.2.2.3.d.ii. Syllabe pré-focale

[68] Test t de comparaison à 1.
[69] Test t de comparaison à 1.

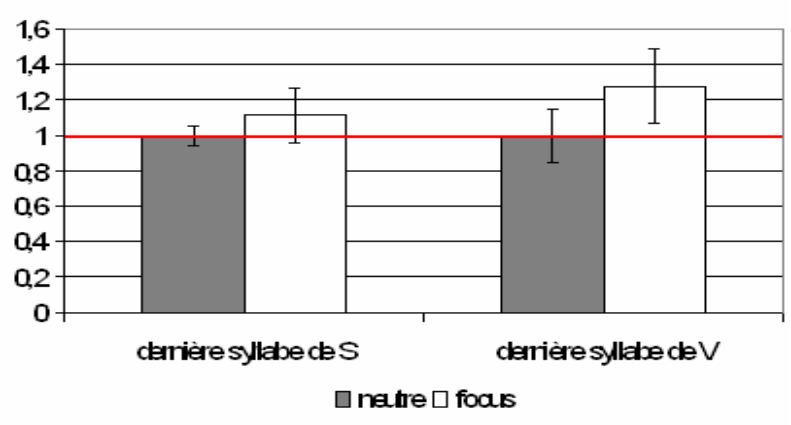


FIGURE III.13 – *Moyennes des durées de la dernière syllabe des constituants S et V lorsque le constituant suivant est ou n'est pas focalisé (syllabe pré-focale).*

La durée de la syllabe pré-focale (dernière syllabe du constituant précédant le constituant focalisé) a donc été analysée de façon plus précise. Les résultats de cette analyse sont consignés dans le graphique de la figure III.13. La figure permet de constater un allongement de la dernière syllabe d'un constituant lorsque le constituant suivant est focalisé. Cet allongement est d'en moyenne 19,3% (10,9% pour la dernière syllabe de S et FV et 27,7% pour la dernière syllabe de V et FO) et est globalement significatif[70] : t=5,472 (p<0,001). Cependant, en raffinant l'analyse statistique on se rend compte que cet allongement n'est significatif que pour la dernière syllabe du syntagme verbal (tests t de comparaison à 1 : S+FV : t=2,884 p=0,11 ; V+FO : t=5,260 p<0,001).

A.2.2.3.d.iii. Premier segment focalisé

En parole délexicalisée, on avait constaté qu'il existait une fermeture marquée à l'initiale du constituant focalisé (cf. section A.2.2.2.e.ii du présent chapitre). Afin de tenter de trouver un équivalent à cette fermeture initiale en parole lexicalisée, nous étudierons maintenant le premier phonème du constituant focalisé ou premier segment. La durée de ce segment a ainsi été mesurée et comparée à celle du même segment dans le cas neutre. Les résultats sont consignés dans le graphique b de la figure III.14. Il semble que lorsqu'un syntagme est focalisé son premier segment est allongé de façon très importante par rapport à la durée de ce même segment dans le cas neutre (en moyenne 53,2% : 35,8% pour le premier segment de S et FS, 61,6% pour le premier segment de V et FV et 62,3% pour le premier segment de O et FO). Cet allongement est à la fois globalement significatif (test t de comparaison à 1 avec toutes les données : t=8,861 p<0,001) et individuellement significatif (test t de comparaison à 1 avec chaque sous-ensemble de données : S : t=3,698 p=0,002, V : t=5,91 p<0,001, O : t=6,036 p<0,001). Il est de plus significativement plus important que l'allongement focal global (test t de comparaison des durées moyennes des syllabes focales et du premier segment du constituant focalisé avec toutes les données : l'hypothèse d'égalité des moyennes doit être rejetée : t=-2,782 p=0,007).

[70] Test t de comparaison à 1.

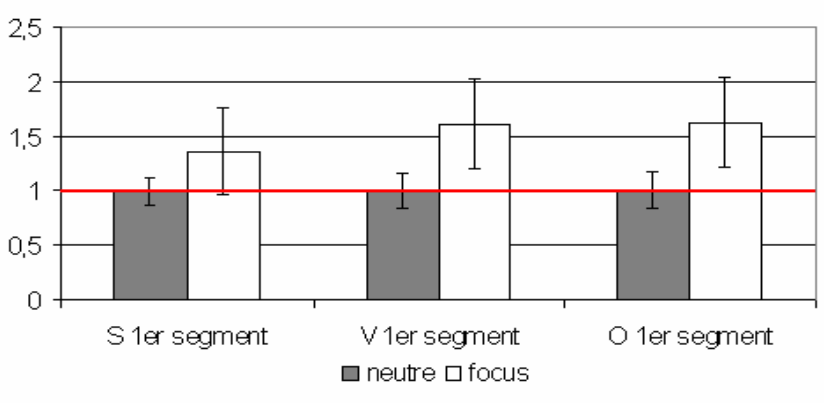


FIGURE III.14 – *Durées des premiers segments (premier phonème) des constituants (S,V et O) dans les cas où le syntagme auquel ils appartiennent est ou non focalisé.*

A.2.2.3.e. Analyse articulatoire

A.2.2.3.e.i. Aire intéro-labiale

Comme il a été expliqué plus haut, l'aire sous la courbe d'aire intéro-labiale, normalisée par la durée, nommée ci-après « amplitude moyenne de l'aire intéro-labiale » a été calculée. Le graphique a. de la figure III.15 permet de constater que, globalement, lorsqu'un syntagme est focalisé, l'amplitude moyenne de l'aire intéro-labiale correspondant à ce syntagme est plus importante que celle correspondant aux autres syntagmes de l'énoncé (les barres correspondant à S&FS, V&FV et O&FO sont plus hautes que les autres).

Le contraste intra-énoncé moyen est de 32,1% (38,6% pour FS, 43,1% pour FV et 14,6% pour FO) et est significatif puisqu'on note un effet significatif du facteur congruence[71] de l'ANOVA[72] (cf. table III.5) : $F(1,15)=202,165$ ($p<0,001$). On note de plus un effet significatif du type de focalisation[73] (cf. table III.5) : $F(2,30)=4,313$ ($p=0,023$) ce qui est dû au fait que l'amplitude moyenne de l'aire intéro-labiale est significativement plus importante lorsque la focalisation porte sur le verbe ou l'objet que lorsqu'elle porte sur le sujet (contraste de l'ANOVA significatif à $p=0,013$). Bien que l'interaction[74] ne soit pas significative ($F(1,241,30)=3,735$ $p=0,061$) on remarquera néanmoins de nouveau que le contraste intra-énoncé est le plus important lorsque le verbe est focalisé.

[71] congruence : facteur intra-sujet à deux niveaux : cas congruents (S&FS, V&FV et O&FO) et cas incongruents (S&FV et S&FO, V&FS et V&FO et O&FS et O&FV).
[72] ANOVA à deux facteurs intra-sujets : congruence (deux niveaux) et type de focalisation (trois niveaux) ; pour une explication complète et détaillée de l'ANOVA voir la section A.2.1.2.d du chapitre III.
[73] Type de focalisation : facteur intra-sujet à trois niveaux : FS, FV et FO.
[74] Interaction congruence×type de focalisation

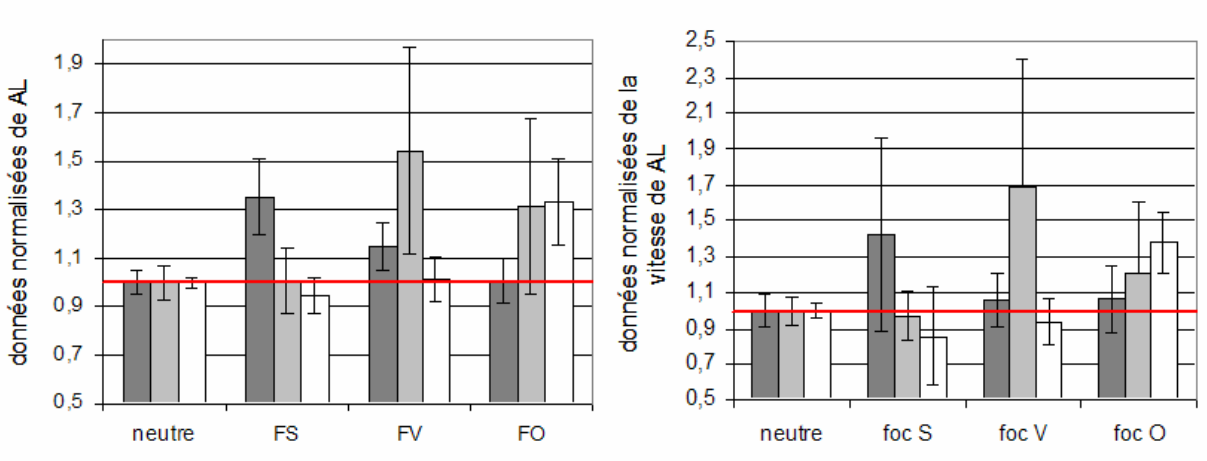


FIGURE III.15 – *a. (gauche) moyennes des valeurs normalisées d'aire sous la courbe d'aire intéro-labiale pour chaque syntagme puis pour chaque type de focalisation ; b. (droite) moyennes des moyennes des pics de vitesse de l'aire intérolabiale normalisés pour chaque syntagme et pour chaque type de focalisation.*

mesure	effet congruence	effet type de focalisation	interaction	test t congruence	test t post-foc	test t pré-foc	
aire intéro-labiale	F(1,15)=202,165 p<0,001	F(2,30)=4,313 p=0,023	F(1,241,30)=3,755 p=0,061	t=9,745 p<0,001	t=-0,766 p=0,447	t=4,304 p<0,001	t=0,363 p=0,722
							t=4,757 p<0,001
vitesse de variation	F(1,15)=47,482 p<0,001	F(2,50)=1,883 p=0,17	-	t=6,616 p<0,001	t=-2,888 p=0,006	t=2,816 p=0,007	t=1,414 p=0,178
							t=2,454 p=0,02

TABLE III.5 - *Résultats des tests statistiques menés sur les données normalisées d'aire intéro-labiale selon la méthode décrite à la section A.2.1.2.d du chapitre III.*

Le graphique a. de la figure III.15 montre que lorsqu'un syntagme est focalisé, l'aire intéro-labiale moyenne sur ce syntagme augmente par rapport au cas neutre (les barres S&FS, V&FV et O&FO sont au-dessus de 1). Cette augmentation est systématiquement observée *i.e.* dans 100% des cas. Elle est d'en moyenne 41% (35,3% pour S&FS, 54,5% pour V&FV et 33,1% pour O&FO) et est significative[75] (cf. table III.5) : t=9,745 (p<0,001). On note donc pour la parole lexicalisée, comme il avait été observé pour la parole délexicalisée, une hyper-articulation focale. On notera que l'hyper-articulation est beaucoup plus nette lorsque la focalisation porte sur le verbe (la barre V&FV est plus haute que les barres S&FS et O&FO).

De la même manière que pour la durée (cf. ci-dessus), le graphique de la figure III.16 représente l'aire intéro-labiale des constituants focalisés en fonction du nombre de syllabes qu'ils contiennent et du type de focalisation considéré. On voit bien sur ce graphique que c'est pour les verbes monosyllabiques que l'on observe cette hyper-articulation, supérieure à celle pour le reste du corpus. De même que pour la durée (cf. ci-dessus), il est probable que, le locuteur ne disposant que d'une syllabe pour signaler la focalisation, les indices et corrélats soient renforcés sur cette unique syllabe.

[75] Test t de comparaison à 1.

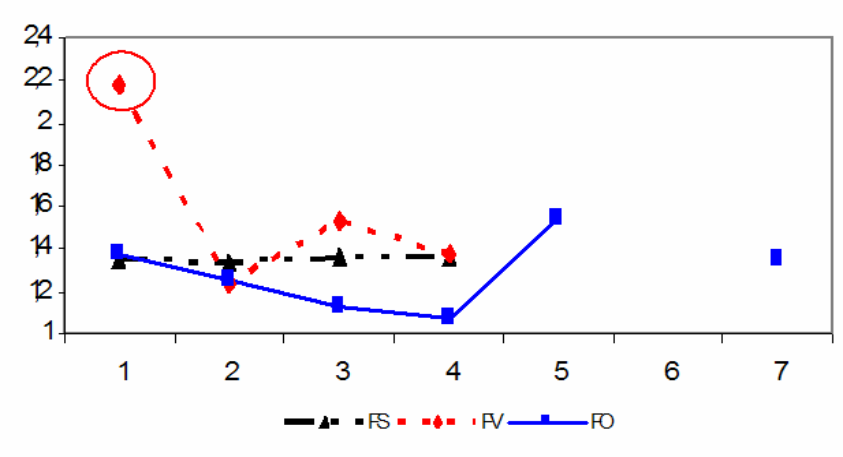


FIGURE III.16 – *Moyennes des données normalisées d'aire sous la courbe d'aire intéro-labiale sur le syntagme focalisé pour chaque type de focalisation et en fonction du nombre de syllabes du constituant considéré.*

Le graphique de la figure III.15 permet aussi de noter une augmentation de l'aire intéro-labiale moyenne des éléments directement pré-focaux et ce dans 96,9% des cas (93,8% des cas pour S&FV et 100% des cas pour V&FO). Cette augmentation est d'en moyenne 23% (14,6% pour S&FV et 31,4% pour V&FO) et elle est significative[76] (cf. table III.5) : t=4,757 (p<0,001). Or cette hyper-articulation n'est observée de nouveau que sur l'élément directement pré-focal puisque pour S&FO on n'observe pas de variation significative[77] (+0,8% ; t=0,363 p=0,722). Cette hyper-articulation pré-focale fait écho à la fois à ce qui avait été observé en parole délexicalisée et à ce qui a été décrit dans l'analyse des durées (cf. section A.2.2.3.d du présent chapitre). La figure III.17 que l'on pourra comparer à la figure III.8 (mêmes énoncés dans les deux cas mais en parole délexicalisée pour la figure III.8 et en parole lexicalisée pour la figure III.17) montre bien qu'il y a anticipation de l'hyper-articulation focale. Dans ce cas la focalisation porte sur l'objet, or on voit nettement que l'hyper-articulation commence dès la fin du verbe. Il semble donc qu'en parole lexicalisée on retrouve la stratégie d'anticipation de la focalisation mise en évidence en parole délexicalisée.

[76] Test t de comparaison à 1.
[77] Test t de comparaison à 1.

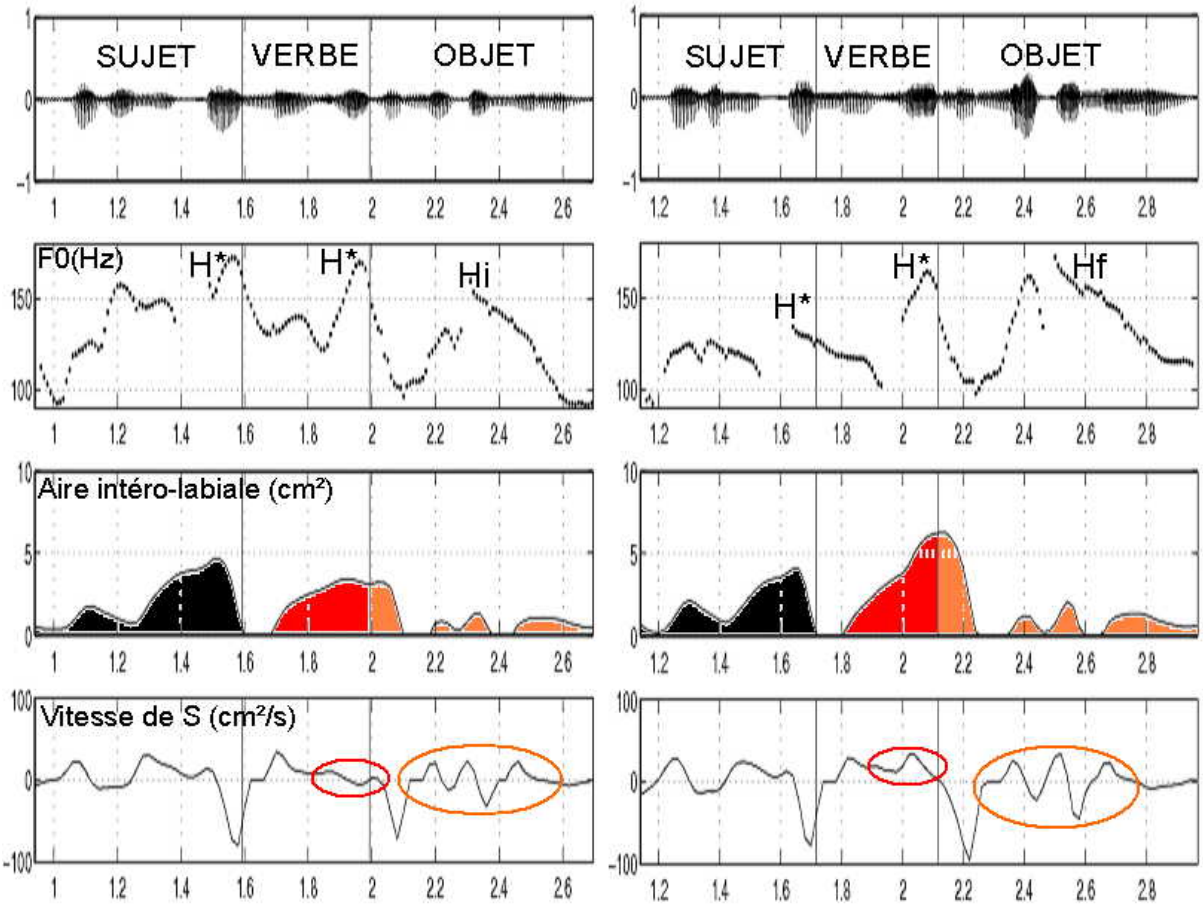


FIGURE III.17 – *a.(haut) signal acoustique ; b.(deuxième en partant du haut) suivi de F0 (en Hz) ; c.(troisième) aire intéro-labiale (cm²) et d. vitesse de variation de l'aire intéro-labiale (cm²/s) en fonction du temps ; même énoncé en version neutre (gauche) et avec focalisation sur l'objet (droite) : illustration du phénomène d'anticipation de la focalisation.*
L'énoncé prononcé était [Véroniqua][mangeait][les mauvais melons.].

Enfin, on observe une très faible diminution de l'amplitude de l'aire intéro-labiale des séquences post-focales dans 61,3% des cas et d'en moyenne 0,5%, mais celle-ci n'est globalement pas significative[78] (cf. table III.5: t=-0,766 p=0,447).

A.2.2.3.e.ii. Vitesse de variation de l'aire intéro-labiale

Comme on l'a déjà vu pour la parole délexicalisée, les pics de vitesse de variation (*i.e.* de la dérivée première) d'un paramètre articulatoire peuvent être de meilleurs corrélats que le paramètre lui-même. Les pics de vitesse de variation de l'aire intéro-labiale ont donc été détectés puis normalisés et moyennés sur tous les syntagmes identiques dans un type de focalisation (tous les S en FS par exemple) et le graphique b. de la figure III.15 résume les résultats obtenus. Il apparaît ainsi que lorsqu'un syntagme est focalisé, les pics de la vitesse de variation de l'aire intéro-labiale correspondant à ce syntagme sont plus importants que ceux correspondant aux autres syntagmes de l'énoncé (les barres correspondant à S&FS, V&FV et O&FO sont plus hautes que les autres). Le contraste intra-énoncé moyen est de 48,9% (55,8% pour FS, 69,5% pour FV et 21,5% pour FO) et est

[78] Test t de comparaison à 1.

significatif puisqu'on note un effet significatif du facteur congruence[79] de l'ANOVA[80] (cf. table III.5) : F(1,15)=47,482 (p<0,001). On ne note pas d'effet significatif du type de focalisation[81] (cf. table III.5) : F(2,30)=1,883 (p=0,17). On constate que c'est de nouveau lorsque la focalisation porte sur le verbe que le contraste intra-énoncé est le plus important.

Le graphique b. de la figure III.15 montre que lorsqu'un syntagme est focalisé, l'aire intéro-labiale sur ce syntagme augmente par rapport au cas neutre (les barres S&FS, V&FV et O&FO sont au-dessus de 1). Cette augmentation est d'en moyenne 49,3% (42,3% pour S&FS, 69% pour V&FV et 38,2% pour O&FO) et est significative[82] (cf. table III.5 : t=6,616 p<0,001). On note donc pour la parole lexicalisée, comme il avait été observé pour la parole délexicalisée, que lorsqu'un syntagme est focalisé, le locuteur met plus de « force articulatoire » dans son geste. On notera encore une fois que l'augmentation de cette « force » est plus nette lorsque la focalisation porte sur le verbe (la barre V&FV est plus haute que les barres S&FS et V&FV).

Le graphique de la figure III.18 montre en fait que c'est pour les verbes monosyllabiques que l'on observe cette accentuation plus nette. Il est probable, comme expliqué plus haut, que le locuteur ne disposant que d'une syllabe pour signaler la focalisation, les indices et corrélats soient renforcés sur cette unique syllabe.

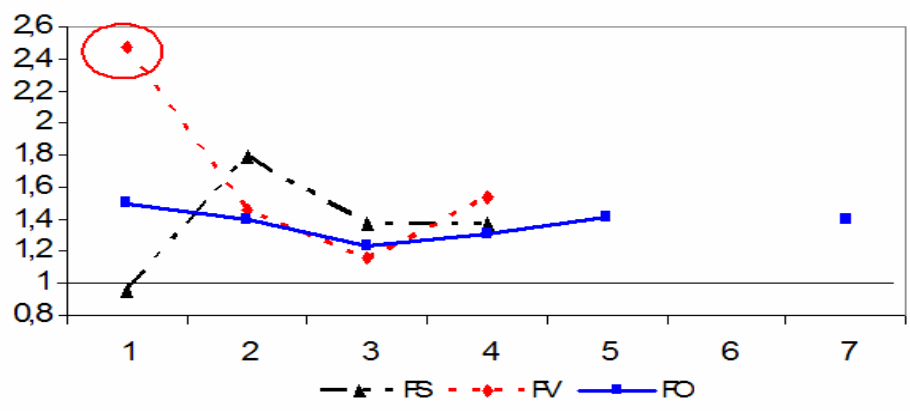


FIGURE III.18 – *Moyennes des données normalisées des pics de vitesse de l'aire intéro-labiale sur le syntagme focalisé pour chaque type de focalisation et en fonction du nombre de syllabes du constituant considéré.*

Le graphique b. de la figure III.15 permet de noter une augmentation de l'amplitude des pics de vitesse de variation de l'aire intéro-labiale des constituants directement pré-focaux et ce dans 65,6% des cas (62,5% des cas pour S&FV et 68,8% des cas pour V&FO). Cette augmentation est d'en moyenne 13,3% (5,7% pour S&FV et 20,9% pour V&FO) et elle est significative[83] (cf. table III.5) : t=2,454 (p=0,02). Elle n'est que de 6,6% sur S&FO et n'est pas significative[84] : t=1,414 (p=0,178).

[79] Congruence : facteur intra-sujet à deux niveaux : cas congruents (S&FS, V&FV et O&FO) et cas incongruents (S&FV et S&FO, V&FS et V&FO et O&FS et O&FV).
[80] ANOVA à deux facteurs intra-sujets : congruence (deux niveaux) et type de focalisation (trois niveaux) ; pour une explication complète et détaillée de l'ANOVA voir la section A.2.1.2.d du chapitre III.
[81] Type de focalisation : facteur intra-sujet à trois niveaux : FS, FV et FO.
[82] Test t de comparaison à 1.
[83] Test t de comparaison à 1.
[84] Test t de comparaison à 1.

Donc elle concerne bien uniquement l'élément directement pré-focal. Sur l'exemple donné en figure III.17, on remarque qu'il y a en effet anticipation de la focalisation. Dans ce cas, la focalisation porte sur l'objet, or on voit nettement que les pics de vitesse de variation de l'aire intéro-labiale commencent à être plus importants dès la fin du verbe. Il semble donc qu'en parole lexicalisée on retrouve la stratégie d'anticipation de la focalisation mise en évidence en parole délexicalisée, à la fois pour l'aire intéro-labiale et pour sa vitesse de variation.

A.2.2.3.f. Conclusion : vers un premier modèle de la production d'indices visibles de la focalisation contrastive prosodique en français

A.2.2.3.f.i. Constituant focalisé

Il apparaît donc un allongement significatif des syllabes focales (intra : +29,8%, inter : +33,3%) et ce dans 100% des cas. Cet ***allongement focal*** correspond à ce qui a été observé pour la parole délexicalisée et les ordres de grandeur sont les mêmes. On trouve que le premier segment focalisé est allongé de façon encore plus significative (+53,2%) et ce dans 97,9% des cas. Ceci marque ainsi peut-être le début du constituant focalisé de façon nette et précise.

L'***allongement du premier segment focal*** est beaucoup plus marqué pour les cas de focalisation sur le verbe et sur l'objet. Comme il a été suggéré plus haut, il existe déjà en effet peut-être un allongement à l'initiale de l'énoncé (et donc à l'initiale du sujet) de façon inhérente dans tous les cas de figure et ceci peut-être pour signaler le début de l'énoncé. L'allongement supplémentaire dû à la focalisation est peut-être ainsi moins important parce qu'il y a déjà un allongement dans le cas neutre (inhérent au début de l'énoncé).

En outre ce résultat d'allongement du premier segment fait écho à la fermeture initiale allongée observée en parole délexicalisée qui était d'ailleurs du même ordre de grandeur.

On note de plus une augmentation de l'aire intéro-labiale (intra : +32,1%, inter : +41%) dans 97,9% des cas. Les pics de vitesse de variation de l'aire intéro-labiale sont eux aussi accentués (intra : +48,9%, inter : +49,3%) dans 93,5% des cas. Ceci correspondrait à une « force articulatoire » accrue pour les gestes articulatoires correspondant à un constituant focalisé. Cette augmentation à la fois de l'amplitude moyenne et des pics de vitesse de variation de l'aire intéro-labiale correspond à ce qu'on appelle d'***hyper-articulation focale***.

L'hyper-articulation et l'allongement que l'on note sur le constituant focalisé sont les plus nets pour la focalisation sur le verbe et les moins nets pour la focalisation sur l'objet. Or, dans ce corpus, les objets sont en moyenne plus longs que les sujets, eux-mêmes plus longs que les verbes (2,9 syllabes pour les sujets, 2,6 pour les verbes et 4,25 pour les objets). Il apparaît ainsi que plus le constituant à focaliser est long (plus il a de syllabes) moins le marquage de chacune des syllabes est net. Cette observation sera reprise et analysée dans la section A.2.4.3 du présent chapitre.

A.2.2.3.f.ii. Séquence pré-focale

En ce qui concerne la durée, on note un allongement assez important de la syllabe pré-focale (+19,3%). Cet allongement est plus marqué pour la dernière syllabe du verbe pré-focal (focalisation sur l'objet) que pour la dernière syllabe du sujet pré-focal (focalisation sur le verbe).

On note de plus une hyper-articulation du constituant directement pré-focal, soit une augmentation de l'aire intéro-labiale (+23%) dans 93,8% des cas et des pics de vitesse de variation de l'aire intéro-

labiale (+13,3%) dans 65,7% des cas. Cette hyper-articulation est nettement plus marquée sur le verbe pré-focal (focalisation sur l'objet) que sur le sujet pré-focal (focalisation sur le verbe).

On note ainsi en parole normale une ***stratégie d'anticipation de la focalisation*** identique à celle observée en parole délexicalisée. Le locuteur semble donc préparer la focalisation dès la fin du constituant précédant le constituant à focaliser. Ce phénomène a déjà été illustré grâce à la figure III.17. Cependant il pourrait être objecté dans ce cas de figure précis que l'objet focalisé (dans ce cas là : [les mauvais melons]) ne commençait pas par une fermeture des lèvres et que donc la coarticulation du verbe pré-focal vers l'objet focalisé pouvait entraîner cette impression d'hyper-articulation pré-focale. La figure III.19 donne donc un exemple d'hyper-articulation pré-focale correspondant à un constituant focalisé avec fermeture initiale des lèvres (annulation de l'aire intéro-labiale) et donc pour lequel aucun argument de coarticulation ne peut être mis en avant. Or on observe encore ici une hyper-articulation sur la fin du sujet pré-focal.

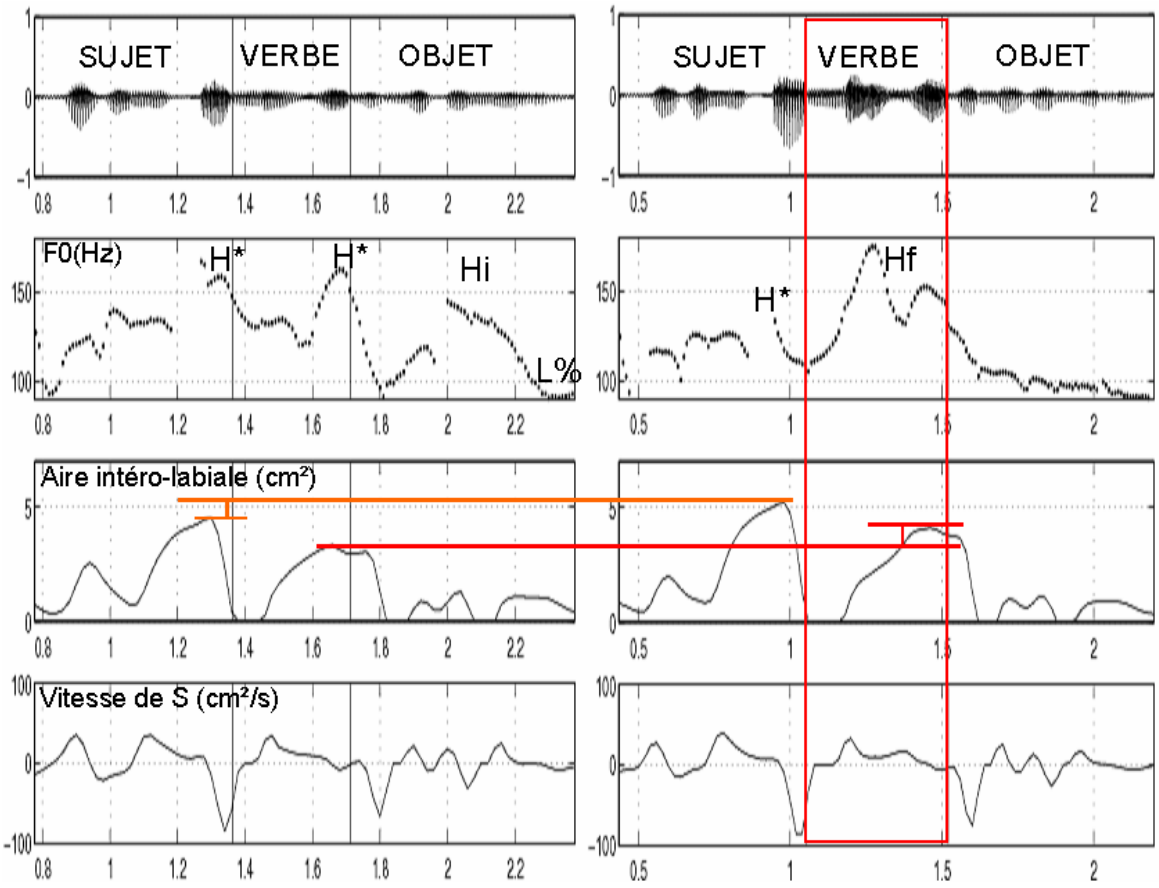


FIGURE III.19 – *a.(haut) signal acoustique ; b.(deuxième graphique en partant du haut) suivi de F0 (en Hz) ; c.(troisième) aire intéro-labiale (cm²) ; d.(quatrième) vitesse de variation de l'aire intéro-labiale (cm²/s) en fonction du temps ; même énoncé en version neutre (gauche) et avec focalisation sur le verbe (droite) ; illustration du phénomène d'anticipation de la focalisation avec un constituant focalisé pour lequel on observe une fermeture initiale (annulation de l'aire intéro-labiale). L'énoncé prononcé était [Véroniqua][mangeait][les mauvais melons.].*

A.2.2.3.f.iii. Séquence post-focale

Pour ce locuteur, aucune variation n'a été observée sur la séquence post-focale aussi bien en ce qui concerne les durées que l'articulation. Les tendances articulatoires observées vont vers une très légère baisse de l'amplitude des mouvements mais cette baisse n'est pas significative.

A.2.2.3.f.iv. Comparaison des stratégies acoustiques et articulatoires

On se souviendra que les productions acoustiques focalisées du locuteur A avaient été étudiées en détail dans le chapitre II. Il était apparu que celui-ci signale la focalisation par un allongement des syllabes focales et un maximum de F0 sur le constituant focalisé, que l'on peut mettre en parallèle à l'hyper-articulation focale. Nous avions également observé une atténuation de F0 sur l'élément directement pré-focal puis une désaccentuation de la séquence post-focale, que l'on pourrait qualifier d'hypo-variation laryngée pré- et post-focale. Cette stratégie acoustique est donc différente de celle utilisée sur le plan articulatoire puisque le locuteur A n'hypo-articule pas la séquence post-focale et a tendance à hyper-articuler la syllabe pré-focale. Il apparaît donc que le contrôle lié à la focalisation est différent pour les variations laryngées et supra-laryngées.

A.2.3. Étude des productions du locuteur B

L'étude pilote réalisée sur le locuteur A a permis de conclure, chez ce locuteur particulier, à l'existence de corrélats articulatoires potentiellement visibles de la focalisation contrastive en français. La question est maintenant de savoir s'il existe aussi des corrélats chez d'autres locuteurs et si c'est le cas, si ces corrélats sont identiques à ceux identifiés pour le locuteur A. En d'autres termes, le but est de déterminer si l'étude précédente est généralisable et dans quelle mesure. C'est pourquoi il a été décidé d'effectuer une étude similaire chez un autre locuteur.

A.2.3.1. Mise en œuvre expérimentale

A.2.3.1.a. Corpus[85]

Pour cette étude avec un nouveau locuteur, le corpus AV1[86] n'a pas été réutilisé tel quel. Il a été légèrement modifié après la détection de certains inconvénients lors des analyses précédentes. Certaines consonnes (telles que [g] ou [s]) posaient problème notamment pour la segmentation acoustique. Certains constituants « objets » étaient trop longs (7 syllabes pour les phrases (1) et (3)) et rendaient parfois la production de la focalisation compliquée pour le locuteur (tendance à diviser en deux sous-groupes par exemple). De plus, il n'y avait pas suffisamment d'occurrences de voyelles protruses ce qui avait rendu inutile l'analyse du paramètre de protrusion labiale. Or la protrusion labiale est sans aucun doute visible et devrait aussi être affectée par le phénomène d'hyper-articulation. Les phrases du corpus AV1 ont donc été remaniées voire totalement abandonnées au profit de nouvelles. Les phrases (2), (4), (6) et (7) du corpus AV1 ont néanmoins été conservées telles quelles et ce en vue de mener par la suite un test perceptif multi-locuteurs (locuteurs A et B) en audio-visuel avec de la parole chuchotée (voir section B du chapitre IV). Le nouveau corpus utilisé pour cet

[85] Le corpus décrit ici sera de nouveau utilisé dans la suite pour d'autres travaux et sera ci-après nommé corpus AV2.
[86] Le corpus AV1 est décrit en détails dans la section 3.1.1 du chapitre II.

enregistrement est ainsi composé de 13 phrases, toujours de structure SVO. Les syllabes sont également des syllabes CV et chaque phrase est susceptible d'avoir une réalisation prosodique du type {[LHiLH*]$_S$ [LHiLH*]$_V$ [LHiLL%]$_O$}. Les consonnes sonores ont été privilégiées afin de faciliter le suivi de F0. Les 13 phrases du corpus sont[87] :

(1) [Romain]$_{S2}$ [ranima]$_{V3}$ [la jolie maman]$_{O5}$

(2) [Véroniqua]$_{S4}$ [mangeait]$_{V2}$ [les mauvais melons]$_{O5}$

(3) [Mon mari]$_{S3}$ [veut ranimer]$_{V4}$ [Romain]$_{O2}$

(4) [Les loups]$_{S2}$ [suivaient]$_{V2}$ [Marilou]$_{O3}$

(5) [La nounou]$_{S3}$ [mariera]$_{V3}$ [Li]$_{O1}$

(6) [Le lama lent]$_{S4}$ [lut]$_{V1}$ [Marinella]$_{O4}$

(7) [Marinella]$_{S4}$ [va laminer]$_{V4}$ [Numu]$_{O2}$

(8) [Lou]$_{S1}$ [mima]$_{V2}$ [le lama]$_{O3}$

(9) [Le nominé]$_{S4}$ [lut]$_{V1}$ [les longs mots]$_{O3}$

(10) [La nounou]$_{S3}$ [vit]$_{V1}$ [Lou]$_{O1}$

(11) [Les loups]$_{S2}$ [mimaient]$_{V2}$ [Marilou]$_{O3}$

(12) [Lou]$_{S1}$ [ramena]$_{V3}$ [Manu]$_{O2}$

(13) [Li]$_{S1}$ [ralluma]$_{V3}$ [les moulinets]$_{O4}$

A.2.3.1.b. Enregistrement

Le corpus a été enregistré par le locuteur B[88] à l'aide de la plate-forme expérimentale de l'ICP décrite dans la section A.2.1.1 du présent chapitre. L'enregistrement a été réalisé dans la chambre sourde de l'ICP. Chaque phrase a été enregistrée pour quatre conditions : focalisation sur le sujet, sur le verbe, sur l'objet et cas neutre. Les énoncés ont tous été enregistrés en parole normale lexicalisée. Les phrases (1), (2), (3) et (4) ont de plus été enregistrées en mode chuchotée pour tous les types d'énonciation. Deux répétitions ont été enregistrées pour chaque type d'énoncé. Au total 104 énoncés ont donc été enregistrés en parole normale lexicalisée et 32 en parole normale chuchotée.

La méthode d'obtention de la focalisation utilisée ici était légèrement différente de celle utilisée pour les enregistrements avec le locuteur A. Le locuteur entendait un prompt audio[89] dans lequel deux interlocuteurs conversaient. Le premier interlocuteur prononçait la phrase à produire et le deuxième interlocuteur n'ayant pas bien compris un des éléments (S, V ou O) de la phrase répétait celle-ci comme il l'avait perçue et sous forme interrogative. L'enregistrement se déroulait ainsi comme dans l'exemple (III.1) dans lequel les majuscules signalent la focalisation. La tâche du locuteur était donc de répéter la phrase initiale en corrigeant l'élément mal compris par le deuxième interlocuteur.

[87] Les [ ] indiquent le découpage en composants sujet, verbe et objet. Les indices en bas à droite de chaque composant donnent son rôle au sein de la phrase (S, V ou O) et le nombre de syllabes qui le composent.
[88] Voir la section D du chapitre *Notes et indices de lecture* pour un récapitulatif des locuteurs.
[89] Prompt audio enregistré à l'avance par deux locutrices de langue maternelle française et diffusé par des hauts-parleurs dans le but de déclencher une production spécifique chez l'interlocuteur.

(III.1) Le locuteur entend :

Interlocuteur 1 : Les loups suivaient Marilou.

Interlocuteur 2 : Les loups suivaient Aurélie ?

Le locuteur dit : Les loups suivaient MARILOU.

Aucune indication n'a été donnée au locuteur sur la manière de produire la focalisation (*e.g.* quelle(s) syllabe(s) devai(en)t être accentuée(s)).

A.2.3.1.c. Mesures et traitements

A.2.3.1.c.i. Durées

Les durées de toutes les syllabes du corpus ont été mesurées gâce à la segmentation aoustique éffectuée suivant la méthode décrite à la section C.1.4.1 du chapitre II. Nous tenterons ici de déterminer si on observe chez ce locuteur un allongement des syllabes focales et plus particulièrement du premier segment focal. La séquence pré-focale et, de façon plus précise, la durée de la syllabe pré-focale seront aussi analysées.

A.2.3.1.c.ii. Données articulatoires

De façon à établir des comparaisons avec les analyses faites pour le locuteur A, l'aire intéro-labiale et sa vitesse de variation seront analysées de façon précise. Les techniques de mesure seront les mêmes que celles décrites dans la section A.2.2.3.c.ii du présent chapitre. La protrusion de la lèvre supérieure qui n'avait pas pu être analysée pour le locuteur A, faute de segment suffisamment protrus dans le corpus, le sera ici grâce au nouveau corpus. Les données de protrusion seront ramenées à une référence afin de ne conserver que la protrusion par rapport à la position de repos. Cette référence sera évaluée pour chaque énoncé à l'aide de la portion de silence et donc de la position de repos qui existe au début de chaque enregistrement avant que le locuteur ne commence à parler. Suite à cette mise à référence on obtiendra des valeurs positives et négatives (lorsque la lèvre supérieure est rentrée). Puisque c'est ici une différence visible par rapport au cas neutre qui nous intéressera, c'est la valeur absolue de ces données mises à référence qui sera prise en compte pour l'étude. La technique de normalisation employée sera exactement la même que celle utilisée pour l'étude de l'aire intéro-labiale.

A.2.3.1.c.iii. Normalisation des données

Les données seront toutes normalisées selon la méthode décrite à la section A.2.2.3.c.iii du présent chapitre.

A.2.3.1.d. Validation acoustique

Un total de 104 énoncés a été enregistré en parole normale lexicalisée pour ce locuteur. Une validation acoustique des données a été effectuée selon la méthode décrite à la section A.2.1.2.b du présent chapitre. Celle-ci a permis de mettre en évidence quatre énoncés pour lesquels la focalisation avait été produite de façon non satisfaisante tant du point de vue de l'analyse acoustique que du point de vue de l'analyse perceptive. Parmi ces quatre énoncés, deux correspondaient à des cas de focalisation sur le sujet, un autre à un cas de focalisation sur l'objet et le dernier à un énoncé neutre. Après validation acoustique, il restait donc un total de 100 énoncés à analyser.

A.2.3.2. Analyse de la durée

A.2.3.2.a. Syllabes focales

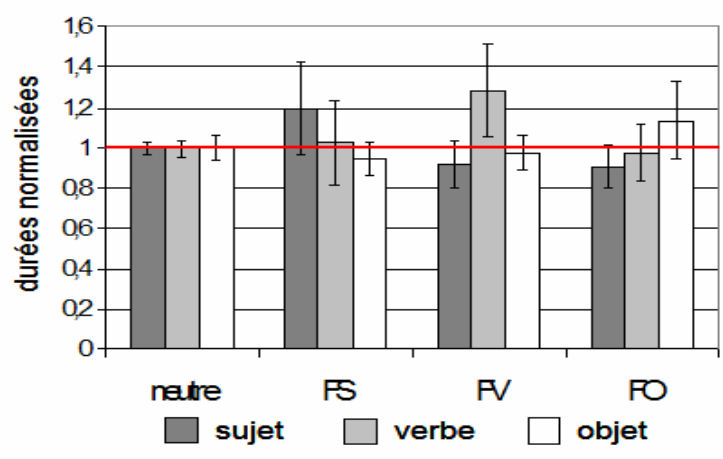


FIGURE III.20 – *Moyennes des durées normalisées des syllabes de chaque type de syntagme et pour chaque type de focalisation (le cas neutre correspond à la valeur 1).*

test	effet congruence	effet type de focalisation	interaction	test t congruence	test t post-foc	test t pré-foc	
durées	F(1,25)=180,702 p<0,001	F(2,50)=3,561 p=0,036	F(2,50)=2,238 p=0,117	t=8,212 p<0,001	t=-1,086 p=0,281	t=-4,792 p<0,001	t=-4,68 p<0,001
							t=-2,748 p=0,008

TABLE III.6 – *Résultats des test statistiques menés sur les données de durée normalisées pour le locuteur B et selon la méthode décrite à la section A.2.1.2.d du chapitre III.*

Les résultats des mesures de durée sont consignés dans le graphique de la figure III.20. On constate que de façon générale, la durée moyenne d'une syllabe est plus importante si celle-ci appartient à un syntagme focalisé que si elle appartient à un autre syntagme du même énoncé. Le contraste intra-énoncé est en moyenne de 25,9% (21,3% pour FS, 35,7% pour FV et 20,6% pour FO) et il est significatif (facteur congruence[90] de l'ANOVA[91], cf. table III.6) : F(1,25)=180,702 (p<0,001). Le type de focalisation[92] a lui aussi un effet significatif (F(2,50)=3,561 p=0,036) car la durée des syllabes est significativement plus importante quel que soit le constituant considéré lorsque la focalisation porte sur le sujet ou le verbe (contraste de l'ANOVA significatif à p=0,011). L'interaction congruence×type de focalisation[93] est quant à elle non significative (F(2,50)=2,238 p=0,117) mais on constate que c'est de nouveau lorsque la focalisation porte sur le verbe que le contraste intra-énoncé est le plus important.

[90] Congruence : facteur intra-sujet à deux niveaux : cas congruents (S&FS, V&FV et O&FO) et cas incongruents (S&FV et S&FO, V&FS et V&FO et O&FS et O&FV).
[91] ANOVA à deux facteurs intra-sujets : congruence (deux niveaux) et type de focalisation (trois niveaux) ; pour une explication complète et détaillée de l'ANOVA voir la section A.2.1.2.d du chapitre III.
[92] Type de focalisation : facteur intra-sujet à trois niveaux : FS, FV et FO.
[93] Les effets principaux (congruence et type de focalisation) n'ont pas tous les deux un effet significatif.

Le graphique de la figure III.20 permet de plus de voir que lorsqu'un syntagme est focalisé, les durées de ses syllabes augmentent par rapport au cas neutre (les barres S&FS, V&FV et O&FO sont au-dessus de 1) et ce pour 69,2% des sujets focalisés, 96,2% des verbes focalisés et 80% des objets focalisés. L'allongement est d'en moyenne 20,6% (19,7% pour FS, 28,6% pour FV et 13,4% pour FO). Cet allongement est significatif[94] (t=8,259 p<0,001) et est supérieur au seuil de perception auditif de l'allongement qui est de 20% (cf. section C.2 du chapitre II).

Le graphique de la figure III.21 représente les durées normalisées des syllabes des constituants focalisés en fonction du nombre de syllabes qu'ils contiennent (en abscisse) et du type de focalisation considéré (différentes courbes). On constate grâce à ce graphique que quel que soit le type de focalisation considéré, l'allongement focal est nettement plus prononcé dans les cas où le constituant ne possède qu'une seule syllabe.

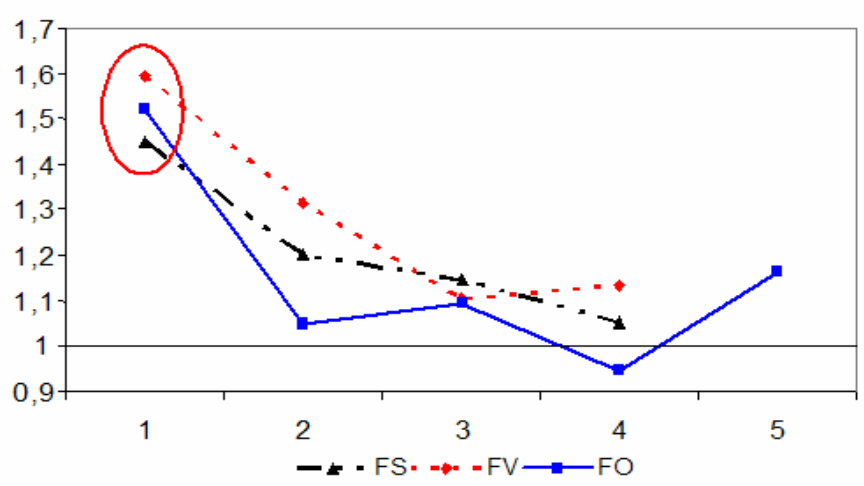


FIGURE III.21 – *Moyennes des durées normalisées des syllabes de chaque syntagme focalisé pour chaque type de focalisation (différentes courbes) et en fonction du nombre de syllabes du constituant considéré (en abscisse).*

On constate grâce à la figure III.20 que lorsqu'un syntagme est focalisé, il semblerait que la séquence qui le suit voit une très légère réduction de la durée de ses syllabes (les barres V&FS, O&FS et O&FV sont en dessous de 1) et ce dans 61,5% des cas pour V&FS, 84,6% des cas pour O&FS et 76,9% des cas pour O&FV. Cette réduction est d'en moyenne 1,8% ce qui est peu mais elle est néanmoins significative[95] : t=-2,748 (p=0,008).

On constate une baisse de la durée de toutes les syllabes pré-focales (les barres S&FV, S&FO et V&FO sont en dessous de 1) dans 80,8% des cas pour S&FV, 80% des cas pour S&FO et 60% des cas pour V&FO. Cette baisse est d'en moyenne 7,1% (-8,2% pour S&FV, -9,3% pour S&FO et –2,7% pour V&FO) et est globalement significative[96] : t=-4,309 (p<0,001).

[94] Test t de comparaison à 1.
[95] Test t de comparaison à 1.
[96] Test t de comparaison à 1.

A.2.3.2.b. Syllabe pré-focale

Puisqu'il est apparu chez le locuteur A que la durée de la syllabe directement pré-focale était affectée, cette même syllabe a été étudiée chez le locuteur B. Le graphique de la figure III.22 montre ainsi qu'il y a très peu de variation de la durée de la dernière syllabe d'un constituant, selon que cette syllabe précède ou non un syntagme focalisé. D'ailleurs il n'y a pas de différence significative entre les cas neutres et les cas où le syntagme suivant est focalisé[97] : t=-0,912 (p=0,366).

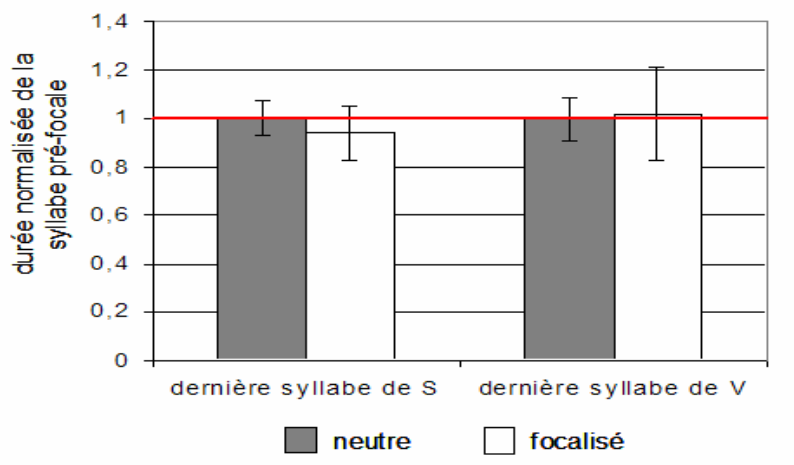


FIGURE III.22 – *Moyennes des durées normalisées de la dernière syllabe des constituants S et V lorsque le constituant suivant n'est ou n'est pas focalisé (syllabe pré-focale).*

A.2.3.2.c. Premier segment focalisé

On avait également constaté pour le locuteur A, un allongement du premier segment focalisé significativement plus important que l'allongement focal global. Le premier segment focalisé a donc été étudié pour le locuteur B et les résultats sont consignés dans le graphique de la figure III.23. Il semble ainsi que, mis à part le cas sujet, lorsqu'un syntagme est focalisé, son premier segment est allongé de façon très importante par rapport à la durée de ce même segment dans le cas neutre (+34,6% pour FV et +51,1% pour FO). Cet allongement est globalement significatif (test t de comparaison à 1 avec toutes les données : t=4,359 p<0,001). Il est de plus significativement plus important que l'allongement focal global (test t de comparaison des durées moyennes des syllabes focales et du premier segment du constituant focalisé avec toutes les données : l'hypothèse d'égalité des moyennes doit être rejetée : t=1,108 p=0,27).

[97] Test t de comparaison à 1.

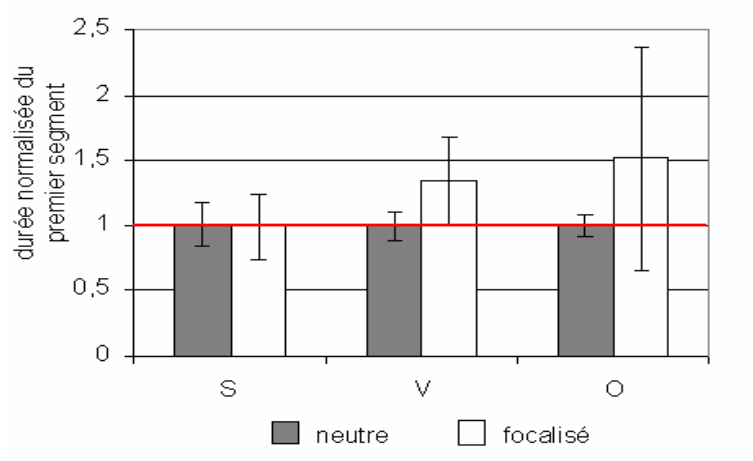


FIGURE III.23 – *Moyennes des durées normalisées du premier phonème des constituants S, V et O lorsque le constituant auquel il appartient est ou n'est pas focalisé.*

A.2.3.3. Analyse articulatoire

A.2.3.3.a. Aire intéro-labiale

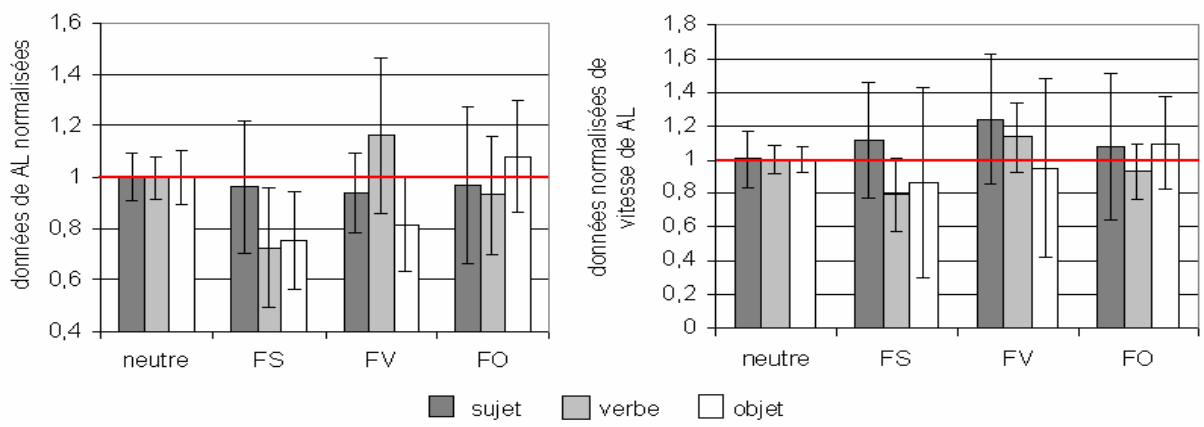


FIGURE III.24 – *a.(gauche) moyennes des données normalisées d'aire intéro-labiale pour chaque type de syntagme et chaque type de focalisation ; b.(droite) moyennes des pics de vitesse normalisés pour chaque syntagme et chaque type de focalisation ; résultats du locuteur B.*

mesure	effet congruence	effet type de focalisation	interaction	test t congruence	test t post-foc	test t pré-foc	
aire intéro-labiale	F(1,25)=53,376 p<0,001	F(2,50)=8,582 p=0,001	F(2,50)=2,285 p=0,112	t=2,268 p=0,026	t=-10,404 p<0,001	t=-2,016 p=0,047	t=-0,489 p=0,629 t=-2,457 p=0,017
vitesse	-	-	-	t=3,789 p<0,001	t=-2,611 p=0,011	-	

TABLE III.7 – *Résultats des tests statistiques menés sur les données normalisées d'aire intéro-labiale pour le locuteur B selon la méthode décrite à la section A.2.1.2.d du chapitre III.*

On notera grâce au graphique a de la figure III.24 que lorsqu'un syntagme est focalisé, l'aire intéro-labiale lui correspondant, qui représente comme on l'a vu l'amplitude moyenne des mouvements, est

plus importante que pour les autres syntagmes du même énoncé (les barres S&FS, V&FV et O&FO sont plus grandes que les autres barres). Le contraste intra-énoncé moyen est de 25,3% (29,9% pour FS, 32,3% pour FV et 13,6% pour FO) et est significatif (effet congruence[98] de l'ANOVA[99], cf. table III.7 : $F(1,25)=53,376$ $p<0,001$). De plus, on note un effet significatif du type de focalisation[100] : $F(2,50)=8,582$ ($p=0,001$) ceci étant dû au fait que l'aire intéro-labiale est plus importante lorsque la focalisation porte sur le verbe ou l'objet (contraste de l'ANOVA significatif à $p=0,001$). Bien que l'interaction entre ces deux facteurs ne soit pas significative ($F(2,50)=2,285$ $p=0,112$), on notera que c'est de nouveau lorsque c'est le verbe qui est focalisé que le contraste intra-énoncé est le plus important.

On se souvient que pour le locuteur A, l'aire intéro-labiale était non seulement plus importante que celle des autres syntagmes du même énoncé mais aussi que sa valeur pour le même syntagme dans le cas neutre. Pour le locuteur B, on observe cette augmentation uniquement pour 45,8% des sujets focalisés, 73,1% des verbes focalisés et 72% des objets focalisés. On notera d'ailleurs que la barre S&FS du graphique a. de la figure III.24 est en dessous de 1 ce qui signifie qu'en moyenne lorsque le sujet est focalisé l'aire intéro-labiale lui correspondant n'augmente pas par rapport au cas neutre et même elle diminue (-3,7%). En moyenne, on n'observe une augmentation par rapport au cas neutre que pour V&FV et O&FO, celle-ci est de 12,1% (16,2% pour V&FV et 8% pour O&FO). De façon générale, l'augmentation par rapport au cas neutre est tout de même tout juste significative[101] : $t=2,268$ ($p=0,026$).

On constate grâce au graphique a. de la figure III.24 que l'aire intéro-labiale de la séquence pré-focale diminue par rapport au cas neutre (les barres S&FV et V&FO sont en dessous de 1) et ce dans 65,4% des cas pour S&FV et 76% des cas pour V&FO. Cette diminution est en moyenne de 6,5% (6% pour S&FV et 6,9% pour V&FO) et est significative[102] : $t=-2,457$ ($p=0,017$). On constate que cette diminution n'est présente que sur le constituant directement pré-focal puisqu'on n'observe pas de variation significative pour S&FO (-2,9%, test t de comparaison à 1 : $t=-0,489$ $p=0,629$). Il semblerait donc que le locuteur B tende à hypo-articuler le constituant qui précède directement le constituant focalisé.

En ce qui concerne la séquence post-focale, le graphique a. de la figure III.24 permet de constater que l'aire intéro-labiale y est plus faible que dans le cas neutre (les barres V&FS, O&FS et O&FV sont en dessous de 1) et ce dans 91,7% des cas pour V&FS, 87,5% des cas pour O&FS et 84,5% des cas pour O&FV. Cette diminution est d'en moyenne 22,1% (27,5% pour V&FS, 24,3% pour O&FS et 18,3% pour O&FV) et elle est nettement significative (test t de comparaison à 1[103]) : $t=-10,404$ ($p<0,001$). Le locuteur B hypo-articule donc la séquence post-focale.

On pourra donc conclure que ce locuteur tend à répartir son effort entre l'hyper-articulation de l'élément focal et l'hypo-articulation des séquences pré- et post-focales et plus particulièrement de la séquence post-focale et ce afin de créer un contraste articulatoire au sein de l'énoncé entre ce qui est focalisé et ce qui ne l'est pas.

---

[98] Cas congruents : S&FS, V&FV et O&FO ; cas incongruents : V&FS et O&FS, S&FV et O&FV et S&FO et V&FO.

[99] ANOVA à deux facteurs intra-sujets : congruence (deux niveaux) et type de focalisation (trois niveaux) ; pour une explication complète et détaillée du test voir la section A.2.1.2.d du chapitre III.

[100] Type de focalisation : facteur intra-sujet à trois niveaux : FS, FV et FO.

[101] Test t de comparaison à 1 : les données sont normalisées et le cas neutre correspond donc à la valeur 1.

[102] Test t de comparaison à 1 : les données sont normalisées et le cas neutre correspond donc à la valeur 1.

[103] Les données sont normalisées et le cas neutre correspond donc à la valeur 1.

Le graphique de la figure III.25 représente les données normalisées d'aire intéro-labiale des constituants focalisés en fonction du nombre de syllabes qu'ils contiennent (en abscisse) et du type de focalisation considéré (différentes courbes). On constate que globalement l'hyper-articulation de l'aire intéro-labiale lors de la focalisation est plus importante pour les constituants monosyllabiques et qu'elle décroît avec l'augmentation du nombre de syllabes du constituant focalisé. Cette observation suggère que l'hyper-articulation est répartie entre les syllabes du constituant à focaliser *i.e.* plus il y a de syllabes moins chacune d'entre elles est hyper-articulée. Cette observation ne peut cependant pas être faite pour le cas de focalisation sur le sujet. En fait, les sujets monosyllabiques sont 'Li' et 'Lou'. En ce qui concerne 'Lou', la focalisation provoque l'augmentation de la protrusion labiale et donc la *diminution* de l'aire intéro-labiale correspondante. En ce qui concerne 'Li', la focalisation provoque une augmentation de l'étirement qui peut induire une diminution de l'ouverture et ainsi aboutir à une non variation voire à une diminution de l'aire intéro-labiale correspondante. En moyennant ces deux cas de figure, on aboutit à une tendance générale à la baisse de l'aire intéro-labiale.

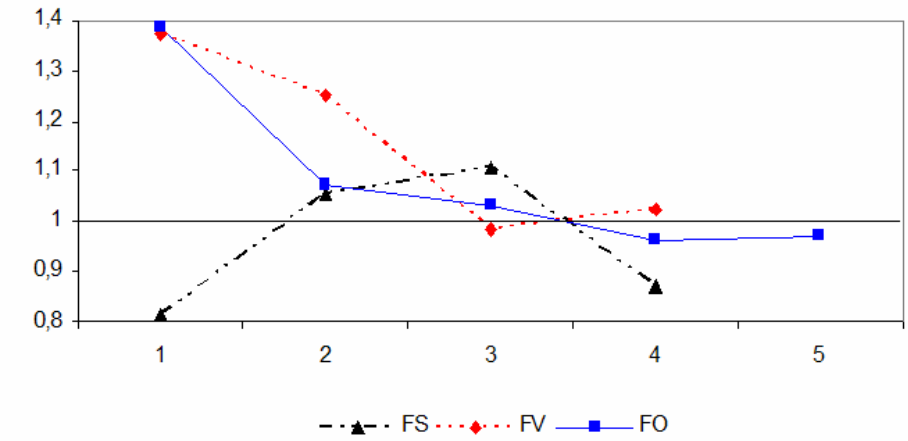


FIGURE III.25 – *Moyennes des données normalisées d'aire intéro-labiale sur le syntagme focalisé pour chaque type de focalisation et en fonction du nombre de syllabes du constituant focalisé considéré : résultats pour le locuteur B.*

On se souvient que pour le locuteur A, les pics de vitesse de variation de l'aire intéro-labiale étaient de bons corrélats de la focalisation contrastive puisque lorsqu'un syntagme était focalisé on notait un augmentation significative de l'amplitude de ces pics sur ce même syntagme à la fois par rapport au reste de l'énoncé focalisé et par rapport au cas neutre. Les pics de vitesse de variation de l'aire intéro-labiale ont donc également été détectés chez le locuteur B et les résultats de ces mesures sont consignés dans le graphique b. de la figure III.24. On constate rapidement que les tendances sont moins nettes que pour l'aire intéro-labiale. En fait, il n'existe pas de contraste réel au sein de l'énoncé focalisé entre ce qui est focalisé et ce qui ne l'est pas sauf lorsque la focalisation porte sur le sujet. (les barres V&FV et O&FO ne sont pas au-dessus des autres barres correspondant respectivement à FV et FO). Ceci est dû au fait qu'il semble que les pics de vitesse soient toujours plus importants sur le sujet et ce quel que soit le type de focalisation (les barres S&FV et S&FO sont aussi très hautes). Les pics de vitesse sont en effet plus hauts que pour le cas neutre dans 68% des cas pour S&FV et dans 56% des cas pour S&FO. Bien qu'il semble par conséquent ne pas exister de contraste intra-énoncé, on note cependant que les pics de vitesse correspondant au constituant focalisé sont tout de

même en général augmentés par rapport au cas neutre (les barres S&FS, V&FV et O&FO sont toutes au-dessus de 1. Cette augmentation existe pour 58,3% des sujets focalisés, 68% des verbes focalisés et 60% des objets focalisés et est en moyenne de 11,5% (11,2% pour S&FS, 13,3% pour V&FV et 10% pour O&FO). Elle est globalement significative[104] (t=3,789 p<0,001).

On note de plus une tendance à la diminution des pics de vitesse par rapport au cas neutre pour la séquence post-focale (dans 87,5% des cas pour V&FS et O&FS et dans 76% des cas pour O&FV). Cette baisse est d'en moyenne 11,2% (20,9% pour V&FS, 13,3% pour O&FS et 5,3% pour O&FV) et est globalement significative[105] (t=-2,611 p=0,011).

Enfin, on notera que dans 72% des cas, le verbe pré-focal (focalisation sur l'objet) voit l'amplitude de ses pics de vitesse diminuer (6,9% en moyenne).

En ce qui concerne la vitesse de variation de l'aire intéro-labiale, il semblerait donc qu'il y ait bien une tendance à l'augmentation des amplitudes des pics sur le constituant focalisé par rapport au cas neutre ainsi qu'une diminution des ces mêmes pics sur la séquence post-focale. Néanmoins, les contrastes intra-énoncés sont masqués par le fait qu'on observe systématiquement une augmentation des pics de vitesse du sujet qu'il soit focalisé ou non. Cette augmentation pourrait faire partie des marques de début d'énoncé déjà évoquées plus haut.

Le graphique de la figure III.26 représente les données normalisées de pics de la vitesse de variation de l'aire intéro-labiale des constituants focalisés en fonction du nombre de syllabes qu'ils contiennent (en abscisse) et du type de focalisation considéré (différentes courbes). Les mêmes commentaires que ceux qui avaient été effectués pour la figure III.25, pourront aussi être fait pour cette figure.

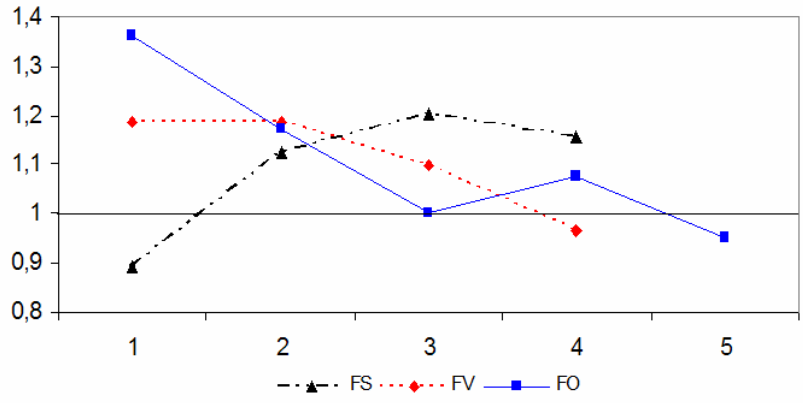


FIGURE III.26 – *Moyennes des données normalisées de vitesse de variation de l'aire intéro-labiale sur le syntagme focalisé pour chaque type de focalisation et en fonction du nombre de syllabes du constituant focalisé considéré : résultats pour le locuteur B.*

A.2.3.3.b. Protrusion de la lèvre supérieure

Le paramètre de protrusion de la lèvre supérieure n'avait pas pu être étudié pour le locuteur A car le corpus ne contenait pas suffisamment d'occurrences de voyelles protruses. C'est en partie pourquoi

[104] Test t de comparaison à 1.
[105] Test t de comparaison à 1.

un nouveau corpus avait été créé. Cette fois le paramètre de protrusion a donc pu être étudié de façon précise. La figure III.27 fournit les résultats des mesures de protrusion.

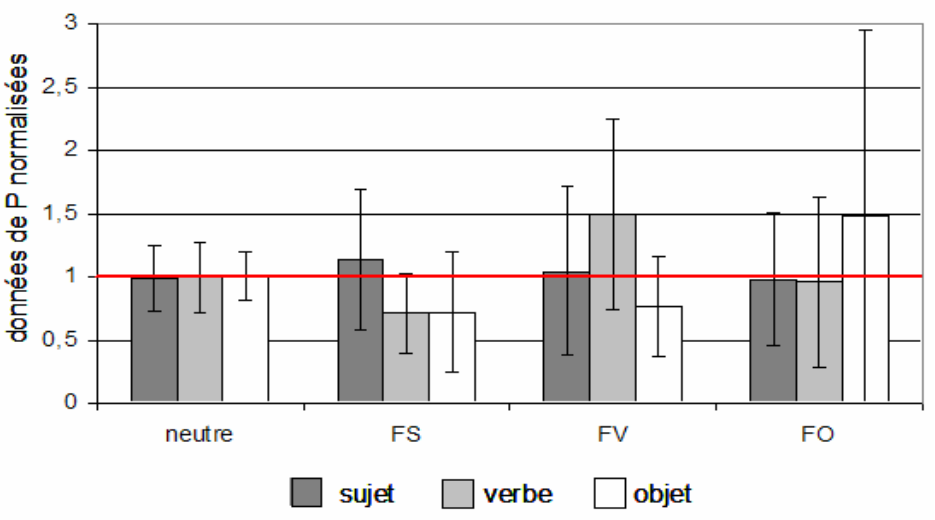


FIGURE III.27 – *Moyennes des données normalisées de protrusion de la lèvre supérieure (P1) pour chaque type de syntagme et chaque type de focalisation.*

mesure	effet congruence	effet type de focalisation	interaction	test t congruence	test t post-foc	test t pré-foc	
protrusion	F(1,25)=19,798 p<0,001	F(2,50)=2,714 p=0,076	-	t=3,307 p=0,001	t=-6,164 p<0,001	t=-0,069 p=0,945	t=-0,21 p=0,835
							t=-0,039 p=0,969

TABLE III.8 – *Résultats des tests statistiques menés sur les données normalisées de protrusion de la lèvre supérieure pour le locuteur B selon la méthode décrite à la section A.2.1.2.d du chapitre III.*

On notera ainsi (cf. figure III.27) que lorsqu'un syntagme est focalisé, sa protrusion est plus importante que pour les autres syntagmes du même énoncé (les barres S&FS, V&FV et O&FO sont plus grandes que les autres barres). Le contraste intra-énoncé moyen est ainsi de moyen 59,3% (59,1% pour S&FS, 64% pour V&FV et 54,7% pour O&FO) et est significatif (facteur congruence[106] de l'ANOVA[107], cf. table III.8 : F(1,25)=19,798 p<0,001). On ne note aucun effet significatif du type de focalisation[108] : F(2,50)=2,714 (p=0,076). Bien qu'on ne puisse donc pas s'intéresser à l'interaction congruence×type de focalisation[109], on notera tout de même que c'est de nouveau lorsque c'est le verbe qui est focalisé que le contraste intra-énoncé est le plus important.

On observe également grâce à la figure III.27 que la protrusion est plus importante sur un syntagme lorsqu'il est focalisé que sur le même syntagme dans le cas neutre (les barres S&FS, V&FV et O&FO sont au-dessus de 1) et ce pour 54,2% des sujets focalisés, 73,1% des verbes focalisés et

[106] Cas congruents : S&FS, V&FV et O&FO ; cas incongruents : V&FS et O&FS, S&FV et O&FV et S&FO et V&FO.
[107] ANOVA à deux facteurs intra-sujets : congruence (deux niveaux) et type de focalisation (trois niveaux) ; pour une explication complète et détaillée du test voir la section 1.2.1.2.4 du chapitre II.
[108] Type de focalisation : facteur intra-sujet à trois niveaux : FS, FV et FO.
[109] Les effets principaux n'étant pas tous les deux significatifs.

56% des objets focalisés. Elle est d'en moyenne 36,9% (13,5% pour S&FS, 49,1% pour V&FV et 47,9% pour O&FO) et est significative[110] : t=3,307 (p=0,001).

En ce qui concerne les éléments pré-focaux, la tendance générale est à la baisse de la protrusion par rapport au cas neutre (baisse vérifiée dans 61,5% des cas pour S&FV, 64% des cas pour S&FO et 72% des cas pour V&FO). Elle est d'en moyenne 2,1% pour S&FO et 4,1% pour V&FO mais est inexistante en moyenne pour S&FV (+4,8%). Elle n'est d'ailleurs globalement pas significative[111] : t=-0,069 (p=0,945).

Lorsqu'un constituant est focalisé, il semblerait que la séquence qui le suit voit une baisse importante de sa protrusion (les barres V&FS, O&FS et O&FV sont en dessous de 1). Ceci se vérifie dans 75% des cas V&FS, 87,5% des cas O&FS et 73,1% des cas O&FV. La baisse est d'en moyenne 25,8% (-29,2% pour V&FS, -28,1% pour O&FS et -22,9% pour O&FV) et est significative[112] : t=-6,164 (p<0,001).

En ce qui concerne la protrusion, il semble donc que le locuteur B ait tendance à hyper-articuler le constituant focal alors qu'il hypo-articule la séquence post-focale créant ainsi un contraste visible important au sein de l'énoncé entre ce qui est focalisé et ce qui ne l'est pas. Il ne marque cependant aucune variation de la protrusion sur la séquence pré-focale.

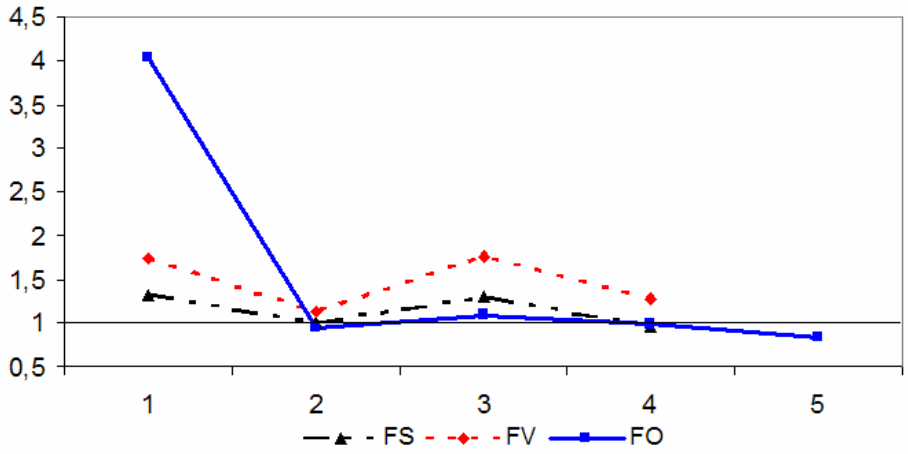


FIGURE III.28 – *Moyennes des données normalisées de protrusion de la lèvre supérieure sur le syntagme focalisé pour chaque type de focalisation et en fonction du nombre de syllabes du constituant focalisé considéré : résultats pour le locuteur B.*

Le graphique de la figure III.28 représente les données normalisées de protrusion de la lèvre supérieure des constituants focalisés en fonction du nombre de syllabes qu'ils contiennent (en abscisse) et du type de focalisation considéré (différentes courbes). On constatera que l'échelle est beaucoup plus étendue que pour les figures du même type qui ont été exposées précédemment (figures III.21, III.25 et III.26). Ceci s'explique par le fait que pour deux énoncés focalisés sur l'objet, on mesure une augmentation de la protrusion beaucoup plus importante que pour tous les autres énoncés. Il s'agit de la phrase (5) dans le cas où l'objet 'Li' est focalisé. On s'étonnera de cette

[110] Test t de comparaison à 1.
[111] Test t de comparaison à 1.
[112] Test t de comparaison à 1.

augmentation puisque lorsque 'Li' est focalisé c'est l'étirement des lèvres qui augmente et donc éventuellement l'aire intéro-labiale et non la protrusion. Cependant, chez ce locuteur, lorsque l'étirement augmente la lèvre supérieure rentre et la valeur absolue de la protrusion (mesure qui a été effectuée, cf. section A.2.3.1.c du présent chapitre) augmente ainsi. Or dans le cas neutre, la protrusion est quasi nulle sur 'Li'. Lorsque les valeurs de protrusion correspondant aux cas focalisés sont normalisées, dans ce cas précis elles sont divisées par une valeur très proche de 0. Donc bien que l'augmentation absolue de la protrusion liée à la lèvre supérieure qui rentre soit faible, suite à l'opération de normalisation, elle devient très importante.

A.2.3.4. Conclusion : stratégie de focalisation du locuteur B

A.2.3.4.a. Constituant focalisé

Les syllabes focales sont significativement allongées dans 81,8% des cas (intra : +25,9% inter : +20,6%). Cet ***allongement focal*** est encore plus important pour le premier segment du constituant focalisé (+28,4%). Cet ***allongement du premier segment focal*** semble mettre encore plus en relief l'endroit où commence le syntagme focalisé et est beaucoup plus marqué lorsque la focalisation porte sur l'objet que lorsqu'elle porte sur le verbe ou le sujet. L'explication est certainement la même que celle qui avait été donnée dans la section A.2.2.3.f.i du présent chapitre pour le locuteur A, *i.e.* que le sujet étant en position initiale dans l'énoncé, il subit peut-être un allongement inhérent de son premier segment qu'il soit focalisé ou non, ce qui correspondrait à une stratégie de marquage du début de l'énoncé.

Il y a de plus hyper-articulation focale de l'aire intéro-labiale (intra : +25,3% inter : +6,8%) et ce dans 63,6% des cas et de la protrusion (intra :+59,3% inter : +36,9%) dans 61,1% des cas. On n'observe pas de contraste intra-énoncé en ce qui concerne la vitesse de variation de l'aire intéro-labiale mais ceci est dû au fait que les pics de vitesse du sujet sont toujours augmentés, peut-être dans le but d'indiquer le début de l'énoncé. En réalité on observe bien une augmentation de l'amplitude des pics de vitesse du constituant focalisé par rapport au cas neutre (+11,5%) dans 62,1% des cas.

Le graphique a. de la figure III.29 fournit les données de durées, d'aire intéro-labiale et de protrusion du constituant focalisé pour chaque type de syntagme (S, V et O). Il permet de constater que l'hyper-articulation (en aire intéro-labiale et en protrusion) et l'allongement sont plus marqués lorsque la focalisation porte sur le verbe. Or dans ce corpus les syntagmes objets sont en moyenne plus longs que les sujets, eux-mêmes plus longs que les verbes (2,6 syllabes en moyenne pour les sujets, 2,4 pour les verbes et 2,9 pour les objets). Le graphique b. de la figure III.29 donne les résultats de durées, d'aire intéro-labiale et de protrusion du constituant focalisé en fonction du nombre de syllabes du constituant en question. On constate grâce à ce graphique que, en général, plus le nombre de syllabes est important, moins l'allongement (resp. l'hyper-articulation en aire intéro-labiale et en protrusion) est important. Le point correspondant à la protrusion pour les constituants monosyllabiques est anormalement élevé car, dans ce corpus, les constituants monosyllabiques étaient les constituants pour lesquels il y avait le plus de voyelles protruses. On constate donc que plus il y a de syllabes dans un constituant, moins le marquage de la focalisation sur celui-ci est net. Les verbes étant ainsi les plus courts constituants (en moyenne), ils sont globalement plus marqués. Cette observation sera reprise et développée dans la section A.2.4.3 du chapitre II.

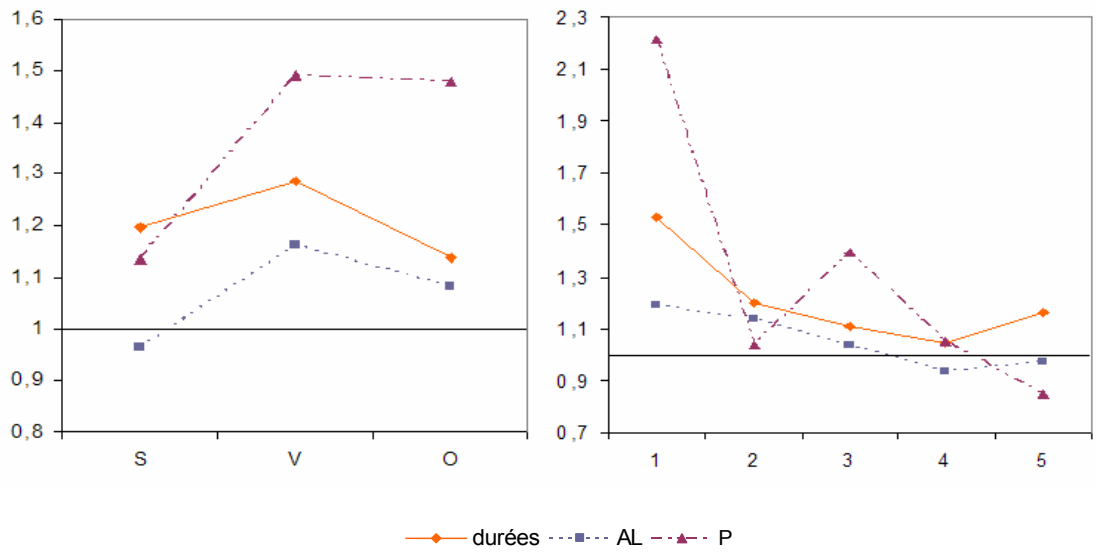


FIGURE III.29 – *Données normalisées de durées, d'aire intéro-labiale (S) et de protrusion (P) corespondant aux constituants focalisés en fonction a.(gauche) du type de syntagme (S, V ou O) et b.(droite) du nombre de syllabes du constituant.*

Les graphiques de la figure III.30 représentent les trois données de durées, d'aire intéro-labiale et de protrusion des constituants focalisés les unes en fonction des autres. Chaque point correspond ainsi à un constituant focalisé. Si ce point se trouve dans le cadran en haut à droite (les deux données sont supérieures à 1) c'est que les deux paramètres sont saillants lors de la focalisation. Si le point se trouve soit dans le cadran en haut à gauche soit dans celui en bas à droite c'est qu'un seul des deux paramètres a été marqué et s'il se trouve dans le cadran en bas à gauche c'est qu'aucun des deux paramètres du graphe n'a été utilisé. Globalement, on constate que, dans la majorité des cas, au moins deux des trois paramètres représentés sont augmentés. En effet, le plus souvent lorsqu'un point se trouve dans le cadran en bas à gauche sur l'un des graphiques, il n'y est plus pour les deux autres graphiques. Cependant il existe quelques cas pour lesquels aucun des paramètres n'est augmenté mais ces cas représentent 3,8% des données seulement. On pourra ainsi conclure que dans la plupart des cas, le locuteur B se sert au moins d'un des paramètres parmi l'allongement, l'hyper-articulation de l'aire intéro-labiale ou celle de la protrusion pour mettre en valeur le constituant focalisé. Notons, qu'ainsi qu'il a été décrit plus haut, d'autres paramètres pourraient être examinés, tel que le contraste entre syntagme focal et syntagmes pré- ou post-focaux ou les pics de vitesse de l'aire intéro-labiale. La combinaison des trois paramètres décrits dans les graphes ci-dessous avec ces paramètres supplémentaires pourrait ainsi peut-être expliquer les 3,8% de données restantes.

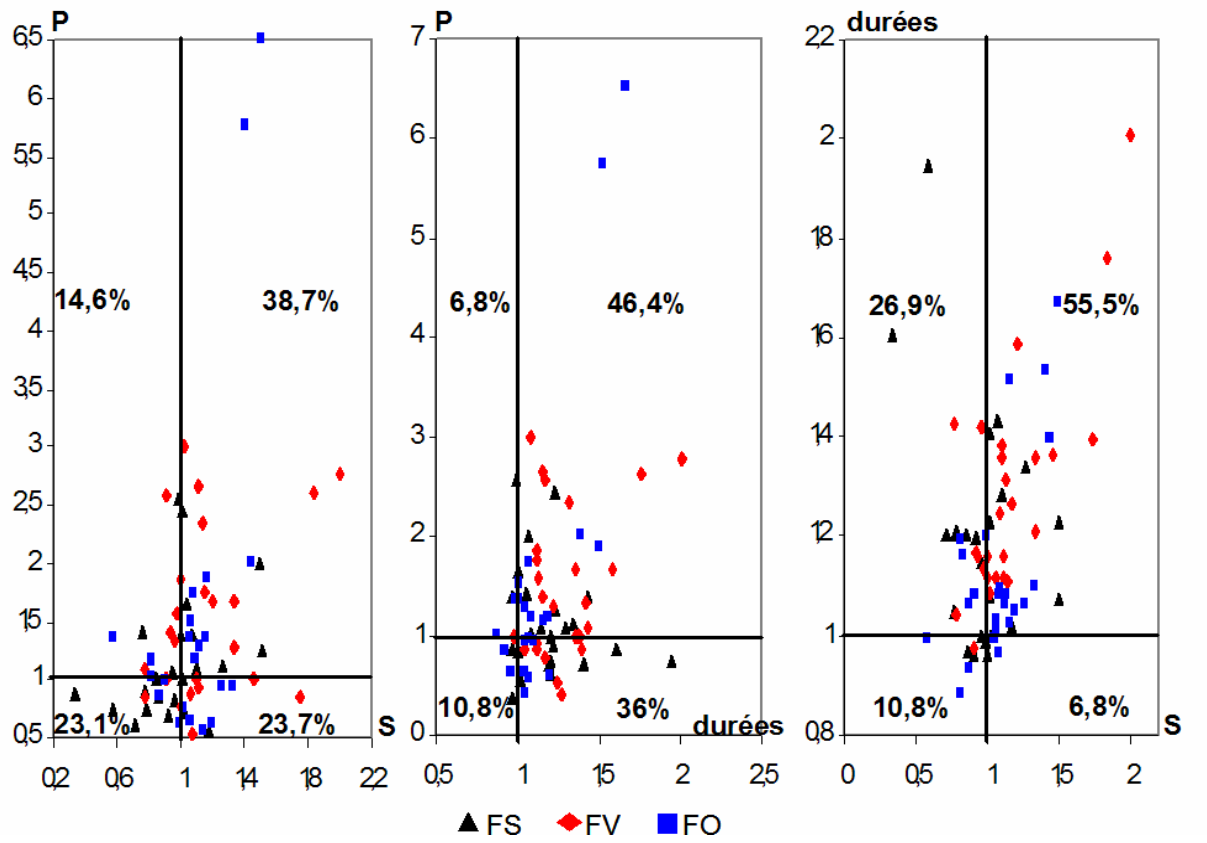


FIGURE III.30 – *a. (gauche) données normalisées de protrusion en fonction de celles d'aire intéro-labiale pour chaque énoncé focalisé enregistré en fonction du constituant focalisé (S, V ou O) ; b. (centre) même graphique pour les données normalisées de protrusion en fonction de celles de durées ; c. (droite) même graphique pour les données normalisées de durées en fonction de celles d'aire intéro-labiale. Les pourcentages correspondent au pourcentage des énoncés contenu dans chaque cadran. Par exemple, 14,6% des énoncés focalisés correspondent à une augmentation de la protrusion (donnée supérieure à 1) mais à une diminution de l'aire intéro-labiale (donnée inférieure à 1).*

A.2.3.4.b. Séquence pré-focale

On note une nette diminution de la durée des syllabes de la séquence pré-focale dans 73,6% des cas (-7,1%). Cette diminution est plus marquée pour la séquence pré-focale correspondant à une focalisation sur le verbe. On note également une diminution significative de l'aire intéro-labiale (-6,5%) correspondant au constituant directement pré-focal.

A.2.3.4.c. Séquence post-focale

Les syllabes de la séquence post-focale voient une très légère (mais significative) diminution de leur durée dans 74,3% des cas (-1,8%). Elles sont de plus nettement moins articulées aussi bien au niveau de l'aire intéro-labiale (en moyenne, dans 87,9% des cas, de 22,1%) qu'au niveau de la protrusion (en moyenne, dans 78,5% des cas, de 25,8%). On note ainsi une ***hypo-articulation post-focale***. Celle-ci est plus importante après le sujet focalisé qu'après le verbe.

A.2.3.4.d. Bilan

La figure III.31 résume les données d'aire intéro-labiale, de protrusion et de durées en fonction de la séquence considérée : focalisée, pré-focale ou post-focale pour le locuteur B. La droite correspondant à la valeur 1 signale la position du cas neutre. Cette figure permet de résumer ce qui a été décrit en

détails ci-dessus. Il semblerait que lorsque le locuteur B focalise un constituant, il l'allonge et accentue son aire intéro-labiale et sa protrusion. Néanmoins, le constituant focalisé n'est pas le seul affecté puisque la séquence pré-focale voit sa durée et son aire intéro-labiale diminuer. La séquence post-focale quant à elle voit son aire intéro-labiale et sa protrusion considérablement diminuer aussi. L'énoncé dans son intégralité est donc affecté par la focalisation. Il semblerait ainsi que ce locuteur répartisse son effort entre la mise en relief de ce qui est focalisé (hyper-articulation et allongement) et l'atténuation de ce qui n'est pas focalisé (hypo-articulation essentiellement post-focale). Cette constatation est intéressante puisque les études sur la focalisation se sont souvent intéressées uniquement au constituant focalisé lui-même. Pourtant il semblerait que tout l'énoncé soit affecté comme l'ont d'ailleurs déjà suggéré Summers [1987] et Erickson [1998] pour l'articulation et Lehiste [1970] pour la durée.

La figure III.31 permet de plus de constater que c'est la protrusion qui est la plus augmentée lors de la focalisation tant par rapport au cas neutre (droite à 1) que par rapport au reste de l'énoncé (foc vs. pre-foc et post-foc).

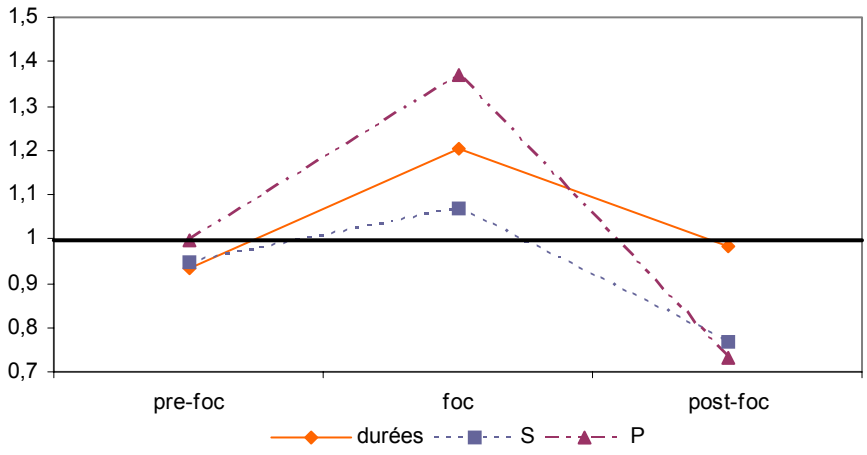


FIGURE III.31 – *Données de durées, d'aire intéro-labiale et de protrusion pour le locuteur B en fonction de la séquence considérée de l'énoncé comportant un constituant focalisé (foc : constituant focalisé, pre-foc : séquence pré-focale et post-foc : séquence post-focale).*

A.2.4. Comparaisons inter-locuteurs et conclusions

A.2.4.1. Remarque générale sur les deux locuteurs

On soulignera ici le fait que le locuteur A était habitué et entraîné à être enregistré pour des études en production de la parole. Ses productions se révèlent ainsi très nettes voire exagérées. Le locuteur B était quant à lui « naïf » et n'avait jamais participé à de tels enregistrements.

A.2.4.2. Bilan des stratégies mises en place : variance et invariance

La figure III.32 résume les comportements des deux locuteurs (A et B) lors de la focalisation, à la fois sur les éléments pré-focal, focal et post-focal. Pour le locuteur A, on avait observé une nette augmentation par rapport au cas neutre pour l'élément focal alors que cela n'est pas le cas pour le locuteur B. Par contre, on n'avait pas observé de baisse significative de l'aire intéro-labiale pour la

séquence post-focale chez le locuteur A alors qu'on en observe une nettement significative chez le locuteur B. On peut ainsi conclure que la stratégie du locuteur A est d'hyper-articuler nettement l'élément focalisé afin de le mettre en relief à la fois par rapport au reste de l'énoncé et par rapport à la façon dont il serait articulé dans le cas neutre. Quant au locuteur B, il semblerait qu'il répartisse son effort en hyper-articulant un peu l'élément focalisé mais surtout en hypo-articulant ce qui suit créant ainsi un contraste avec le reste de l'énoncé. On constatera ainsi que, bien qu'ils le fassent de façons différentes, les deux locuteurs créent un contraste articulatoire au sein de l'énoncé. Ce contraste est plus important chez le locuteur A (38,2%) que chez le locuteur B (25%). Cependant on notera que, comme il a été expliqué au début de cette section, le locuteur A est un locuteur entraîné puisqu'il a déjà participé à de nombreuses expériences pour la recherche sur la parole : il a ainsi sans doute pris l'habitude d'articuler très clairement et le fait maintenant spontanément. Quant au locuteur B, il s'agit d'un locuteur complètement naïf qui n'avait au préalable jamais participé à des enregistrements. En outre, il semble, d'après les témoignages de son entourage, qu'il ait tendance à peu articuler de façon générale. Le compromis entre l'hypo-articulation pré- ou post-focale et l'hyper-articulation focale est donc probablement compris entre ces deux extrêmes pour un locuteur moyen et c'est ce que nous tenterons d'analyser pour les autres enregistrements effectués.

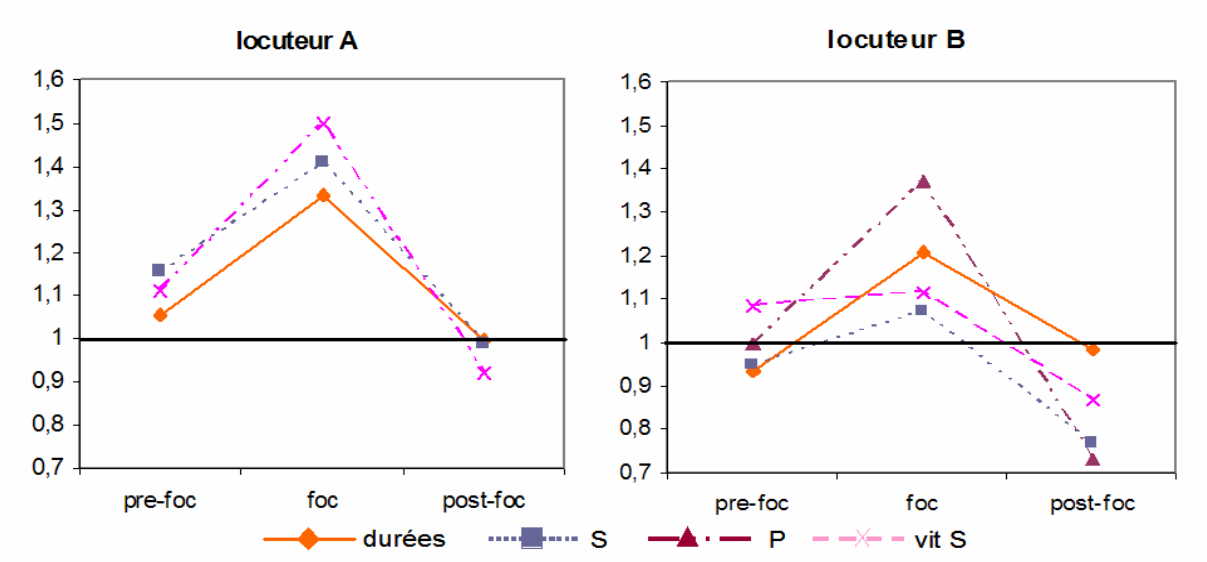


FIGURE III.32 – *Données normalisées de durées, d'aire intéro-labiale, de vitesse de variation de l'aire intéro-labiale et de protrusion pour les locuteurs A (gauche) et B (droite) en fonction de la séquence considérée de l'énoncé comportant un constituant focalisé (foc : constituant focalisé, pre-foc : séquence pré-focale et post-foc : séquence post-focale). NB : les données de protrusion ne sont pas disponibles pour le locuteur A.*

A.2.4.3. Discussion sur l'intensité du marquage « visible »

On a constaté pour les deux locuteurs que plus le nombre de syllabes du constituant à focaliser était important moins le marquage « visuel » de la focalisation était net (allongement et hyper-articulation moins forts). On peut proposer deux explications à ce phénomène.

La première est celle d'une répartition d'un même « effort » entre un nombre de syllabes plus ou moins grand. Le locuteur disposerait ainsi d'une certaine quantité d'énergie à utiliser et serait obligé de répartir celle-ci entre toutes les syllabes à focaliser. Il est en effet probable que l'hyper-articulation

soit difficile à produire sur une durée prolongée. Il s'agirait donc d'une stratégie « dépendante du locuteur » (selon la terminologie de Lindblom [1990]) .

La seconde serait liée au fait que moins il y a de syllabes dans le constituant focalisé, moins il y aura de syllabes marquées par la focalisation. On peut ainsi penser que le locuteur cherche à mieux marquer ces syllabes pour qu'elles soient mieux mises en valeur. Dans le cas où le constituant focalisé contient suffisamment de syllabes, il suffirait d'un léger marquage sur l'ensemble de ses syllabes pour qu'il soit mis en relief par rapport au reste de l'énoncé et que la focalisation soit visible. Il s'agirait donc là d'une stratégie « dépendante du spectateur ».

A.3. Analyse de données Optotrak

Il apparaît maintenant de façon claire qu'il existe des corrélats articulatoires visibles de la focalisation contrastive en français. Néanmoins les études précédentes montrent aussi qu'il existe différentes stratégies de focalisation d'un locuteur à l'autre donnant ainsi lieu à la production de corrélats qui peuvent être différents. C'est pourquoi il semble nécessaire d'approfondir l'étude en analysant les productions articulatoires d'autres locuteurs. De façon à avoir des données à la fois de même nature mais aussi complémentaires (certains phénomènes articulatoires n'avaient pas pu être correctement mesurés avec le système labiométrique), il a été décidé d'utiliser un autre système de mesure. C'est un système de suivi tridimensionnel de points de l'espace qui a été choisi : l'Optotrak.

A.3.1. Mise en œuvre expérimentale

A.3.1.1. Dispositif de mesures tridimensionnelles : l'Optotrak

L'Optotrak de Northern Digital est un système de suivi de capteurs infra-rouge (IR). Il capte la lumière infrarouge émise par des diodes IRED (Infra-Red Emitting Diods) placées sur le sujet dont on souhaite mesurer les mouvements. Il est composé de trois caméras sensibles à l'infrarouge et permet d'obtenir la position (coordonnées tri-dimensionnelles) des diodes au cours du temps. Les coordonnées tri-dimensionnelles sont exprimées dans un repère absolu arbitraire.

La position des caméras de suivi est pré-calibrée. Le système étant optique, il faut veiller à ce qu'aucun des marqueurs ne soit caché pendant les mesures puisque les caméras ne le détecteraient alors plus. Il peut arriver qu'un ou plusieurs marqueurs « sortent » du champ des caméras (par exemple quand le locuteur tourne trop la tête d'un côté ou de l'autre).

Un dispositif comprenant quatre marqueurs (marqueurs 1-4 sur la figure III.34 et voir figure III.33) est fixé à la tête du locuteur afin d'obtenir une approximation correcte des mouvements de ce qui est appellé le *« rigid body »* (corps rigide) et qui correspond en fait ici à la tête du sujet. Les mouvements de la tête peuvent ainsi être soustraits aux mouvements des marqueurs placés sur le visage. Dans notre cas, le calcul des mouvements du corps rigide a permis d'obtenir les mouvements de la tête du sujet dans le repère absolu. Ces mouvements sont caractérisés par trois angles de rotation selon les trois axes du repère et par trois translations. Le logiciel de l'Optotrak calcule ensuite les coordonnées tri-dimensionnelles des marqueurs dans le repère relatif de la tête ce qui permet d'obtenir les mouvements relatifs des marqueurs placés sur le visage par rapport à la tête du sujet. L'avantage des

marqueurs utilisés avec l'Optotrak est qu'ils ne sont pas sensibles aux interférences provenant des réflexions ou de l'éclairage ambiant. De plus, les coordonnées tri-dimensionnelles des marqueurs dans le repère absolu sont déterminées en temps réel.

Les photos de la figure III.33 donne une idée de l'appareillage et de la mise en place expérimentale.

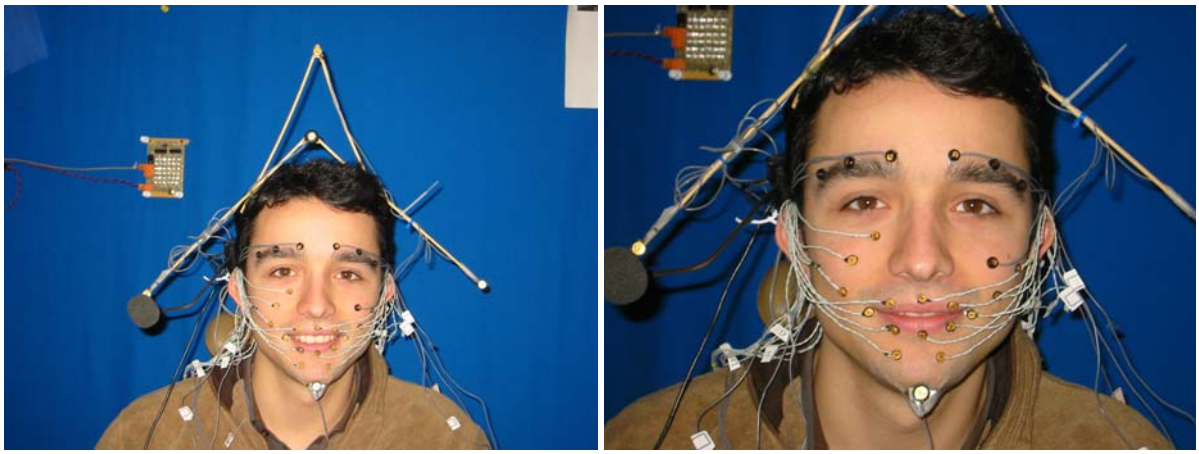


FIGURE III.33 – *Photo donnant un aperçu du dispositif d'enregistrement des données Optotrak.*

A.3.1.2. Protocole expérimental

A.3.1.2.a. Données techniques

Les enregistrements Optotrak ont été effectués aux laboratoires Advanced Telecommunications Research (ATR) au Japon avec la collaboration de Harold Hill. Disposant de deux Optotrak identiques (Optotrak 3020 de Northern Digital), nous les avons associé afin d'obtenir un total de 6 caméras. Grâce à la mise en place d'un tel dispositif, il a été possible de limiter énormément la « perte » d'un marqueur, due par exemple à une trop forte rotation de la tête du sujet. Les positions des marqueurs ont été échantillonnées à 60Hz.

Le signal acoustique a été enregistré simultanément aux données spatiales sur deux canaux différents. L'enregistrement avec le microphone intégré à l'Optotrak était trop bruité et un enregistrement simultané avec un microphone directionnel très proche de la bouche du locuteur a donc été effectué. L'Optotrak peut en effet enregistrer de façon synchrone deux canaux audio indépendants. L'échantillonnage des signaux audio s'est fait à 22kHz.

A.3.1.2.b. Positionnement des diodes IRED

Un total de 28 marqueurs (dont les quatre sur le dispositif de la tête, cf. section A.3.1.1 du présent chapitre) a été placé sur le visage des locuteurs grâce à un adhésif double face spécialement conçu à cet effet. Leurs positions sont référencées sur le schéma de la figure III.34. Huit marqueurs ont ainsi été disposés sur les lèvres, trois sur chaque sourcil, un sur la partie semi-rigide en dessous du menton (pour obtenir les mouvements de la mandibule) et les autres sur les joues.

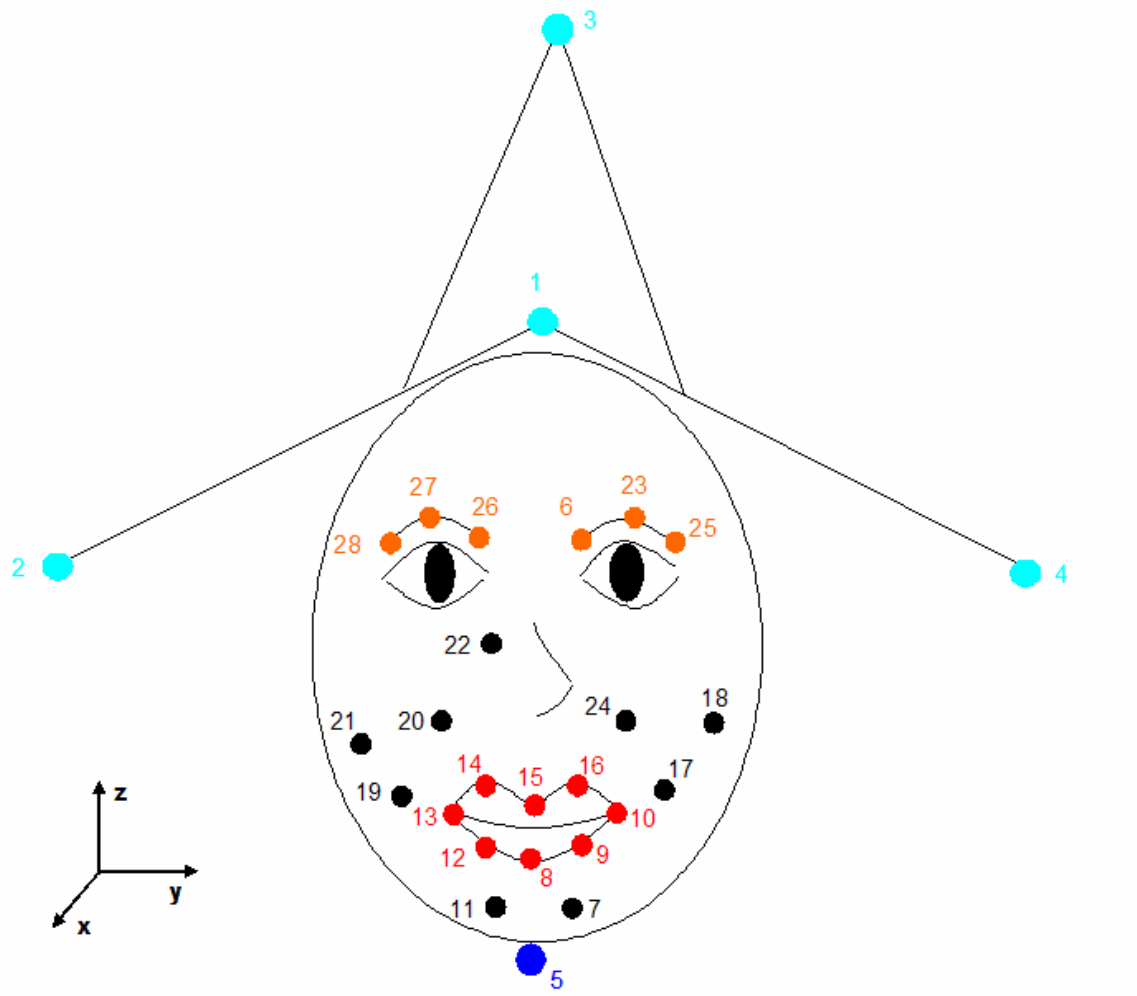


FIGURE III.34 - *Schéma de principe des positions des 28 marqueurs tels qu'ils ont été disposés sur les visages des cinq locuteurs.*

A.3.1.2.c. Corpus et enregistrements

Le corpus qui a été utilisé pour les enregistrements est le corpus AV2 qui a été décrit dans la section A.2.3.1.a du présent chapitre. Chacune des treize phrases du corpus a été enregistrée sous quatre conditions de focalisation (sujet, verbe, objet et version neutre) et à chaque fois deux répétitions ont été effectuées. Au total 104 énoncés ont donc été enregistrés. Les enregistrements ont été effectués dans une salle isolée du bruit extérieur. Avant le début de l'enregistrement, une séance de calibration a été effectuée pour l'Optotrak.

La méthode d'obtention de la focalisation utilisée est celle décrite à la section A.2.3.1.b du présent chapitre. Les enregistrements ont été effectués en trois phases. La première a permis d'enregistrer toutes les versions des phrases (5), (6), (11), (12), (13) réparties dans un ordre aléatoire. La deuxième était constituée des phrases (7), (8), (9) et (10) et enfin la troisième des phrases (1), (2), (3) et (4) toujours dans des ordres aléatoires. Ce découpage en trois phases a été décidé afin de permettre aux locuteurs de se reposer entre chaque phase et par exemple de boire si besoin était.

A.3.1.2.d. Locuteurs

Cinq locuteurs français de langue maternelle française ont été enregistrés sur une période de cinq jours. Ces locuteurs sont les locuteurs B, C, D, E et F[113]. On note donc que le locuteur B avait déjà été enregistré pour le même corpus avec la technique de suivi automatique du contour des lèvres (cf. section A.2.3 du présent chapitre). Ceci nous permettra donc par la suite d'effectuer des comparaisons entre les deux techniques de mesures.

A.3.2. Méthodologie d'analyse

A.3.2.1. Mesures

Les analyses effectuées avec le système de détection automatique du contour des lèvres et décrites dans la section A.2 du présent chapitre, ont montré que la focalisation contrastive était signalée par une hyper-articulation de l'élément focalisé (augmentation de l'amplitude des gestes articulatoires) et une augmentation de la durée de ce même élément. On observe de plus chez certains locuteurs une hypo-articulation (réduction de l'amplitude des gestes articulatoires) de la séquence pré- ou post-focale. On observe aussi parfois encore une hyper-articulation de l'élément directement pré-focal accompagnée d'un allongement de ce même élément correspondant vraisemblablement à la mise en place d'une stratégie d'anticipation de la focalisation contrastive.

Les gestes articulatoires analysés précédemment étaient l'aire aux lèvres (et donc l'ouverture et l'étirement des lèvres), les mouvements de la mandibule et la protrusion. Avec les données Optotrak nous n'avons pas accès à l'aire aux lèvres. Nous pouvons cependant extraire des données donnant une bonne estimation de l'ouverture et de l'étirement des lèvres. L'ouverture des lèvres correspond ainsi à la différence des coordonnées z des marqueurs 8 et 15 (voir figure III.34). L'étirement des lèvres correspond quant à lui à la différence des coordonnées y des marqueurs 10 et 13 (voir figure III.34). Les données Optotrak permettent également d'accéder à la protrusion de la lèvre supérieure. La protrusion correspond en effet à la coordonnée x du marqueur 15 (voir figure III.34). L'intérêt de l'Optotrak du point de vue articulatoire est que la mesure de la protrusion est, selon nos propres observations, plus précise que celle effectuée par le système de suivi automatique du contour des lèvres. Une assez bonne approximation des mouvements de la mandibule peut également être obtenue grâce à la coordonnée z du marqueur 5 (voir figure III.34). Dans l'étude articulatoire à partir des données vidéo, le suivi des mouvements de la mandibule était réalisé par le suivi d'une pastille bleue collée au milieu du menton du locuteur (cf. section A.2.1.1.b du présent chapitre). Si cette technique est la seule réalisable à partir de données vidéo de face, elle est toutefois, comme il a déjà été suggéré dans la section A.2.2.2.d.i du présent chapitre, imparfaite. En effet, il a été montré que les mouvements du menton ne reflètent pas adéquatement les mouvements de la mandibule. Dans la séquence /iba/ par exemple, la mandibule commence à baisser pendant que la lèvre inférieure monte pour le /b/, entraînant le menton vers le haut (Badin, communication personnelle). Dans cette nouvelle étude à l'aide de l'Optotrak, il m'a été possible de suivre la cinématique d'un point situé sous le menton (marqueur 5) et reflétant donc de façon probablement plus adéquate les mouvements de la mandibule. D'où un autre intérêt d'utiliser cette technique (Optotrak) pour faire des mesures articulatoires.

[113] Voir la section D du chapitre *Notes et indices de lecture* pour un récapitulatif des locuteurs.

Dans la suite, les abréviations suivantes, déjà exposées auparavant, seront à nouveau utilisées :

- A : étirement des lèvres ;
- B : ouverture des lèvres ;
- P : protrusion ;
- MM : mouvements de la mandibule.

Les durées de toutes les syllabes ont également été mesurées grâce à la segmentation acoustique effectuée selon la méthode décrite à la section C.1.4.1 du chapitre II.

A.3.2.2. Mise en forme des données

A.3.2.2.a. Extraction des données

Comme il a été expliqué plus tôt, après traitement par le logiciel de Northern Digital fonctionnant avec l'Optotrak, on obtient les coordonnées tridimensionnelles de toutes les diodes IRED placées sur le visage du locuteur. Ces coordonnées tridimensionnelles sont exprimées dans le repère de la tête, il s'agit donc des mouvements relatifs des diodes (et donc des points du visage considérés) par rapport à la tête. Le logiciel donne également les rotations et les translations du corps rigide qui correspond approximativement à la tête. Ces rotations et translations sont exprimées dans le repère absolu arbitraire évoqué plus haut.

Pour calculer les paramètres A et B, nous avons donc calculé des différences de coordonnées. Les paramètres P et MM étaient obtenus directement par extraction de la coordonnée en question (cf. section A.3.2.1 du présent chapitre).

A.3.2.2.b. Mise à référence

L'évolution temporelle de chaque paramètre pouvait ensuite être visualisée avec le signal acoustique et la segmentation acoustique grâce à l'application Trap (cf. section A.2.1.2.c du présent chapitre). Pour chaque paramètre, c'est la variation de celui-ci par rapport à une position de référence qui nous intéressait. Une détection manuelle de cette référence a été effectuée pour chaque énoncé. Il s'agissait en fait de déterminer les instants entre lesquels une moyenne était ensuite calculée. Cette valeur moyenne correspondait ensuite à la référence qui était retranchée à toutes les valeurs du paramètre pour l'énoncé considéré. Pour le paramètre d'ouverture des lèvres (B), les valeurs ainsi obtenues étaient toujours positives, une valeur nulle correspondant à une fermeture des lèvres. Cependant les valeurs obtenues après mise à référence pour les paramètres A, P et MM pouvaient êtres alternativement positives ou négatives (lèvre supérieure qui « rentre » ou qui avance, mandibule qui monte ou qui descend, lèvres qui s'étirent ou se rétractent). Ce qui nous intéressait de façon générale était de savoir si l'articulation était augmentée ou au contraire réduite, peu importe dans quel sens elle s'effectuait. Or pour tous ces paramètres, aussi bien des valeurs positives qui augmentent que des valeurs négatives qui diminuent représentent une articulation augmentée. Ce sont donc les valeurs absolues des mesures (après soustraction de la valeur de référence) pour les paramètres A, P et MM qui ont été analysées.

A.3.2.2.c. Mise en forme

C'est ici l'aire sous la courbe d'évolution temporelle des paramètres obtenus après mise en forme qui a été mesurée selon la technique décrite à la section A.2.2.3.c.ii du présent chapitre. C'est ainsi une

donnée représentant l'amplitude moyenne des mouvements articulatoire qui est obtenue. Cette aire a été calculée sur chacun des syntagmes (S, V et O) de chacun des énoncés et pour chaque paramètre. Après normalisation par la durée, on obtient finalement une valeur d'aire normalisée pour chaque syntagme de chaque énoncé (trois valeurs par énoncé). Une normalisation globale a ensuite été effectuée selon le principe exposé à la section A.2.2.3.c.iii du présent chapitre avec comme base les valeurs d'aires mesurées pour les énoncés neutres. Comme il a été expliqué à la section A.2.2.3.c.iii du présent chapitre, le but de cette normalisation est de pouvoir comparer les valeurs mesurées à travers tout le corpus. En effet, il existe une variabilité intrinsèque importante due à l'utilisation d'un grand nombre de syllabes différentes dans le corpus. Après normalisation, les cas neutres correspondront à la valeur 1, une valeur supérieure à 1 correspondra à une augmentation par rapport au cas neutre et une valeur inférieure à 1 correspondra à une diminution par rapport au cas neutre.

A.3.2.3. Attentes a priori et présentation des résultats

Les premières études menées grâce au système de détection automatique du contour des lèvres qui ont été décrites dans la section A.2 du présent chapitre, permettent de se faire une idée de ce que sera l'évolution des données mesurées grâce à l'Optotrak. Les hypothèses formulées grâce à ces études sont décrites ci-dessous. Les résultats des mesures seront ensuite présentés selon le même schéma afin de voir clairement pour chaque locuteur et pour chaque mesure effectuée si les hypothèses sont vérifiées ou non.

A.3.2.3.a. Durées

On s'attend à obtenir :

- contrastes intra-énoncés[114] :
  - → une durée moyenne normalisée significativement plus grande pour les syllabes de l'élément focalisé que pour les autres syllabes du même énoncé probablement chez tous les locuteurs ;
- comparaison des énoncés focalisés avec les mêmes énoncés dans le cas neutre :
  - → une augmentation significative de la durée moyenne des syllabes focales probablement chez tous les locuteurs ;
  - → pas de variation de la durée des syllabes post-focales probablement chez tous les locuteurs ;
  - → un allongement éventuel de la durée de la syllabe pré-focale seulement chez certains locuteurs.

[114] Contraste intra-énoncé : contraste entre ce qui est focalisé et ce qui ne l'est pas au sein d'un énoncé.

A.3.2.3.b. Mesures articulatoires

On s'attend à obtenir :

- contrastes intra-énoncés[115] :
  - → une amplitude moyenne normalisée des gestes articulatoires significativement plus importante sur l'élément focalisé que sur le reste du même énoncé probablement chez tous les locuteurs ;
- comparaison des énoncés focalisés avec les mêmes énoncés dans le cas neutre :
  - → une hyper-articulation (augmentation significative de l'amplitude moyenne des gestes articulatoires) de l'élément focalisé appelée ci-après ***hyper-articulation focale***, probablement chez tous les locuteurs ;
  - → une hypo-articulation (réduction significative de l'amplitude des gestes articulatoires) éventuelle de la séquence post-focale appelée ci-après ***hypo-articulation post-focale***, seulement chez certains locuteurs ;
  - → hyper- ou hypo-articulation éventuelle directement avant l'élément focalisé appelée ci-après ***hyper-articulation pré-focale*** ou ***anticipation pré-focale***, seulement chez certains locuteurs.

A.3.2.3.c. Analyses statistiques

Les analyses statistiques effectuées sont conformes à celles décrites à la section A.2.1.2.d du présent chapitre et seront présentées sous la même forme (tableaux de résultats).

A.3.3. Bilan des résultats

Afin de ne pas surcharger la lecture, les résultats et analyses statistiques sont décrits en détail en annexe 3. Ils sont aussi consignés et résumés dans le tableau III.9 ci-dessous. Cette section a pour but de donner une description synthétisée des indices « visibles » de la focalisation contrastive pour chaque paramètre et de présenter la stratégie « visible » de focalisation de chaque locuteur. Une généralisation globale sera également effectuée.

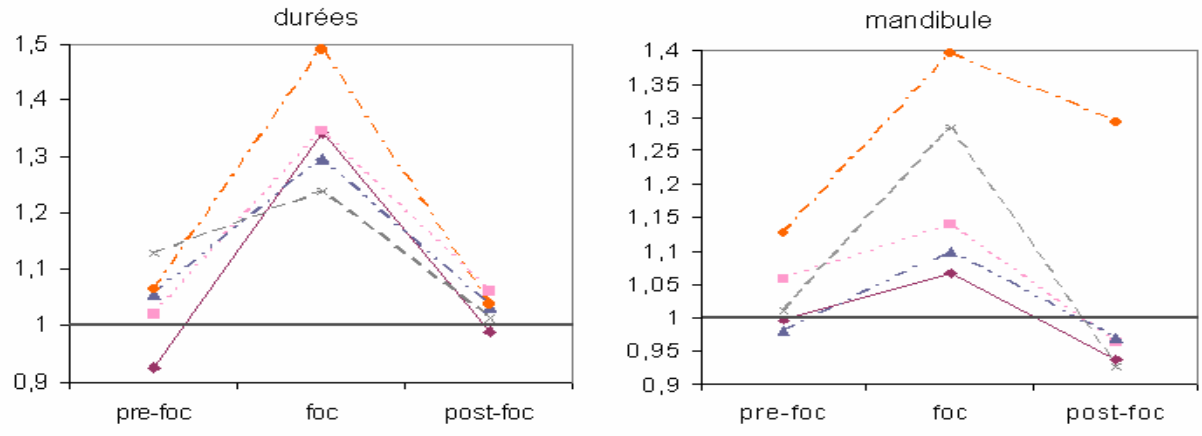
		Locuteur B		**Locuteur C**		**Locuteur D**		**Locuteur E**		**Locuteur F**	
séquence pré-focale	**D**	-4,6%	FO>FV	+4,7%	FO>>FV	+8,1%	FV>FO A	+16,4%	FV>>FO	+9%	FV>>FO A
	MM	(-0,2%)		+6%	FV>FO	(+0,6%)	FV>>FO(X)	(+2,8%)	FO>>FV(X) A	(+11,1%)	FO>>FV(X)
	B	(-0,2%)		+6,8%	FV>FO	+5,7%	FV>FO A	-3,4%	FV>FO	+2,2%	FO>>FV(X)
	A	(+2,4%)	FV>FO(X)	(+3,7%)	FO>FV	(+0,5%)	FV>FO(X)	(+2,7%)	FO>>FV(X) A	(-3,5%)	FV>>FO(X)
	P	+25,1%	FO>>FV A	(+10,9%)	FO>>FV	(+4,4%)	FV>>FO(X)	+24,4%	FV>>FO	+53,7%	FV>FO
constituant focal		intra	inter	intra	inter	intra	inter	intra	inter	intra	inter
	D	+38,7%	+34,3%	+30,5%	+34,8%	+25,3%	+29,8%	+16,8%	+23,9%	+43,8%	+49%

[115] Contraste intra-énoncé : contraste entre ce qui est focalisé et ce qui ne l'est pas au sein d'un énoncé.

		FV>FO>FS	FV>FO>FS	FV>FO>FS	FV>FO>FS	FV>FS>>FO	FV>FS>>FO	FS>>FV>>FO	FS>FV>FO	FS>FV>>FO	FV>FS>>FO
	MM	+9,9%	+6,6%	+13,9%	+14%	+12,5%	+9,9%	+31,5%	+28,5%	+18,5%	+39,5%
		FV>FO>FS	FO>FV>FS	FS>>FV>>FO	FS>FV>FO	FS>>FO~FV	FS>FV>>FO	FO>FV>FS	FO>FV>FS	FO>>FV>>FS	FO>FS>>FV
	B	+13,8%	+8,7%	+11,1% (pas FO)	+13,9%	+8,4% (FO <1%)	+7,6%	+17,4%	+13,3%	+13,5%	+13,6%
		FV>FO>FS	FO>FV>FS	FS>>FV>>FO	FS>FV>FO	FS>>FV>>FO	FS>>FV>>FO	FS>FO>FV	FO>FV>FS	FS>FV>>FO	FS>FV>FO
	A	(+12,6%)	(+1,6%)	+17%	+17,3%	(+11%)	(+10,2%)	+19,4%	+14,5% (pas FS)	+23,2%	+30%
		FS>>FO>FV	FS>>FO>>FV	FS>>FO~FV	FS~FO>>FV	(FO>>FS~FV)	(FV>>FO~FS)	FV>FO>>FS	FV~FO>>FS	FO>FS>>FV	FS>FO>>FV
	P	+96,1%	+98,4%	+25%	+30,9%	+37,4%	+32%	+39,4%	+46%	+69,1% (pas FV)	+112,4%
		FO>FS>FV	FO>FS>FV	FS>FO>>FV	FO>>FS>FV	FS>FV>>FO	FV>FS>>FO	FS>>FV>>FO	FS>FV>>FO	FS>FO>>FV	FS>FO>>FV
séquence post-focale	**D**	invariance		+3,3%	FV~FS	+3,3%	FS>FV	invariance		+5,6%	FV>>FS
	MM	-6,5%	FV~FS	-4,4%	FV>FS	(-2%)	FS>>FV	-6,2%	FS>>FV	+23,8%	FS>>FV
	B	-8,9%	FV~FS	-2,2%	FS>>FV	-4,5%	FS>>FV	-8%	FS>>FV(X)	-2,3%	FV>FS
	A	-15,6%	FV>FS	(-8,1%)	FV>>FS	(+4,9%)	FV>>FS	-5,8%	FS>>FV(X)	(+8,4%)	FS>>FV(X)
	P	-8,1%	FS>>FV(X)	(-0,6%)	FS>>FV(X)	-1%	FV>FS	-12,9%	FS>FV	+41,6%	FS>FV

TABLE III.9 – *Résumé des résultats obtenus pour chaque paramètre et pour chaque locuteur (les résultats entre parenthèses sont non significatifs).*

A.3.3.1. Bilan inter-locuteurs



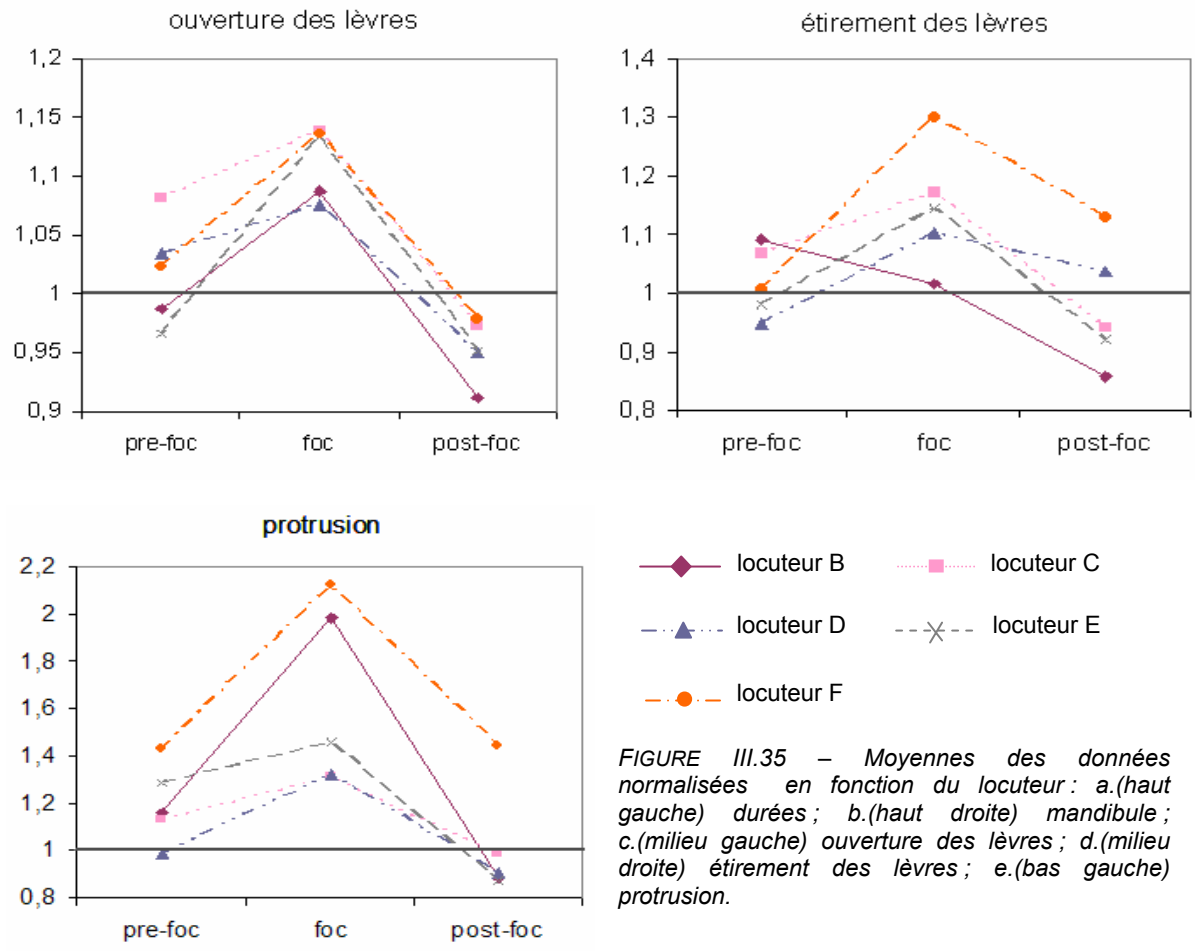


FIGURE III.35 – Moyennes des données normalisées en fonction du locuteur : a.(haut gauche) durées ; b.(haut droite) mandibule ; c.(milieu gauche) ouverture des lèvres ; d.(milieu droite) étirement des lèvres ; e.(bas gauche) protrusion.

A.3.3.1.a. Durées

En ce qui concerne la durée, on note un comportement assez homogène chez tous les locuteurs *i.e.* allongement des syllabes focales par rapport :

- au reste de l'énoncé : de +16,8% (E) à 43,8% (F) ; voir graphique a. de la figure III.35 : les points correspondant au constituant focalisé (foc) sont au-dessus des autres points des courbes ;
- au cas neutre : de +23,9% (E) à +49% (F) ; voir graphique a. de la figure III.35 : ces mêmes points sont largement au-dessus de 1.

Le locuteur F est celui qui allonge le plus les syllabes focales (contrastes intra- et inter-énoncés les plus importants) et le locuteur E celui qui les allonge le moins. On note que les plus grands contrastes sont observés pour les cas de focalisations sur le verbe et parfois aussi sur le sujet.

La plupart des locuteurs (tous sauf B) allonge aussi le constituant directement pré-focal, cet allongement variant de +4,7% (C) à +16,4% (E). Le seuil de perception audio (20%, cf. section C.2 du chapitre II) n'est cependant atteint par aucun des locuteurs. On peut penser que ceci est dû au fait qu'il s'agit ici de la moyenne des durées de toutes les syllabes du constituant pré-focal et non de la syllabe directement pré-focale uniquement. Or celle-ci est peut-être la seule qui soit allongée si on se base sur les résultats obtenus avec les données vidéos pour le locuteur A (cf. section A.2.2.3.d.ii du

présent chapitre). On notera que le locuteur B se comporte de façon totalement différente puisqu'il réduit de façon significative les durées des syllabes pré-focales (-4,6%).

En ce qui concerne la séquence post-focale, les résultats tendent vers une invariance des durées (locuteurs B et E) ou une très faible augmentation, certes significative, mais nettement inférieure au seuil de perception audio (de 3,3% à 5,6% pour les locuteurs C, D et F).

A.3.3.1.b. Mouvements de la mandibule

Concernant les mouvements de la mandibule, bien que les comportements diffèrent de façon quantitative, les tendances d'évolution sont les mêmes : on observe une hyper-articulation focale par rapport :

- au reste de l'énoncé : de +9,9% (B) à +31,5% (E) ; graphique b. de la figure III.35 : les points correspondant au constituant focalisé (foc) sont au-dessus des autres ;
- au cas neutre : de +6,6% (B) à 39,5% (F) ; voir graphique b. de la figure III.35 : ces mêmes points sont largement au-dessus de 1.

L'hyper-articulation la plus forte par rapport au cas neutre est relevée pour le locuteur F (+39,5%) mais le contraste intra-énoncé le plus net est noté chez le locuteur E (+31,5%). L'hyper-articulation la moins nette correspond au locuteur B.

On n'observe aucune variation significative concernant la séquence pré-focale sauf pour le locuteur C pour lequel on observe une hyper-articulation significative de toute la séquence pré-focale (+6%). Bien que celle-ci ne soit pas significative et ce certainement à cause des grands écarts-types observés, on notera tout de même qu'en moyenne il existe une forte hyper-articulation de la séquence pré-focale chez le locuteur F.

On observe enfin une hypo-articulation significative de la séquence post-focale chez les locuteurs B, C et E (de –4,4% (C) à –6,5% (B)), cette hypo-articulation est également notée chez le locuteur D mais elle n'est pas significative. De façon assez étonnante, on relève une très forte hyper-articulation significative de la séquence post-focale chez le locuteur F (+23,8%). Toutefois, cette hyper-articulation post-focale est moindre que celle du constituant focalisé.

A.3.3.1.c. Ouverture des lèvres

En ce qui concerne l'ouverture des lèvres, les comportements sont assez homogènes : on observe chez tous les locuteurs une hyper-articulation focale par rapport :

- au reste de l'énoncé : de 8,4% (D) à 17,4% (E) ; graphique c. de la figure III.35 : les points correspondant au constituant focalisé (foc) sont au-dessus des autres ;
- au cas neutre : de +7,6% (D) à +13,9% (C et F) ; graphique c. de la III.35 : ces mêmes points sont largement au-dessus de 1.

On notera que chez les locuteurs C et D, le contraste intra-énoncé est inexistant lorsque la focalisation porte sur l'objet. Chez le locuteur C, ceci est dû au fait que l'énoncé dans son intégralité semble être hyper-articulé et l'objet focalisé ne ressort donc pas par rapport au reste de l'énoncé. Chez le locuteur D, on notera qu'il y a très peu d'hyper-articulation focale. De façon générale, on relève les plus grands contrastes intra-énoncé pour les cas de focalisation sur le sujet. La quantité d'hyper-articulation est assez homogène, elle est d'en moyenne +12,8% pour les contrastes intra-énoncé et de +11,4% pour les contrastes inter-énoncés.

On observe deux tendances globales pour la séquence pré-focale : hyper-articulation de toute la séquence pré-focale (significative chez C et D et non significative chez F) et hypo-articulation (significative chez E et non significative chez B).

On note enfin une hypo-articulation significative de la séquence post-focale (de –2,2% (C) à –8,9% (B)) chez tous les locuteurs sauf F, chez qui on observe bien en moyenne une hypo-articulation mais non significative.

A.3.3.1.d. Étirement des lèvres

Concernant l'étirement des lèvres, aucune tendance nette n'est réellement observée. Rappelons que l'étirement avait été calculé comme la différence des coordonnées y des marqueurs placés à l'extérieur des commissures des lèvres. Or ces marqueurs sont très délicats à placer étant donné la conformation de la zone du visage en question. Etant placés à l'extérieur des lèvres, ils ne reflètent peut-être pas très précisément l'étirement des lèvres. On notera aussi que ces marqueurs se détachent relativement souvent et qu'il faut ainsi les repositionner : ce faisant, on change légèrement leur position.

Toutefois, on peut tenter d'extraire quelques caractéristiques pour ce paramètre. En règle générale, le constituant focalisé est hyper-articulé à la fois par rapport au reste de l'énoncé (graphique d de la figure III.35 : les points correspondant au constituant focalisé (foc) sont au-dessus des autres) et par rapport au cas neutre (figure III.35.d : ces mêmes points sont largement au-dessus de 1). L'hyper-articulation n'est significative que chez C, E et F. On observe la plus nette hyper-articulation chez le locuteur F(intra : +23,2% inter +30%) et la moins nette chez B.

En ce qui concerne la séquence pré-focale, deux tendances sont observées mais aucune n'est significative : hyper-articulation (chez B, C, D et E) et hypo-articulation (chez F). La séquence post-focale est soit hypo-articulée (significatif chez B et E et non significatif chez C) ou hyper-articulée (non significatif chez D et F).

A.3.3.1.e. Protrusion de la lèvre supérieure

Concernant la protrusion de la lèvre supérieure, les comportements sont assez homogènes. On observe une hyper-articulation du constituant focalisé à la fois par rapport :

- au reste de l'énoncé : de +25% (C) à 96,1% (B) ; graphique e. de la figure III.35 : les points correspondant au constituant focalisé (foc) sont au-dessus des autres ;
- au cas neutre : de 30,9% (C) à 112,4% (F) ; graphique e. de la figure III.35 : ces mêmes points sont largement au-dessus de 1.

On notera que chez les locuteurs E et F, le contraste intra-énoncé est inexistant lorsque la focalisation porte sur l'objet et sur le verbe. L'hyper-articulation est la plus forte chez le locuteur F par rapport au cas neutre mais le contraste intra-énoncé le plus important est observé chez le locuteur B.

La séquence pré-focale est hyper-articulée chez tous les locuteurs (hyper-articulation significative seulement chez B, E et F : de +24,4% (E) à +53,7% chez (F) ; anticipation chez B). La séquence post-focale est hypo-articulée chez tous (significatif seulement chez B, D et E, de –1% (D) à –12,9% (E)) sauf F chez qui on observe de façon assez étonnante une hyper-articulation très forte (+41,6%), mais moins forte que l'hyper-articulation focale.

A.3.3.2. Bilan intra-locuteur

Les graphiques de la figure III.36 permettent de résumer l'évolution pour chaque locuteur des données focales, pré- et post-focales pour chaque paramètre par rapport au cas neutre.

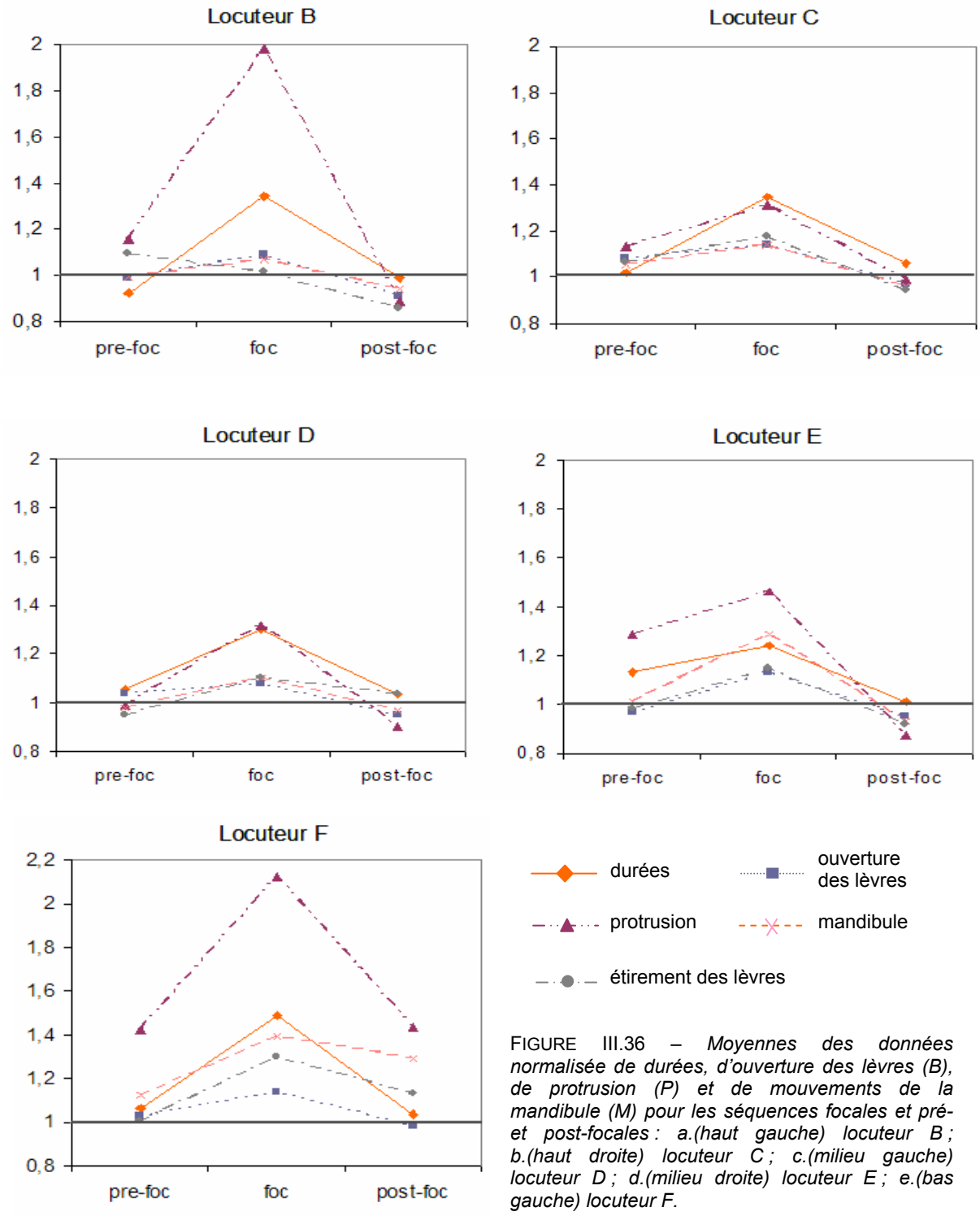


FIGURE III.36 – *Moyennes des données normalisée de durées, d'ouverture des lèvres (B), de protrusion (P) et de mouvements de la mandibule (M) pour les séquences focales et pré- et post-focales : a.(haut gauche) locuteur B ; b.(haut droite) locuteur C ; c.(milieu gauche) locuteur D ; d.(milieu droite) locuteur E ; e.(bas gauche) locuteur F.*

A.3.3.2.a. Locuteur B

Le graphique a. de la figure III.36 permet d'avoir un aperçu global de la stratégie de focalisation de ce locuteur concernant chacun des paramètres étudiés.

Le locuteur B allonge systématiquement les syllabes focales (intra : +38,7% inter : +34,3%). Cet allongement focal est le plus marqué lorsque la focalisation porte sur le verbe et le moins marqué lorsque celle-ci porte sur le sujet. Il ne modifie pas la durée des syllabes de la séquence post-focale et diminue la durée de toutes les syllabes de la séquence pré-focale de façon significative (-4,6%).

Il semblerait que ce locuteur hyper-articule significativement tous les paramètres étudiés sauf l'étirement des lèvres lorsqu'il focalise et ceci à la fois par rapport au reste de l'énoncé focalisé et par rapport à l'énoncé neutre. On voit clairement que le point du milieu sur toutes les courbes du graphique a. de la figure III.36 est à la fois au-dessus des autres points et au-dessus de 1 (sauf pour l'étirement). L'indice articulatoire « visible » le plus marqué, c'est-à-dire pour lequel on observe les plus fortes hyper-articulations intra- et inter-énoncés, est de loin la protrusion (intra : +96,1% et inter : +98,4%), les mouvements de la mandibule et l'ouverture des lèvres sont assez peu hyper-articulés (intra : 11,9% en moyenne et inter : 7,7%). Le contraste moyen (tous paramètres confondus) est de 39,9%, l'hyper-articulation moyenne par rapport au cas neutre (tous paramètres confondus) est de 49,6%.

Il y a toujours chez ce locuteur une hypo-articulation post-focale (atténuation de 9,8% en moyenne). Les points correspondant à la séquence post-focale sur le graphique a. de la figure III.36 (points correspondant à « post-foc ») sont en dessous de 1. C'est grâce à cette hypo-articulation post-focale que le contraste visible entre ce qui est focalisé et ce qui ne l'est pas dans un énoncé est assez important. Il semble en effet que ce locuteur répartisse son effort en hyper-articulant un peu ce qui est focalisé et en hypo-articulant un peu ce qui suit. Il crée ainsi un contraste visible au sein de l'énoncé focalisé.

On n'observe jamais chez ce locuteur d'hyper-articulation pré-focale sauf pour la protrusion. Cette dernière hyper-articulation pré-focale semble de plus être liée à la mise en place d'une stratégie d'anticipation de la focalisation puisqu'elle n'est pas présente sur le sujet lorsqu'il s'agit d'une focalisation verbale. Il semblerait donc que ce locuteur mette en place une stratégie d'anticipation uniquement pour la protrusion. On peut penser que cela est dû au fait que l'hyper-articulation étant beaucoup plus importante pour ce paramètre, pour atteindre des valeurs aussi importantes sur l'élément focalisé, il faut une transition et donc une phase d'anticipation pendant laquelle l'amplitude des gestes de protrusion augmente progressivement.

Chez ce locuteur, la focalisation affecte donc l'énoncé dans son intégralité et essentiellement à la fois le constituant focalisé et ce qui le suit. Le locuteur allonge et hyper-articule le constituant focalisé puis hypo-articule ce qui suit, accentuant ainsi le contraste visible au sein de l'énoncé entre ce qui est focalisé et ce qui l'est pas. Il peut lui arriver d'anticiper la focalisation lorsque l'hyper-articulation focale est très forte et ce sûrement pour effectuer une transition articulatoire. Les indices « visibles » les plus affectés par la focalisation sont la protrusion et la durée.

On rappellera ici que les productions articulatoires focalisées du locuteur B avaient été analysées avec une autre technique de mesure (cf. section A.2.3 du présent chapitre) : le système labiométrique de l'ICP. L'enregistrement avait été effectué pour le même corpus et il serait donc intéressant de comparer les résultats de l'analyse Optotrak avec ces autres résultats. Le tableau III.10 résume les résultats obtenus avec ces deux systèmes de mesure. On notera que la même stratégie générale *i.e.*

hyper-articulation de la séquence focale et hypo-articulation de la séquence post-focale a été observée. En ce qui concerne les quantités mesurées, si l'on compare la somme des données relatives correspondant aux paramètres A et B obtenues avec l'Optotrak, aux données relatives correspondant à AL obtenues avec le systéme labiométrique, les ordres de grandeur obtenus sont les mêmes pour l'hyper-articulation focale et l'hypo-articulation post-focale. Des différences sont cependant observées en ce qui concerne la séquence pré-focale mais les résultats statistiques n'étaient pas très nets. En ce qui concerne la protrusion, les tendances d'évolution focales et post-focales sont les mêmes. On notera que les différences d'ordre de grandeur sont peut-être liées aux différences de précision des deux systèmes pour cette mesure. Globalement, les résultats obtenus avec les deux systèmes de mesure sont en accord sauf en ce qui concerne la séquence pré-focale pour laquelle nous rappelons qu'il a été difficile de déterminer une stratégie définie.

		Optotrak				Système labiométrique	
		ΔA	ΔB	Δ(AxB)	P	AL	P
séquence pré-focale		(+2,4%)	(-0,2%)	(+2,2%)	+25,1%	-6,5%	(-2,1%)
séquence focale	intra	+12,6%	+13,8%	+26,4%	+96,1%	+25,3%	+59,3%
	inter	+1,6%	+8,7%	+10,3%	+98,4%	+12,1%	+36,9%
séquence post-focale		-15,6%	-8,9%	-24,5%	-8,1%	-22,1%	-25,8%

TABLE III.10 – *Bilan des données articulatoires obtenues pour le locuteur B avec les systèmes Optotrak et labiométrique (les résultats entre parenthèses ne sont pas significatifs).*

A.3.3.2.b. Locuteur C

Le graphique b. de la figure III.36 permet d'avoir un aperçu global de la stratégie de focalisation de ce locuteur concernant chacun des paramètres étudiés.

Le locuteur C allonge systématiquement les syllabes focales (intra : +30,5% inter : +34,8%). Il modifie de plus légèrement mais de façon significative la durée des syllabes de la séquence post-focale (allongement : +3,3% par rapport au cas neutre). Ce locuteur allonge également les syllabes de l'élément pré-focal (+4,7%) probablement dans le cadre de la mise en place d'une stratégie d'anticipation de la focalisation.

Ce locuteur hyper-articule tous les paramètres étudiés lorsqu'il focalise, à la fois par rapport au reste de l'énoncé focalisé et par rapport à l'énoncé neutre. L'indice articulatoire « visible » le plus marqué est la protrusion (intra : +25% et inter : +30,9%), les mouvements de la mandibule et l'ouverture et l'étirement des lèvres sont assez peu hyper-articulés (intra : 14% en moyenne et inter : 15,1%). Le contraste moyen (tous paramètres confondus) est de 16,8%, l'hyper-articulation moyenne (tous paramètres confondus) est de 19%.

Il n'y a, chez ce locuteur, une hypo-articulation post-focale significative que pour les mouvements de la mandibule et l'ouverture des lèvres (respectivement –4,4% et –2,2%). Cette hypo-articulation est aussi présente en moyenne pour l'étirement labial et elle est même assez forte (-8,1%) bien qu'elle ne soit pas statistiquement significative. La stratégie de ce locuteur consiste donc essentiellement à hyper-articuler ce qui est focalisé. Pour les mouvements de la mandibule et l'ouverture des lèvres (et

l'étirement labial), il y a cependant une hypo-articulation post-focale faible, sans doute pour compenser une hyper-articulation relativement faible de l'élément focal (respectivement 14% et 13,9% (et 17%)) et ainsi renforcer le contraste entre ce qui est focalisé et ce qui ne l'est pas. Ce renforcement n'est pas nécessaire pour la protrusion qui est déjà bien marquée sur le constituant focalisé.

On n'observe chez ce locuteur une hyper-articulation pré-focale significative que pour les mouvements de la mandibule et l'ouverture des lèvres, une nouvelle fois (respectivement +6% et +6,8%). Cependant, bien qu'elle ne soit pas statistiquement significative, cette hyper-articulation est également présente pour la protrusion et l'étirement des lèvres et elle est même en moyenne assez importante pour la protrusion (+10,9%). Elle ne semble cependant pas être liée à la mise en place d'une stratégie d'anticipation de la focalisation, et ce pour aucun des paramètres étudiés, puisqu'elle est également présente sur le sujet lorsqu'il s'agit d'une focalisation sur l'objet et donc sur un élément non directement pré-focal. On peut penser qu'il s'agirait d'une conséquence du fait que, dès le début de l'énoncé, le locuteur veut signaler le fait qu'un élément va être focalisé et donc commence déjà à hyper-articuler. Cette hyper-articulation de toute la séquence pré-focale est certainement la cause du fait que les contrastes intra-énoncés sont ici moins importants que les augmentations par rapport au cas neutre.

Globalement, ce locuteur adopte donc la stratégie suivante : quand il focalise un constituant, il commence à hyper-articuler légèrement dès le début de l'énoncé (les points correspondant à la séquence pré-focale (pre-foc) sur la figure III.36.b sont d'ailleurs au-dessus de 1), de plus, il allonge les syllabes de la séquence directement pré-focale (anticipation temporelle de la focalisation). Quand il atteint l'élément focalisé, il hyper-articule encore plus et allonge les syllabes. La séquence post-focale est quant à elle légèrement hypo-articulée (les points correspondant à la séquence post-focale (post-foc) sur la figure III.36.b sont d'ailleurs au-dessus de 1) ce qui renforce le contraste visible au sein de l'énoncé (sauf pour la protrusion). Les indices « visibles » les plus affectés par la focalisation sont la durée et la protrusion.

A.3.3.2.c. Locuteur D

Le graphique c. de la figure III.36 permet d'avoir un aperçu global de la stratégie de focalisation de ce locuteur concernant chacun des paramètres étudiés.

Le locuteur D allonge systématiquement les syllabes focales (intra : +25,3% inter : +29,8%). Il modifie de plus légèrement la durée des syllabes de la séquence post-focale (allongement : +3,3%) ainsi que les syllabes de l'élément pré-focal (+8,1%) probablement dans le cadre de la mise en place d'une stratégie d'anticipation de la focalisation.

Il semblerait que ce locuteur hyper-articule de façon significative tous les paramètres étudiés, sauf l'étirement des lèvres, lorsqu'il focalise et ceci à la fois par rapport au reste de l'énoncé focalisé et par rapport à l'énoncé neutre. En ce qui concerne l'étirement des lèvres, l'hyper-articulation existe mais elle n'atteint pas le niveau de signification. L'indice articulatoire « visible » le plus marqué est de loin la protrusion (intra : +37,4% et inter : +32%), les mouvements de la mandibule et l'ouverture des lèvres sont moins hyper-articulés (intra : 10,5% en moyenne et inter : 8,8%). Le contraste moyen (tous paramètres confondus) est de 19,4%, l'hyper-articulation moyenne (tous paramètres confondus) est de 16,5%.

Il n'y a chez ce locuteur une hypo-articulation post-focale significative que pour les paramètres d'ouverture des lèvres et de protrusion. Cette hypo-articulation est assez faible (-2,8% en moyenne).

On observe également une tendance moyenne à l'hypo-articulation pour les mouvements de la mandibule (-2%) mais celle-ci n'est pas statistiquement significative. Pour l'ouverture des lèvres, il s'agit peut-être là de compenser une faible hyper-articulation focale (+7,6% seulement) dans le but de *renforcer* le contraste intra-énoncé. Pour la protrusion, l'hypo-articulation post-focale est certes significative mais très faible (-1% seulement). En ce qui concerne l'étirement des lèvres on note une hyper-articulation (+4,9%) mais celle-ci n'étant pas statistiquement significative, on peut penser qu'il s'agit là d'un artéfact causé par un ou plusieurs énoncés isolés pour lesquels on a fait une mesure de forte hyper-articulation alors que pour les autres il n'y en avait pas. On notera que c'est certainement aussi à cause de cette hyper-articulation post-focale moyenne assez forte que les contrastes intra-énoncés ne sont pas statistiquement significatifs).

On n'observe chez ce locuteur une hyper-articulation pré-focale que pour le paramètre d'ouverture des lèvres (+5,7%). Cette hyper-articulation pré-focale semble de plus être liée à la mise en place d'une stratégie d'anticipation de la focalisation puisqu'elle n'est pas présente sur le sujet lorsqu'il s'agit d'une focalisation objet, elle n'a donc lieu que sur l'élément directement pré-focal. On notera que pour les autres paramètres cette hyper-articulation existe aussi en moyenne bien qu'elle ne soit pas statistiquement significative. Elle n'affecte en fait quasiment toujours que le sujet pré-focal dans le cas de la focalisation sur le verbe et c'est sûrement pour cela qu'elle n'est pas significative en moyenne. Il semblerait donc que ce locuteur mette en place une stratégie d'anticipation uniquement lorsque c'est le verbe qui est focalisé. On pourra d'ailleurs noter que, de façon générale, l'objet est le constituant qui est le moins hyper-articulé, il est donc naturel que le locuteur n'ait pas besoin d'anticiper.

On peut donc conclure que la plupart du temps, ce locuteur tend à hyper-articuler et allonger le constituant focalisé (sauf pour l'étirement labial). Il met en place une stratégie d'anticipation articulatoire mais uniquement lorsque la focalisation porte sur le verbe, il allonge de plus les syllabes de la séquence directement pré-focale (anticipation temporelle de la focalisation). La séquence post-focale est quant à elle légèrement hypo-articulée (les points correspondant à la séquence post-focale (post-foc) sur la figure III.36.c sont d'ailleurs au-dessus de 1) ce qui renforce le contraste visible au sein de l'énoncé (sauf pour l'étirement labial). Les indices « visibles » les plus affectés par la focalisation sont la protrusion et la durée.

A.3.3.2.d. Locuteur E

Le graphique d. de la figure III.36 permet d'avoir un aperçu global de la stratégie de focalisation de ce locuteur concernant chacun des paramètres étudiés.

Le locuteur E allonge systématiquement les syllabes focales (intra : +16,8% inter +23,9%). Il ne modifie pas la durée des syllabes de la séquence post-focale. Par contre, il allonge les syllabes de l'élément pré-focal (+16,4%), sans doute dans le cadre de la mise en place d'une stratégie d'anticipation de la focalisation.

Il semblerait que ce locuteur hyper-articule tous les paramètres étudiés lorsqu'il focalise, ceci à la fois par rapport au reste de l'énoncé focalisé et par rapport à l'énoncé neutre. L'indice articulatoire « visible » le plus marqué, c'est-à-dire pour lequel on observe les plus fortes hyper-articulations intra- et inter-énoncés, est de loin la protrusion (intra : +39,4% et inter : +46%), les mouvements de la mandibule, l'ouverture et l'étirement des lèvres sont assez peu hyper-articulés (intra : 22,8% en moyenne et inter : 18,8%). Le contraste moyen (tous paramètres confondus) est de 26,9%, l'hyper-articulation moyenne (tous paramètres confondus) est de 25,6%.

Il y a toujours chez ce locuteur une hypo-articulation post-focale (atténuation de 8,2% en moyenne). Grâce à cette hypo-articulation post-focale, le contraste entre ce qui est focalisé et ce qui ne l'est pas dans un énoncé est renforcé. Il semble en effet que ce locuteur répartisse son effort en hyper-articulant ce qui est focalisé et en hypo-articulant ce qui suit. Il crée ainsi un contraste visible au sein de l'énoncé focalisé.

On n'observe jamais chez ce locuteur d'hyper-articulation pré-focale sauf pour la protrusion (+24,4%). Cette dernière hyper-articulation pré-focale très forte semble ne pas être liée à la mise en place d'une stratégie d'anticipation de la focalisation puisqu'elle est aussi présente sur le sujet lorsqu'il s'agit d'une focalisation sur l'objet. Il est possible qu'étant donné le niveau d'hyper-articulation de la protrusion sur le syntagme focalisé, le locuteur commence à hyper-articuler dès le début de l'énoncé. Il y a également une hyper-articulation faible de la séquence pré-focale pour les mouvements de la mandibule et l'étirement labial (respectivement +2,8% et +2,7%). Celle-ci n'est cependant pas significative. En ce qui concerne l'ouverture des lèvres on constate une hypo-articulation significative de la séquence pré-focale (-3,4%).

Il semble ainsi que ce locuteur hyper-articule et allonge le constituant focalisé. Mais les corrélats de la focalisation s'étendent au-delà du syntagme directement focalisé. On observe en effet une hypo-articulation post-focale qui renforce le contraste visible entre ce qui est focalisé et ce qui ne l'est pas au sein de l'énoncé. Il existe également une hyper-articulation faible de la séquence pré-focale sauf pour l'ouverture des lèvres (hypo-articulation). Les indices « visibles » les plus affectés par la focalisation sont la protrusion, les mouvements de la mandibule et la durée.

A.3.3.2.e. Locuteur F

Le graphique e. de la figure III.36 permet d'avoir un aperçu global de la stratégie de focalisation de ce locuteur concernant chacun des paramètres étudiés.

Le locuteur F allonge systématiquement les syllabes focales (intra : +43,8% inter : +49%). Il modifie de plus légèrement la durée des syllabes de la séquence post-focale (allongement : +5,6%). Ce locuteur allonge également les syllabes de l'élément pré-focal (+9%) probablement dans le cadre de la mise en place d'une stratégie d'anticipation de la focalisation.

Il semblerait que ce locuteur hyper-articule tous les paramètres étudiés lorsqu'il focalise et ceci à la fois par rapport au reste de l'énoncé focalisé et par rapport à l'énoncé neutre. L'indice articulatoire « visible » le plus marqué c'est-à-dire pour lequel on observe les plus fortes hyper-articulations intra- et inter-énoncés est de loin la protrusion (intra : +69,1% et inter : +112,4%), les mouvements de la mandibule et l'ouverture et l'étirement des lèvres sont moins hyper-articulés (intra : 18,4% en moyenne et inter : 27,8%). Le contraste moyen (tous paramètres confondus) est de 31,1%, l'hyper-articulation moyenne (tous paramètres confondus) est de 65,6%.

Il n'y a jamais chez ce locuteur d'hypo-articulation post-focale par rapport au cas neutre. Et on constate même une hyper-articulation post-focale par rapport au cas neutre très importante pour les paramètres de mouvements de la mandibule et de protrusion (respectivement +23,8% et +41,6%). Cependant le constituant post-focal est nettement moins hyper-articulé que le constituant focal. Un contraste intra-énoncé important existe donc bien. De plus, le locuteur ne peut peut-être pas produire une variation articulatoire suffisante pour atteindre une hypo-articulation post-focale par rapport au cas neutre. D'ailleurs on notera que quand l'hyper-articulation focale n'est pas très importante (ouverture des lèvres), on observe une hypo-articulation moyenne (-2,3%) bien que celle-ci ne soit pas significative.

On n'observe jamais chez ce locuteur d'hyper-articulation pré-focale sauf pour la protrusion (+53,7%). Cette hyper-articulation pré-focale très importante est peut-être liée à la mise en place d'une stratégie d'anticipation de la focalisation puisqu'elle n'est pas présente sur le sujet lorsqu'il s'agit d'une focalisation sur l'objet et donc elle n'est présente que sur l'élément directement pré-focal. Il semblerait donc que ce locuteur mette en place une stratégie d'anticipation uniquement pour la protrusion. On peut penser que cela est dû au fait que l'hyper-articulation étant beaucoup plus importante pour ce paramètre, pour atteindre des valeurs aussi importantes sur l'élément focalisé, il faut une transition et donc une phase d'anticipation pendant laquelle l'amplitude des gestes de protrusion augmente progressivement.

On remarque globalement chez ce locuteur que, bien que l'hyper-articulation par rapport au cas neutre soit très importante (65,6% en moyenne), le contraste intra-énoncé l'est beaucoup moins (31,1%) même s'il demeure tout de même relativement important par rapport aux autres locuteurs. Ceci est dû au fait qu'il y a bien souvent, chez ce locuteur, hyper-articulation d'une autre partie de l'énoncé et notamment forte hyper-articulation parfois de la séquence post-focale. On peut penser qu'après avoir hyper-articulé l'élément focalisé ce locuteur « continue sur sa lancée » et hyper-articule aussi le reste de l'énoncé. Le contraste articulatoire à produire est peut-être trop grand. L'indice « visible » le plus affecté par la focalisation est la protrusion. On notera d'ailleurs le marquage excessivement fort de la protrusion d'autant que l'échelle du graphique e. de la figure III.36 est plus grande que celle des graphiques correspondant aux autres locuteurs.

A.3.3.3. Synthèse et discussion

A.3.3.3.a. Synthèse et résumé des résultats

Les analyses articulatoires et temporelles détaillées ci-dessus permettent de tirer quelques conclusions générales. Il apparaît ainsi que tous les locuteurs allongent les syllabes focales de façon significative. Cet allongement est le plus important chez le locuteur F et le moins important chez le locuteur E aussi bien en intra- qu'en inter-énoncés. Bien que l'allongement soit plus ou moins grand d'un locuteur à l'autre, il est toujours supérieur au seuil de perception auditif de l'allongement (cf. section C.2 du chapitre II). Les modifications temporelles s'étendent souvent à tout l'énoncé mais de façon différente selon les locuteurs. Le locuteur B diminue les durées syllabiques pré-focales alors qu'il ne modifie rien à la séquence post-focale. Quant aux autres locuteurs, ils allongent tous aussi bien les syllabes de la séquence pré-focale que celles de la séquence post-focale, l'allongement le plus fort étant noté sur la séquence pré-focale. L'allongement pré-focal ne concerne cependant que le constituant directement pré-focal et il peut ainsi être avancé qu'il s'agit là d'une anticipation de la focalisation. En ce qui concerne la séquence post-focale, il est possible que le locuteur soit entraîné par l'allongement du constituant focalisé et continue ainsi à allonger, mais de façon beaucoup moins nette, afin de signaler que cette séquence n'est plus focalisée. On notera que, bien que ces allongements pré- et post-focaux atteignent le niveau de signification statistique, ils n'atteignent pas le seuil de perception auditif de l'allongement (cf. section C.2 du chapitre II).

Les résultats sont beaucoup moins homogènes au niveau articulatoire. Pour résumer, en disant qu'on observe principalement deux types de stratégies. Pour les deux stratégies, on observe une hyper-articulation du constituant focalisé à la fois par rapport au reste de l'énoncé focalisé (intra) et par rapport au cas neutre (inter). Cette hyper-articulation est plus ou moins importante selon les locuteurs. Les locuteurs B, C, D et E hypo-articulent le plus souvent de façon plus ou moins importante la séquence post-focale, accentuant ainsi le contraste entre ce qui focalisé et ce qui ne

l'est pas dans l'énoncé. Quant à la séquence pré-focale le comportement est moins systématique puisqu'on observe parfois pour certains paramètres une anticipation voire une hyper-articulation dès le début de l'énoncé. Il semblerait en effet que quand l'hyper-articulation focale est particulièrement forte, celle-ci commence parfois dès le début de l'énoncé même si le constituant focalisé n'est pas le premier de l'énoncé. Le locuteur F adopte une stratégie différente qui consiste à continuer l'hyper-articulation focale sur la séquence post-focale, bien que de manière beaucoup moins importante. Dans son cas l'hyper-articulation focale est souvent mieux marquée (plus importante) que pour les autres locuteurs et on peut penser que le contraste au sein de l'énoncé étant déjà suffisamment important, il n'y a pas besoin de rajouter des indices en dehors du constituant focal. On n'observe en effet pas d'anticipation de la focalisation sauf pour certains paramètres pour lesquels l'hyper-articulation est tellement forte qu'elle nécessite probablement une phase de transition.

Pour tous les locuteurs, il semblerait que les indices « visibles » les plus marqués soient la protrusion et la durée.

A.3.3.3.b. Discussion sur l'intensité du marquage « visible »

Rappelons-nous de l'observation qui avait été faite à l'issue de l'analyse des données vidéo (cf. section A.2 du présent chapitre). Il avait en effet été noté que plus le nombre de syllabe d'un constituant était important, moins les indices « visibles » étaient intensément marqués. Nous tenterons dans cette section de déterminer si la même observation peut être faite à partir des données Optotrak.

Le graphique de la figure III.37 donne les valeurs moyennes des durées normalisées des syllabes focales en fonction du nombre de syllabes du constituant focalisé (en abscisse) et du locuteur (différentes courbes). Ce graphique permet de voir que pour tous les locuteurs, plus un constituant comporte de syllabes moins l'allongement moyen est important. On retrouve ce qui avait été discuté pour les données vidéo (cf. section A.2.4.3 du présent chapitre) *i.e.* que le marquage « visible » de l'allongement est moins marqué lorsqu'il y a plusieurs syllabes à focaliser. On notera d'ailleurs que pour la plupart des locuteurs les contrastes les plus marqués sont relevés lorsque la focalisation porte sur le verbe, qui est en moyenne dans le corpus le syntagme le moins long (2,6 syllabes en moyenne pour S, 2,4 pour V et 2,9 pour O).

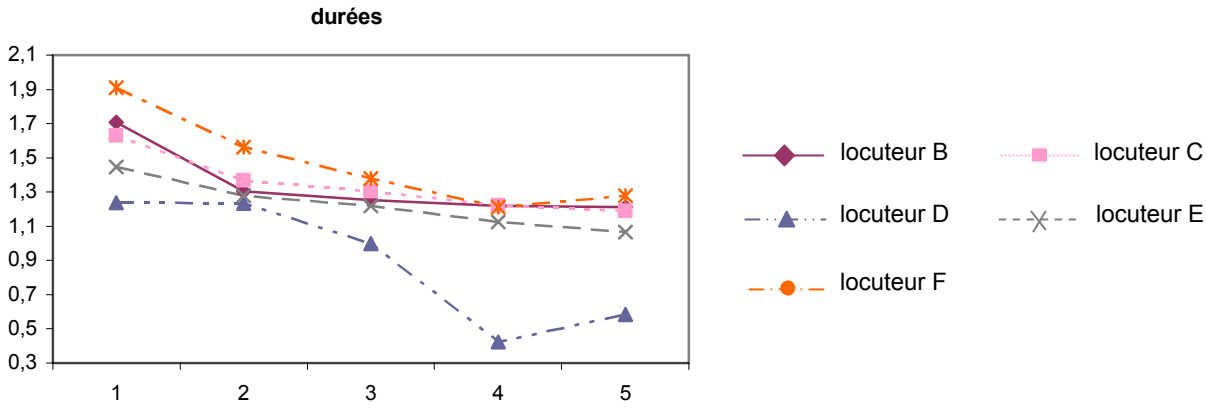


FIGURE III.37 – *Données de durées normalisées des syllabes focales en fonction du nombre de syllabes du constituant focalisé (en abscisse) et du locuteur (différentes courbes).*

Les mêmes analyses ont été conduites pour les paramètres articulatoires (mandibule, ouverture et étirement des lèvres, protrusion de la lèvre supérieure). Bien que l'on observe une tendance à la diminution de l'intensité du marquage avec l'augmentation du nombre de syllabes, les résultats sont moins nets que pour la durée. Ceci est certainement dû à la variabilité intrinsèque du corpus. Aucune des variations des paramètres articulatoires ne permet de décrire l'intégralité des données. Parfois, en fonction de la nature phonétique de la syllabe en question, un paramètre spécifique sera amplifié alors que les autres ne le seront pas voire diminueront. Ainsi, par exemple, lorsque la syllabe /li/ appartient à un syntagme focalisé, on s'attend à ce que l'étirement des lèvres soit amplifié mais que l'ouverture des lèvres et de la mandibule diminuent. Pour tenir compte de cette variabilité intrinsèque, nous avons tracé un graphique complémentaire (graphique de la figure III.39) qui donne le nombre de voyelles censées impliquer chaque paramètre en fonction du nombre de syllabes du constituant. On note que les constituants de 3 et 4 syllabes comportent beaucoup plus de cas de voyelles articulées à l'aide des paramètres d'ouverture des lèvres et de protrusion que les autres constituants (de taille 1, 2 et 5). Les graphiques b. et d. de la figure III.38 deviennent ainsi interprétables. La décroissance escomptée est interrompue par un plateau sur les constituants à trois ou quatre syllabes précisément parce que ces constituants renferment un plus grand pourcentage de voyelles ouvertes et protruses. En ce qui concerne l'étirement des lèvres, les résultats sont plus difficiles à interpréter. Rappelons à ce sujet qu'il est probable que les mesures de ce paramètre soient imprécises.

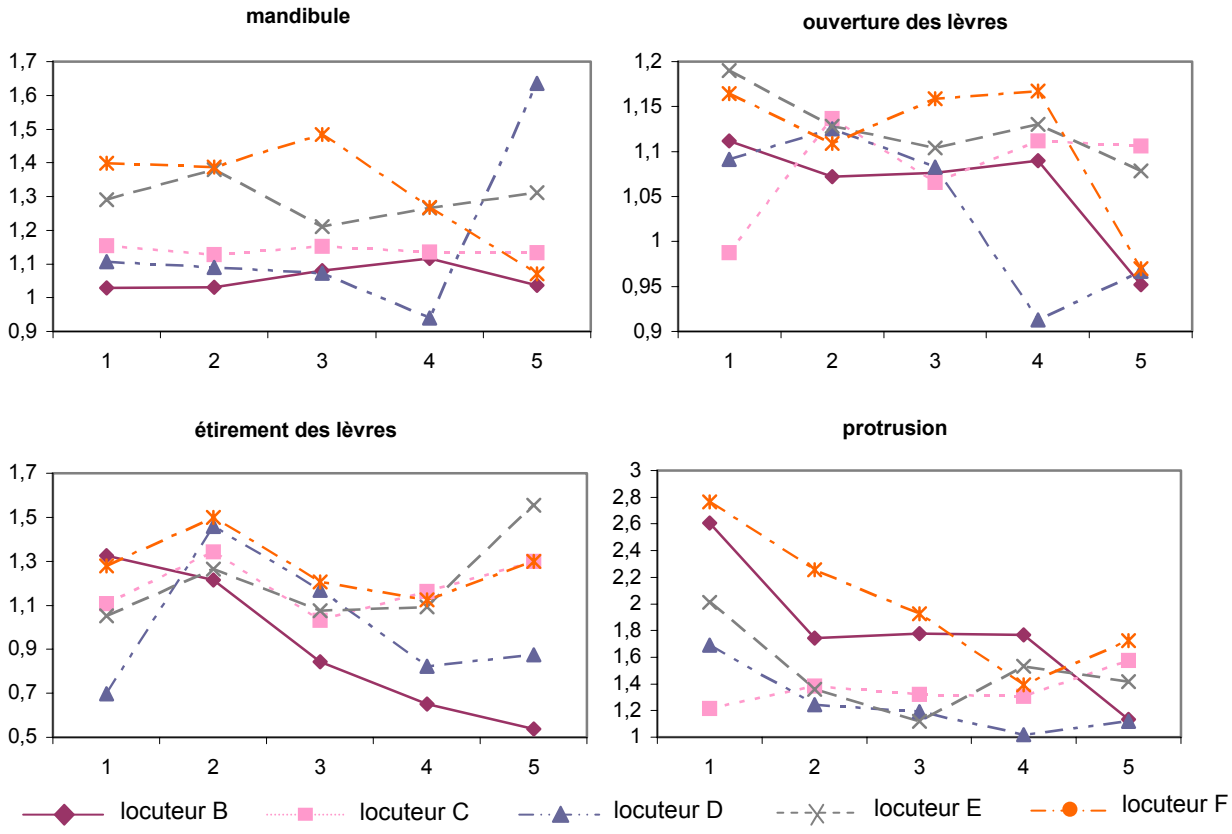


FIGURE III.38 – *Données normalisées pour chaque paramètre articulatoire en fonction du nombre de syllabes du constituant focalisé (en abscisse) et du locuteur (différentes courbes).*

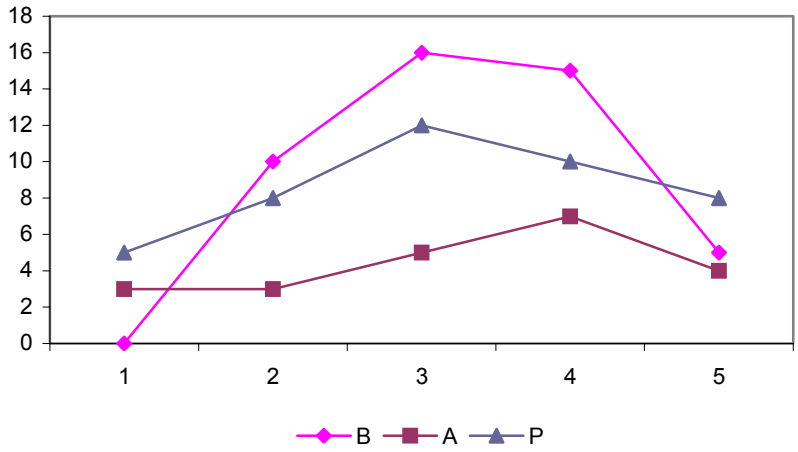


FIGURE III.39 – *Prédiction du nombre de voyelles du corpus impliquant chaque paramètre articulatoire (A : étirement, B : ouverture et P : protrusion) en fonction du nombre de syllabes du constituant auquel elles apparatiennent.*

A.4. Synthèse : vers un modèle de la production d'indices articulatoires « visibles » de la focalisation contrastive en français

Les études décrites précédemment menées avec plusieurs systèmes de mesures et pour plusieurs locuteurs ont montré qu'il existait bien des indices potentiellement visibles de la focalisation contrastive en français. Certains gestes articulatoires visibles sont en effet affectés par la présence d'une focalisation dans l'énoncé ainsi que par sa localisation. Ces indices existent pour tous les locuteurs bien que la façon dont ils sont marqués puisse varier d'un locuteur à l'autre. Les études sur les données vidéos (voir section A.2 du présent chapitre) menées sur deux locuteurs ont permis d'avoir une première idée des paramètres impliqués. Puis les études avec le système Optotrak, menées sur cinq locuteurs, ont permis d'analyser le problème de façon plus approfondie et de tenter d'extraire des invariants et des variations par rapport à ceux-ci. Le but de cette section sera de tenter de donner une première description multi-locuteurs des indices visibles de la focalisation contrastive en français.

Les indices « visibles » étudiés ici ont principalement été la durée et les indices articulatoires fournis par la bouche et la mandibule. On a pu constater que ces indices « visibles » étaient affectés « visiblement » par la focalisation. La focalisation contrastive possède donc des corrélats qui peuvent être visibles par l'interlocuteur lors des processus de perception de la parole.

Les études menées avec le système Optotrak sur cinq locuteurs, ont permis d'établir une première classification des comportements articulatoires liés à la focalisation. Les résultats des analyses menées sur deux locuteurs avec la technique labiométrique peuvent aussi rentrer dans cette classification. En faisant une synthèse des résultats obtenus avec les deux techniques de mesure, on est ainsi en mesure de décrire principalement deux stratégies lesquelles permettent de représenter de façon satisfaisante les comportements de tous les locuteurs :

- **Stratégie de signalisation visuelle absolue :** le constituant focalisé est allongé et hyper-articulé dans une large mesure, par rapport au cas neutre : les amplitudes de tous les paramètres articulatoires (aire intéro-labiale, protrusion, positions de la mandibule) sont amplifiées et les pics de vitesse sont plus importants reflétant une augmentation de la force articulatoire du geste sous-jacent. Cette amplification se fait à la fois par rapport au cas neutre (contraste inter-énoncés) et par rapport au reste de l'énoncé (contraste intra-énoncé). Les locuteurs adoptant cette stratégie concentrent donc leurs efforts sur l'hyper-articulation du constituant focalisé. On note aussi parfois une anticipation de la focalisation c'est-à-dire que l'amplitude des gestes articulatoires, leur vitesse et la durée syllabique commencent à augmenter sur la syllabe précédant directement l'élément focalisé. La figure III.40 représente cette stratégie de façon schématique.

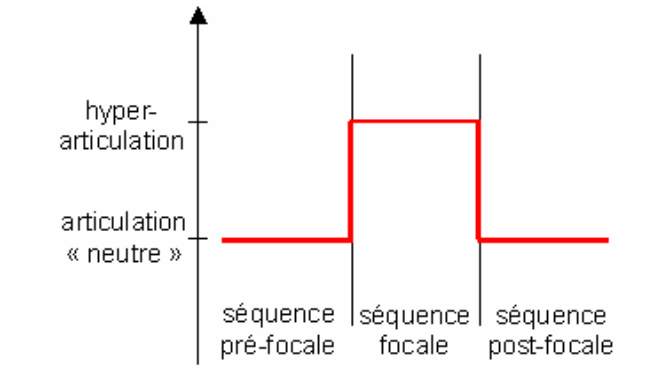


FIGURE III.40 – *Représentation schématique de la stratégie de signalisation visuelle absolue de la focalisation contrastive prosodique.*

- **Stratégie de signalisation visuelle différentielle :** dans ce cas, le constituant focalisé est aussi allongé et hyper-articulé. Cette hyper-articulation est parfois anticipée. La différence par rapport à la stratégie précédente est que la séquence post-focale est hypo-articulée par rapport au cas neutre, c'est-à-dire que l'amplitude des gestes articulatoires y est significativement moins importante. Un contraste important est ainsi créé au sein de l'énoncé focalisé (contraste intra-énoncé). L'hyper-articulation focale n'est pas très importante mais elle est renforcée par l'hypo-articulation post-focale. La figure III.41 représente cette stratégie de façon schématique.

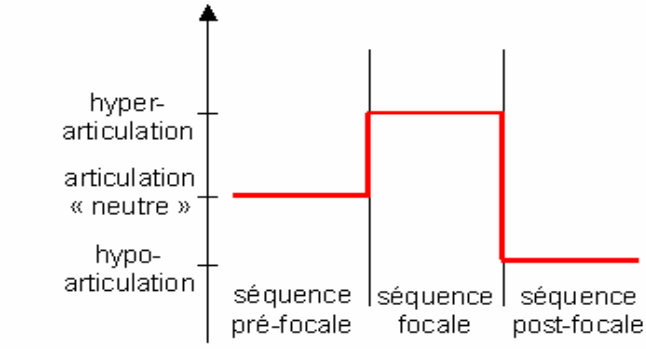


FIGURE III.41 – *Représentation schématique de la stratégie de signalisation visuelle différentielle de la focalisation contrastive prosodique.*

Il conviendra de souligner les fortes variabilités inter-locuteurs en matière de signalisation visible de la focalisation contrastive. Même lorsque les locuteurs utilisent la même stratégie globale, on observe beaucoup de variations dans le degré d'hyper- ou hypo-articulation ou même dans la nature des paramètres utilisés et surtout leur importance relative au niveau du marquage global. Il apparaît également que certains locuteurs anticipent la signalisation visuelle de la focalisation en commencant à hyper-articuler dès la syllabe directement pré-focale. L'observation la plus constante qui puisse être faite est l'importance du marquage du paramètre de protrusion. Bien que l'intensité du marquage de ce paramètre varie d'un locuteur à l'autre, il apparaît que c'est dans la majeure partie des cas, le paramètre qui est les plus marqué. De façon générale, on observe aussi que la stratégie de signalisation différentielle est plus fréquemment utilisée par les locuteurs que la stratégie de signalisation absolue (quatre locuteurs utilisent la stratégie de signalisation différentielle contre deux qui utilisent la stratégie de signalisation absolue). On notera, surtout dans le cadre de la stratégie de signalisation différentielle, que la focalisation affecte l'énoncé dans son intégralité au niveau articulatoire comme l'avaient suggéré Summers [1987] et Erickson [1998]. La stratégie de signalisation différentielle semble d'ailleurs être la plus naturelle sur le plan l'effort biomécanique puisqu'elle permet de répartir cet effort et la dépense d'énergie sur l'énoncé dans son intégralité. Le locuteur a ainsi moins d'effort articulatoire à fournir pour créer le même contraste intra-énoncé. Cette stratégie est en accord avec le modèle H&H (*pour « Hyperspeech & Hypospeech »*) de Lindblom [1990]. Selon Lindblom, la production de la parole est adaptative. *« Speakers can, and typically do, tune their performance according to communicative and situational demands, controlling the interplay between production-oriented factors on the one hand, and output-oriented constraints on the other. »* (Les locuteurs peuvent, et en général le font, accorder leur performance en tenant compte des exigences communicatives et situationnelles, en contrôlant l'interaction entre les facteurs liés aux contraintes de production d'une part et ceux liés aux contraintes de réception d'autre part.) Il existerait donc une interaction locuteur/auditeur qui détermine en partie les productions. *« Speakers are expected to vary their output along a continuum of hyper- and hypospeech »* (On s'attend à ce que les productions des locuteurs varient le long d'un continuum d'hyper- à hypo-articulation). Lorsque les contraintes perceptives sont dominantes, l'articulation du locuteur s'efforce d'être précise et soignée et donc des hyper-formes apparaissent tandis que lorsque ce sont les contraintes de production qui prédominent le locuteur cherche à minimiser son effort et des hypo-formes sont observées. Un contraste suffisant (*« sufficient contrast »*) est donc visé par le locuteur. La stratégie de signalisation différentielle correspond justement à cette quête d'un contraste suffisant en produisant l'effort minimum.

B. Autres gestes faciaux et focalisation : une étude préliminaire

Comme il a été exposé dans le chapitre I, la littérature permet de penser qu'il existerait un lien entre certains gestes faciaux et la prosodie en général et la focalisation prosodique en particulier. Il apparaît néanmoins que ces liens sont loin d'être systématiques et que les études menées jusqu'ici ne permettent pas pour l'instant de conclure de façon tranchée. Ces études montrent cependant qu'en dehors de la bouche, les principaux gestes observés en lien avec la focalisation sont les mouvements

des sourcils et de la tête. C'est pourquoi nous avons décidé d'étudier précisément ces deux types de mouvements.

B.1. Données expérimentales

Afin de tenter de mettre en relation les mouvements des sourcils et de la tête avec la focalisation constrastive prosodique en français, nous avons utilisé les données Optotrak enregistrées aux laboratoires ATR. L'enregistrement de ces données a été décrit de façon détaillée dans la section A.3.1 du présent chapitre et le lecteur pourra s'y reporter pour trouver des informations sur le corpus utilisé et les méthodes d'enregistrement employées.

B.1.1. Mesure des mouvements des sourcils

Les mouvements des sourcils ont été mesurés à l'aide des marqueurs 23 et 27 (cf. figure III.36). La plupart des études décrites dans la littérature ayant observé que les principaux mouvements de sourcils étaient des haussements, nous nous sommes intéressée à la coordonnée z de chacun de ces marqueurs. Les mouvements du sourcil gauche ont ainsi été représentés par le marqueur 23 et ceux du sourcil droit par le marqueur 27. Une référence correspondant à la position de repos des sourcils, a été soustraite aux données de mouvement des sourcils afin d'obtenir les mouvements relatifs à cette position de « repos ». Cette référence a été évaluée pour chaque énoncé enregistré comme correspondant à la position du marqueur en question pendant les 500 premières millisecondes.

B.1.2. Mesure des mouvements de la tête

Les mouvements de la tête ont été évalués grâce à un dispositif comprenant quatre marqueurs (marqueurs 1-4) et placé sur la tête du locuteur (cf. figure III.36). Ce dispositif a permis d'obtenir une approximation correcte des mouvements du corps rigide qui correspond à la tête du locuteur. Ces mouvements sont décrits par trois translations (selon les trois axes du repère) et trois angles de rotations (autour de ces mêmes axes). Les études décrites dans la littérature ayant mis en avant une forte tendance à bouger la tête selon un hochement, il a été décidé de s'intéresser principalement à la rotation autour de l'axe y (cf. figure III.36). Une position de référence a été déterminée à partir de la position de la tête au « repos » pour chaque énoncé (500 premières millisecondes de l'énoncé). Celle-ci a ensuite été soustraite au signal correspondant à l'énoncé dans son intégralité afin d'obtenir les mouvements par rapport à cette position au « repos ». Les mouvements ont été séparés en deux types : hochement vers le haut (correspondant à un angle positif) et hochement vers le bas (correspondant à un angle négatif). Seuls les mouvements vers le haut ont été analysés, les mouvements vers le bas étant quasiment inexistants.

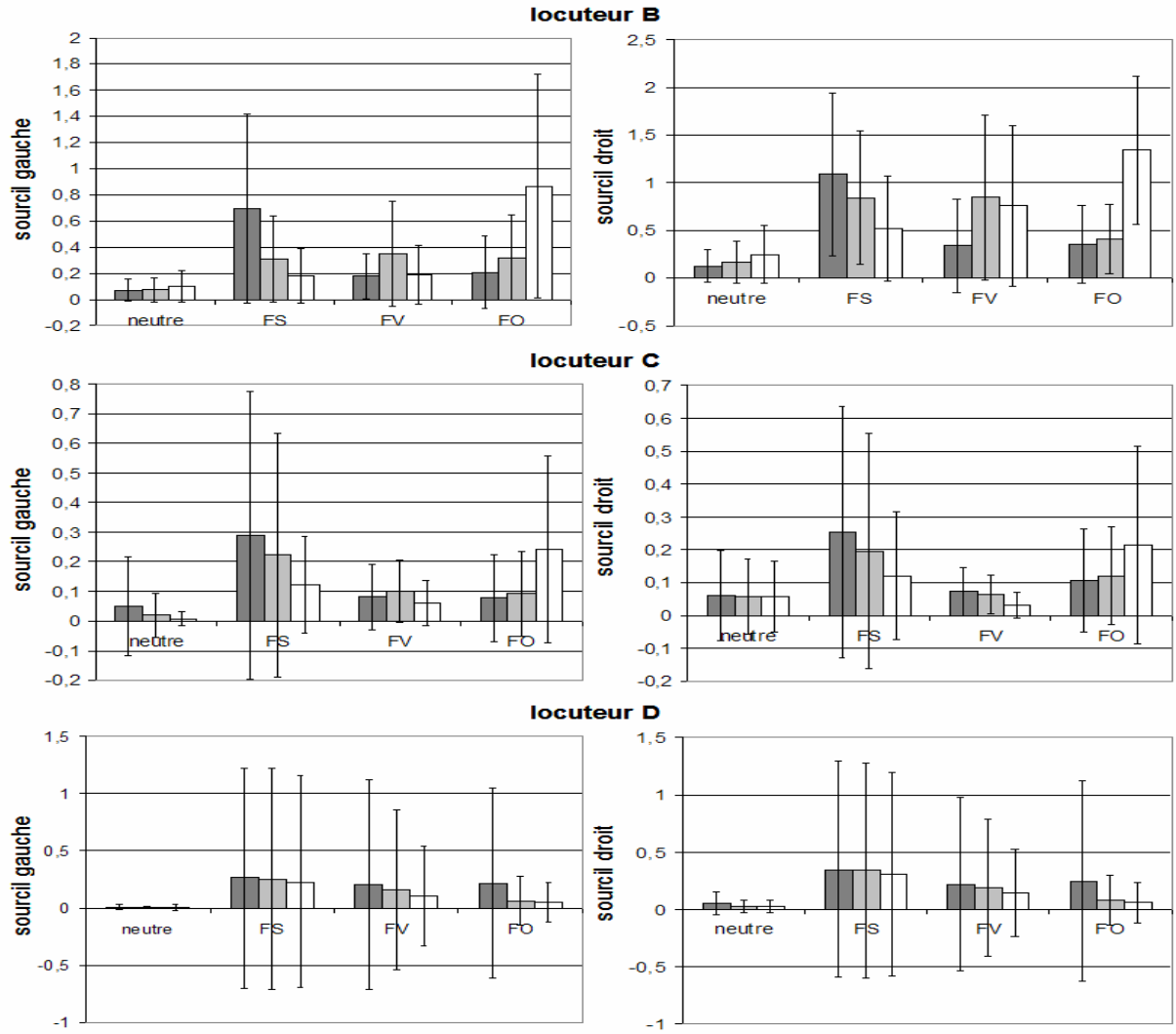
B.1.3. Traitement des mesures

Les données obtenues selon les méthodes décrites ci-dessus seront traitées de la façon suivante : l'aire sous la courbe correspondant à l'évolution temporelle des paramètres sera calculée pour chaque

syntagme de chaque énoncé produit. Des moyennes seront ensuite calculées pour chaque type de syntagme sous un type de focalisation. Ce traitement sera effectué indépendamment pour les données correspondant à chaque locuteur. Rappelons qu'il s'agit d'une étude préliminaire et qu'il est donc difficile de faire des hypothèses *a priori* sur les comportements qui seront observés.

B.2. Résultats

B.2.1. Mouvements des sourcils



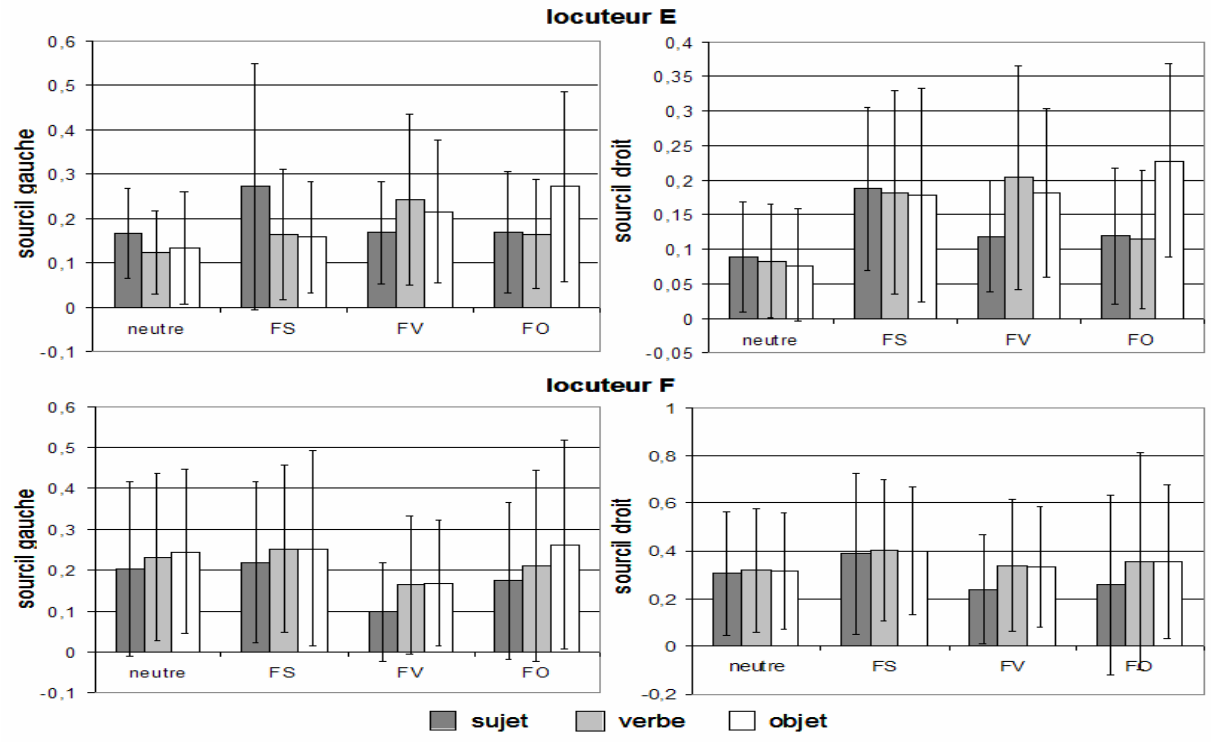


FIGURE III.42 – *Aires moyennes sous les courbes d'évolution temporelle des sourcils droit (à droite) et gauche (à gauche) moyennées pour chaque type de syntagme sous chaque type de focalisation et pour chaque locuteur.*

B.2.1.1. Locuteur B

Les graphiques a. et b. de la figure III.42 permettent de constater que, lorsque le locuteur B bouge les sourcils, c'est le plus souvent d'une part lorsque l'un des constituants de l'énoncé est focalisé (quasiment pas de mouvement dans le cas neutre), et d'autre part, précisément sur le constituant focalisé lui-même (les barres S&FS, V&FV et O&FO sont plus hautes que toutes les autres). En fait, on observe un mouvement du sourcil droit (respectivement gauche) supérieur au niveau du cas neutre dans 92% (respectivement 88%) des cas pour la focalisation sur le sujet, dans 65,4% (respectivement 69,2%) des cas pour la focalisation sur le verbe et dans 94,7% (respectivement 89,5%) des cas pour la focalisation sur l'objet. Il semblerait donc que ce locuteur bouge quasiment systématiquement ses sourcils lorsqu'il focalise. Cependant on notera une grande variation dans l'amplitude moyenne de ces mouvements comme le suggèrent les écarts-types importants des graphiques a. et b. de la figure III.42. Pour le sourcil droit, les amplitudes moyennes des mouvements varient de 0,3 à 4,15 mm (moyenne : 1,28 mm) et, pour le sourcil gauche, elles varient de 0,09 à 3,59 (moyenne : 0,75 mm). Or on pourra intuitivement penser qu'un mouvement de faible amplitude (0,09 mm par exemple) sera certainement peu perceptible.

De plus, on constate grâce aux graphiques a. et b. de la figure III.42 que les mouvements sont parfois amorcés avant le début du constituant focalisé (barres S&FV, S&FO et V&FO plus hautes que celles correspondant au cas neutre) et se terminent après la fin de ce même constituant (barres V&FS, O&FS et O&FV plus hautes que celles correspondant au cas neutre). On constate également que ce locuteur bouge plus le sourcil droit que le sourcil gauche.

B.2.1.2. Locuteur C

Les graphiques c. et d. de la figure III.42 permettent de faire les mêmes observations que pour le locuteur B : lorsque le locuteur C bouge les sourcils, c'est le plus souvent d'une part lorsque l'un des constituants de l'énoncé est focalisé (quasiment pas de mouvement dans le cas neutre), et d'autre part, précisément sur le constituant focalisé lui-même (les barres S&FS, V&FV et O&FO sont plus hautes que toutes les autres). Le sourcil droit (respectivement gauche) bouge sur le constituant focalisé plus que dans le cas neutre dans 64% (respectivement 72%) des cas pour la focalisation sur le sujet, dans 39,1% (respectivement 69,6%) des cas pour la focalisation sur le verbe et dans 76,9% (respectivement 88,5%) des cas pour la focalisation sur l'objet. Ce locuteur bouge donc souvent ses sourcils sur le constituant focalisé mais on constatera que les amplitudes moyennes des mouvements sont très faibles puisqu'elles varient de 0,07 à 1,77 mm (moyenne : 0,26mm) pour le sourcil droit et de 0,02 à 2,31 mm (moyenne : 0,27 mm) pour le sourcil gauche. On peut donc s'interroger sur la possibilité de percevoir des mouvements d'amplitudes si faibles.

Les graphiques c. et d. de la figure III.42 montrent aussi que lorsque le constituant focalisé est accompagné d'un mouvement des sourcils, celui-ci est parfois amorcé avant le début du constituant focalisé (barres S&FV, S&FO et V&FO plus hautes que celles correspondant au cas neutre) et se prolonge après la fin de ce même constituant (barres V&FS, O&FS et O&FV plus hautes que celles correspondant au cas neutre). Ce locuteur bouge les deux sourcils de façon à peu près équivalente.

B.2.1.3. Locuteur D

Les graphiques e. et f. de la figure III.42 illustrent le comportement du locuteur D quant aux mouvements de sourcil. Il semblerait que ce locuteur bouge plus les sourcils lorsque l'énoncé est focalisé par rapport au cas neutre (barres correspondant au cas neutre quasiment à 0 alors que les autres sont plus grandes). Pourtant il n'apparaît aucun lien entre le type de focalisation et la localisation des mouvements des sourcils sauf peut-être que les mouvements sont plus souvent exécutés en début d'énoncé.

B.2.1.4. Locuteur E

Les graphiques g. et h. de la figure III.42 permettent de faire les mêmes observations que pour les locuteurs B et C : il semble que lorsque le locuteur D bouge les sourcils, ce soit le plus souvent d'une part lorsque l'un des constituants de l'énoncé est focalisé (quasiment pas de mouvement dans le cas neutre), et d'autre part, précisément sur le constituant focalisé lui-même (les barres S&FS, V&FV et O&FO sont plus hautes que toutes les autres). Le sourcil droit (respectivement gauche) bouge sur le constituant focalisé plus que dans le cas neutre dans 76,9% (respectivement 57,7%) des cas pour la focalisation sur le sujet, dans 69,2% (respectivement 69,2%) des cas pour la focalisation sur le verbe et dans 88% (respectivement 72%) des cas pour la focalisation sur l'objet. Les amplitudes moyennes des mouvements des sourcils sont très faibles : elles varient de 0,1 à 0,57 mm (moyenne : 0,26mm) pour le sourcil droit et de 0,16 à 1,15 mm (moyenne : 0,37 mm) pour le sourcil gauche.

Les graphiques g. et h. de la figure III.42 montrent aussi que lorsque le constituant focalisé est accompagné d'un mouvement des sourcils, celui-ci est parfois amorcé avant le début du constituant focalisé (barres S&FV, S&FO et V&FO plus hautes que celles correspondant au cas neutre) et se prolonge après la fin de ce même constituant (barres V&FS, O&FS et O&FV plus hautes que celles

correspondant au cas neutre). Les amplitudes moyennes des mouvements du sourcil gauche sont plus importantes.

B.2.1.5. Locuteur F

Les graphiques i. et j. de la figure III.42 montrent que le locuteur F bouge très souvent les sourcils que l'énoncé soit focalisé ou non. Néanmoins l'amplitude moyenne de ces mouvements est toujours assez faible puisque le mouvement le plus ample observé est de 1,48 mm. Il est donc fortement probable que, chez ce locuteur, les mouvements produits soient liés à la parole en générale et non exclusivement des corrélats de la focalisation

B.2.1.6. Conclusion

On notera qu'il semble y avoir une correspondance entre mouvements de sourcils et focalisation chez seulement trois des locuteurs (B, C et E). Les autres locuteurs soit ne bougent pas les sourcils du tout, soit les bougent légèrement mais de façon aléatoire par rapport à l'énoncé.

Chez les locuteurs chez qui on observe une correspondance, celle-ci n'est pas systématique *i.e.* on n'observe pas systématiquement un mouvement de sourcil sur le constituant focalisé. Le locuteur B produit cependant très souvent un mouvement de sourcil pour accompagner la focalisation. Chez les deux autres locuteurs, les fréquences sont plus faibles.

On notera cependant qu'hormis chez le locuteur B, les amplitudes moyennes des mouvements des sourcils lors de la focalisation sont très faibles. Il est possible que ces mouvements ne soient d'ailleurs d'aucune utilité pour la perception.

Enfin, on notera que, seul les mouvements des deux sourcils du locuteur C ont des amplitudes moyennes similaires. Le locuteur B bouge plus son sourcil droit alors que le locuteur E bouge plus le gauche.

B.2.2. Mouvements de la tête

La figure III.43 fournit les résultats des mesures des mouvements de la tête pour chaque type de syntagme sous chaque type de focalisation et pour chaque locuteur.

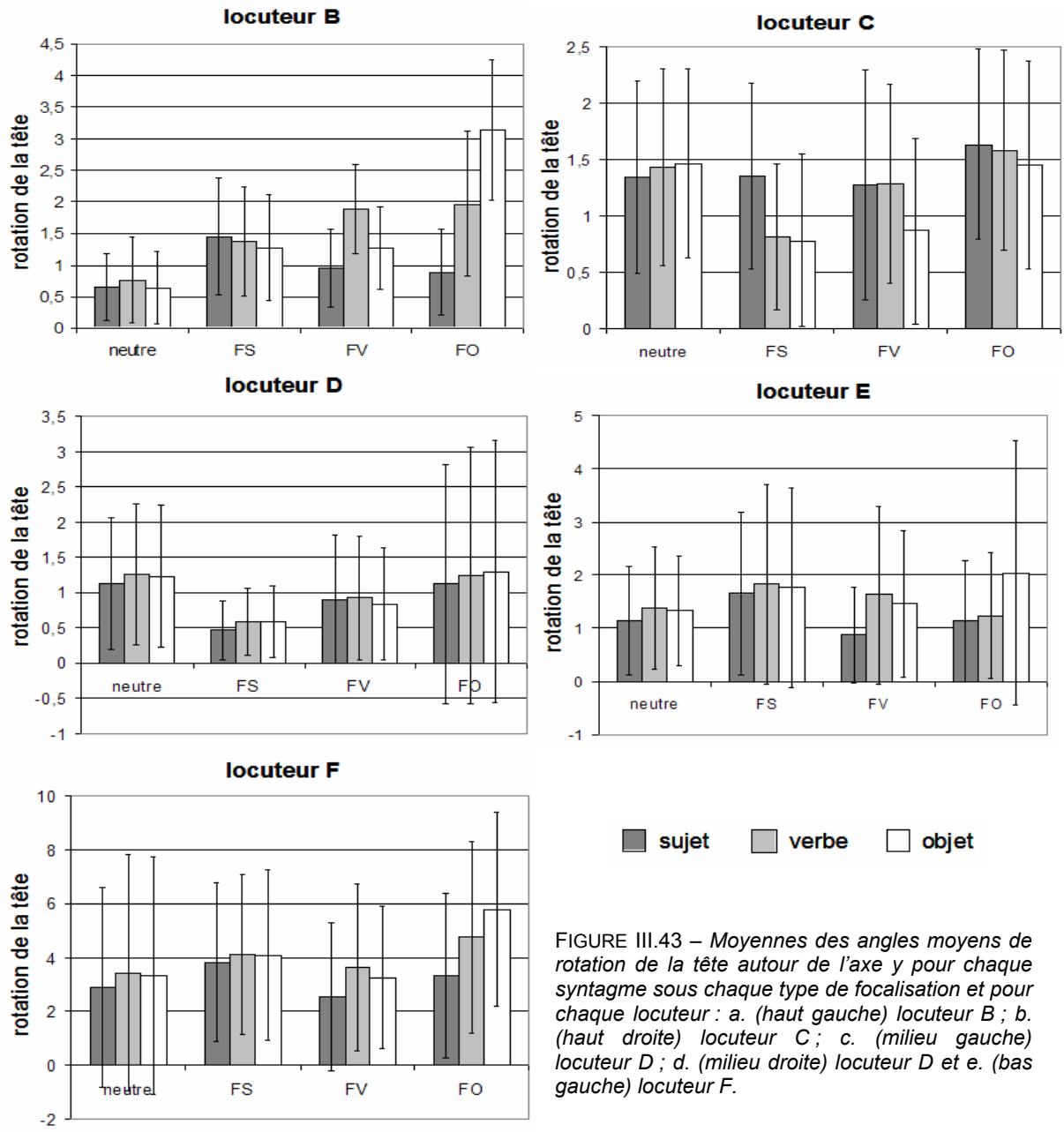


FIGURE III.43 – *Moyennes des angles moyens de rotation de la tête autour de l'axe y pour chaque syntagme sous chaque type de focalisation et pour chaque locuteur : a. (haut gauche) locuteur B ; b. (haut droite) locuteur C ; c. (milieu gauche) locuteur D ; d. (milieu droite) locuteur D et e. (bas gauche) locuteur F.*

On constate que seul le locuteur B semble utiliser les mouvements de la tête en correspondance avec la focalisation (barres S&FS, V&FV et O&FO plus hautes que toutes les autres barres). Ce locuteur semble ainsi produire un mouvement de hochement de la tête vers le haut sur le constituant focalisé. Chez les autres locuteurs, on observe aussi des mouvements de tête mais ils ne semblent pas être liés à la focalisation, il en existe d'ailleurs autant sur le cas neutre que sur les cas focalisés, voire plus. La validité de ces résultats peut être partiellement remise en cause par les conditions de mesure. Bien que les mouvements de la tête ne soient pas impossibles avec l'Optotrak contrairement au systéme de mesure labiométrique, il est possible que les locuteurs ne se soient pas sentis très libres de leur mouvements craignant de perturber les mesures en faisant tomber les marqueurs par exemple.

B.3. Conclusions : sourcils, tête et focalisation : qu'en est-il donc ?

Il apparaît donc qu'il pourrait y avoir un lien entre gestes faciaux et focalisation prosodique. Ce lien est cependant loin d'être systématique surtout en ce qui concerne les mouvements de la tête. Il semble de plus qu'il soit très dépendant du locuteur tant au niveau de sa réalisation qu'au niveau de l'amplitude du geste et de l'alignement temporel de celui-ci avec l'audio. On soulignera néanmoins que cette étude n'était que préliminaire puisqu'elle n'envisageait que les mouvements de haussements des sourcils et de hochement (rotation autour de l'axe y) de la tête vers le haut. Il est possible que d'autres composantes de ces gestes jouent un rôle. Il est également possible que d'autres gestes faciaux interviennent. Il faudra donc approfondir cette étude.

– Chapitre IV –

Intervention des indices « visibles » dans la perception de la deixis prosodique

Nous savons maintenant grâce aux études menées en production (cf. chapitre précédent), qu'il existe des indices visibles de la deixis ou focalisation contrastive prosodique en français. La focalisation peut donc en principe être vue. Cependant, on ne sait pas si ces indices « visibles » sont en effet « vus » c'est-à-dire utilisés lors du processus de perception de la parole. Le but de ce chapitre sera d'analyser cet aspect du problème grâce à plusieurs tests perceptifs en modalité visuelle puis audiovisuellle.

A. Perception visuelle de la deixis prosodique

C'est avant tout à la perception des indices « visibles » identifiés dans le chapitre III qu'on s'intéressera ici. C'est pourquoi les premiers tests menés furent des tests perceptifs en modalité visuelle seule. Le but est en premier lieu de savoir si une telle perception de la focalisation prosodique est possible *i.e.* est-il possible, lorsqu'on n'entend pas du tout le locuteur, et qu'on le voit simplement parler, de détecter si une focalisation a été produite et éventuellement sur quel constituant elle l'a été ? En plus de savoir si une telle perception est possible, le défi sera de tenter d'identifier les indices la permettant et de voir s'il s'agit de ceux repérés lors de l'étude en production.

Comme pour les études en production, il a été décidé de commencer par effectuer un test préliminaire fondé sur une tâche la plus simple et claire possible et surtout sur une tâche pour laquelle on fait en sorte qu'aucune tâche parasite ne vienne distraire les participants. Le test préliminaire portera ainsi sur la perception de la parole délexicalisée. Ce test ne pouvant être que préliminaire, la parole délexicalisée ayant ses limites et ne permettant pas toujours de tirer des conclusions sur la parole en général, des tests perceptifs utilisant la parole réelle (lexicalisée) ont ensuite été menés et ce en utilisant les productions de plusieurs locuteurs.

A.1. Test A : Étude préliminaire : parole délexicalisée

L'intérêt d'utiliser ici la parole délexicalisée est de centrer la tâche perceptive sur l'unique détection de la focalisation et d'avoir le moins de difficultés ou de distractions extérieures possibles pour les participants. Toutes les syllabes sont identiques et donc les participants n'auront pas, par exemple, à effectuer de tâche de discrimination syllabique visuelle (lecture labiale).

A.1.1. Description du test

A.1.1.1. Corpus

Quatre phrases du corpus AV1[116], ainsi que tous les enregistrements en parole délexicalisée leur correspondant, ont été extraits à partir de l'enregistrement décrit dans la section A.2.2.1 du chapitre III. Si le corpus n'a pas été utilisé dans son intégralité, c'était par souci de simplification pour les

[116] Le corpus AV1 a été décrit en détails dans la section 3.1.1 du chapitre I et est résumé en annexe 1.

participants. Les phrases choisies sont en effet celles pour lesquelles les structures syllabiques sont les plus équilibrées du corpus. Ces phrases ne comportent pas d'items monosyllabiques et sont composées de syntagmes ne différant pas de façon excessive par leur nombre de syllabes. Le but du test est en effet d'analyser la capacité des participants à détecter et localiser la focalisation *i.e.* à dire s'il y en a et sur quel constituant elle porte : sujet, verbe ou objet. Le fait d'utiliser des phrases à peu près équilibrées permettait de simplifier la tâche. La focalisation portant toujours au moins sur deux syllabes et les syntagmes n'étant pas trop complexes à localiser (correspondance approximative avec initial, médian et final), la tâche de localisation d'un syntagme n'était pas trop complexe bien qu'il s'agisse de parole délexicalisée. Les phrases utilisées furent les phrases (2), (4), (6) et (7) du corpus AV1 :

(2) [Romain]$_{S2}$ [ranima]$_{V3}$ [la jolie maman]$_{O5}$

(4) [Véroniqua]$_{S4}$ [mangeait]$_{V2}$ [les mauvais melons]$_{O5}$

(6) [Mon mari]$_{S3}$ [veut ranimer]$_{V4}$ [Romain]$_{O2}$

(7) [Les loups]$_{S2}$ [suivaient]$_{V2}$ [Marilou]$_{O3}$

A.1.1.2. Protocole expérimental

L'expérience a été menée dans une pièce calme de l'ICP dans laquelle les participants étaient isolés des expérimentateurs. Les participants étaient assis devant un moniteur vidéo avec des haut-parleurs disposés de chaque côté. Les images vidéo présentées aux participants au test étaient celles qui avaient été enregistrées par le locuteur A et qui ont servi pour l'étude décrite dans la section A.2.2.1 du chapitre III. La figure III.1 de cette même section donne ainsi un exemple des images vues par les participants au test perceptif. Ces images peuvent paraîtrent non naturelles puisque les yeux sont cachés et que le locuteur est maquillé en bleu. Cependant, le but étant de tester la perception des corrélats décrits lors de l'étude en production, c'est-à-dire les corrélats articulatoires de la bouche et de la mandibule, ces images conviennent parfaitement. Elles permettent en effet de contrôler le fait que les corrélats utilisés pour la perception visuelle pourraient être d'une autre nature (mouvements des sourcils par exemple). De plus, il a été suggéré que les marqueurs spécifiques placés sur le visage d'un locuteur ne perturbaient pas de façon significative les processus perceptifs.

Pendant le test, les participants entendaient d'abord un prompt audio[117] constitué d'une des quatre phrases énoncée en parole lexicalisée et de façon neutre (cas neutre). Les participants entendaient donc les « vrais » mots et non les /mamama/. Ils avaient été informés au préalable que, soit le sujet, soit le verbe, soit l'objet de cette phrase était erroné. Ils voyaient ensuite sans rien entendre la vidéo d'un locuteur qui corrigeait la phrase en remplaçant le constituant erroné par le constituant correct. Ce faisant, le locuteur produisait une focalisation sur le constituant corrigé. Le constituant erroné avait exactement le même nombre de syllabes que le constituant correct. La correction était effectuée en parole délexicalisée (toutes les syllabes étaient remplacées par la syllabe unique /ma/). Les participants voyaient les vues de face et de profil du locuteur produisant la correction mais n'entendaient rien. L'exemple (IV.1) permet de comprendre comment s'est déroulé le test, les lettres majuscules symbolisent la focalisation :

[117] Prompt audio : stimulus audio enregistré à l'avance visant à déclencher une production spécifique chez l'interlocuteur.

(IV.1) Le participant entend un premier locuteur dire :

Romain ranima la jolie maman.

La phrase correcte est (mais le participant ne le sait pas) :

Denis ranima la jolie maman.

Le participant voit un second locuteur (locuteur A) dire :

MAMA mamama ma mama mama.

Avant de passer le test, les participants avaient été avertis du fait que parfois le second locuteur ne produisait aucune correction. Dans ce cas, l'énoncé était neutre et le premier locuteur n'avait en fait commis d'erreur sur aucun constituant.

La tâche était de déterminer quel constituant (S, V, O ou aucun) avait été corrigé par le second locuteur (locuteur A). Les participants disposaient d'une feuille de réponse sur laquelle les phrases prononcées par le premier locuteur étaient présentées comme ci-dessous :

[Romain] [ranima] [la jolie maman.]

Les participants surlignaient le constituant qu'ils percevaient comme ayant été corrigé s'ils pensaient qu'il y avait eu correction et rien s'ils pensaient qu'aucun des constituants n'avait été corrigé par le second locuteur (locuteur A). L'utilité du prompt audio en parole lexicalisée était que les participants aient une référence auditive afin d'effectuer une tâche linguistique. En effet, si l'énoncé focalisé en parole délexicalisée avait été présenté directement aux participants, la tâche aurait pu être réduite à une simple tâche rythmique. Si le prompt n'a été présenté que selon la modalité auditive, c'est pour éviter que les participants n'effectuent une tâche de correspondance visuelle entre les deux énoncés. Ceci aurait en effet pu altérer le processus naturel de perception de la focalisation.

Un total de 32 paires d'énoncés (1 paire étant constituée de la phrase erronée en audio et de la même phrase corrigée en visuel) était disponible (4 phrases, 4 conditions de focalisation et 2 répétitions de chaque énoncé dans chaque condition de focalisation[118]). Cinq sous-tests ont été mis au point, chacun constitué d'une combinaison aléatoire des 32 paires. Le test s'est déroulé en cinq phases au cours desquelles chaque participant passait les cinq sous-tests. Les participants ont ainsi tous passé les mêmes cinq sous-tests mais dans des ordres différents (*e.g.* lors de la phase 1, le participant 1 a passé le sous-test 2 mais le participant 2 le sous-test 3 ; lors de la phase 2, le participant 1 a passé le sous-test 5 mais le participant 2 le sous-test 1 …). Un total de 160 paires d'énoncés a donc été présenté à chaque participant.

Avant de commencer le test en lui-même les participants se sont entraînés brièvement à la tâche sur cinq paires de phrases différentes de celles utilisées dans le test lui-même. Cette phase d'entraînement avait essentiellement pour but de familiariser les participants avec la tâche et surtout avec la parole délexicalisée qui peut être assez déstabilisante. Cette phase est cependant restée très brève de façon à ce que les participants ne s'habituent pas trop à la tâche avant même de commencer le test. Les participants n'ont de plus obtenu aucun retour sur leurs performances (*i.e.* s'ils avaient identifié correctement ou non le constituant focalisé). Les phrases utilisées lors des phases d'entraînement étaient différentes de celles utilisées pendant le test.

[118] Pour plus de détails sur l'enregistrement des données, voir la section A.2.2 du chapitre III.

A.1.1.3. Les participants

Un total de 28 participants de langue maternelle française a passé le test (11 hommes et 17 femmes). Ces participants étaient âgés de 18 à 52 ans et étaient originaires de diverses régions de France. A leur connaissance, aucun d'entre eux n'avait de problèmes auditifs ou visuels. Une fois que chaque participant avait passé le test, il lui était demandé d'expliquer de façon précise ce qu'il/elle avait fait. Ceci a permis de vérifier que la tâche accomplie était bien la bonne. Trois participants avaient ainsi mal compris la tâche à accomplir et leurs résultats ont donc été éliminés pour l'analyse.

A.1.1.4. Méthode d'analyse des résultats

Une analyse préliminaire consistera à avoir une *vue d'ensemble* des résultats. Elle visera ainsi à analyser la moyenne globale des résultats et à comparer celle-ci au niveau de hasard[119]. A l'issue du test, tous les participants avaient passé les mêmes cinq sous-tests mais dans des ordres différents. Les cinq sous-tests étaient constitués des mêmes stimuli mais dans des ordres aléatoires différents. Il faudra donc vérifier que les scores sont significativement indépendants de l'ordre dans lequel les stimuli étaient présentés *i.e.* que les scores ne dépendent pas du sous-test considéré. Si les résultats sont globalement significativement plus élevés que le niveau de hasard, les analyses décrites ci-dessous seront menées.

La deuxième étape visera à tester l'*influence de l'entraînement i.e.* à déterminer si l'entraînement a une influence sur les résultats. Le test s'est déroulé en cinq phases successives durant lesquelles les participants ont tous passé les mêmes cinq sous-tests (un sous-test par phase) mais dans des ordres différents. Le but sera de déterminer si les participants ont progressé au cours du test, c'est-à-dire s'ils ont obtenu de meilleurs résultats pour la cinquième phase qu'ils ont passée que pour la première par exemple. Il est en effet possible que l'entraînement à la tâche joue un rôle et qu'une fois que les participants sont habitués à l'exercice, leurs performances s'améliorent. Les phases peuvent êtres comparées entre elles puisqu'elles sont constituées exactement des mêmes stimuli mais dans des ordres différents.

Les étapes suivantes viseront à analyser les résultats de façon approfondie. La troisième étape visera ainsi à étudier les résultats en fonction de la *condition de focalisation* considérée (FS, FV, FO ou cas neutre). Le but sera de déterminer si un (ou plusieurs) des types de focalisation a (ont) été plus facile(s) à identifier que les autres. La quatrième étape consistera à analyser la *matrice de confusion* c'est-à-dire à analyser les tendances dans les erreurs commises par les participants. Ont-ils fait plus souvent des erreurs dans un sens que dans un autre ? La cinquième et dernière étape consistera à analyser les *résultats obtenus pour chaque stimulus*. Cette étape visera notamment à comparer les scores perceptifs obtenus pour chaque stimulus à la présence et à l'intensité du marquage des indices « visibles » identifiés lors de l'étude en production détaillée à la section A.2.2.2 chapitre III.

[119] Niveau de hasard : 25% car il y a quatre réponses possibles : neutre, FS, FV ou FO.

A.1.2. Résultats

A.1.2.1. Vue d'ensemble

La figure IV.1 donne les pourcentages de réponses correctes (condition de focalisation identifiée correctement) pour chaque participant. Chaque barre correspond à un participant. Le pourcentage moyen de réponses correctes pour tous les participants est de 86% ce qui est significativement supérieur au niveau de hasard qui est de 25%[120] (t=22,82 p<0,001). On peut ainsi conclure qu'il existe une information visuelle véhiculant la focalisation contrastive et que les participants y ont été très sensibles. Les scores obtenus sont étonnement élevés compte tenu du fait que les participants ont rapporté avoir trouvé le test difficile. Ceci suggère que les indices prosodiques visuels doivent être utilisés sans que les participants en aient pleinement conscience.

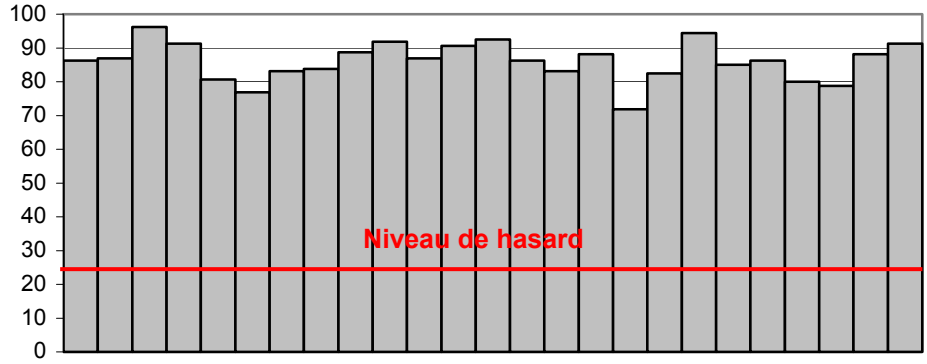


FIGURE IV.1 – *Pourcentages de réponses correctes (condition de focalisation identifiée correctement) pour chaque participant (chaque barre correspond à un participant).*

Afin d'examiner un à un les facteurs dans le but d'éliminer d'abord ceux qui ne sont pas significatifs, plusieurs ANOVA successives ont été effectuées avec exactement le même jeu de données. Il a ainsi été vérifié que les scores étaient significativement indépendants de l'ordre de présentation des stimuli. Une analyse de la variance à un facteur (sous-test considéré) montre en effet que l'hypothèse nulle d'égalité des moyennes pour les cinq sous-tests ne peut pas être rejetée (F(4,120)=0,077 p=0,989).

A.1.2.2. Influence de l'entraînement

On a vu que l'ordre de présentation, c'est-à-dire le sous-test considéré, n'avait pas d'effet significatif sur les résultats. Le but est maintenant de déterminer s'il y a eu progression au cours des cinq phases du test. Le graphique de la figure IV.2 donne les moyennes des pourcentages de réponses correctes en fonction de la phase de test considérée. Il permet de voir qu'il y a en moyenne une légère progression absolue (+7,8%) au cours du test de la phase 1 à la phase 5. Une analyse de la variance à un facteur permet de mettre en évidence l'existence d'un effet significatif de l'entraînement sur les performances des participants (F(4,120)=4,115 p=0,004). L'analyse des contrastes montre en fait que la progression globale (entre les phases 1 et 5) est significative (p<0,001). Il apparaît que la

[120] Niveau de hasard : 25% car il y a quatre réponses possibles (neutre, FS, FV ou FO).

progression se fait essentiellement entre les phases 1 et 2 (p=0,008) puisqu'il n'y a plus de progression significative entre les phases 2 et 5 (p=0,313). Il existe donc bien un effet de l'entraînement mais celui-ci n'est significatif qu'entre les phases 1 et 2. On pourra donc supposer qu'il est la conséquence de l'existence d'un temps d'adaptation à la tâche. Les participants trouvaient en effet que la tâche était un peu complexe et ont ainsi certainement eu besoin d'une phase du test (la première) pour bien se familiariser avec la tâche. On notera tout de même que l'effet de l'entraînement est faible puisque, entre les résultats pour la phase 1 et la moyenne des résultats pour les autres phases, on observe une progression absolue de 6,2% seulement. Les performances lors de la première phase de test étaient en effet déjà très bonnes (81,1% de réponses correctes) et bien supérieures au niveau de hasard (25%). On peut donc penser que les indices utiles à la détection visuelle de la focalisation sont utilisés de façon partiellement non consciente et ne dépendent pas vraiment de l'entraînement.

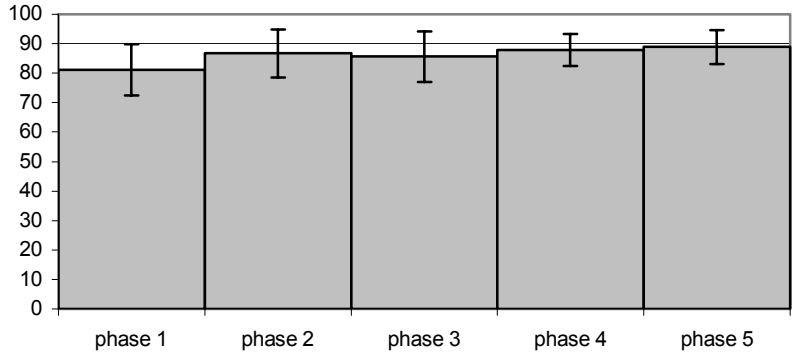


FIGURE IV.2 – *Moyennes des pourcentages de réponses correctes pour chacune des cinq phases du test.*

A.1.2.3. Différences entre les types de focalisation (cas neutre, FS, FV et FO)

Le but de cette analyse est d'étudier les performances des participants en fonction de la condition de focalisation à identifier (cas neutre, focalisation sur le sujet, sur le verbe ou sur l'objet). Ces performances sont données dans le graphique de la figure IV.3. Une analyse de la variance à un facteur (type de focalisation[121]) permet de mettre en avant un effet significatif de la condition de focalisation sur les résultats ($F(3,96)=6,457$ $p<0,001$). L'analyse des contrastes de l'ANOVA permet de montrer que les performances sont significativement meilleures lorsque la focalisation porte sur le sujet (p=0,003). Le cas neutre est significativement plus facile à détecter que les cas de focalisation sur le verbe ou sur l'objet (p=0,002) mais on ne relève pas de différence significative entre les performances correspondant aux cas neutres et celles correspondant aux cas de focalisation sur le sujet (p=0,542). On peut ainsi conclure que la focalisation sur le sujet et le cas neutre sont plus faciles à identifier que les autres conditions.

Il convient d'approfondir ces observations en analysant les types d'erreurs commises par les participants grâce à l'étude de la matrice de confusion.

[121] Type de focalisation : facteur intra-sujet à quatre niveaux : FS, FV, FO et cas neutre.

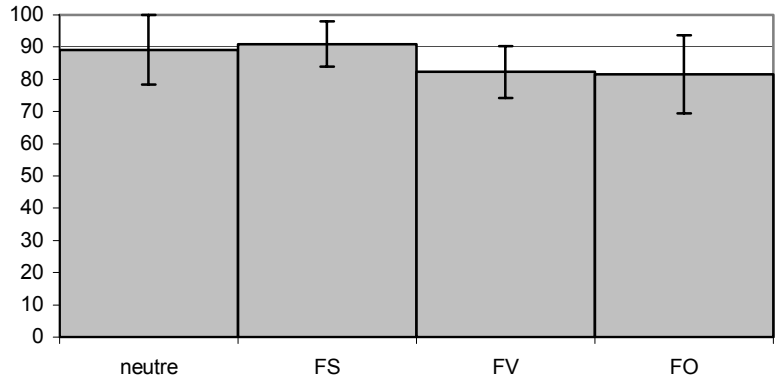


FIGURE IV.3 – *Moyennes des pourcentages de réponses correctes pour chacune des conditions de focalisation (cas neutre, FS, FV et FO).*

A.1.2.4. Analyse de la matrice de confusion

stimulus \ réponse fournie	neutre	FS	FV	FO	
neutre	**88,8**	7,8	2	1,4	100
FS	6,9	**91**	2,1	0,1	100
FV	16	2,2	**81**	0,5	100
FO	8,2	0,2	10	**82**	100
	30	25,3	23,7	21	

TABLE IV.1 – *Matrice de confusion donnant les fréquences (en %) de chaque type d'association faite par les participants. Par exemple, 91% des stimuli correspondant à une focalisation sur le sujet ont été identifiés comme étant focalisés sur le sujet (et donc identifiés correctement). La colonne de droite correspond à la somme des pourcentages de tous les types de réponses pour un type de stimulus (cette somme doit être égale à 100%). La ligne du bas correspond à la moyenne des pourcentages de chaque type de réponse (pourcentage du nombre de réponse total).*

La table IV.1 donne les fréquences (en %) de chaque type d'association. On retrouve que globalement ce sont les associations correctes qui sont les plus fréquentes c'est-à-dire celles correspondant à la diagonale. Néanmoins, ce sont les erreurs commises et le sens dans lequel elles l'ont le plus fréquemment été qui vont principalement attirer ici notre attention. On constate ainsi que lorsqu'elle est mal identifiée, la focalisation sur le sujet est le plus souvent confondue avec le cas neutre (6,9% des stimuli FS sont confondus avec le cas neutre contre seulement 2,1% confondus avec FV et 0,1% avec FO ce qui représente 75,8% des erreurs commises sur les stimuli FS). Les stimuli focalisés sur le verbe sont eux aussi le plus souvent confondus avec le cas neutre (16% des stimuli FV sont confondus avec le cas neutre contre seulement 2,2% avec FS et 0,5% avec FO, soit 85,6% des erreurs commises sur les stimuli FV). Quant aux stimuli focalisés sur l'objet, ils sont le plus fréquemment confondus soit avec une focalisation sur le verbe soit avec le cas neutre (10,1% des stimuli FO sont confondus avec FV soit 54,6% des erreurs commises et 8,2% avec le cas neutre soit 44,3% des erreurs commises sur les stimuli FO). Lorsqu'ils ont été mal détectés, les cas neutres ont

le plus souvent été confondus avec une focalisation sur le sujet (7,8% des stimuli neutres ont été confondus avec des énoncés focalisés sur le sujet ce qui représente 69,6% des erreurs commises sur les stimuli neutres).

La première tendance qui émerge donc est que les confusions se font le plus souvent vers le cas neutre. Les participants jugent ainsi plus souvent qu'il n'y a pas de focalisation lorsqu'il y en a, plutôt que de confondre un type de focalisation avec un autre. Ceci paraît assez naturel puisqu'on peut penser que la focalisation peut-être plus ou moins bien marquée visuellement selon les stimuli et qu'un énoncé focalisé peut ainsi parfois apparaître comme neutre. On notera de plus que globalement les participants ont plus souvent donné une réponse « neutre »[122] plutôt que n'importe quel autre type de réponse ce qui peut d'ailleurs expliquer le fait que les performances soient meilleures pour la condition neutre. La dernière ligne de la matrice de confusion de la table IV.1 donne les résultats des calculs des moyennes des pourcentages correspondant à un type de réponse que l'identification soit correcte ou non. Cette somme est la plus élevée pour le type de réponse « neutre » (30% contre 25,3% pour FS, 23,8% pour FV et 21% pour FO). Il y a donc une tendance à sur-percevoir le cas neutre.

Il apparaît que lorsque la focalisation sur le sujet est confondue avec un autre type de focalisation, il s'agit presque toujours d'une focalisation sur le verbe. La confusion se fait donc plutôt avec le constituant voisin du constituant focalisé.

On notera de plus que les réponses « focalisation sur l'objet » sont les moins fréquentes (voir la dernière ligne de la table IV.1). Il semble ainsi qu'il y ait une tendance à la sous-perception de la focalisation en fin de phrase (*i.e.* sur l'objet ici). Ce résultat avait d'ailleurs déjà été mis en évidence par Thompson [1934].

Le tableau IV.1 permet également de voir que les stimuli focalisés sur l'objet sont assez souvent confondus avec des cas de focalisation sur le verbe. Ce même tableau montre que lorsqu'une focalisation sur le verbe est confondue avec un autre type de focalisation, il s'agit le plus souvent d'une confusion vers la focalisation sujet. Les confusions faites par les participants se font ainsi plutôt vers le constituant directement pré-focal. Une explication possible de cette tendance est la présence d'une hyper-articulation pré-focale chez ce locuteur.

A.1.2.5. Analyse approfondie des résultats pour chaque stimulus

Globalement, les résultats de ce test perceptif suggèrent que des indices visuels peuvent intervenir lors des processus de perception et de compréhension de la focalisation contrastive prosodique en français. Cependant, ces résultats ne permettent pas de conclure que ce sont les indices « visibles » qui ont été identifiés dans le chapitre III qui sont en effet utilisés en perception. Le but est donc maintenant d'analyser de façon détaillée les réponses fournies pour chacun des 32 stimuli et d'essayer, si possible, d'établir une relation entre la qualité de la perception et la présence et l'intensité des indices décrits dans la section A.2.2.2 du chapitre III.

Les pourcentages de bonnes réponses ont été calculés pour chaque stimulus et sont fournis dans l'annexe 4. Il est ainsi apparu que seulement trois stimuli correspondaient à des pourcentages

[122] On rappellera également qu'il n'y avait pas de case à cocher pour les interprétations « neutre ». Dans l'analyse des résultats, les cas d'identification « neutre » sont donc confondus avec les cas où les sujets n'avaient tout simplement pas répondu. Cette faille dans le protocole a été corrigée par la suite sans que les performances pour le cas neutre ne s'en trouvent changées.

d'identification faibles (autour de 35%) alors que huit stimuli correspondaient à des pourcentages d'identification très élevés (supérieurs à 95%). Tous les autres stimuli correspondaient à des scores au-delà de 80% et donc à de très bons résultats perceptifs. Une analyse croisée a été réalisée entre les données perceptives et les données en production décrites à la section A.2.2.2 du chapitre III. Le but était de voir si les indices manquaient ou étaient faiblement marqués pour les trois stimuli très mal perçus et particulièrement bien marqués pour les huit stimuli très bien perçus. Les données de durées moyennes des syllabes focalisées, de mouvements de la mandibule et de vitesse de ces mouvements correspondant au constituant focalisé, ont été relevées et moyennées par catégories (*mauvaise perception i.e.* scores d'environ 35%, *bonne perception i.e.* scores de 80 à 85%, *très bonne perception i.e.* scores de 85 à 90%, *perception excellente i.e.* scores de 90 à 95% et *perception quasi-parfaite i.e.* scores de 95 à 100%) Ces moyennes ont ensuite été normalisées entre 0 et 1 pour chaque type de données (*e.g.* les données de durées entre elles …). Une valeur de 1 correspond ainsi à la catégorie perceptive pour laquelle l'indice en question est le plus marqué et une valeur de 0 à celle pour laquelle l'indice est le moins marqué. Les résultats de ces calculs sont donnés dans le graphique de la figure IV.4.

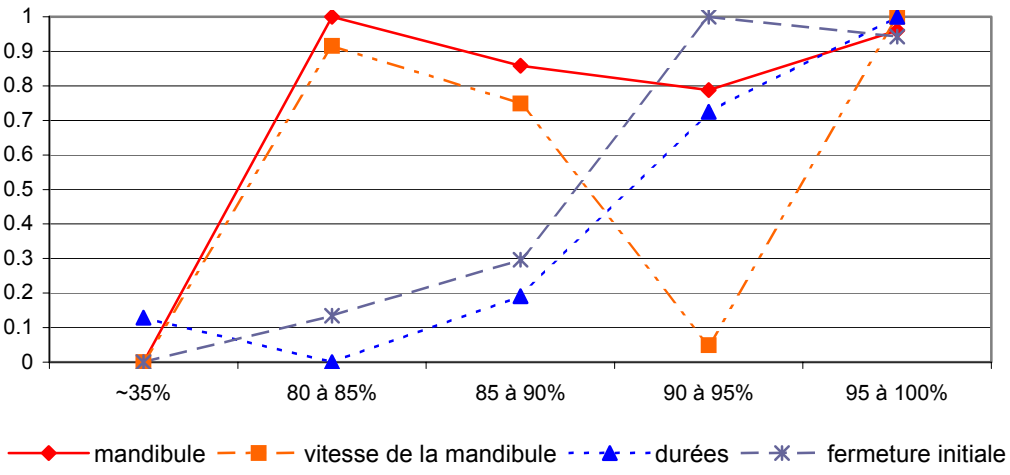


FIGURE IV.4 – *Valeurs normalisées entre 0 et 1 des données pour les différents indices visibles (mouvements de la mandibule, vitesse de ces mouvements, durée moyenne des syllabes et durée de la fermeture initiale pour l'élément focalisé, voir section A.2.2.2 du chapitre III pour plus de détails) pour chaque catégorie perceptive (« ~35% » : mauvaise perception ; « 80 à 85% » : bonne perception, « 85 à 90% » : très bonne perception, « 90 à 95% » : perception excellente et « 95 à 100% » : perception quasi-parfaite).*

En ce concerne les stimuli mal perçus (catégorie perceptive « ~35% »), il apparaît que les indices visibles sont bien souvent les moins bien marqués (tous les points correspondent à une valeur faible). Les mouvements de la mandibule et leur vitesse sont beaucoup moins amplifiés sous la focalisation pour ces stimuli que pour ceux correspondant à de bons ou très bons scores perceptifs. La fermeture initiale est également très faiblement marquée.

Quant aux stimuli perçus de façon quasiment parfaite (catégorie perceptive « 95 à 100% »), tous les indices visibles apparaissent très marqués (les valeurs correspondants à cette catégorie sont assez élevées). L'ouverture de la mandibule et sa vitesse pour le constituant focalisé sont très amplifiées sous la focalisation. La fermeture initiale des lèvres est elle aussi bien marquée. Tous les

indices visibles sont donc nettement marqués et surtout bien souvent les mieux marqués en comparaison avec les autres catégories perceptives (valeurs à 1).

Il apparaît donc que les stimuli très bien perçus correspondent aux stimuli pour lesquels les indices « visibles » sont les mieux marqués et que les moins bien perçus sont ceux pour lesquels ces mêmes indices sont les moins bien marqués. On pourra donc conclure que les indices visibles identifiés lors de l'étude en production (section A.2.2.2 du chapitre III) c'est-à-dire l'allongement et l'hyper-articulation mandibulaire focales et l'anticipation (temporelle et articulatoire) sont très probablement ceux qui sont utilisés pour la perception visuelle de la focalisation du moins quand seule la partie inférieure du visage du locuteur est visible.

En ce qui concerne les stimuli correspondants aux trois catégories perceptives intermédiaires, on remarquera que le marquage des indices est globalement assez bons (les scores de perception sont d'ailleurs très bons puisqu'ils sont tous supérieurs à 80%). Cependant, il apparaît que pour chacune de ces catégories au moins l'un des indices est peu marqué. Pour la catégorie « 80 à 85% », il s'agit des indices de durées et de fermeture initiale (indices temporels). Pour la catégorie « 85 à 90% », les indices articulatoires sont très légèrement moins bien marqués que pour la catégorie « 80 à 85% », mais les indices temporels sont mieux marqués. On notera ainsi qu'il apparaît que la perception est meilleure si tous les indices sont utilisés même s'ils sont un peu moins bien marqués. Pour la catégorie « 90 à 95% », on remarquera que les corrélats sont tous assez bien marqués sauf celui de vitesse de variation de la mandibule. En fait, une analyse détaillée des données de vitesse des stimuli correspondant à cette catégorie, montre que la moyenne est fortement abaissée à cause d'un stimulus en particulier pour lequel l'indice de vitesse n'est pas du tout utilisé. Les autres stimuli de cette catégorie sont quant à eux bien marqués en ce qui concerne la vitesse et, sans ce stimulus particulier, la moyenne serait bien plus élevée.

Il semble donc qu'il soit plus facile de détecter la focalisation visuellement lorsque tous les indices visibles sont présents même s'ils ne sont que moyennement marqués par rapport aux autres catégories perceptives que lorsque certains indices visibles sont très marqués mais les autres pas du tout.

A.1.3. Conclusion

Le but de ce test perceptif était double. Il s'agissait d'abord de déterminer si la détection de la focalisation était possible à partir de la modalité visuelle seule et donc si des indices visuels pouvaient être utilisés pour cette détection. Le second objectif était de déterminer dans la mesure du possible quels indices visuels étaient utilisés.

Il est apparu que la détection de la focalisation à partir de la modalité visuelle seule est non seulement possible mais aussi très bonne. Les résultats obtenus sont en effet nettement supérieurs au niveau de hasard. Bien que les participants aient besoin d'une phase d'adaptation, il semble que par la suite il n'y ait aucun effet d'entraînement. Il apparaît que les erreurs les plus souvent commises conduisent à confondre un cas de focalisation avec un cas neutre. Lorsque les participants confondent deux cas de focalisation, ils ont tendance à assimiler la focalisation à un constituant voisin plutôt qu'à un constituant plus éloigné de l'énoncé.

Une analyse détaillée des résultats pour chaque stimulus a permis de montrer que les stimuli mal perçus correspondent en moyenne à des cas où les corrélats visibles identifiés lors de l'étude en production sont les moins marqués. Les stimuli très bien perçus correspondent quant à eux aux cas

où les indices visibles sont en moyenne les mieux marqués. On peut ainsi penser que les indices utilisés en perception sont au moins en partie ceux identifiés dans l'étude en production. On remarque de plus que la focalisation est bien détectée quand au moins un des indices est présent. Il apparaît que la perception est meilleure quand tous les corrélats sont moyennement marqués plutôt que lorsque certains sont très marqués mais les autres seulement faiblement.

De façon générale, on pourra conclure que la modalité visuelle peut intervenir dans le processus de perception de la focalisation contrastive en français. Il faut néanmoins se souvenir que le présent test a été mené pour la parole délexicalisée et qu'il ne faut donc pas tirer de conclusions trop hâtives sur la parole en général.

A.2. Étude de la parole réelle (lexicalisée)

Grâce à l'étude préliminaire réalisée sur la parole délexicalisée et présentée ci-dessus, on peut penser qu'il existe sûrement des indices visuels pour percevoir la focalisation contrastive en français. Il semblerait de plus que les indices utilisés correspondent au moins en partie à ceux identifiés au cours de l'étude en production. Il reste cependant à confirmer ces résultats pour la parole réelle.

A.2.1. Test B : Test perceptif avec les données du locuteur A

Ce test a essentiellement pour but de confirmer les résultats obtenus précédemment pour la parole délexicalisée. Il sera donc basé sur les productions du même locuteur (locuteur A) mais cette fois-ci en parole lexicalisée.

A.2.1.1. Description du test

A.2.1.1.a. Corpus

Les enregistrements en parole lexicalisée des mêmes quatre phrases que celles utilisées pour l'étude préliminaire[123] ont été extraits à partir des enregistrements décrits dans la section A.2.2 du chapitre III.

A.2.1.1.b. Protocole expérimental

L'expérience a eu lieu dans une pièce calme de l'ICP afin d'isoler les participants à la fois des bruits extérieurs et des expérimentateurs. Les vidéos étaient montrées sur un moniteur vidéo avec des haut-parleurs disposés de chaque côté. Les images vidéos présentées aux participants du test perceptif étaient celles qui avaient été enregistrées pour le locuteur A et qui ont servis pour l'étude décrite dans la section A.2.2.3 du chapitre III. La figure III.1 de la section A.2.1.1 du chapitre III donne ainsi un exemple des images vues par les participants.

Il a été indiqué aux participants qu'ils suivraient une conversation entre deux interlocuteurs. Ils entendraient d'abord un locuteur prononçant un prompt audio[124]. Un des constituants de cette phrase

[123] C'est-à-dire les phrases (2), (4), (6) et (7) du corpus AV1.
[124] Prompt audio : stimulus auditif enregistré à l'avance et ayant pour but de déclencher une production spécifique chez l'interlocuteur.

(S, V ou O) serait mal compris par son interlocuteur, qui produirait la phrase comprise sur un mode interrogatif en vue de questionner le premier interlocuteur sur sa compréhension de la phrase (les participants ne verraient ni n'entendraient cette question). Le premier interlocuteur répéterait ensuite la phrase qu'il avait produite initialement en insistant sur l'élément mal compris par son interlocuteur (*i.e.* en le focalisant). Les participants verraient une vidéo de profil et de face de ce locuteur en train d'effectuer la correction, sans son. L'expérience se déroulait ainsi comme dans l'exemple (IV.2) :

(IV.2) Locuteur 1 (le participant entend) :

Romain ranima la jolie maman.

Locuteur 2 (mais le participant n'entend et ne voit rien) :

Denis ranima la jolie maman ?

Locuteur 1 (le participant voit mais n'entend rien) :

ROMAIN ranima la jolie maman.

Comme pour le test A, les participants étaient prévenus du fait que, parfois, il n'y aurait pas de correction (*i.e.* cas neutre). La tâche était la même que celle du test A. La grille de réponse a cependant été légèrement modifiée par rapport à celle du test A. Elle comportait ainsi à droite une case vide que les participants devaient surligner s'ils pensaient qu'il n'y avait pas eu de correction (*i.e.* identification d'un cas neutre). Le but était de discriminer les cas d'identifications volontaires d'un cas neutre par les participants, des cas d'omission de réponse ou de confusion. Répondre « cas neutre » nécessitait ainsi une action volontaire. La grille de réponse était du type de celle présentée ci-dessous :

Romain	ranima	la jolie maman.	

Pour le prompt audio, qui permet au participant de connaître la phrase sur laquelle portera la focalisation, nous avons utilisé les enregistrements correspondant aux cas neutres. L'utilité de ce prompt est la même que celle évoquée dans la description du protocole expérimental du test A (cf. section A.1.1.2 du présent chapitre).

Comme pour le test A, nous disposions de 32 paires de phrases au total (les mêmes que celles du test A mais dans leurs versions lexicalisées). Le test a été divisé en cinq sous-tests de la même façon que pour le test A. Les participants ont tous passé les cinq sous-tests au cours de cinq phases de test mais pas dans le même ordre. Chaque participant a donc évalué au total 160 paires de phrases. La phase d'entraînement s'est déroulée de la même manière que pour le test A.

A.2.1.1.c. Les participants

Un total de 33 participants de langue maternelle française a passé le test (8 hommes et 25 femmes). Ces participants étaient âgés de 18 à 52 ans et étaient originaires de diverses régions de France. Aucun d'entre eux n'avaient de problèmes auditifs ou visuels connus. A l'issue du test, chaque participant devait expliquer de façon précise ce qu'il/elle avait fait. Ceci a permis de vérifier que la tâche accomplie était bien la bonne. Les résultats d'un des participants ont ainsi dû être écartés parce qu'il s'est avéré qu'il n'avait pas bien compris la tâche.

A.2.1.1.d. Méthode d'analyse des résultats

Etant donné que le protocole expérimental mis en œuvre pour ce test perceptif est quasiment identique à celui du test A (cf. section A.1.1.2 du présent chapitre), la méthode d'analyse des résultats sera exactement la même. Pour plus d'informations, le lecteur pourra ainsi se reporter à la section A.1.1.4 du présent chapitre pour une description détaillée de cette méthode.

A.2.1.2. Résultats

A.2.1.2.a. Vue d'ensemble

La figure IV.5 donne les pourcentages de réponses correctes (condition de focalisation correctement identifiée) pour chaque participant. La moyenne des pourcentages de réponses correctes pour les 32 participants s'élève à 71,4%. Ce score est à comparer au niveau de hasard[125] auquel il est significativement supérieur ($t=32{,}111$ $p<0{,}001$). On pourra ainsi conclure à la fois qu'il existe des corrélats visuels à la focalisation contrastive prosodique pour ce locuteur en parole lexicalisée et que les participants y sont sensibles. Ce score élevé est une fois encore assez surprenant, étant donné que la plupart des participants a rapporté que le test leur avait paru difficile. Ceci suggère que les indices visuels seraient utilisés de façon partiellement non consciente.

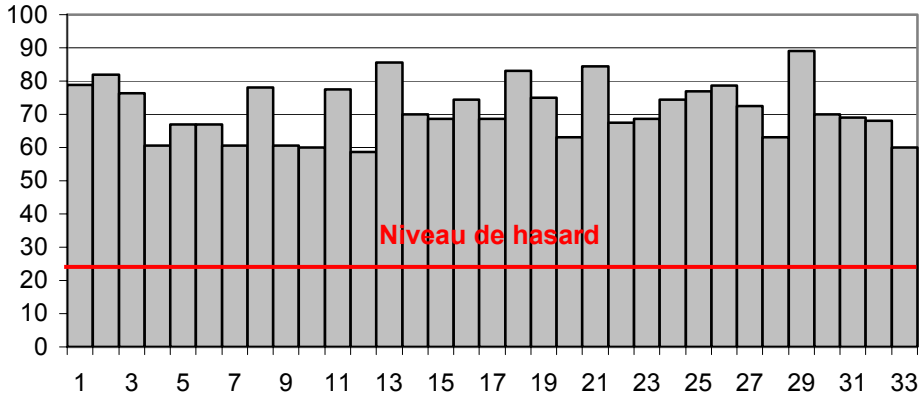


FIGURE IV.5 – *Pourcentages de réponses correctes (condition de focalisation identifiée correctement) pour chaque participant (chaque barre correspond à un participant).*

Afin d'examiner un à un les facteurs dans le but d'éliminer d'abord ceux qui ne sont pas significatifs, plusieurs ANOVA successives ont été effectuées avec exactement le même jeu de données. Il a été vérifié que les scores étaient significativement indépendants de l'ordre de présentation des stimuli. Une analyse de la variance à un facteur (sous-test[126]) montre en effet que l'hypothèse nulle d'égalité des moyennes pour les cinq sous-tests ne peut pas être rejetée ($F(4,160)=0{,}176$ $p=0{,}95$).

[125] Niveau de hasard : 25% car il y a quatre réponses possibles : neutre, FS, FV ou FO.
[126] Facteur intra-sujet à cinq niveaux : sous-tests 1, 2, 3, 4 et 5.

A.2.1.2.b. Influence de l'entraînement

Le graphique de la figure IV.6 donne les moyennes des pourcentages de réponses correctes en fonction de la phase de test considérée. Une analyse de la variance à un facteur (phase du test[127]) permet de montrer que les cinq moyennes ne sont pas significativement différentes (F(4,160)=0,548 p=0,701). On constate ainsi que les scores des participants ne se sont pas améliorés avec l'entraînement, l'habitude à la tâche ne leur a en effet pas permis d'obtenir de meilleurs résultats.

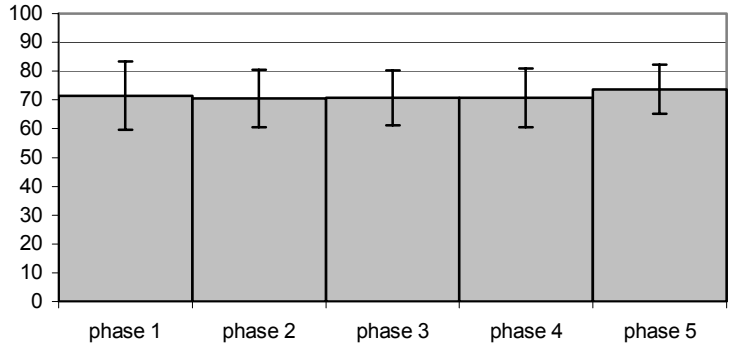


FIGURE IV.6 – *Moyennes des pourcentages de réponses correctes pour chacune des cinq phases du test.*

A.2.1.2.c. Différences entre les types de focalisation (FS, FV, FO et cas neutre)

Le graphique de la figure IV.7 donne les moyennes des pourcentages de réponses correctes en fonction de la condition de focalisation. Une analyse de la variance à un facteur (type de focalisation[128]) montre qu'il existe un effet significatif du type de focalisation (F(3,128)=11,297 p<0,001). L'analyse des contrastes de l'ANOVA permet de constater que les performances sont significativement meilleures lorsque la focalisation porte sur le sujet que pour tous les autres types de focalisation (p<0,001). Le cas neutre est quant à lui plus facile à identifier que les cas de focalisation sur le verbe ou sur l'objet (p=0,001). Les résultats obtenus pour les conditions de focalisation sur le verbe et sur l'objet ne sont pas statistiquement différents (p=0,83). Les résultats obtenus ici sont ainsi sensiblement les mêmes que ceux qui avaient été obtenus pour le test préliminaire en parole délexicalisée.

Ces observations peuvent être approfondies par l'étude de la matrice de confusion.

[127] Facteur intra-sujet à cinq niveaux : phases 1, 2, 3, 4 et 5.
[128] Type de focalisation : facteur intra-sujet à quatre niveaux : FS, FV, FO et cas neutre.

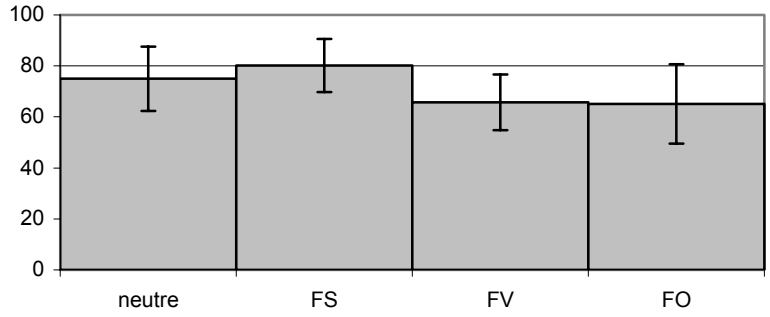


FIGURE IV.7 – *Moyennes des pourcentages de réponses correctes pour chacune des conditions de focalisation (cas neutre, FS, FV et FO).*

A.2.1.2.d. Analyse de la matrice de confusion

stimulus \ réponse fournie	neutre	FS	FV	FO	
neutre	**74,9**	9,0	11,5	4,5	100
FS	15,8	**80,2**	3,9	0,2	100
FV	26,1	7,4	**65,7**	0,8	100
FO	9,0	0,6	25,4	**65,0**	100
	31,5	24,3	26,6	17,6	

TABLE IV.2 – *Matrice de confusion donnant les pourcentages de chaque type d'association faite par les participants. Par exemple : 80,2% des stimuli correspondant à une focalisation sur le sujet ont été identifiés comme étant focalisés sur le sujet (et donc identifiés correctement). La colonne de droite correspond à la somme des pourcentages de tous les types de réponses pour un type de stimulus (cette somme doit être égale à 100%). La ligne du bas correspond à la moyenne des pourcentages de chaque type de réponse (pourcentage du nombre de réponse total).*

La table IV.2 fournit la matrice de confusion des erreurs observées. On remarquera en premier lieu que les associations les plus fréquentes sont évidemment celles correspondant à des réponses correctes (diagonale de la matrice). On constate que lorsqu'une erreur est commise sur un stimulus focalisé sur le sujet, c'est le plus souvent vers le cas neutre (15,8% des stimuli FS sont confondus avec le cas neutre contre seulement 3,9% confondus avec FV et 0,2% avec FO ce qui représente 79,4% des erreurs commises sur les stimuli FS). Les erreurs commises sur les stimuli focalisés sur le verbe se font également le plus souvent vers le cas neutre (26,1% des stimuli FV sont confondus avec le cas neutre contre seulement 7,4% confondus avec FS et 0,8% avec FO ce qui représente 76,1% des erreurs commises sur les stimuli FV). Les stimuli focalisés sur l'objet qui ont été mal identifiés ont le plus souvent été confondus avec une focalisation sur le verbe (25,4% des stimuli FO sont confondus avec FV contre seulement 9% confondus avec le cas neutre et 0,6% avec FS soit 72,6% des erreurs commises sur les stimuli FO). Quand un stimulus neutre est interprété comme étant focalisé, c'est le plus souvent sur le verbe ou sur le sujet (11,5% des stimuli neutres sont confondus avec FV soit 46% des erreurs commises et 9% sont confondus avec FS soit 36% des erreurs contre seulement 4,5% confondus avec FO).

La première tendance qui émerge de nouveau, est que les confusions sont le plus souvent faites avec le cas neutre. Ceci signifie que les participants donnent une interprétation neutre alors que le stimulus était focalisé. Ce résultat est le même que celui qui avait été obtenu pour le test mené avec la parole délexicalisée (voir section A.1.2.4 du présent chapitre). On constate également en parole lexicalisée que les réponses « neutre » sont les plus fréquentes[129] (cf. la dernière ligne du tableau IV.2 où ont été calculées les moyennes des colonnes de la matrice de confusion : 31,5 pour « neutre » contre 24,3 pour FS, 26,6 pour FV et 17,6 pour FO).

Il apparaît, comme pour le test A, que lorsque la focalisation sur le sujet est confondue avec un autre type de focalisation, il s'agit presque toujours d'une focalisation sur le verbe. De façon similaire, lorsque la focalisation sur l'objet est confondue avec un autre type de focalisation, il s'agit presque toujours d'une focalisation sur le verbe. La confusion se fait donc plutôt avec le constituant voisin du constituant focalisé qu'avec un constituant plus éloigné.

On notera qu'en parole lexicalisée, les réponses « focalisation sur l'objet » sont également les moins fréquentes (voir la dernière ligne de la table IV.2). Il semble ainsi qu'il y ait une tendance à la sous-perception de la focalisation en fin de phrase (*i.e.* sur l'objet ici).

Le tableau IV.2 permet également de voir que les stimuli focalisés sur l'objet sont très souvent confondus avec des cas de focalisation sur le verbe. Ce même tableau montre que lorsqu'une focalisation sur le verbe est confondue avec un autre type de focalisation, il s'agit le plus souvent d'une confusion vers la focalisation sur le sujet. Les confusions faites par les participants se font donc plutôt vers le constituant directement pré-focal. On observe ainsi une asymétrie des confusions avec les constituants pré- et post-focaux. La focalisation sur le sujet n'est en effet confondue avec une focalisation sur le verbe que dans 3,9% des cas alors qu'une focalisation sur le verbe est confondue avec une focalisation sur le sujet dans 7,4% des cas. De même, on remarque que la focalisation sur le verbe n'est confondue avec une focalisation sur l'objet que dans 0,8% des cas alors qu'une focalisation sur l'objet est confondue avec une focalisation sur le verbe dans 25,4% des cas. Une explication possible de cette tendance est la présence d'une hyper-articulation pré-focale chez ce locuteur (locuteur A) qui conduirait donc plutôt les participants à se tromper en croyant que c'est le constituant pré-focal qui est focalisé.

On constatera que les résultats obtenus ici sont totalement en accord avec ceux qui avaient été obtenus en parole délexicalisée (cf. section A.1.2.4 du présent chapitre).

A.2.1.2.e. Analyse approfondie des résultats pour chaque stimulus

Les pourcentages moyens de réponses correctes ont été calculés sur tous les locuteurs et ce pour chaque stimulus, le tableau de l'annexe 5 fournit le détail de ces pourcentages. Il est ainsi apparu que deux stimuli correspondaient à des scores très faibles (~5%), que deux autres correspondaient à des scores proches du niveau de hasard (~25%) alors que tous les autres correspondaient à des scores de plus de 60% (de bons à très bons). Une analyse croisée a été réalisée entre les données perceptives recueillies ici et les données en production décrites à la section A.2.2.3 du chapitre III. Le but était ainsi de déterminer s'il y avait une corrélation entre les résultats perceptifs et la présence/saillance des indices articulatoires « visibles ». Les données de durées, d'aire intéro-labiale et de vitesse de variation de l'aire intéro-labiale ont été relevées puis normalisées entre 0 et 1 pour

[129] Notons que, suite au changement du type de grille de réponse utilisée, les réponses « neutre » ne peuvent plus correspondre à des omissions de réponse puisque les sujets devaient explicitement cocher une case pour répondre « neutre ».

chaque type de données (*e.g.* les données de durées entre elles …) et enfin moyennées par catégories perceptives (« ~5% » : scores perceptifs d'environ 5% *i.e. très mauvaise perception*, « ~25% » : stimuli correspondant à des scores perceptifs d'environ 25% *i.e. mauvaise perception*, « 60 à 75% » : scores perceptifs compris entre 60 et 75% *i.e. bonne perception,* « 75 à 90% » : stimuli correspondant à des scores perceptifs compris entre 75 et 90% *i.e. très bonne perception*, « 95 à 100% » : scores perceptifs compris entre 95 et 100% *i.e. perception excellente*). Les résultats de ces calculs sont donnés dans le graphique de la figure IV.8.

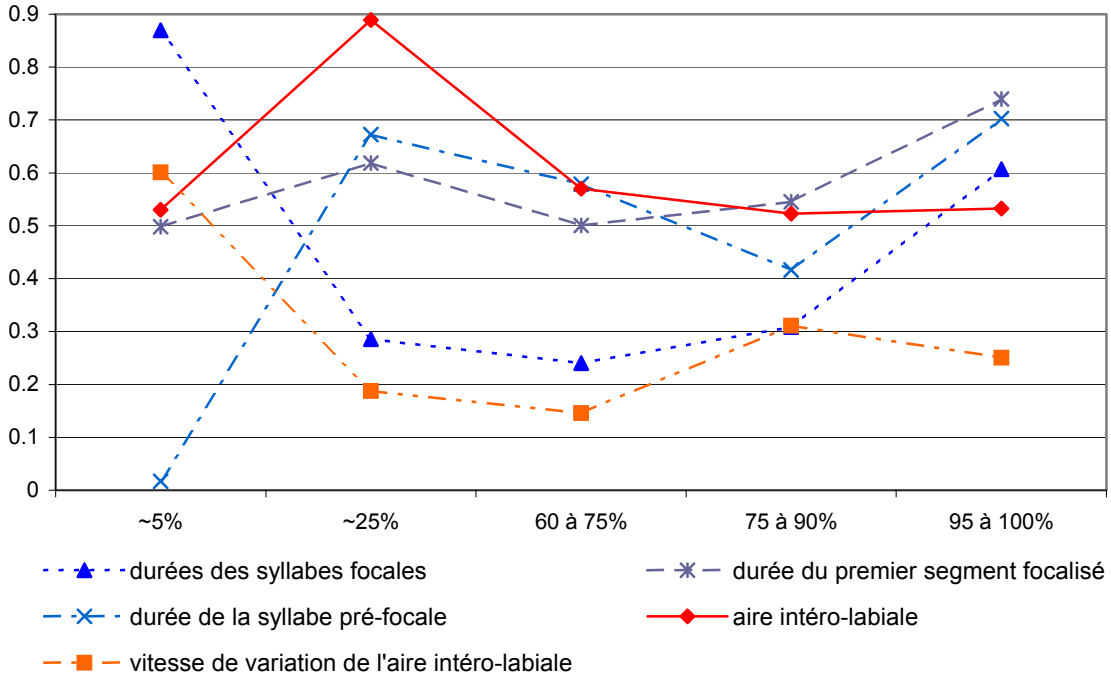


FIGURE IV.8 – *Moyennes des données de durées des syllabes focales, de durée du premier segment focalisé, de durée de la syllabe pré-focale, d'aire intéro-labiale et de vitesse de variation de l'aire intéro-labiale du constituant focalisé pour tous les stimuli de chaque catégorie perceptive (« ~5% » : très mauvaise perception, « ~25% » : mauvaise perception, « 60 à 75% » : bonne perception, « 75 à 90% » : très bonne perception et « 95 à 100% » : perception excellente).*

Les deux stimuli correspondant à des scores autour de 5% correspondent aux deux versions de la phrase (7) avec focalisation sur le verbe (*i.e. Les loups SUIVAIENT Marilou.*). On pourra s'étonner du fait qu'ils soient si mal perçus alors que la figure IV.8 montre qu'ils correspondent à des indices « visibles » bien marqués (sauf l'allongement de la syllabe pré-focale). Ce résultat peut être expliqué de par la nature des syllabes composant le constituant à détecter ici comme étant focalisé. La syllabe [syi] correspond à une très faible aire intéro-labiale pour [y]. Pour des raisons expliquées dans Cathiard *et al.* [2004], la transition [y-i] donne naissance à un glide intermédiaire et peut donc être prononcée [syɥi]. Lors de la production du glide, Cathiard *et al.* ont montré, chez ce même locuteur, que l'aire intéro-labiale peut devenir aussi faible que 0,5 mm². Lorsque cette aire très faible augmente avec la focalisation, elle augmente de façon très nette par rapport au cas neutre certes, mais partant d'une valeur basse, le pourcentage d'augmentation est forcément très important. Ce pourcentage d'augmentation très important ne signifie donc pas que la valeur d'aire intéro-labiale correspondant à l'élément focalisé est forte. Celle-ci reste en fait assez faible dans l'absolu. Il est ainsi possible que l'augmentation par rapport au cas neutre soit difficile à détecter parmi d'autres syllabes de l'énoncé

qui correspondront, de par leur nature, à une aire intéro-labiale nettement plus importante (comme le [mɑ] de *Marilou* par exemple). On notera aussi que, bien que l'aire intéro-labiale soit un peu plus importante pour l'autre syllabe du constituant focalisé ([vɛ]), la même remarque peut néanmoins être formulée. La combinaison de deux syllabes aux aires intéro-labiales faibles bien que nettement marquées par la focalisation, peut ainsi donner lieu à une mauvaise perception. On retiendra donc que les deux stimuli correspondant à des scores autour de 5% correspondent à un cas de figure tout à fait particulier.

Intéressons-nous maintenant aux cas des deux stimuli correspondant à des scores proche du niveau de hasard (25%). Ces deux stimuli correspondent aux deux versions de la phrase (4) avec focalisation sur l'objet (*i.e. Véroniqua mangeait LES MAUVAIS MELONS.*). Il paraît également étonnant que ces stimuli soient si mal perçus puisqu'ils correspondent à des indices « visibles » bien marqués (sauf la vitesse de variation de l'aire aux lèvres et la durée). En fait, on peut expliquer ce phénomène en remarquant que les autres constituants de l'énoncé (non-focalisés) correspondent de façon inhérente à de grandes ouvertures des lèvres et de la mandibule. L'objet quant à lui contient trois syllabes peu ouvertes ([mo], [mø] et [lɔ̃]) donc même lorsqu'il est focalisé et hyper-articulé, il doit apparaître moins ouvert que les autres constituants de l'énoncé. Le contraste intra-énoncé n'étant pas suffisant, les participants ne parviennent pas à détecter la focalisation.

Tous les autres stimuli correspondent à de bons scores de perception et à des indices « visibles » assez bien marqués. On constate que la perception est la meilleure (catégorie « 95 à 100% ») lorsque tous les indices, sauf la vitesse de variation de l'aire intéro-labiale, sont bien marqués. Notons que le fait que la vitesse de variation de l'aire intéro-labiale soit faiblement marquée pour toutes les catégories autres que la catégorie « ~5% » peut certainement s'expliquer par le marquage « artificiel » excessif de cet indice pour cette dernière catégorie. Souvenons-nous en effet qu'une normalisation avait été effectuée pour chaque indice entre les catégories.

On observe d'autre part que la perception est un peu moins bonne quand les paramètres sont moins bien marqués ou quand l'un d'entre eux n'est pratiquement pas marqué (catégories « 60 à 75% » et « 75 à 90% »). La comparaison des données correspondant aux catégories « 60 à 75% » et « 75 à 90% » permet de conclure que la perception est meilleure lorsque tous les indices sont moyennement marqués plutôt que quand certains sont bien marqués mais d'autres non.

On pourra ainsi conclure que les stimuli les mieux perçus correspondent à ceux qui sont le plus intensément marqués par la focalisation, au moins en ce qui concerne les indices visibles étudiés en production, c'est-à-dire l'allongement des syllabes focales et tout particulièrement du premier segment du constituant focalisé, l'allongement de la syllabe pré-focale et l'hyper-articulation focale en aire intéro-labiale (cf. section A.2.2.3 du chapitre III). Cette observation permet de penser que les indices visuels utilisés pour la perception incluent ceux qui ont été décrits en production. Les participants s'appuient apparemment au moins en partie sur ces indices pour détecter la focalisation puisque, lorsque ceux-ci sont bien marqués, la perception est meilleure. Ces indices aident donc à la perception.

A.2.1.3. Conclusion

Les résultats de ce test perceptif visuel sur la parole lexicalisée permettent de confirmer les premières conclusions qui avaient été faites à l'issue du test préliminaire sur la parole délexicalisée (test A). Les résultats montrent en effet que les participants perçoivent avec succès la focalisation visuellement et

ce sans entraînement. Il apparaît donc qu'il existe des indices visuels à la focalisation contrastive prosodique en français et que ceux-ci sont utilisés lors de la perception visuelle. Comme pour le test préliminaire, lorsque les participants se trompent, c'est le plus souvent en faveur d'une interprétation neutre. On remarque aussi qu'il existe une tendance à la sous-perception de la focalisation sur l'objet qui est de plus souvent confondue avec une focalisation sur le verbe. Lorsqu'ils se trompent, les participants ont tendance à percevoir une focalisation sur un constituant voisin du constituant en effet focalisé. Les confusions s'orientent plus souvent vers le constituant pré-focal que vers le constituant post-focal. Il apparaît enfin que les stimuli les mieux perçus sont également ceux pour lesquels les indices visibles identifiés en production (cf. section A.2.2.3 du chapitre III) sont les mieux marqués. On peut donc penser que ces indices sont, au moins en partie, ceux qui sont utilisés pour la perception visuelle de la focalisation contrastive prosodique en français.

A.2.2. Test C : Test perceptif avec les données du locuteur B

Le test B a permis de confirmer les résultats obtenus après l'étude perceptive préliminaire sur la parole délexicalisée. Ce test a en effet permis de montrer qu'il existait dans la parole réelle (lexicalisée) des indices visuels permettant d'extraire l'information de focalisation contrastive et sa localisation. Néanmoins, ce test a été mené sur la base des données vidéo enregistrées pour le locuteur A. Or il a déjà été avancé dans le chapitre III que les indices « visibles » chez ce locuteur étaient très marqués et notamment plus que chez d'autres locuteurs. Le locuteur A est en effet un locuteur habitué à être enregistré et donc à articuler de façon très nette. Il est donc très intéressant d'utiliser ses productions dans le cadre d'études préliminaires, puisqu'on parvient ainsi à extraire relativement facilement les caractéristiques principales. Il est cependant nécessaire d'effectuer ensuite des tests avec d'autres locuteurs plus « naïfs » afin de confirmer les résultats et de tirer des conclusions plus générales. C'est pourquoi il a été décidé de mener un test perceptif avec les données vidéo enregistrées pour le locuteur B.

A.2.2.1. Description du test

A.2.2.1.a. Corpus

Neuf phrases du corpus AV2[130] ont été sélectionnées sur la base de la clarté des indices « visibles » de la focalisation. Le chapitre III a en effet montré que les indices « visibles » étaient bien moins marqués chez le locuteur B que chez le locuteur A. Afin que le test perceptif en visuel seul ne soit pas trop complexe, les vidéos ont toutes été visionnées et les phrases pour lesquelles la détection était susceptible d'être trop complexe ont été éliminées. Neuf phrases ont ainsi été retenues (les phrases (5)-(13)) :

(5) $[\text{La nounou}]_{S3}$ $[\text{mariera}]_{V3}$ $[\text{Li}]_{O1}$

(6) $[\text{Le lama lent}]_{S4}$ $[\text{lut}]_{V1}$ $[\text{Marinella}]_{O4}$

(7) $[\text{Marinella}]_{S4}$ $[\text{va laminer}]_{V4}$ $[\text{Numu}]_{O2}$

(8) $[\text{Lou}]_{S1}$ $[\text{mima}]_{V2}$ $[\text{le lama}]_{O3}$

(9) $[\text{Le nominé}]_{S4}$ $[\text{lut}]_{V1}$ $[\text{les longs mots}]_{O3}$

[130] Le corpus AV2 a été décrit en détail dans la section A.2.3.1.a du chapitre III et est résumé en annexe 2.

(10) [La nounou]$_{S3}$ [vit]$_{V1}$ [Lou]$_{O1}$

(11) [Les loups]$_{S2}$ [mimaient]$_{V2}$ [Marilou]$_{O3}$

(12) [Lou]$_{S1}$ [ramena]$_{V3}$ [Manu]$_{O2}$

(13) [Li]$_{S1}$ [ralluma]$_{V3}$ [les moulinets]$_{O4}$

Une fois les phrases sélectionnées, tous les enregistrements leur correspondant ont été extraits (quatre conditions de focalisation et deux répétitions pour chacune). Les données utilisées ici correspondent à celles dont l'enregistrement a été décrit dans la section A.2.3.1.b du chapitre III c'est-à-dire à celles qui ont été analysées dans l'étude en production décrite dans la section A.2.3 du chapitre III.

A.2.2.1.b. Protocole expérimental

Les conditions expérimentales ainsi que le protocole mis en place étaient quasiment identiques à ceux du test B. La tâche perceptive était donc une tâche de détection de la correction (*i.e.* indirectement de la focalisation). La grille de réponse fournies aux participants était du même type que celle fournie aux participants au test B (cf. section A.2.1.1.b du présent chapitre). Les seules différences par rapport au test B sont décrites ci-dessous.

Nous disposions cette fois de 72 paires de phrases au total (9 phrases, 4 conditions de focalisation, 2 répétitions). Le test a donc été divisé en deux sous-tests correspondant à deux combinaisons aléatoires différentes des mêmes 72 paires. Les participants ont chacun passé une des deux phases avec la vue de face et l'autre avec la vue de profil. La moitié des images était ainsi cachée puisque celles-ci comprenaient les deux vues en même temps (cf. figure III.1 qui donne un exemple des images vidéo enregistrées). Si les deux vues ont été séparées, c'était dans le but, d'une part, de comparer les résultats perceptifs pour chacune des vues et, d'autre part, pour ne pas fournir trop d'informations simultanément aux participants, la tâche étant déjà relativement complexe. Le test avait donc lieu en deux phases au cours desquelles les participants ont tous passé les deux sous-tests mais pas dans le même ordre et pas forcément avec la même vue. Certains passaient ainsi d'abord l'ordre 1 en vue de face, puis l'ordre 2 en vue de profil alors que d'autres passaient d'abord l'ordre 1 en vue de profil, puis l'ordre 2 en vue de face ou encore d'abord l'ordre 2 en vue de face, puis l'ordre 1 en vue de profil. Il y avait ainsi quatre combinaisons possibles. Chaque participant a ainsi évalué 144 paires de phrases au total.

Avant chaque phase de test, les participants se sont brièvement entraînés à la tâche. Deux courtes phases d'entraînement (huit paires de phrases pour chacune d'elle) avaient lieu avant de passer le test correspondant à chacune des vues. Lors de la première phase, au lieu de se contenter de voir l'énonciation de la correction comme dans le test lui-même, les participants l'entendaient aussi afin de se familiariser avec le locuteur et avec la tâche de façon progressive. La seconde phase d'entraînement se déroulait exactement de la même façon que le test lui-même (énonciation de la correction en modalité visuelle seule). Les phrases utilisées lors des phases d'entraînement étaient différentes de celles utilisées pendant le test. Les participants n'ont obtenu aucun retour sur leurs performances (*i.e.* s'ils avaient identifié correctement ou non le constituant focalisé).

A.2.2.1.c. Les participants

Un total de 27 participants francophones de naissance a passé le test (4 hommes et 23 femmes). Ces participants étaient âgés de 18 à 31 ans et étaient originaires de diverses régions de France. Aucun

d'eux n'a rapporté de troubles auditifs ou visuels connus. Une fois que chaque participant eut passé le test, il lui était demandé d'expliquer de façon précise ce qu'il/elle avait fait. Ceci a permis de vérifier que la tâche accomplie était bien la bonne. Suite à cette vérification, il s'est avéré qu'aucun des participants n'avaient mal compris la tâche. Tous les résultats ont ainsi pu être conservés pour analyse.

A.2.2.2. Résultats

A.2.2.2.a. Vue d'ensemble

La figure IV.9 donne les pourcentages de réponses correctes (condition de focalisation correctement identifiée) pour chaque participant. La moyenne des pourcentages de réponses correctes pour les 27 participants s'élève à 43,2%. Ce score est significativement supérieur au niveau de hasard [131] (t=16,339 p<0,001). Il apparaît donc que les participants sont également sensibles aux informations visuelles sur la focalisation contrastive pour ce locuteur (locuteur B). Ces résultats confirment ainsi ceux qui avaient été obtenus pour le locuteur A bien que l'on constate, conformément aux attentes, que les résultats ne sont pas aussi bons. Le locuteur A ayant tendance à articuler de façon très nette probablement à cause de son habitude d'être enregistré, il est assez logique qu'il soit mieux perçu visuellement. Les données en production avaient en effet montré que les indices « visibles » étaient moins nettement marqués pour le locuteur B que pour le locuteur A. Ce score reste cependant relativement élevé et est tout de même surprenant, étant donné que la plupart des participants a rapporté que le test était très difficile. Ceci suggère une fois de plus que les indices visuels seraient utilisés de façon partiellement non consciente.

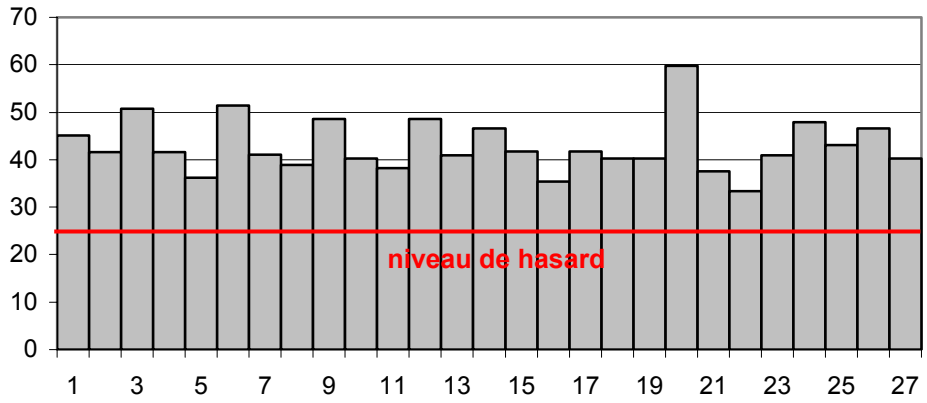


FIGURE IV.9 – *Pourcentages de réponses correctes (condition de focalisation correctement identifiée) pour chaque participant (chaque barre correspond à un participant).*

Il a été vérifié que les scores étaient indépendants de l'ordre de présentation des stimuli c'est-à-dire que les résultats étaient indépendants du sous-test considéré. Les deux sous-tests correspondent en effet à des ordres de présentation différents des stimuli. Une analyse de la variance à un facteur

[131] Niveau de hasard : 25% car il y a quatre réponses possibles : neutre, FS, FV ou FO.

(sous-test[132]) montre en effet que l'hypothèse nulle d'égalité des moyennes pour les deux sous-tests ne peut pas être rejetée (F(1,52)=0,012 p=0,913).

L'influence de l'entraînement a également été analysée. Le test s'est déroulé en deux phases successives, il est ainsi possible qu'on note une progression au cours de ces deux phases. Celle-ci pourrait être due à l'entraînement. Les participants s'étant habitués et entraînés à la tâche pourraient ainsi obtenir de meilleurs résultats lors de la seconde phase que lors de la première. Les phases peuvent êtres comparées entre elles puisqu'elles sont constituées exactement des mêmes stimuli mais présentés dans des ordres différents et qu'il vient d'être montré que l'ordre n'avait pas d'influence sur les résultats. Une analyse de la variance à un facteur (phase de test[133]) permet de montrer que l'entraînement n'a aucun effet significatif sur les performances des participants puisque l'hypothèse nulle d'égalité des moyennes pour les deux phases de test ne peut pas être rejetée (F(1,52)=0,456 p=0,503).

A.2.2.2.b. Influence de la vue : face ou profil

Comme il a été décrit dans le protocole expérimental (voir section A.2.2.1.b du présent chapitre), les participants ont passé deux phases de test au cours desquelles ils ont vu successivement, dans un ordre ou dans l'autre, deux vues différentes du locuteur : vue de face et vue de profil. Il se peut que ces vues aient eu une influence sur les performances des participants, par exemple que ceux-ci aient mieux réussi à détecter la focalisation pour l'une des deux vues. Le graphique de la figure IV.10 fournit les pourcentages de réponses correctes en fonction de la vue considérée. Il permet de voir que les résultats sont légèrement meilleurs pour la vue de profil que pour la vue de face. Cependant, une analyse de la variance à un facteur (vue[134]) permet de montrer qu'en fait la vue selon laquelle le locuteur est observé n'a pas d'influence sur les résultats obtenus. L'hypothèse nulle d'égalité des moyennes pour les deux vues ne peut en effet pas être rejetée (F(1,52)=1,415 p=0,24).

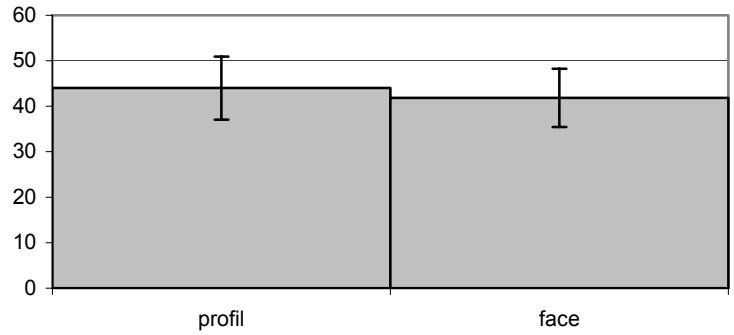


FIGURE IV.10 – *Moyennes des pourcentages de réponses correctes pour chacune des vues (profil et face).*

A.2.2.2.c. Influence du type de focalisation (neutre, FS, FV ou FO)

Le but est ici de déterminer si les performances dépendent du type de focalisation à identifier (cas neutre, FS, FV ou FO). Le graphique de la figure IV.11 donne les pourcentages de réponses correctes

[132] Facteur intra-sujet à deux niveaux : deux sous-tests.
[133] Facteur intra-sujet à deux niveaux : deux phases de test.
[134] Facteur intra sujet à deux niveaux : vue de face et vue de profil.

en fonction de la condition de focalisation. Une analyse de la variance à un facteur (type de focalisation [135]) montre qu'il y a un effet significatif du type de focalisation sur les résultats ($F(3,104)=11,682$ $p<0,001$). L'analyse des contrastes de l'ANOVA permet de noter que les résultats sont significativement meilleurs lorsque la focalisation porte sur le verbe ou sur le sujet ($p<0,001$). Les résultats correspondant à la focalisation sur l'objet sont significativement inférieurs à ceux obtenus pour tous les autres types de focalisation ($p<0,001$). Ces observations seront approfondies grâce à une analyse de la matrice de confusion.

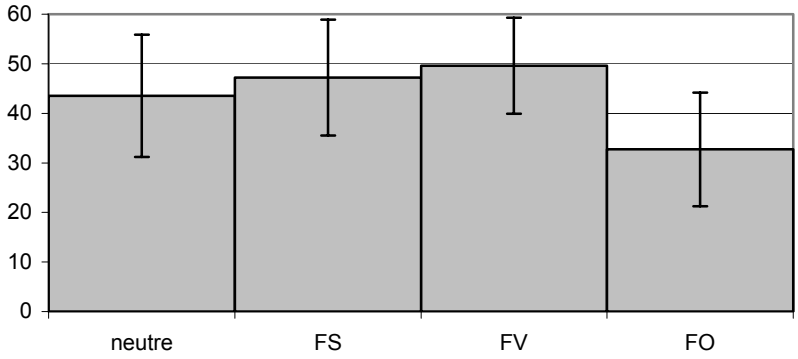


FIGURE IV.11 – *Moyennes des pourcentages de réponses correctes pour chacune des conditions de focalisation (cas neutre, FS, FV et FO).*

A.2.2.2.d. Analyse de la matrice de confusion

réponse fournie / stimulus	neutre	FS	FV	FO	
neutre	**43,5**	25,1	18,9	12,4	100
FS	36,0	**47,2**	13,2	3,6	100
FV	32,9	14,5	**49,6**	3,0	100
FO	32,0	12,8	22,5	**32,7**	100
	36,1	24,9	26,1	12,9	

TABLE IV.3 – *Matrice de confusion donnant les pourcentages de chaque type d'association faite par les participants. Par exemple : 47,2% des stimuli correspondant à une focalisation sur le sujet ont été identifiés comme étant focalisés sur le sujet (et donc identifiés correctement). La colonne de droite correspond à la somme des pourcentages de tous les types de réponses pour un type de stimulus (cette somme doit être égale à 100%). La ligne du bas correspond à la moyenne des pourcentages de chaque type de réponse (pourcentage du nombre de réponse total).*

Le tableau IV.3 fournit les pourcentages de chaque type d'association (matrice de confusion). Cette matrice montre que, la plupart du temps, les participants ont fourni des réponses correctes puisque les associations les plus fréquentes sont celles qui correspondent aux associations correctes (diagonale de la matrice). Intéressons-nous cependant plus particulièrement aux erreurs commises

[135] Facteur intra-sujet à quatre niveaux : neutre, FS, FV et FO.

(associations incorrectes) et aux tendances dans ces erreurs. Le tableau IV.3 permet de voir que lorsque la focalisation sur le sujet est incorrectement identifiée, elle est le plus souvent confondue avec le cas neutre (36% des stimuli FS sont confondus avec le cas neutre contre 13,2% confondus avec FV et 3,6% avec FO, ce qui représente 68,2% des erreurs commises sur les stimuli FS). La focalisation sur le verbe est, elle aussi, le plus souvent confondue avec le cas neutre (32,9% des stimuli FV sont confondus avec le cas neutre contre 14,5% confondus avec FS et 3% avec FO, ce qui représente 65,3% des erreurs commises sur les stimuli FV). Lorsque les participants identifient mal une focalisation sur l'objet c'est le plus souvent en faveur d'une interprétation neutre ou focalisation sur le verbe (32% des stimuli FO sont confondus avec le cas neutre soit 47,5% des confusions faites et 22,5% soit 33,4% des erreurs commises avec FV contre 12,8% avec FS). Lorsqu'ils sont identifiés comme étant focalisés, les cas neutres sont le plus souvent identifiés comme correspondant à des cas de focalisation sur le sujet ou sur le verbe (25,1% des cas neutres mal identifiés sont confondus avec FS soit 44,5% des erreurs commises et 18,9% avec FV soit 33,5% des confusions contre seulement 12,4% avec FO).

Il apparaît donc de façon nette, comme pour les test menés avec les données du locuteur A, que lorsque les participants font une erreur de détection c'est le plus souvent vers une interprétation neutre. C'est-à-dire que lorsque les participants se trompent, c'est souvent parce qu'ils ne détectent pas que l'énoncé est focalisé. Les réponses « neutre » sont d'ailleurs globalement les plus fréquentes (cf. dernière ligne du tableau IV.3 où ont été calculées les moyennes des pourcentages correspondant à chaque type de réponse : 36,1% contre 24,9% pour FS, 26,1% pour FV et 12,9% pour FO). On peut penser qu'en effet, si les indices visuels ne sont pas assez marqués, la tendance naturelle sera de ne pas se rendre compte qu'il y a focalisation.

Il apparaît aussi, avec les données du locuteur B, que lorsque la focalisation sur le sujet est confondue avec un autre type de focalisation, il s'agit beaucoup plus fréquemment d'une confusion avec un cas de focalisation sur le verbe. De façon similaire on observe que les focalisations sur l'objet sont confondues le plus fréquemment avec une focalisation sur le verbe. La confusion se fait donc plutôt avec le constituant voisin au constituant focalisé qu'avec un constituant plus éloigné.

On observe aussi une asymétrie des confusions avec les constituants pré- et post-focaux. La focalisation sur le verbe n'est confondue avec une focalisation sur l'objet que dans 3% des cas alors qu'une focalisation sur l'objet est confondue avec une focalisation sur le verbe dans 22,5% des cas. Bien que beaucoup moins nette, cette tendance asymétrique est aussi observé pour la focalisation sujet qui est confondue avec la focalisation sur le verbe dans 13,2% des cas alors que la focalisation sur le verbe est confondue avec une focalisation sur le sujet dans 14,5% des cas. Une explication possible de cette tendance est la présence d'une hypo-articulation post-focale chez le locuteur B qui empêcherait les participants de croire que le constituant post-focal est focalisé.

On notera que les observations faites sur les confusions pour ce test mené avec les productions du locuteur B vont dans le même sens que ce qui avait été observé pour le test perceptif B mené avec les productions du locuteur A (voir section A.2.1.2.e du présent chapitre).

A.2.2.2.e. Analyse approfondie des résultats pour chaque stimulus

Les pourcentages de réponses correctes ainsi que les pourcentages correspondant à chaque type de confusion ont été calculés pour chaque stimulus séparément et sont fournis en annexe 6. À la suite de ces calculs, six catégories perceptives ont pu être établies :

- **très mauvaise perception** : 30,6% des stimuli correspondent à des scores inférieurs au hasard (25%) dont 36,4% à des scores excessivement mauvais (inférieurs à 10% de réponses correctes) ;
- **mauvaise perception** : 13,9% des stimuli correspondent à des scores proches du hasard mais légèrement supérieurs (de 25 à 35% de réponses correctes) ;
- **perception moyenne** : des scores moyens (de 35 à 55% de réponses correctes) ont été obtenus pour 20,8% des stimuli ;
- **bonne perception** : de bons scores (de 55 à 75%) ont été obtenus pour 15,3% des stimuli ;
- **très bonne perception** : 12,5% des stimuli correspondent à de très bons scores perceptifs (de 75 à 85% de réponses correctes) ;
- **excellente perception** : 6,9% des stimuli correspondent à des scores perceptifs excellents (de 85 à 100% de réponses correctes).

Les scores perceptifs calculés pour chaque stimulus ont été combinés aux données correspondant aux indices visibles identifiés lors de l'étude en production (voir la section A.2.3 du chapitre III). Le but était d'étudier le marquage visible de la focalisation contrastive comparativement aux scores perceptifs obtenus et surtout de déterminer s'il existe des différences entre les indices visibles correspondant aux stimuli mal perçus et ceux correspondant aux stimuli bien perçus.

En ce qui concerne la catégorie perceptive des stimuli très mal perçus, on séparera l'analyse en deux sous-parties : les perceptions qui vont vers le cas neutre c'est-à-dire les stimuli pour lesquels ce sont des interprétations « neutre » qui ont été données le plus fréquemment au lieu de la réponse correcte et les perceptions qui vont vers un autre type de focalisation (FS, FV ou FO), c'est-à-dire les stimuli pour lesquels c'est un des constituants non focalisés qui a été détecté comme l'étant. Les confusions avec le cas neutre dans cette catégorie correspondent le plus souvent à des stimuli pour lesquels tous les indices visibles sont très faiblement marqués ou même souvent inexistants. On peut donc conclure que les stimuli très mal perçus correspondent le plus souvent à des indices « visibles » peu ou pas marqués. On observe cependant des cas (4 stimuli) pour lesquels, bien que la perception soit très mauvaise, il semble que les indices visibles soient bien, voire très bien, marqués. Intéressons-nous à ces cas particulier.

Le premier correspond à la phrase (5) avec focalisation sur le sujet (*i.e.* LA NOUNOU mariera Li.). Dans ce cas, il n'y a pas d'allongement focal. On note par contre un fort contraste intra-énoncé pour la protrusion (paramètre articulatoire prépondérant pour [la nunu]). Cependant, il est tout naturel que ce contraste soit important puisque les verbe et objet de cet énoncé ne sont pas du tout protrus. Les spectateurs peuvent donc s'attendre à ce que ce contraste intra-énoncé soit important même dans le cas neutre. Le contraste inter-énoncés est existant mais on constate qu'il n'est pas très fort. Finalement on pourra ainsi conclure que cet énoncé ne correspondait pas à des indices visuels très marqués.

Les deuxième et troisième cas correspondent à la phrase (10) avec focalisation sur l'objet (*i.e.* La nounou vit LOU.). Cet objet, bien que fortement marqué visuellement, ne comportait qu'une seule syllabe en fin d'énoncé. Il est possible que la prégnance visuelle soit ainsi insuffisante.

Le quatrième cas correspond à la phrase (12) focalisée sur le verbe (*i.e.* Lou RAMENA Manu.). L'interprétation de ce cas est la même que celle donnée pour le premier cas, le paramètre articulatoire prépondérant dans ce cas étant l'aire intéro-labiale.

Pour les stimuli ayant été confondus avec un autre type de focalisation, on note deux cas de figure. Le premier correspond aux cas où le constituant focalisé n'était pas très marqué et où un autre constituant était légèrement ou fortement marqué d'où la confusion. Le second cas correspond à des constituants monosyllabiques qui ont apparemment été regroupés avec un constituant voisin.

Quant à la catégorie perceptive des stimuli mal perçus, les stimuli le plus souvent confondus avec le cas neutre correspondent à des indices visibles très faiblement marqués. Les stimuli pour lesquels il y a confusion avec un autre type de focalisation sont aussi en plus relativement souvent confondus avec le cas neutre. Ils correspondent eux aussi le plus souvent à des indices visibles faiblement ou pas du tout marqués sur le constituant focalisé alors qu'il semble qu'un ou plusieurs des autres constituants du même énoncé soient marqués par l'un des paramètres normalement « réservés » au constituant focalisé (*e.g.* un allongement ou une hyper-articulation légère). La combinaison de ces deux phénomènes a pour conséquence soit de mener vers une interprétation « neutre » (pas de contraste intra-énoncé) soit de mener à une confusion avec un autre type de focalisation.

A quelques exceptions près, on notera donc que, globalement, les stimuli mal perçus correspondent à des cas où les indices visibles (allongement focal, hyper-articulation focale de l'aire intéro-labiale et de la protrusion et hypo-articulation post-focale de ces mêmes paramètres) sont très faiblement voire pas du tout marqués.

Les stimuli très bien perçus correspondent quant à eux le plus souvent à des corrélats tous très bien marqués. On note un cas pour lequel un seul corrélat est marqué. Ce cas est cependant exceptionnel puisque le marquage est anormalement élevé. Il existe cependant un stimulus pour lequel, bien qu'aucun des corrélats ne soient bien marqués (certains ne sont même pas marqués du tout), la perception est bonne.

Quant aux stimuli correspondant à une excellente perception, on note que les indices visibles leur correspondant sont tous très fortement marqués.

On pourra ainsi conclure, que de façon générale, les stimuli très bien perçus correspondent à un fort marquage des indices visibles.

On a donc noté que les stimuli mal perçus correspondent à des corrélats non ou très faiblement marqués alors que les stimuli bien perçus correspondent à des indices visibles très marqués. On pourra donc conclure que les indices visuels utilisés par les participants pour percevoir la focalisation contrastive prosodique comprennent les indices visibles qui avaient été identifiés lors de l'étude en production (voir section A.2.3 du chapitre III) c'est-à-dire un allongement et une hyper-articulation focale ainsi qu'une hypo-articulation focale (aire intéro-labiale et protrusion). Il existe cependant quelques exceptions intéressantes à relever *i.e.* d'un côté des stimuli mal perçus correspondant à des indices visibles très marqués et de l'autre des stimuli bien perçus alors que les indices visibles ne sont pas marqués. Ces derniers suggèrent que les indices visibles mis en avant ici ne sont sûrement pas les seuls à intervenir puisque s'ils ne sont pas présents, la perception peut tout de même être bonne. Les premiers peuvent laisser penser que des indices contradictoires peuvent peut-être parfois perturber la perception puisque, bien que les indices soient là, la perception est mauvaise.

A.2.2.3. Conclusion

Ce test perceptif a permis de confirmer les résultats obtenus pour le locuteur A, c'est-à-dire qu'il est possible de percevoir la focalisation contrastive prosodique à partir de la modalité visuelle seule. Il a en effet été trouvé pour le locuteur B qu'il existe des indices visibles à la focalisation contrastive

prosodique en français et que ceux-ci permettent d'extraire visuellement l'information de focalisation et ce, semble-t-il, de façon partiellement non consciente. L'entraînement n'a pas d'influence sur les résultats. Le test a été mené à la fois pour une vue de face du locuteur et pour une vue de profil mais il s'avère que cette vue n'a pas d'effet significatif sur les résultats. La focalisation sur le verbe est la plus facile à détecter alors que la focalisation sur l'objet est la moins facile. Les erreurs commises par les participants tendent à être souvent en direction d'une identification neutre en présence de focalisation. On observe une sous-perception du cas de focalisation sur l'objet qui est relativement fréquemment confondu avec un cas de focalisation sur le verbe. Globalement il apparaît que lorsqu'un type de focalisation est confondu avec un autre, il s'agit le plus souvent d'une confusion vers une focalisation qui porterait sur un constituant voisin plutôt qu'un autre constituant plus éloigné dans l'énoncé. Les participants tendent plus à considérer comme focalisé un constituant pré-focal qu'un constituant post-focal. On constate enfin qu'il existe une corrélation entre la présence et l'intensité des indices « visibles » identifiés dans l'étude en production (section A.2.3 du chapitre III) et le score perceptif obtenu. Il semble donc que les indices visibles identifiés jouent un rôle au niveau de la perception visuelle et fassent partie des indices visuels aidant à cette perception. Quelques exceptions permettent cependant de penser que ce ne sont pas les seuls.

A.3. Bilan : la perception visuelle de la deixis prosodique en français

Il a ainsi été montré pour deux locuteurs aux stratégies de focalisation « visuelle » assez différentes (voir section A.2.4 du chapitre III pour plus de détails), qu'il est possible de percevoir la focalisation contrastive prosodique à partir de la modalité visuelle seule. Ceci permet de conclure qu'il existe des indices visibles à la focalisation contrastive prosodique en français et que ceux-ci permettent d'extraire visuellement l'information de focalisation. Il semble que cette perception soit essentiellement non consciente pour les deux locuteurs puisque les participants réussissent très bien à accomplir la tâche (scores perceptifs nettement supérieurs au hasard) alors qu'ils la jugent difficile.

L'entraînement n'a pas d'effet sur les performances suggérant que la perception visuelle est inconsciente et ne résulte pas d'une procédure explicite. Ce n'est pas parce que les participants apprennent à accomplir la tâche qu'ils y parviennent si bien.

Globalement, les résultats sont nettement meilleurs pour le test mené avec les productions du locuteur A (71,4% de réponses correctes en moyenne pour le test B/locuteur A et 43,2% pour le test C/locuteur B). Ceci était prévisible puisque les résultats en production montraient déjà que les indices visibles de la focalisation étudiés (durée, aire intéro-labiale et protrusion) étaient nettement plus marqués chez le locuteur A. Néanmoins, comme il a été avancé dans la section A.2.4.1 du chapitre III, ces résultats sont quelque peu biaisés par le fait que le locuteur A est habitué à être enregistré et à articuler de façon très nette et claire. Le locuteur B quant à lui est « naïf » et est d'ailleurs souvent jugé comme articulant assez peu par ses proches. Les résultats obtenus représentent ainsi deux extrêmes et la moyenne se situe sûrement entre les performances obtenues pour chacun de ces locuteurs.

De façon détaillée, on notera que la focalisation sur l'objet est toujours la moins facile à détecter visuellement. Il apparaît d'ailleurs qu'elle est sous-perçue. Cette observation est en accord avec les

résultats décrits par d'autres chercheurs qui avaient constaté que la détection visuelle de la focalisation était moins bonne en fin de phrase (*e.g.* Thompson [1934]).

Les erreurs commises sont souvent de ne pas détecter la focalisation (réponses « neutre ») alors qu'il y en a.

On constate de plus pour les deux locuteurs une tendance assez nette à confondre une focalisation sur l'objet avec une focalisation sur le verbe et de façon générale à plutôt confondre une focalisation sur un constituant avec une focalisation sur un constituant immédiatement voisin. De plus, on note une asymétrie des confusions. Lorsque deux types de focalisation sont confondus, les participants ont plutôt tendance à détecter une focalisation sur le constituant pré-focal que sur le constituant post-focal. Chez le locuteur A, il s'agit peut-être là d'une conséquence de l'existence d'une hyper-articulation pré-focale rehaussant ce constituant. Chez le locuteur B, il s'agirait plutôt d'une conséquence de l'hypo-articulation post-focale inhibant la confusion vers ce constituant.

Il est ainsi apparu clairement qu'il existe des indices visuels à la focalisation constrastive prosodique en français puisqu'il est possible de la détecter à partir de la modalité visuelle seule. Néanmoins, nous avons également tenté de déterminer quels sont ces indices. Les études en production (sections A.2.2.3 et A.2.3 du chapitre III) avaient en effet permis d'identifier un certain nombre d'indices visibles à la focalisation *i.e.* un allongement de toutes les syllabes focales (encore plus net sur le premier segment du constituant focalisé) et une hyper-articulation focale ainsi qu'une anticipation de la focalisation (temporelle et articulatoire) chez le locuteur A et une hypo-articulation post-focale chez le locuteur B. Des études détaillées ont permis de mettre en évidence qu'une bonne perception correspondait bien souvent à un bon marquage de ces mêmes indices alors qu'une mauvaise perception correspondait à un marquage faible voire inexistant de ces indices. Ces observations permettent de penser que les indices visibles étudiés font sans doute partie des indices visuels utilisés lors de la perception visuelle de la focalisation. La présence de quelques exceptions permet de plus de penser que ce ne sont pas les seuls indices qui interviennent. On a également pu observer que pour une perception visuelle optimale, il vaut mieux que tous les indices soient présents même moyennement marqués plutôt qu'un seul mais fortement marqué. En fait, ceci permet de penser que les spectateurs ont des attentes sur la coordination inter-articulateurs et ce par rapport à leur pratique de la parole. Lorsqu'un seul paramètre est marqué, il est ainsi possible qu'ils jugent la situation non naturelle.

Maintenant qu'on sait qu'il existe des indices visuels à la deixis prosodique en français et au moins en partie lesquels ils sont, il serait intéressant de voir comment la modalité visuelle interagit avec la modalité auditive lors de la perception audiovisuelle de la parole.

B. Perception audiovisuelle : apport de la modalité visuelle lorsque la modalité auditive est dégradée

Les tests perceptifs décrits précédemment (section A du présent chapitre) ont permis de démontrer que la focalisation contrastive prosodique pouvait être bien perçue même lorsque seule la modalité visuelle était disponible. La focalisation contrastive prosodique est donc visible en français. Cependant, il serait intéressant de déterminer dans quelle mesure la perception de cette focalisation peut-être audiovisuelle, c'est-à-dire dans quelle mesure la modalité visuelle y contribue. Il va donc s'agir de mettre en place une expérience de perception audiovisuelle de façon à déterminer si la modalité visuelle apporte une information permettant d'améliorer les performances lors de la perception du *locus* de la focalisation contrastive en français.

B.1. Problématique : perception auditive vs. perception audiovisuelle

On s'attend à ce que la perception auditive de la focalisation contrastive soit très bonne et donc qu'il n'y ait pas d'amélioration significative des performances perceptives lorsque la modalité visuelle devient disponible. Il a donc fallu dégrader l'information prosodique acoustique de focalisation dans le signal afin de diminuer considérablement les performances perceptives en audio seul. Il deviendrait ainsi possible de déterminer si l'information visuelle permet de recouvrer l'information de focalisation. Comme il a été montré en introduction, les paradigmes mis en place lors des tests de perception audiovisuelle de la parole utilisent souvent le bruit pour dégrader le signal audio. Cependant, il apparaîtrait que le fait de bruiter un signal audio, bien que réduisant son intelligibilité lexicale, ne détériore pas le contour global de fréquence fondamentale (composante prosodique importante pour la focalisation). Miller & Nicely [1955] ont en effet par exemple montré que le trait de voisement est le plus résistant au bruit. Il fallait donc parvenir à dégrader le signal de fréquence fondamentale. Pour ce faire, nous avons eu l'idée d'utiliser la parole chuchotée, en effet, lorsque l'on chuchote, l'information de fréquence fondamentale est effacée puisqu'il n'y a plus de vibration des cordes vocales. De plus, la parole chuchotée sert à parler à quelqu'un dans le but que lui seul comprenne ce qu'on dit, sans que personne autour ne puisse entendre. La consigne donnée aux locuteurs a ainsi été de se faire comprendre par un interlocuteur situé en face d'eux et non de lui chuchoter à l'oreille. On peut donc penser que, dans ce cas, les locuteurs vont d'autant plus se servir du visuel pour communiquer ce qu'ils ne peuvent plus transmettre par le signal acoustique et donc qu'ils vont par exemple bien hyper-articuler sur les constituants importants.

B.2. Méthodologie expérimentale

B.2.1. Élaboration des stimuli

B.2.1.1. Données de base

Ce sont les données en parole chuchotée qui avaient été enregistrées précédemment pour les locuteurs A et B qui ont été utilisées. Pour des détails sur les enregistrements, le lecteur pourra se reporter aux sections A.2.2.1 et A.2.3.1 du chapitre III. Seules les phrases (2), (4), (6) et (7) ont été utilisées[136]. Comme il a été expliqué dans les sections A.2.2.1 et A.2.3.1 du chapitre III, deux répétitions de chaque phrase dans chaque type de focalisation (neutre, FS, FV et FO) ont été enregistrées en parole chuchotée. La répétition pour laquelle la focalisation avait été la mieux produite d'un point de vue audiovisuel, a été conservée. Un total de 16 énoncés par locuteur était ainsi disponible, soit 32 énoncés en tout.

B.2.1.2. Analyse : perception auditive de la focalisation contrastive pour la parole chuchotée

Après enregistrement, un test perceptif audio informel a été mené afin de vérifier que la perception auditive de la focalisation contrastive prosodique était bien dégradée avec la parole chuchotée. C'est en effet le but qui était recherché et la raison pour laquelle ce type de parole avait été choisi. Or ce test a révélé que cela n'était pas le cas. Il est ainsi apparu que la parole chuchotée permet de s'affranchir de l'information purement intonative comme on l'avait prévu (plus de fréquence fondamentale). Cependant, il semble qu'à l'écoute des signaux d'abord puis à leur analyse, l'information d'intensité prenne en quelque sorte le relais. Cette information est déjà présente en parole « normale » voisée mais elle semble avoir une moins grande importance puisque la F0 véhicule également l'information. Lorsque l'information de F0 n'est plus présente, l'information d'intensité semble être renforcée et parvient à véhiculer à elle seule le contraste. On observe ce qu'on pourrait appeler une adaptation acoustique à la perte de la fréquence fondamentale. De fortes variations d'intensité entre éléments focalisés et non focalisés sont ainsi mesurées. Il fallait donc parvenir à s'affranchir de cette information.

B.2.1.3. Pondération de l'intensité

Afin d'éviter que l'intensité ne fournisse trop d'information acoustique sur la focalisation, il a été décidé de la pondérer. Le but était ainsi de ramener l'intensité des énoncés focalisés au plus proche de celle de ces mêmes énoncés en version neutre. Or, lorsqu'un constituant est focalisé, son intensité est plus importante que dans la version neutre de la même phrase. Les constituants qui suivent le constituant focalisé ont quant à eux une intensité réduite par rapport à ce même cas neutre. Il fallait donc à la fois diminuer l'intensité des constituants focaux et augmenter l'intensité des constituants post-focaux. Ceci a été réalisé à l'aide d'un script Matlab.

[136] Pour une description détaillée, se reporter à la section C.1.1 du chapitre II ou à l'annexe 1.

Un premier traitement a consisté à mesurer l'intensité moyenne de chacun des constituants des énoncés dans leurs versions neutres, puis de ramener la valeur moyenne de l'intensité des constituants de tous les autres énoncés de cette phrase (énoncés focalisés) au niveau de l'intensité moyenne de la version neutre. Cette routine a été appliquée à tous les énoncés mais les stimuli audio obtenus étaient toujours trop nettement « focalisés ». Les coefficients de pondération ont donc été assignés empiriquement, en fonction du résultat obtenu à l'écoute et à l'analyse acoustique des signaux. Ce procédé empirique a été appliqué à chaque énoncé focalisé. Il a permis une adaptation aux caractéristiques particulières de chaque énoncé.

A la suite de ce traitement, un autre test perceptif audio informel a été mené et a montré que la focalisation contrastive était cette fois beaucoup moins bien perçue. La perception était tout de même supérieure au niveau du hasard ce qui était attendu puisqu'il subsistait encore un indice acoustique de la focalisation contrastive : la durée. On sait aussi maintenant que la focalisation s'accompagne d'une hyper-articulation. Or, lorsque l'on hyper-articule, les formants acoustiques sont mieux réalisés. Cette meilleure réalisation des formants pourrait être un autre indice acoustique qui subsiste.

Bien que certains indices acoustiques soient toujours présents, après traitement, les signaux audio étaient suffisamment dégradés et le test perceptif audiovisuel pouvait donc être mené.

B.2.1.4. Finalisation

Des films ont donc été montés[137] avec les images enregistrées pour chaque locuteur et les signaux audio obtenus après manipulation. On se souviendra que les enregistrements effectués pour l'étude de production avaient également été réalisés en parole chuchotée pour les deux locuteurs (cf. sections A.2.2.1 et A.2.3.1 du chapitre III). La figure III.1 de la section A.2.1.1.a du chapitre III donne un exemple des images enregistrées et donc des images vues par les participants au test perceptif. Trois films ont été réalisés : un avec les stimuli en audiovisuel, un avec les stimuli en visuel seul et un avec les stimuli en audio seul. Chaque film était composé de deux séquences comprenant chacune les 32 stimuli dans des ordres différents et destinées l'une à être vue de face et l'autre de profil. Pour chaque séquence de chaque film, les stimuli ont été classés dans des ordres aléatoires à chaque fois différents.

B.2.2. Paradigme expérimental

Le principe général du test était d'expliquer aux participants qu'ils allaient voir, entendre ou voir et entendre des énoncés extraits de conversations entre deux interlocuteurs. Lors de ces conversations, le premier interlocuteur prononçait d'abord une phrase (extraite du corpus). Son interlocuteur ayant mal compris un des constituants de cette phrase (S, V ou O), répétait ce qu'il avait compris sous forme interrogative. Le premier interlocuteur répétait donc la phrase initiale en corrigeant le constituant mal compris par son interlocuteur et, ce faisant, produisait une focalisation sur ce constituant. Il était dit aux participants qu'ils allaient, soit voir, soit entendre, soit voir et entendre cette correction pendant le test. Leur tâche serait alors de déterminer quel constituant avait été corrigé par le premier interlocuteur (*i.e.* focalisé). Bien entendu, le terme de *focalisation contrastive* n'a jamais été employé avec les participants, les seuls termes employés étaient *correction* ou *insistance*. Il était donc indirectement demandé aux participants de déterminer la partie de la phrase (sujet, verbe ou objet)

[137] Les films ont été montés grâce au logiciel *Velocity* de chez DPS (www.dps.com.fr).

qui était focalisée. En fonction de la phase du test dans laquelle le participant se trouvait, l'énoncé corrigé était présenté en audio-visuel, en audio seul ou en visuel seul. Les participants n'entendaient (ou ne voyait) donc en fait qu'un des deux interlocuteurs de la conversation (le locuteur 1). L'exemple (IV.3) permet de comprendre comment le test se déroulait (les lettres majuscules symbolisent la focalisation, le locuteur 1 peut être soit le locuteur A soit le locuteur B).

(IV.3) Le locuteur 1 dit (le participant ne l'entend pas et ne le voit pas mais lit la phrase) :

Romain ranima la jolie maman.

Le locuteur 2 dit (mais le participant ne l'entend ni le voit) :

Denis ranima la jolie maman ?

Le participant entend et voit, ou entend seulement, ou voit seulement le locuteur 1 dire en chuchotant :

ROMAIN ranima la jolie maman.

Pour répondre, les participants disposaient d'une grille du même type que celle des tests B et C décrits dans la section A du présent chapitre. Ils étaient avertis de la possibilité que l'énoncé à « juger » soit en fait neutre tout comme pour les tests B et C décrits dans la section A du présent chapitre.

Deux tests ont été mis au point. Chaque participant passait un des deux tests. Lors du test 1, les participants étaient d'abord testés en audiovisuel puis en audio seul et enfin en visuel seul et lors du test 2 les participants étaient d'abord testés en audio seul puis en audiovisuel et enfin en visuel seul. Le but étant d'analyser l'apport de la modalité visuelle, ce protocole permet de comparer les performances en audiovisuel avec les performances en audio seul. En effet, pour le test 1, on pourrait craindre qu'il y ait un apprentissage pendant la phase de perception audiovisuelle. Ceci pourrait alors influencer les performances en audio seul (les augmentant de manière significative) ce qui biaiserait la comparaison. C'est pour cette raison que la mesure des performances avec le test 2 était indispensable. La condition de perception en visuel seul représente un contrôle.

Un total de 32 stimuli prononcés chacun par deux locuteurs a ainsi été présenté aux participants sous 3 conditions (audio seul, visuel seul et audiovisuel). Ceci représente donc un total de 192 stimuli.

B.2.3. Les participants

Un total de 13 participants francophones de naissance a passé le test (8 hommes et 5 femmes). Ces participants étaient âgés de 19 à 57 ans et étaient originaires de diverses régions de France. Aucun d'eux n'a rapporté de troubles auditifs ou visuels connus. Une fois que chaque participant eut passé le test, il lui était demandé d'expliquer de façon précise ce qu'il/elle avait fait. Ceci a permis de vérifier que la tâche accomplie était bien la bonne. Suite à cette vérification, il s'est avéré qu'aucun des participants n'avait mal compris la tâche. Tous les résultats ont ainsi pu être conservés pour analyse.

B.2.4. Attentes a priori

On s'attend à obtenir de meilleurs résultats pour la condition audiovisuelle que pour la condition audio seul. Les résultats en audio seul devraient être moyens puisqu'il ne reste plus beaucoup d'indices

acoustiques à la focalisation. En audiovisuel, la tâche devrait être plus facile, les participants disposeront d'indices supplémentaires pour détecter la focalisation, lesquels indices peuvent aider à la perception de la focalisation comme on l'a vu lors des tests perceptifs en visuel seul (cf. section A du présent chapitre). Les résultats en audiovisuel devraient aussi être meilleurs que les résultats en visuel seul. Quant aux résultats en audio seul par rapport aux résultats en visuel seul, il n'y a pas de raison particulière d'avoir une attente a priori.

On notera de plus qu'on s'attend à trouver une différence significative en fonction du locuteur. Les résultats des tests en visuel seul suggèrent en effet que le locuteur A est plus intelligible visuellement que le locuteur B, au moins en ce qui concerne la focalisation. Les scores perceptifs obtenus en visuel seul pour le locuteur A (voir la section A.2.1.2 du présent chapitre) étaient en effet nettement plus élevés que ceux obtenus pour le locuteur B (voir la section 1.2.2.2 du présent chapitre). Les résultats en production (voir la section A.2 du chapitre III) confirment cette observation puisque les indices « visibles » sont plus marqués chez le locuteur A.

B.3. Résultats

B.3.1. Analyse générale

Le graphique de la figure IV.12 donne les pourcentages de réponses correctes (condition de focalisation correctement identifiée) pour chaque participant en fonction de la condition considérée : audiovisuel, audio seul ou visuel seul. Les moyennes sur tous les participants pour chaque condition sont données par les rectangles rouges. Comme on s'y attendait, on constate rapidement que les performances sont meilleures en audiovisuel que pour les autres conditions (AV : 74% ; A : 64,9% ; V : 61,8%). Il apparaît donc globalement que quand les indices prosodiques acoustiques manquent, les indices visuels permettent de recouvrer une partie de l'information. Quelle que soit la condition considérée, on remarquera que tous les résultats obtenus sont significativement meilleurs que le niveau de hasard[138] (AV : $t=31,478$ $p<0,001$; A : $t=13,369$ $p<0,001$; V : $t=13,374$ $p<0,001$). Ceci était tout à fait prévisible puisque, pour la condition audio seul, il subsiste quelques indices acoustiques : les indices de durée (allongement des syllabes focales et surtout du premier segment du constituant focalisé), sûrement aussi des indices d'intensité qui n'ont pas pu être « gommés » totalement et des indices formantiques liés à l'hyper-articulation. Pour la condition 'visuel seul', les participants disposent de tous les indices visibles qui se révèlent, comme on l'a vu avec les tests perceptifs en visuel seul (cf. section A du présent chapitre), assez utiles pour la perception de la focalisation.

[138] Niveau de hasard : 25% car il y a quatre réponses possibles : neutre, FS, FV ou FO.

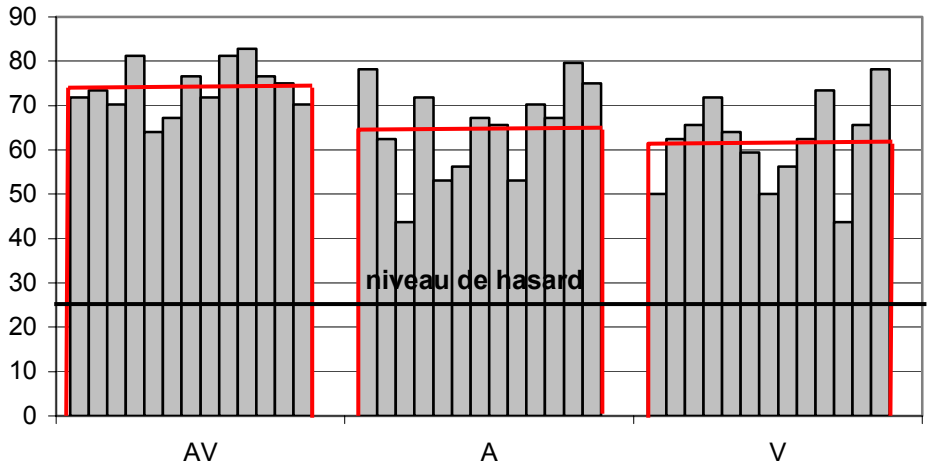


FIGURE IV.12 – *Pourcentages de réponses correctes pour chaque participant (chaque barre représente un participant) et pour chaque condition (AV : audiovisuel, A : audio seul, V : visuel seul) et moyennes sur tous les participants (rectangles rouges).*

Le graphique a. de la figure IV.13 donne les moyennes des pourcentages de réponses correctes en fonction de la condition et du locuteur. On constate que conformément aux résultats des tests perceptifs en visuel seul, les performances sont globalement meilleures pour le locuteur A. Il apparaît que la progression entre les conditions 'audio seul' et 'audiovisuel' est moins nette chez le locuteur A (+3,9% seulement chez le locuteur A contre 14,4% chez le locuteur B). On constate en effet que, pour le locuteur A, les résultats en audio seul sont déjà très bons (80,8% de réponses correctes pour le locuteur A contre 49% chez le locuteur B). En fait, les résultats en audio seul (80,8%) et en visuel seul (68,3%) du locuteur A sont mêmes largement supérieurs à ceux du locuteur B en audiovisuel (63,5% de réponses correctes en audiovisuel chez le locuteur B). On s'attendait à ce résultat étant donné les différences de marquage de la focalisation entre les deux locuteurs. Cependant, comme il a déjà été expliqué précédemment, le locuteur A étant habitué à être enregistré, a tendance à toujours articuler de façon très nette, ce qui peut sembler parfois moins naturel bien que plus intelligible. Même si les performances globales sont moins bonnes, les résultats obtenus pour le locuteur B, dont les productions paraissent plus naturelles, seront considérés comme plus conformes à celles des locuteurs en situation de parole ou de conversation naturelle.

Le graphique b. de la figure IV.13 donne les moyennes des pourcentages de réponses correctes en fonction de la condition et de la vue (face ou profil). On constate qu'il n'y a pas de différences entre les résultats correspondant à la vue de face et ceux correspondant à la vue de profil. Ce résultat est en accord avec ce qui avait été obtenu pour le test C en visuel seul (cf. section A.2.2 du présent chapitre).

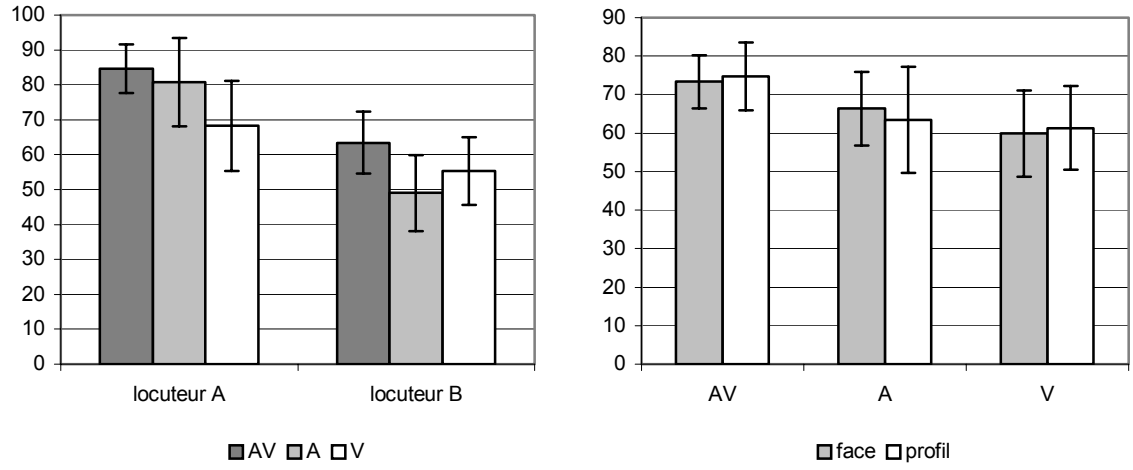


FIGURE IV.13 – *Moyennes des pourcentages de réponses correctes a.(gauche) en fonction du locuteur et de la condition (AV : audiovisuel, A : audio seul ou V : visuel seul) et b.(droite) en fonction de la vue (face ou profil) et de la condition (AV, A ou V).*

B.3.2. Analyse statistique

Une analyse de variance à quatre facteurs intra-sujets a été menée. Les quatre facteurs étaient : la modalité (trois niveaux : audiovisuel, audio seul et visuel seul), le locuteur (deux niveaux : locuteurs A et B), la vue (deux niveaux : face et profil) et le type de focalisation (quatre niveaux : neutre, FS, FV et FO). Les résultats de cette analyse sont consignés dans le tableau IV.4.

facteur	modalité	locuteur	vue	type de focalisation	modalité*locuteur
F	F(2,24)=7,232	F(1,12)=121,384	F(1,12)=0,224	F(1,716,36)=2,715	F(2,24)=11,231
signification	0,003	<0,001	0,63	0,096	<0,001

TABLE IV.4 – *Résultats de l'ANOVA à 4 facteurs menée sur les résultats du test perceptif audiovisuel. Les quatre facteurs sont : la modalité (AV, A ou V), le locuteur (A ou B), la vue (face ou profil) et le type de focalisation (neutre, FS, FV ou FO).*

On note que l'effet de la modalité est significatif (F(2,24)=7,232 p=0,003). L'analyse des contrastes de l'ANOVA permet de montrer que ce sont les résultats en audiovisuel qui sont significativement meilleurs que ceux correspondant aux autres modalités (p<0,001). Les résultats pour les conditions audio seul et visuel seul ne sont pas statistiquement différents (p=0,451). L'amélioration des performances perceptives de l'audio seul à l'audiovisuel est donc statistiquement significative.

L'effet du locuteur est lui aussi nettement significatif (F(1,12)=121,384 p<0,001). On a en effet remarqué que les résultats étaient meilleurs pour le locuteur A quelle que soit la modalité considérée et cette différence apparaît être statistiquement significative.

Les effets de la vue et du type de focalisation ne sont quant à eux pas significatifs (vue : F(1,12)=0,244 p=0,63 ; type de focalisation : F(1,716,36)=11,231 p<0,001).

B.3.3. Analyse approfondie pour chaque stimulus

Les résultats pour chaque stimulus ont été étudiés en détails en fonction de la modalité. Les pourcentages moyens de réponses correctes pour chaque stimulus et pour chaque modalité (AV, A et V) sont donnés en annexe 7. De façon générale, les stimuli peuvent être classés en trois catégories majeures :

- D'abord la catégorie **« AV≥A,V »**, pour laquelle les performances sont les meilleures en audiovisuel. 46,9% des stimuli correspondent à ce cas de figure. Dans ce cas de figure, l'audio a été « bien » dégradé et on peut penser qu'il n'y a plus beaucoup d'indices acoustiques à la focalisation contrastive sauf, bien sûr, les indices de durées. La perception auditive n'est donc pas très bonne bien qu'elle soit tout de même supérieure au hasard. Les indices « visibles » peuvent être plus ou moins marqués (et la perception en visuel seul sera ainsi parfois meilleure parfois moins bonne que celle en audio seul (V<A ou V>A)) mais ces indices existent et en combinant les deux modalités on obtient de bonnes performances.
- Puis la catégorie **« A≥AV>V »**, pour laquelle ce sont les performances en audio seul qui sont les meilleures, suivies de celles en audiovisuel, puis de celles en visuel seul. 28,1% des stimuli correspondent à ce cas de figure. Dans ce cas, il semble assez clair que la modalité visuelle dégrade la perception plutôt que de l'améliorer comme on pouvait s'y attendre. Deux explications sont possibles, elles correspondent à deux cas de figure. Dans le premier cas, il y a peu d'indices visuels et donc les performances en audio seul et en audiovisuel seul sont sensiblement les mêmes (A~AV) : il n'y a pas d'indices visuels et donc quand on ajoute la modalité visuelle, les performances ne changent pas. C'est le cas pour la moitié des stimuli de cette catégorie. Dans le second cas, les performances en audio seul sont meilleures qu'en audiovisuel (A>AV). Dans ce cas, les indices visuels ne sont pas seulement absents mais mènent les participants sur une « mauvaise piste ». Par exemple, une anticipation visible pourrait induire les participants en erreur. C'est le cas pour l'autre moitié des stimuli de cette catégorie.
- Et enfin la catégorie **« V>AV,A »**, pour laquelle ce sont les performances en visuel seul qui sont les meilleures et qui comprend 25% des stimuli. Dans ce cas de figure, les indices « visibles » sont présents puisque la perception en visuel seul est bonne mais il semblerait que l'audio viennent perturber la perception de la focalisation. Si les performances en audio seul sont moins bonnes que celles en audiovisuel (AV>A), c'est qu'il est possible que l'audio ait été mal remanié et qu'il induise le participant en erreur d'où des performances moins bonnes en audiovisuel qu'en visuel seul, c'est le cas pour la moitié des stimuli. Si AV=A, l'audio doit être tellement modifié (*i.e.* lors du traitement préalable pour neutraliser l'indice d'intensité, il est possible que l'on ait rendu un autre constituant saillant) que les indices visibles ne peuvent même pas remettre les participants sur la bonne voie, c'est le cas pour 25% des stimuli de cette catégorie. Seuls deux stimuli correspondent à AV<A. Ces stimuli correspondent à des cas isolés.

B.4. Conclusion

Il apparaît donc que les indices visuels peuvent aider à la perception de la focalisation contrastive prosodique lorsque les indices acoustiques sont dégradés. Les performances perceptives de détection et de localisation de la focalisation sont en effet meilleures en audiovisuel qu'en audio seul lorsque les indices de F0 et d'intensité sont largement dégradés. Comme on s'y attendait, les scores de perception sont significativement supérieurs pour le locuteur A. Cependant, comme il a été suggéré, ces résultats sont certainement biaisés par l'habitude du locuteur A à participer à ce type d'enregistrement. L'avantage audiovisuel est ainsi nettement plus fort chez le locuteur B.

– Discussion & Conclusion –

Si vous avez parcouru le long chemin qui vous amène maintenant à lire ceci, vous aurez, je l'espère, principalement compris deux choses : la focalisation prosodique est « visible » et elle est « vue ». Nous allons d'abord faire le point sur les apports méthodologiques de ce travail. Puis nous exposerons les fondements des modèles que les résultats nous ont permis de construire. Nous tenterons d'analyser ce que ce travail peut apporter aux problématiques plus générales de la nature du contrôle de l'hyper-articulation et de la « visibilité » de la prosodie. Nous décrirons les applications possibles et enfin les perspectives de recherche.

A. Apports méthodologiques

A.1. Discussion sur l'utilisation de la parole délexicalisée

Le lecteur aura sans doute noté l'utilisation par deux fois de la parole délexicalisée comme outil pour des études préliminaires, à la fois en production et en perception (cf. sections A.2.2.3 du chapitre III et A.1 du chapitre IV). Or, comme il a déjà été avancé dans la section A.2.2.2.a du chapitre III, l'emploi de ce type de parole est relativement souvent contesté. L'étude des indices visibles de la focalisation contrastive en français était cependant pionnière. Les résultats qui allaient être obtenus étaient au début peu prévisibles et les hypothèses formulées ne résultaient que de la généralisation de ce qui avait pu être observé dans d'autres langues. C'est pourquoi l'utilisation de la parole délexicalisée s'est révélée être très utile. Ne sachant pas précisément à quoi s'attendre et ce surtout pour les mesures articulatoires, il était en effet difficile de commencer par l'étude directe de la parole naturelle lexicalisée avec toutes les variabilités inhérentes qu'elle comporte. Il se serait en effet avéré relativement complexe de parvenir à isoler les effets articulatoires de la focalisation sans avoir une idée préalable de ce qu'il fallait mesurer. La parole délexicalisée est un paradigme puissant puisqu'elle permet de limiter les variabilités articulatoires inter-syllabiques et d'isoler les effets articulatoires supra-segmentaux. Il s'est ainsi révélé possible d'isoler l'effet articulatoire visible de la focalisation contrastive. En perception, l'utilisation de la parole délexicalisée a permis de mener une étude préliminaire en limitant la charge cognitive mise en œuvre pour une tâche somme toute complexe et dont il était difficile d'évaluer à l'avance la difficulté.

On soulignera que, bien que des réserves avaient été émises quant aux résultats obtenus en utilisant la parole délexicalisée, il s'est avéré aussi bien en production qu'en perception que les études suivantes, menées sur la parole lexicalisée, ont confirmé ce qui avait été observé dans les études préliminaires menées avec la parole délexicalisée. On pourra ainsi remarquer qu'au moins en ce qui concerne la focalisation contrastive, la parole délexicalisée a permis de représenter assez efficacement le phénomène et d'obtenir des résultats très pertinents par rapport à la parole lexicalisée qui serait étudiée ensuite.

Les études menées et décrites dans le présent mémoire permettent ainsi de penser que malgré les contestations fortes quant à l'utilisation de la parole délexicalisée dans l'étude de certains phénomènes suprasegmentaux de la parole, il semble que celle-ci soit réellement efficace. Nous recommandons ainsi d'utiliser le paradigme de la parole délexicalisée pour commencer une étude et

obtenir des résultats préliminaires qui seront une base solide pour l'étude ultérieure de la parole lexicalisée.

A.2. Discussion sur le champ minimal requis pour les études des effets prosodiques

Comme il sera rappelé dans la synthèse sur le modèle de la production visuelle de la focalisation, nous avons observé des effets de la focalisation sur l'énoncé dans son intégralité. Quelques autres chercheurs avaient fait la même observation. Pourtant, la plupart des études portant sur l'analyse des effets de la focalisation se limite à l'analyse du constituant focalisé. Notre observation d'un effet global nous incite à souligner l'importance d'étudier les effets prosodiques sur l'énoncé en entier. Nous pensons ainsi que le champ minimal requis pour l'étude des effets prosodiques est l'énoncé. Un champ plus réduit ne permet en effet pas de mettre en évidence tous les aspects des stratégies mises en place par les locuteurs.

B. Modèles cognitifs élaborés

B.1. Un modèle cognitif de la production audiovisuelle de la focalisation contrastive prosodique en français

Les diverses études décrites dans ce mémoire permettent d'établir un modèle cognitif de la production audiovisuelle de la focalisation contrastive. A la lumière de la littérature et de l'étude complémentaire que nous avons menée pour le locuteur A et qui est décrite au chapitre II, nous rappellerons ici les principales manifestations auditives de la focalisation contrastive prosodique en français.

Lorsqu'un constituant est focalisé, sa fréquence fondamentale et son intensité augmentent alors que celles des éléments voisins (pré- et post-focaux) diminuent par rapport au cas neutre ce qui crée un fort contraste au sein de l'énoncé. La focalisation est essentiellement marquée par un ton de focalisation Hf qui remplace le plus souvent le ton initial Hi ou à la fois les tons initial Hi et final H*. On constate de plus un allongement des syllabes focales, le premier phonème du constituant focalisé étant allongé dans une plus large mesure. La syllabe directement pré-focale est également allongée chez le locuteur A. Enfin, on note une réorganisation prosodique de la séquence pré-focale lorsque la focalisation porte sur l'objet et une désaccentuation de la séquence post-focale, les indices de durées véhiculent néanmoins encore l'information de phrasé prosodique.

L'étude des indices articulatoires fournis par la bouche et la mandibule a permis de montrer que ces indices étaient affectés « visiblement » par la focalisation. La focalisation contrastive possède donc des corrélats qui peuvent être visibles par l'interlocuteur lors des processus de perception de la parole. L'analyse de données recueillies avec plusieurs systèmes de mesure et pour plusieurs locuteurs a permis de mettre en évidence deux stratégies décrivant les comportements des locuteurs :

- **Stratégie articulatoire de signalisation absolue :** le constituant focalisé est allongé et hyper-articulé dans une large mesure, par rapport au cas neutre : les amplitudes de tous les paramètres articulatoires (aire intéro-labiale, protrusion, position de la mandibule) sont amplifiées et les pics de vitesse sont plus importants reflétant une augmentation de la force articulatoire du geste sous-jacent. Cette amplification se fait à la fois par rapport au cas neutre (contraste inter-énoncés) et par rapport au reste de l'énoncé (contraste intra-énoncé). Les locuteurs adoptant cette stratégie concentrent donc leurs efforts sur l'hyper-articulation du constituant focalisé. On note aussi parfois une anticipation de la focalisation c'est-à-dire que l'amplitude des gestes articulatoires, leur vitesse et la durée syllabique commencent à augmenter sur la syllabe précédant directement l'élément focalisé.
- **Stratégie articulatoire de signalisation différentielle :** dans ce cas, le constituant focalisé est aussi allongé et hyper-articulé. Cette hyper-articulation est parfois anticipée. La différence par rapport à la stratégie précédente est que la séquence post-focale est hypo-articulée par rapport au cas neutre, c'est-à-dire que l'amplitude des gestes articulatoires y est significativement moins importante. Un contraste important est ainsi créé au sein de l'énoncé focalisé (contraste intra-énoncé). L'hyper-articulation focale n'est pas nécessairement très importante mais elle est renforcée par l'hypo-articulation post-focale.

On note l'existence de fortes variabilités inter-locuteurs même dans le cadre d'une même stratégie et ce essentiellement dans le degré d'hyper- ou hypo-articulation ou même dans la nature des paramètres utilisés et surtout leur importance relative au niveau du marquage global. L'observation la plus constante est l'importance du marquage du paramètre de protrusion qui, rappelons-le, correspond également au paramètre le plus visible (Benoît *et al.* [1994]). La stratégie de signalisation différentielle est plus fréquemment utilisée par les locuteurs que la stratégie de signalisation absolue. On notera, surtout dans le cadre de la stratégie de signalisation différentielle, que la focalisation affecte l'énoncé dans son intégralité au niveau articulatoire.

L'étude préliminaire d'autres gestes faciaux (mouvements des sourcils et de la tête) a montré que les liens entre ceux-ci et la focalisation existent mais ne sont pas systématiques. Les locuteurs adoptent des stratégies très différentes pour ces gestes faciaux, la variabilité intra-locuteur est aussi très forte.

B.2. Modèle cognitif de la perception audiovisuelle de la focalisation contrastive prosodique en français

Les tests perceptifs décrits dans le chapitre IV ont permis de constater que les indices visuels de la focalisation étaient perçus et qu'à eux seuls, ils permettaient souvent de remonter à l'information de focalisation contrastive prosodique (cf. tests en visuel seul). Les indices visuels utilisés lors de la perception correspondent au moins en partie à ceux mis en évidence lors d'études en production. Trois tendances ont été relevées :

- **Tendance à la sur-perception du cas neutre** : les erreurs les plus fréquemment commises en perception visuelle correspondent à une non-détection de la focalisation (*i.e.* réponse « neutre » alors qu'il y en fait eu focalisation) ;

- **Tendance à la sous-perception en fin de phrase** : la détection visuelle de la focalisation est moins bonne en fin de phrase (*i.e.* sous-perception de la focalisation sur l'objet) ;
- **Tendance à la confusion vers le voisin pré-focal** : les confusions entre types de focalisation s'orientent le plus souvent vers le voisin pré-focal plutôt que vers le voisin post-focal ou encore vers un constituant plus éloigné de l'énoncé.

Il a également été montré que les indices visuels étaient utilisés pour la perception audiovisuelle de la parole particulièrement lorsque l'information acoustique de focalisation est dégradée (test de perception audiovisuelle).

On notera que dans le cadre de la perception audiovisuelle de la focalisation, il est préférable que tous les indices « visibles » soient présents même s'ils ne sont que moyennement marqués, plutôt qu'un seul soit présent même très fortement marqué. Ce phénomène reflète la cohérence inter-articulateurs à laquelle les participants percevant de la parole s'attendent.

En situation de communication, il est tout de même très fréquent que l'on puisse voir son interlocuteur et il est tout aussi fréquent que, pour diverses raisons, on n'entende pas très bien ce qu'il nous dit et que l'on soit obligé de se baser sur des indices non acoustiques pour bien comprendre le message qu'il tente de nous faire passer. Or, comme il avait déjà été suggéré en introduction, la focalisation joue un rôle important dans la communication parlée. Il est donc tout à fait intéressant de savoir que même si elle est mal perçue acoustiquement, l'information pourra éventuellement être extraite visuellement.

C. Discussion sur la nature du contrôle articulatoire de la focalisation

C.1. Effets globaux pour un phénomène localisé : quelles implications ?

Rappelons qu'au cours des analyses décrites dans ce mémoire, nous avons plusieurs fois souligné que la focalisation localisée (sur un constituant) affectait non seulement le constituant concerné mais l'énoncé dans son intégralité. Comme nous l'avons déjà évoqué, d'autres chercheurs avaient également observé ce phénomène. On peut se poser la question de savoir pourquoi une telle stratégie est mise en place et ce qu'elle implique sur le plan méthodologique.

Nous proposons deux explications possibles au fait que les effets de la focalisation portent sur tout l'énoncé. La première consiste à dire qu'une certaine quantité « d'énergie » ou « d'effort » est disponible pour produire un énoncé. Or quand l'un des constituants de cet énoncé doit être focalisé, le locuteur sait qu'il va devoir produire une hyper-articulation qui va nécessiter plus « d'énergie ». N'en disposant que d'une certaine quantité, il va devoir la répartir et l'énergie qu'il devra fournir en plus pour hyper-articuler devra être prise ailleurs. C'est pourquoi certaines parties de l'énoncé seraient hypo-articulées. Il s'agirait donc de la part du locuteur d'une stratégie de *répartition de l'effort*.

La seconde consiste à dire que le contrôle du contraste se fait en ligne. Pour produire un énoncé focalisé, le locuteur commencerait sur un mode neutre. Puis, lorsqu'il atteindrait le constituant à focaliser, il commencerait à hyper-articuler pour signaler le contraste. Après ce constituant, le locuteur renforcerait encore plus le contraste en hypo-articulant. Dans ce cas le locuteur utiliserait plutôt une stratégie de *contrôle d'un contraste*.

Plusieurs arguments nous permettent de favoriser la seconde stratégie. En effet, dans le cas où l'objet est focalisé, on observe bien une hyper-articulation focale mais aucune hypo-articulation. Celle-ci n'est en effet observée que sur les constituants post-focaux. Or si le locuteur devait répartir son effort, il devrait hypo-articuler au moins le constituant pré-focal en prévision de la dépense d'énergie nécessaire pour l'hyper-articulation de l'objet. D'autre part, dans le cas où le sujet est focalisé, on remarque que l'intégralité de la séquence post-focale est hypo-articulée alors que l'hyper-articulation n'est pas plus forte que pour les autres types de focalisation. Or, en hypo-articulant toute la séquence post-focale, le locuteur devrait disposer de plus d'énergie pour hyper-articuler le sujet. Il apparaît donc plus probable que le fait que l'énoncé soit affecté dans son intégralité soit lié à la mise en place d'une stratégie de traitement en ligne.

C.2. Hyper-articulation : conséquence physiologique ou désir d'intelligibilité ?

La synthèse des travaux de la littérature (cf. section D du chapitre I) avait permis de souligner le fait que toutes les études ayant porté sur l'analyse des effets articulatoires de la focalisation envisageaient ceux-ci comme étant des conséquences, ou au mieux, des corrélats des phénomènes acoustiques. Or dans ce mémoire, ces mêmes effets ont été considérés du point de vue de la multimodalité. Ils ont en effet été analysés comme des corrélats visuels de la focalisation. Bien que des études perceptives aient en partie corroboré cette approche, puisqu'il est apparu que les spectateurs étaient sensibles à ces effets articulatoires, les analyses ne permettent pas de conclure de façon nette quant aux processus sous-jacents à ces effets articulatoires.

L'hyper-articulation permet en effet de rendre la focalisation plus intelligible visuellement. Cependant il serait intéressant de parvenir à déterminer si elle est une conséquence physiologique qui finalement se révèle être utile pour la perception audiovisuelle de la parole ou si elle est planifiée dans un désir d'augmenter l'intelligibilité visuelle sur une partie importante de l'énoncé. Dans le premier cas, ce serait en essayant d'augmenter l'intelligibilité acoustique (en augmentant F0 et surtout l'intensité) que l'hyper-articulation serait produite, *i.e.* le locuteur parlerait plus fort pour faire ressortir le constituant focalisé et donc ouvrirait plus grand la bouche et la mâchoire. Dans le second cas, la planification de la focalisation tiendrait compte d'une part de la mise en relief acoustique mais aussi de la mise en relief visuelle *i.e.* le locuteur voudrait faire ressortir le constituant focalisé à la fois acoustiquement et visuellement.

Il est assez compliqué de répondre à cette question de façon tranchée puisque la séparation entre ce qui est planifié et contrôlé, et ce qui est la conséquence d'une contrainte est difficile à effectuer. L'étude de Erickson *et al.* 1998 montre que la perception auditive de ce que les auteurs appellent *« emphasis »* (qui correspond à ce qui est ici appelé *focalisation*), est meilleure lorsque le locuteur a plus ouvert la mâchoire. Or la tâche mise en place simulait une conversation téléphonique et on peut donc penser qu'il n'y avait pas chez les locuteurs de désir d'accroître leur intelligibilité visuelle. Cette

étude pourrait donc plutôt laisser à penser que l'hyper-articulation est une conséquence de l'augmentation de l'intelligibilité auditive. Néanmoins aucune conclusion ne peut être tirée de façon claire puisque cette étude ne comporte hélas pas d'analyse acoustique des stimuli. En outre, les conversations téléphoniques audio ne sont pas exemptes de gestes manuels ni de mimiques faciales bien que le locuteur soit conscient du fait que son interlocuteur ne le voit pas.

Par contre une autre étude (Erickson *et al.* [2000]) compare des données acoustiques et articulatoires et amène les auteurs à conclure que, chez certains locuteurs, on observe une forte corrélation entre F0 et ouverture de la mâchoire alors que chez d'autres cette corrélation est très faible. Cette observation tendrait à montrer que l'acoustique et l'articulatoire peuvent être contrôlés indépendamment l'un de l'autre. Cette étude montre de plus que l'emphase est systématiquement liée à une augmentation de l'ouverture de la mâchoire quelle que soit la nature de la voyelle accentuée (*i.e.* fermée [i] ou ouverte comme [æ]). On pourrait ainsi penser que l'augmentation de F0 ou de l'intensité induit systématiquement une augmentation de l'ouverture mandibulaire. Cependant, les voyelles utilisées dans cette étude n'impliquaient pas la protrusion labiale qui, elle, pourrait induire une ouverture mandibulaire plus faible sous focalisation. Cette étude ne permet donc pas non plus de conclure de façon tranchée.

Il semblerait pourtant qu'il soit possible de contrôler indépendamment l'intelligibilité acoustique et l'intelligibilité visuelle (Beautemps *et al.* [1999]). D'ailleurs, on se rendra compte par simple expérience personnelle (oui oui, ici et maintenant si on veut), qu'il est d'une part possible d'OUVRIR GRAND LA BOUCHE en parlant très peu fort ou qu'au contraire il nous est possible de PARLER FORT en essayant d'ouvrir le moins possible la bouche. Cette remarque générale pourrait permettre de penser que l'hyper-articulation focale provient d'un désir d'intelligibilité visuelle accrue. Néanmoins aucune étude précise ne permet de conclure de façon certaine et il serait intéressant de mener une expérience de production de la focalisation au cours de laquelle aucun son ne serait produit et d'analyser le comportement articulatoire obtenu. Bien que cette étude n'ait pas (encore) été menée, il est possible déjà de trouver quelques pistes complémentaires dans les données présentées dans ce mémoire.

C.2.1. Ce que nous apporte un dernier regard sur les données …

On se souviendra d'une constatation qui avait été faite à l'issue de l'analyse des données en production *i.e.* que la protrusion était nettement plus hyper-articulée que tous les autres paramètres articulatoires et ce chez la grande majorité des locuteurs. Si l'hyper-articulation n'était que la conséquence du fait de parler plus fort donc d'atteindre une pression acoustique plus élevée, ce devrait être l'aire intéro-labiale et l'ouverture de la mâchoire qui augmenteraient le plus pour augmenter la pression en sortie de la bouche. La protrusion devrait quant à elle ne pas changer ou alors très peu. Et pourtant c'est l'inverse qui est observé.

On pourra de plus noter que l'étude de Benoît *et al.* [1994] suggèrent que les voyelles [y] et [u] en français sont beaucoup plus intelligibles visuellement que les voyelles ouvertes comme le [a]. Or nos données montrent que c'est justement les voyelles fermées et protruses qui sont les mieux marquées puisque c'est la protrusion qui augmente le plus avec la focalisation. Ce sont donc justement les voyelles les plus « visibles » qui sont hyper-articulées et on peut ainsi penser que c'est pour les rendre encore plus « visibles » qu'elles le sont. Si elles sont plus intelligibles visuellement c'est en effet elles qui seront les plus efficaces pour véhiculer la focalisation.

Il apparaît donc que c'est le paramètre le plus visible (la protrusion) qui est aussi le plus hyper-articulé et non pas le paramètre articulatoire le plus lié à l'intensité (*i.e.* l'aire intéro-labiale ou l'ouverture de la mandibule). A la lumière de ces observations, on pourra dire qu'il est assez censé de penser que l'hyper-articulation focale relève effectivement d'un désir d'intelligibilité visuelle accrue et pas seulement d'une conséquence du fait de parler plus fort, ce qui renforce l'aspect bimodal de la focalisation prosodique.

On se souviendra de plus de la comparaison entre les stratégies acoustiques et articulatoires de la signalisation de la focalisation prosodique chez le locuteur A (cf. section A.2.2.3.f.iv du chapitre III). Celle-ci avait montré qu'il semblerait que le contrôle lié à la focalisation soit différent pour les articulateurs laryngés et supra-laryngés. En outre, les locuteurs A et B ont la même stratégie acoustique alors que leurs stratégies articulatoires diffèrent. Ces observations vont dans le sens d'un contrôle indépendant et permettent de penser qu'il serait possible que les articulateurs supra-laryngés soient contrôlés dans le cadre de la multisensorialité en tenant compte à la fois de perspectives auditives et visuelles.

Un argument supplémentaire peut être formulé à l'aide des données perceptives. Bien que les divers tests perceptifs n'aient pas été tous menés avec les mêmes phrases ou dans les mêmes conditions, les différentes expériences perceptives menées permettent de comparer qualitativement les performances perceptives visuelles correspondant aux productions en parole « normale » vs « chuchotée ». La consigne donnée aux locuteurs pour produire la parole chuchotée était en effet d'être compris par un interlocuteur placé devant eux. Leur tâche n'était donc pas de chuchoter à l'oreille mais bien de se faire comprendre à distance sans être entendu. La seule façon pour eux d'augmenter leur intelligibilité était donc de fournir à leur interlocuteur plus d'indices visuels puisqu'ils ne devaient pas produire trop d'indices acoustiques. Les tests perceptifs B et C avaient montré que les productions en parole normale étaient perçues correctement dans 71,4% des cas pour le locuteur A et dans 43,2% des cas pour le locuteur B. Le test perceptif en audiovisuel avait montré que les productions en parole chuchotée étaient bien perçues visuellement dans 68,3% des cas pour le locuteur A et dans 55,3% des cas pour le locuteur B. Pour le locuteur A, on peut donc penser que la différence perceptive visuelle entre la parole normale et la parole chuchotée est faible. Rappelons que chez ce locuteur l'hyper-articulation en parole normale était déjà très importante, il est donc possible qu'il y ait un effet plafond : les performances auraient déjà atteint leur seuil maximum en parole normale. Par contre, chez le locuteur B, on constate une nette progression des performances perceptives en visuel seul entre la parole normale et la parole chuchotée. Il apparaît ainsi que lorsque ce locuteur doit devenir plus intelligible visuellement, il est capable de contrôler le degré de saillance des indices visibles produits. Pourtant dans le cas de la parole chuchotée, l'hyper-articulation ne peut pas être une simple conséquence de la production de meilleurs indices acoustiques.

C.2.2. Ce que nous suggère une expérience en neurolinguistique

L'étude de Lœvenbruck *et al.* (2005) a permis de mettre en évidence la spécificité de la focalisation prosodique par rapport à la focalisation syntaxique au niveau des activations cérébrales. Quatre conditions ont été étudiées, consistant en la production silencieuse des phrases isosyllabiques suivantes :

Condition de contrôle : “Madeleine m’amena”.

Focalisation prosodique : “MADELEINE m’amena”.

Focalisation syntaxique : “C’est Mad’leine qui m’am’na”.

Focalisation combinée syntaxique et prosodique : “C’est MAD’LEINE qui m’am’na”.

Un schéma d’activation commun aux deux types de focalisation a été mis en évidence (gyrus frontal inférieur gauche, insula gauche et cortex prémoteur bilatéral ; cf. figure C.1). Néanmoins, il semble que la focalisation prosodique requiert en plus l’intervention du gyrus cingulaire inférieur gauche, de l’aire de Wernicke et surtout du gyrus supramarginal gauche. Or le gyrus supramarginal gauche (SMG) est impliqué dans la proprioception. Lors de l’imitation d’une action, il participe au réseau temporo-pariéto-frontal (Iacoboni [2005]) en élaborant, à partir de la description visuelle de l’action à imiter fournie par le cortex temporal, une information somatosensorielle sur l’action à imiter, qui sera envoyée au cortex frontal inférieur. En ce qui concerne la production de la prosodie, Lœvenbruck *et al.* ont émis l’hypothèse que ce qu’apporte l’activation du SMG c’est la sensation proprioceptive de la façon dont le larynx et les articulateurs doivent être positionnés pour produire la focalisation prosodique - qui correspond à une extension des mouvements laryngés et supra-laryngés (variations de F0 et hyper-articulation évoqués plus haut). La focalisation prosodique peut en fait être considérée comme le pointage vers un constituant avec la voix exactement comme on montrerait un objet dans l’espace avec notre doigt. Il a d’ailleurs été montré que le pointage avec le doigt impliquait aussi le lobe pariétal (le plus souvent de l’hémisphère gauche, cf. Astafiev *et al.* [2003]). L’activation du lobe pariétal pourrait donc signifier que lorsque l’on focalise avec la prosodie, on montre le mot dans la phrase avec le larynx (voix) mais aussi avec les articulateurs *i.e.* que la focalisation est véhiculée acoustiquement mais aussi visuellement parce qu’on a conscience de l’importance de la vision dans la communication. Il y aurait ainsi une planification de l’hyper-articulation dans le but de véhiculer une information réellement multisensorielle, à la fois auditive et visuelle, voire proprioceptive.

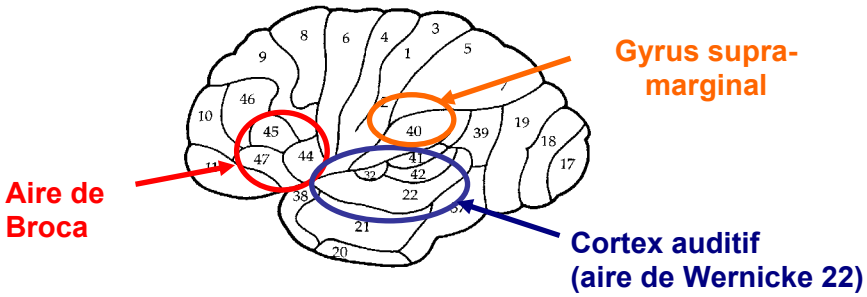


FIGURE C.1 – *Réseau des aires principales impliquées dans la production de la focalisation contrastive prosodique.*

D. Discussion sur la possibilité de généraliser les résultats à la prosodie de façon générale

Le travail décrit dans ce mémoire permet de penser que la focalisation prosodique a des corrélats visibles et que ceux-ci peuvent être utilisés lors de la perception audiovisuelle de la parole. La question qui sera abordée maintenant est de déterminer si ces résultats pourraient être étendus à la prosodie en général. Il a en effet été spécifié très clairement dans l'introduction que la focalisation prosodique était un phénomène tout à fait particulier en ce sens qu'il se rapporte également de façon étroite à la deixis. Le pointage avec la voix, bien qu'il relève sans aucun doute de la prosodie, ne reflète pas entièrement le domaine de l'intonation. Il sera ainsi délicat de conclure à partir de cette étude sur la focalisation prosodique, que de façon générale, la prosodie peut se voir et se voit. Comme il a été évoqué en introduction, il est en effet possible que la focalisation prosodique soit visible parce qu'elle appartient de par sa nature au domaine spatial et non seulement acoustique. Peut-être montre-t-on autant spatialement qu'acoustiquement ?

Bien que nous ne puissions bien entendu pas répondre à cette problématique sans conduire de nouvelles expériences, nous citerons quelques études qui montrent que la focalisation prosodique est « plus » visible que d'autres phénomènes intonatifs. A ce propos, rappelons l'étude de Bernstein *et al.* [1989] décrite au chapitre I. En plus de la discrimination visuelle de focalisation vs non focalisation en anglais, les auteurs avaient aussi mené une étude perceptive en visuel seul sur la discrimination interrogation vs affirmation. Il est apparu que pour les deux tâches les performances étaient supérieures au niveau du hasard mais la différence était nettement plus marquée pour la tâche de localisation de la focalisation (discrimination interrogation/affirmation : 60% de réponses correctes pour un niveau de hasard de 50% ; localisation de la focalisation : 76% de réponses correctes pour un niveau de hasard de 33,3%). Grant *et al.* [1986] avaient également mené une étude perceptive sur la discrimination visuelle interrogation vs affirmation en anglais dans le cadre d'une étude sur les apports des aides tactiles pour la perception de l'intonation chez les sourds et malentendants. Il est apparu que les participants parvenaient à effectuer ce type de discrimination visuellement seulement dans 50% des cas, c'est-à-dire au niveau du hasard.

Ces études suggèrent que la focalisation se verrait « mieux » que d'autres phénomènes pragmatiques signalés par la prosodie. Toutefois nous n'excluons pas que certains événements prosodiques, autres que la focalisation, soient visibles. Il a en effet été constaté par exemple que l'accent primaire en français (LH*, cf. chapitre II) était porteur d'une hyper-articulation par rapport à l'accent secondaire (Lœvenbruck [1999]). Il a d'ailleurs été noté qu'en audio, les auditeurs francophones parviennent probablement à distinguer ces deux types d'accent puisqu'ils sont capables de découper correctement un énoncé en SA, que l'énoncé soit lexicalisé ou non (Rolland & Lœvenbruck [2002]). Il a également été observé que la présence d'un coude de F0 qui fournit un indice du début d'un mot lexical en français (cf. chapitre II) est synchrone avec un pic de vitesse articulatoire (D'Imperio *et al.* [à paraître]). Bien qu'aucun test perceptif ne vienne pour l'instant confirmer que ces indices soient utilisés visuellement, il est important de noter leur existence.

E. Quelles applications directes ?

Comme il a été souligné en introduction, la focalisation contrastive joue un rôle important dans la communication parlée. Or, il est apparu que les auditeurs/spectateurs pouvaient utiliser des indices visuels pour extraire l'information de focalisation. On peut aller plus loin en disant qu'il est possible qu'ils s'attendent à trouver ces indices et que s'ils ne sont pas présents, la perception s'en trouvera perturbée.

On pourra donc penser qu'il est primordial d'inclure ces indices visuels en synthèse audiovisuelle de la parole. Ceci permettra en effet certainement de la rendre plus intelligible et de faciliter la communication homme-machine. D'autre part, l'utilisation d'indices visuels permettra de renforcer l'intelligibilité lors de l'utilisation d'un système de synthèse en environnement bruyant c'est-à-dire la condition la plus fréquente d'utilisation. L'implémentation de ces indices pourrait d'ailleurs être envisagée sur la « tête parlante » de l'ICP (Bailly *et al.* [2003]). Outre l'application directe pour le grand public, les indices visuels pourraient être renforcés dans les systèmes de synthèse audiovisuelle pour permettre aux malentendants de détecter plus facilement l'information de focalisation.

On notera de plus qu'une bonne connaissance de ces indices et de leurs mécanismes de réalisation, mais aussi la compréhension de la façon dont ils interviennent dans la compréhension et le traitement de l'information de focalisation, devrait permettre d'améliorer les systèmes de reconnaissance de la parole. Ces systèmes bénéficieraient en effet d'une meilleure détection de la focalisation. On focalise souvent pour insister sur l'information importante, or, dans le cadre de la communication homme-machine, cette information est primordiale pour que la machine « comprenne » ce que l'utilisateur (humain) attend. Les indices visuels pourraient ainsi permettre aux systèmes de reconnaissance audiovisuelle d'identifier avec plus de précision l'information importante surtout en milieu bruyant comme c'est le cas pour la plupart des applications envisagées. La focalisation est aussi souvent employée lorsqu'on effectue une correction ce qui arrive fréquemment lors de la communication homme/machine. Il serait donc intéressant que les systèmes de reconnaissance audiovisuelle puissent de façon plus précise identifier le lieu de correction.

La technologie de la visiophonie peut également bénéficier des résultats obtenus. En effet, pour la transmission efficace et rapide du signal vidéo, il est nécessaire de compresser de façon importante les images ce qui induit une qualité médiocre à la réception. Or en sachant que certaines zones du visage véhiculent de l'information visuellement, il serait intéressant de favoriser une bonne transmission du signal vidéo pour ces zones précises. On pourrait ainsi utiliser des techniques de compression différentielle afin de privilégier les zones porteuses d'information.

On constate donc que le travail décrit dans ce mémoire peut avoir des applications au niveau des technologies de synthèse et de reconnaissance audiovisuelle de la parole mais aussi pour la visiophonie. Cependant la technologie n'est pas le seul champ d'application possible.

Une meilleure connaissance des indices prosodiques visuels et donc des mouvements articulatoires concomitants pourrait permettre d'optimiser l'enseignement des langues secondes et surtout de leur prononciation et de leur intonation. Un des aspects les plus difficiles à maîtriser lors de l'apprentissage d'une langue étrangère est en effet sa prosodie or celle-ci est importante pour la communication puisque sa mauvaise utilisation est source d'incompréhension (Beckman [1996]). Si l'on comprend mieux comment l'articulation est liée à la prosodie, il sera possible de décrire aux

apprenants les positions et gestes corrects pour la production de celle-ci et ainsi de rendre plus efficace l'enseignement des langues étrangères.

L'orthophonie et la phoniatrie bénéficieraient également d'avancées dans la compréhension des liens entre gestes articulatoires et laryngés afin d'optimiser la rééducation de patients ayant des troubles de la prosodie (aprosodiques et patients parkinsoniens par exemple).

Les applications de ce travail sont donc nombreuses et variées.

F. Et après ?

F.1. Études inter-linguistiques

Nous rappellerons au lecteur que les études décrites dans ce mémoire concernaient le français. Or il a été suggéré que l'intégration de l'information visuelle dans le processus de traitement de la parole perçue puisse varier d'une langue à l'autre**.** En effet, une étude (Sekiyama & Tohkura [1993]) a montré qu'il semblerait que les locuteurs japonais soient moins sensibles à l'information visuelle lorsqu'ils perçoivent la parole. Il apparaît qu'ils sont en effet moins sujets à « l'effet McGurk ». Des hypothèses socioculturelles ont été émises pour expliquer le phénomène (Burnham & Keane [1997]). Une étude récente (Sekiyama & Burnham [2004]) a également permis de montrer que l'intégration de l'information visuelle dans le traitement de la parole se développait avec l'âge pour les anglophones (et donc certainement aussi pour les francophones qui ont une sensibilité au McGurk semblable à celle des anglais) mais moins pour les Japonais.

S'il existe des différences d'intégration de la modalité visuelle entre le français et le japonais, il est possible qu'il y ait une différence de production des indices visuels. On peut supposer que les Japonais hyper-articulent dans une moindre mesure (et peut-être même pas du tout) lorsqu'ils réalisent une focalisation contrastive. Si tel est le cas la focalisation contrastive en japonais sera moins visible et la perception de la focalisation prosodique en japonais essentiellement auditive. On pourra ainsi se demander si la perception visuelle de la focalisation contrastive prosodique est moindre voire absente en japonais.

Ces observations et questionnements mettent en évidence la nécessité de répliquer les études présentées dans ce mémoire pour d'autres langues que le français.

F.2. Vers l'exploration neuronale de la nature de l'information visuelle produite et son mécanisme d'intervention dans la perception

On a à ce jour relativement peu d'informations sur le traitement de la prosodie dans le cerveau. On sait que les capacités langagières (syntaxe, grammaire, sémantique) sont principalement latéralisées à gauche (cf., pour des revues de question, Dronkers *et al.* [2000] et Indefrey & Levelt [2004]), les lobes frontal (aires de Broca : cf figure C.1) et temporal (cortex auditif et notamment aire de

Wernicke : cf figure C.1) en étant les principaux acteurs avec parfois le lobe pariétal (gyrus supramarginal gauche). En 1979, Ross a émis l'hypothèse que les capacités prosodiques étaient quant à elles principalement latéralisées à droite. Cette hypothèse a ensuite était confortée par plusieurs études de patients cérébro-lésés (Ross [1981], Weintraub *et al.* [1981], Klouda *et al.* [1988], Brådvik *et al.* [1991] et Twist *et al.* [1991]). Elle fut longtemps admise comme une hypothèse valide, car elle reflétait la conception traditionnelle de la prosodie comme une simple auxiliaire de la syntaxe et de la sémantique, qui elles recrutent l'hémisphère gauche. Des études en neuroimagerie récentes fournissent d'ailleurs des résultats convergents. En effet, quand les aspects de la prosodie associés au traitement de la *mélodie* sont étudiés, on observe bien des activations de l'hémisphère droit (e.g. Zatorre *et al.* [1992], Tzourio *et al.* [1997] et Mayer *et al.* [2002]).

Cependant, comme il a déjà été expliqué précédemment (cf. section B.2 de l'introduction), les travaux récents en linguistique ont montré que la prosodie est une structure phonologique complexe qui doit être analysée en tant que telle (Beckman [1996]). La prosodie devrait donc recruter l'hémisphère *gauche*, tout comme la syntaxe et la sémantique. Une étude relativement ancienne souvent négligée corrobore cette hypothèse. En effet, l'observation la plus précoce d'un trouble neurologique de la prosodie a été publiée par Monrad-Krohn dès 1947 (Monrad-Krohn [1947]). Monrad-Krohn créa le terme « aprosodie » pour décrire le syndrome « d'accent étranger » chez une patiente qui présentait une lésion de l'hémisphère gauche (aire de Broca). Il suggéra dès lors qu'un lobe frontal *gauche* intact est nécessaire pour la production de la prosodie.

Une revue de la littérature récente en neuropsychologie (ou sur les patients cérébro-lésés) (Baum & Pell [1999]) démontre également que le traitement de la prosodie (en production et perception) ne peut être localisé strictement à droite. Plus précisément, cette revue cite des études de patients cérébro-lésés, sur la production et la perception de l'accent d'emphase (qui est lié comme nous l'avons déjà mentionné à la focalisation prosodique) indiquant que les patients cérébro-lésés à *gauche* (le plus souvent des aphasiques de Broca) peuvent présenter des troubles prosodiques aussi importants, voire plus, que les patients présentant des lésions de l'hémisphère *droit*.

Des études en neuroimagerie très récentes fournissent des résultats allant dans ce même sens. Les données électrophysiologiques de Astésano *et al.* [2004a] indiquent que l'attention à la prosodie (détection d'une incongruence prosodique) recrute essentiellement l'hémisphère *gauche*. Quelques études IRMf sur le traitement perceptif de divers phénomènes prosodiques obtiennent aussi des activations de l'hémisphère *gauche*. En outre, l'étude de Mayer *et al.* [2002] sur la production de schémas prosodiques aux niveaux de la syllabe et du syntagme a aussi révélé une activation de l'hémisphère *gauche*.

Comme il a été expliqué plus haut (cf. C.2.2 du présent chapitre), l'étude de Lœvenbruck *et al.* [2005] a permis de mettre en évidence la spécificité de la focalisation prosodique par rapport à la focalisation syntaxique. Cette étude a permis de décrire le réseau impliqué dans la *production* de la focalisation contrastive. On sait ainsi qu'en production la focalisation contrastive prosodique requiert l'intervention des aires traditionnellement impliquées en production de la parole (Broca et Wernicke) mais aussi une aire spécifique : le gyrus supramarginal gauche. Cette aire pourrait comme il a été expliqué ci-dessus, être recrutée dans le but de produire une information multisensorielle et notamment visuelle. Cette étude permet donc de suggérer que la planification neuronale de la focalisation pourrait être faite dans le cadre de la multisensorialité.

Les tests de perception visuelle de la focalisation décrits dans ce mémoire permettent de suggérer que l'information visuelle fournie par les articulateurs doit être utilisée lors du traitement cérébral de l'information prosodique. En effet, on peut extraire une information prosodique (focalisation

contrastive) à partir de la modalité visuelle seule. Néanmoins, à ce jour, les aires cérébrales recrutées lors de ce processus d'intégration de l'information visuelle pour le traitement de la prosodie n'ont pas encore été identifiées. Or les nouvelles techniques d'imagerie cérébrale paraissent susceptibles d'apporter des connaissances fondamentales sur le traitement de la prosodie par le système nerveux central. L'étude de Calvert *et al.* [1997] a mis en évidence les circuits neuraux impliqués dans la perception visuelle de la parole.

Il serait ainsi très intéressant d'une part de déterminer le réseau cérébral qui est recruté lors de la perception auditive de la focalisation contrastive. Le gyrus supra-marginal intervient-il ? D'autre part, l'étude des aires activées pendant la perception visuelle et audiovisuelle de la focalisation contrastive prosodique permettrait de déterminer les aires qui sont spécifiquement recrutées lors du processus de traitement de l'information prosodique visuelle dans le cerveau et ainsi de tenter de mieux comprendre le processus d'intégration des modalités en parole.

– Bibliographie –

ARGYLE Michael & COOK Mark, 1976. *Gaze and mutual gaze*, Cambridge University Press, Grande-Bretagne.

ASTAFIEV Serguei V., SHULMAN Gordon L., STANLEY Christine M., SNYDER Abraham Z., VAN ESSEN David C. & CORBETTA Maurizio, 2003. « Functional Organization of Human Intraparietal and Frontal Cortex for Attending, Looking, and Pointing », *Journal of Neuroscience*, 23 (11), p. 4689-4699.

ASTÉSANO Corinne, 2001. *Rythme et accentuation en français: invariance et variabilité stylistique*, thèse de doctorat, L'Harmattan édition et diffusion.

ASTÉSANO Corinne, BESSON Mireille & ALTER Kai, 2004a. « Brain potentials during semantic and prosodic processing in French », *Cognitive Brain Research,* 18, p. 172-184.

ASTÉSANO Corinne, MAGNE Cyrille, MOREL Michel, COQUILLON Annelise, ESPESSER Robert, BESSON Mireille & LACHERET-DUJOUR Anne, 2004b. « Marquage acoustique du focus contrastif non codé syntaxiquement en français », dans les *Actes des Journées d'Etude de la Parole 2004, Fès, Maroc,* p. 41-44.

AUDOUY Marc, 2000. *Traitement d'images vidéo pour la capture des mouvements labiaux*, rapport final d'ingénieur, Institut National Polytechnique de Grenoble.

BAHRICK Lorraine E., 2003. « Development of intermodal perception », dans NADEL L. (Ed.), *Encyclopedia of Cognitive Science*, Nature Publishing Group, Londres, vol. 2, p. 614-617.

BAILLY Gérard & BADIN Pierre, 2002. « Seeing tongue movements from outside », dans les *Actes de la conférence ICSLP 2002, Denver, USA,* p. 1913-1916.

BAILLY Gérard, BÉRAR Maxime, ELISEI Frédéric & ODISIO Matthias, 2003. « Audiovisual Speech Synthesis », *International Journal of Speech Technology*, 6, p. 331-346.

BARBOSA Plinio & BAILLY Gérard, 1994. « Characterisation of rhythmic patterns for text-to-speech synthesis », *Speech Communication*, 15, p. 127-137.

BARTELS Christine & KINGSTON John, 1994. « Salient Pitch Cues in the Perception of Contrastive Focus », dans BOSCH P. & VAN DER SANDT R. (Eds.), *Focus & Natural Language Processing, Proceedings of Journal of Semantics Conference on Focus*, *IBM Working Papers, TR-80,* p. 94-106.

BATES Elizabeth & DICK Frederic, 2002. « Language, gesture, and the developing brain », *Developmental Psychobiology,* 40, p. 293-310.

BAUM Shari R., KELSCH DANILOFF J., DANILOFF R. & LEWIS J., 1982. « Sentence Comprehension by Broca's aphasics: effects of some suprasegmental variables », *Brain and Language*, 17, p. 261-271.

BAUM Shari R. & PELL Marc D., 1999. « The neural bases of prosody: Insights from lesion studies and neuroimaging ». *Aphasiology,* 13(8), p. 581-608.

BEAUTEMPS Denis, BOREL Pascal & MANOLIOS Sébastien, 1999. « Hyper-articulated speech: Auditory and visual intelligibilty », dans les *Actes de la conférence Eurospeech 1999, Budapest, Hongrie,* p. 109-112.

BECKMAN Mary E., 1996. « The parsing of prosody », *Language and Cognitive Processes*, 11(1-2), p. 17-67.

BECKMAN Mary E., HIRSCHBERG Julia & SHATTUCK-HUFNAGEL Stefanie, 2005. « The original ToBI system and the evolution of the ToBI framework », dans JUN S.-A. (Ed.), *Prosodic typology: The Phonology of Intonation and Phrasing,* Oxford University Press, chapitre 2.

BENGUEREL André-Pierre, 1971. « Duration of French vowels in unemphatic stress », *Language & Speech*, 14(4), p. 383-391.

BENOÎT Christian, MOHAMADI Tayeb & KANDEL Sonia, 1994. « Effects of Phonetic Context on Audio-Visual Intelligibility of French », *Journal of Speech and Hearing Research*, 37, p. 1195-1203.

BENOÎT Christian, GRICE Martine & HAZAN Valérie, 1996a. « The SUS test: A method for the assessment of text-to-speech synthesis intelligibility using Semantically Unpredictable Sentences », *Speech Communication*, 18, p. 381-392.

BENOÎT Christian, GUIARD-MARIGNY Thierry, LE GOFF Bertrand & ADJOUDANI Ali, 1996b. « Which components of the face do humans and machines best speechread? », dans STORK D. G. & HENNECKE M. E. (Eds.), *Speechreading by Humans and Machines: Models, Systems, and Applications*, Springer-Verlag, New York, p. 315-328.

BERNSTEIN Lynne E., EBERHARDT Silvio P. & DEMOREST Marilyn E., 1989. « Single-channel vibrotactile supplements to visual perception of intonation and stress », *The Journal of the Acoustical Society of America*, 85(1), p. 397-405.

BERNSTEIN Lynne E., AUER Edward T., CHANEY Brian, ALWAN Abeer & KEATING Patricia A., 2000. « Development of a facility for simultaneous recordings of acoustic, optical (3-D motion and video), and physiological speech data », *The Journal of the Acoustical Society of America*, 107, p. 2887.

BESKOW Jonas, 1997. « Animation of Talking Agents », dans les *Actes de la conférence AVSP 1997, ESCA Workshop on Audio-Visual Speech Processing, Rhodes, Grèce,* p. 149-152.

BESLE Julien, FORT Alexandra, DELPUECH Claude & GIARD Marie-Hélène, 2004. « Bimodal speech: early suppressive visual effects in human auditory cortex », *European Journal of Neuroscience*, 20, p. 2225-2234.

BERTRAND Roxanne, BOYER Jacques, CAVÉ Christian, GUAÏTELLA Isabelle & SANTI Serge, 1995. « Relationship between gestures and voice in verbal interaction: Prosodic and kinesic aspects of back-channel signals », dans les *Actes de la conférence ICPhS 1995, Stockholm, Suède,* vol. 2, p. 746-749.

BINNIE Carl A., MONTGOMERY Allen A. & JACKSON Pamela L., 1974. « Auditory and visual contributions to the perception of consonants », *Journal of Speech and Hearing Research*, 17(4), p. 619-630.

BIRCH Stacy & CLIFTON Charles Jr., 1995. « Focus, accent, and argument structure: Effects on language comprehension », *Language and Speech*, 33, p. 365-391.

BIRDWHISTELL Ray L., 1970. « Kinesic stress in American English » dans *Kinesics and context*, University of Pennsylvania Press, Philadelphie (USA).

BOERSMA Paul & WEENINK David, 2005. *PRAAT, doing phonetics by computer,* de la version 3.4 à la version 4.3, [logiciel informatique], obtenu depuis http://www.praat.org/.

BOLINGER Dwight, 1985. *Intonation and its parts*, Edward Arnold, Londres (GB).

BOLINGER Dwight, 1989. *Intonation and its uses,* Stanford University Press, Stanford (USA).

BOYER Jacques, DI CRISTO Albert & GUAÏTELLA Isabelle, 2001. « Rôle de la voix et des gestes dans la focalisation », dans les *Actes du Colloque Oralité et Gestualité 2001*, p. 82-463.

BOYSSON-BARDIES DE Bénédicte, 1996. *Comment la parole vient aux enfants*, éditions Odile Jacob, Paris (France).

BOYSSON-BARDIES DE Bénédicte, SAGART Laurent & DURAND Catherine, 1984. « Discernible differences in the babbling of infants according to target language », *Journal of Child Language*, 11, p. 1-15.

BRÅDVIK Björn, DRAVINS Christina, HOLTÅS S., ROSÉN I., RYDING E. & INGVAR D., 1991. « Disturbances of Speech Prosody Following Right Hemisphere Infarcts », *Acta Neurologica Scandinavica*, 84, p. 114-126.

BRUCE Gösta, 1977. « Swedish word accents in sentence perspective », *Travaux de l'Institut de linguistique de Lund XII*, Gleerups, Lund.

BRUNER Jerome S., 1983. *Child Talk,* Norton, New York (USA).

BRYAN Karen, 1989. « Language Prosody and the Right Hemisphere », *Aphasiology*, 3, p.285-299.

BULL Peter, 2001. « Nonverbal communication », *The Psychologist*, 14(12), p. 644-647.

BURNHAM Denis & KEANE Sheila, 1997. « The Japanese Mcgurk effect: the role of linguistic and cultural factors on auditory-visual speech perception », dans les *Actes de la conférence AVSP-1997, Rhodes, Grèce,* p.93-96.

BUTTERWORTH George, 2003. « Pointing Is the Royal Road to Language for Babies », dans KITA S. (Ed.), *Pointing : Where language, culture, and cognition meet*, Lawrence Erlbaum Associates, Mahwah, (N.J., USA), p. 9-33.

BYRD Dani & SALTZMAN Elliot, 1998. « Intragestural dynamics of multiple prosodic boundaries », *Journal of Phonetics*, 26, p. 173-199.

BYRD Dani, KAUN Abigail, NARAYANAN Shrikanth & SALTZMAN Elliot, 2000. « Phrasal signatures in articulation », dans BROE M.B. & PIERREHUMBERT J.B. (Eds.), *Papers in Laboratory Phonology V, Acquisition and the Lexicon*, Cambridge University Press, p. 70-87.

CALLAN Daniel E., JONES Jeffery A., MUNHALL Kevin, CALLAN Akiko M., KROOS Christian & VATIKIOTIS-BATESON Eric, 2003. « Neural processes underlying perceptual enhancement by visual speech gestures », *NeuroReport*, 14(17), p. 2213-2218.

CALVERT Gemma A., BULLMORE Edward T., BRAMMER Michael J., CAMPBELL Ruth, WILLIAMS Steven C., MCGUIRE Philip K., WOODRUFF Peter W., IVERSEN Susan D. & DAVID Anthony S., 1997. « Activation of auditory cortex during silent lipreading », *Science*, 276, p. 593-596.

CALVERT Gemma A. & CAMPBELL Ruth, 2003. « Reading speech from still and moving faces: the neural substrates of visible speech », *Journal of Cognitive Neuroscience*, 15(1), p. 57-70.

CAMPBELL Ruth, 1988. « Tracing lip movements: Making speech visible », *Visible Language*, 22(1), p. 32-57.

CAMPBELL Ruth, 1992. « The neuropsychology of lipreading », dans BRUCE V., COWEY A., ELLIS A.W. & PERET D.I. (Eds.), *Processing of the facial image*, Clarendon Press, Oxford, p. 39-45.

CAPIRCI Olga, IVERSON Jana M., PIZZUTO Elena & VOLTERRA Virginia, 1996. « Gesture and words during the transition to two-word speech », *Journal of Child Language,* 23, p. 645-673.

CATHIARD Marie-Agnès, 1988/1989. « La perception visuelle de la parole : aperçu de l'état des connaissances », *Bulletin de l'Institut de Phonétique de Grenoble*, 17-18, p. 109-193.

CATHIARD Marie-Agnès, 1994. *La perception visuelle de l'anticipation des gestes vocaliques : cohérence des évènements audibles et visibles dans le flux de la parole*, thèse de Doctorat en Psychologie Cognitive, Université Pierre Mendés France, Grenoble.

CATHIARD Marie-Agnès, GEDZELMAN Séverine, ABRY Christian & LŒVENBRUCK Hélène, 2004. « Naissance de la représentation d'une consonne entre les voyelles : les conditions d'une intégration audiovisuelle », dans les *Actes des XXVèmes Journées d'Études sur la Parole 2004, Fès, Maroc*, p. 117-120.

CAVÉ Christian, GUAÏTELLA Isabelle & SANTI Serge, 1993. « Fréquence fondamentale et mouvements rapides des sourcils : une étude pilote », *Travaux de l'Institut de Phonétique d'Aix*, 15, p. 25-42.

CAVÉ Christian, GUAÏTELLA Isabelle, BERTRAND Roxanne, SANTI Serge, HARLAY Françoise &

ESPESSER Robert, 1996. « About the Relationship between Eyebrow movements and F0 Variation », dans les *Actes de la conférence ICSLP 1996, Philadelphie, USA,* vol. 4, p. 2175-2179.

CERRATO Loredana & SKHIRI Mustafa, 2003. « Analysis and measurement of communicative gestures in human dialogues », dans les *Actes de la conférence AVSP 2003, St Jorioz, France,* p. 251-256.

CHO Taehong, 2005. « Prosodic strengthening and featural enhancement: Evidence from acoustic and articulatory realizations of /ɑ, i/ in English », *The Journal of the Acoustical Society of America*, 117(6), p. 3867-3878.

CHO Taehong, à paraître. « Manifestation of Prosodic Structure in Articulation: Evidence from Lip Kinematics in English », dans *Laboratory Phonology 8,* Mouton de Gruyter, Berlin/New York.

CHOVIL Nicole, 1991/1992. « Discourse-Oriented Facial Displays in Conversation », *Research on Language and Social Interaction*, 25, p. 163-194.

CLECH-DARBON Anne, REBUSCHI Georges & RIALLAND Annie, 1999. « Are there Cleft Sentences in French? », dans TULLER L. & REBUSCHI G. (Eds.), *The Grammar of Focus*, Benjamins, Amsterdam, p. 83-118.

CONDON William S., 1976. « An analysis of behavioural organization », *Sign Language Studies*, 13, p. 285-318.

COOKE J.D., 1980. « The Organization of Simple, Skilled Movements », dans STELMACH G.E. & REQUIN J. (Eds.), *Tutorials in Motor Behavior,* Elsevier Science Publishers B.B., Amsterdam (Pays-Bas).

COSNIER Jacques, 1991. « Les gestes de la question », dans KERBRAT-ORECCHIONI (Ed.), *La question*, Presses Universitaires de Lyon, p. 163-171.

DAHAN Delphine & BERNARD Jean-Marc, 1996. « Interspeaker Variability in Emphatic Accent Production in French », *Language and Speech*, 39(4), p. 341-374.

DE JONG Kenneth, 1995. « The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation », *The Journal of the Acoustical Society of America*, 97(1), p. 491-504.

DE JONG Kenneth, BECKMAN Mary E. & EDWARDS Jan, 1993. « The interplay between prosodic structure and coarticulation », *Language and Speech*, 36(2-3), p. 197-212.

DELAIS-ROUSSARIE Elisabeth, RIALLAND Annie, DOETJES Jenny & MARANDIN Jean-Marie, 2002. « The Prosody of Post Focus Sequences in French », dans les *Actes de la conférence Speech Prosody 2002*, *Aix-en-Provence, France,* p. 239-242.

DELATTRE Pierre, 1966. « Les dix intonations de base du français », *French Review*, 40, p. 1-14.

DELGUTTE Bertrand, 1978. « Technique for the perceptual investigation of F0 contours with application to French », *The Journal of the Acoustical Society of America*, 64, p.1319-1332.

DENG Li, RAMSAY Gordon & SUN Don X., 1997. « Production models as a structural basis for automatic speech recognition », *Speech Communication*, 22(2), p. 93-111.

DE PIJPER J.R., 1980. « A melodical model of British English intonation », *IPO Annual Progress Report,* 15, p. 54-58.

DI CRISTO Albert, 1985. *De la microprosodie à l'intonosyntaxe*, Publications de l'Université de Provence, France.

DI CRISTO Albert, 1998. « Intonation in French », dans HIRST D. & DI CRISTO A. (Eds.), *Intonation Systems: a Survey of Twenty Languages*, Cambridge University Press, p. 195-218.

DI CRISTO Albert, 2000. « Vers une modélisation de l'accentuation du français (deuxième partie) », *Journal of French Language Studies*, 10, p. 27-44.

DI CRISTO Albert & JANKOWSKI Ludovic, 1999. « Prosodic Organisation and Phrasing after Focus in French », dans *les Actes de la conférence ICPhS 1999, San Francisco, USA,* p. 1565-1568.

DIESSEL Holger & TOMASELLO Michael, 2000. « The development of relative clauses in spontaneous child speech », *Cognitive Linguistics*, 11(1/2), p. 131-151.

D'IMPERIO Mariapaola, 2001. « Focus and Tonal Structure in Neapolitan Italian », *Speech Communication*, 33(4), p. 339-356.

D'IMPERIO Mariapaola & HOUSE David, 1997. « Perception of questions and statements in Neapolitan Italian », dans les *Actes de la conférence Eurospeech 1997, Rhodes, Grèce,* p. 251-254.

D'IMPERIO Mariapaola, ESPESSER Robert, LŒVENBRUCK Hélène, MENEZES Caroline, N'GUYEN Noël & WELBY Pauline, à paraître. « Are tones aligned to articulatory events? Evidence from Italian and French », COLE J. & HUALDE J. (Eds.), *Laboratory Phonology IX.*

DITTMAN Allen T., 1974. « The body movement-speech rhythm relationship as a cue for speech encoding », dans WEITZ S. (Ed.), *Nonverbal Communication*, Oxford University Press, New-York (USA), p. 169-181.

DITTMAN Allen T. & LLEWELYN L.G., 1969. « Body movement and speech rhythm in social conversation », *Journal of Personality and Social Psychology*, 11(2), p. 98-106.

DODD Barbara, 1977. « The role of vision in the perception of speech », *Perception,* 6, p. 31-40.

DOHALSKÁ Marie & MEJVALDOVÁ Jana, 2000. « Rôle de la prosodie dans la communication en milieu bruité », dans les *Actes des XXIIIèmes Journées d'Etude de la Parole, Aussois, France,* p. 265-268.

DOHEN Marion, LŒVENBRUCK Hélène, CATHIARD Marie-Agnès & SCHWARTZ Jean-Luc, 2003a. « Potential audiovisual correlates of contrastive focus in French », dans les *Actes de la conférence Eurospeech 2003, Genève, Suisse,* p. 145-148.

DOHEN Marion, LŒVENBRUCK Hélène, CATHIARD Marie-Agnès & SCHWARTZ Jean-Luc, 2003b. « Audiovisual perception of contrastive focus in French », dans les *Actes de la conférence AVSP 2003, St. Jorioz, France*, p. 245-250.

DRONKERS Nina F., PINKER Steven & DAMASIO Antonio, 2000. « Language and the aphasias », dans KANDEL E.R., SCHWARTZ J.H. & JESSEL T.M. (Eds.), *Principles of neural science*, Mc Graw-Hill, New York (USA), p. 1169-1187.

DUCEY Virginie, 2002. *Monstration et interrogation dans la naissance du langage*, mémoire de DEA de Sciences du Langage, université Stendhal Grenoble III.

EKMAN Paul, 1976. « Movements with precise meanings », *Journal of Communication*, 26, p. 14-26.

EKMAN Paul, 1999. « Facial Expressions », dans DALGEISH T. & POWER M. (Eds.), *Handbook of Cognition and Emotion*, John Wiley & Sons Ltd., Sussex (GB), chap. 16.

ENGWALL Olov & BESKOW Jonas, 2003. « Resynthesis of 3D tongue movements from facial data », dans les *Actes de la conférence Eurospeech 2003, Genève, Suisse,* p. 2261-2264.

ERBER Norman P., 1969. « Interaction of audition and vision in the recognition of oral speech stimuli », *Journal of Speech and Hearing Research*, 12(2), p. 423-425.

ERBER Norman P., 1975. « Auditory-visual perception of speech », *Journal of Speech and Hearing Disorders*, 40(4), p. 481-492.

ERICKSON Donna, 1998. « Effects of Contrastive Emphasis on Jaw Opening », *Phonetica*, 55, p. 147-169.

ERICKSON Donna, 2004. « Perception of contrastive emphasis by American English and Japanese listeners », dans les *Actes de la conférence Speech Prosody 2004, Nara, Japon,* p. 701-704.

ERICKSON Donna, LENZO Kevin & FUJIMURA Osamu, 1994. « Manifestations of contrastive emphasis in jaw movement », *The Journal of the Acoustical Society of America*, 95(5), p. 2822.

ERICKSON Donna & LEHISTE Ilse, 1995. « Contrastive emphasis in elicited dialogue: durational compensation », dans les *Actes de la conférence ICPhS 1995, Stockholm, Suède,* vol. 4, p. 352-355.

ERICKSON Donna & FUJIMURA Osamu, 1996. « Maximum jaw displacement in contrastive emphasis », dans les *Actes de la conférence ICSLP 1996, Philadelphie, USA,* vol. 1, p. 141-144.

ERICKSON Donna & HONDA Kiyoshi, 1996. « Jaw displacement and F0 in contrastive emphasis », *The Journal of the Acoustical Society of America*, 99(4), p. 2494.

ERICKSON Donna, FUJIMURA Osamu & PARDO Bryan, 1998. « Articulatory Correlates of Prosodic Control: Emotion and Emphasis », *Language and Speech*, 41(3-4), p. 399-417.

ERICKSON Donna, MAEKAWA Kikuo, HASHI Michiko & DANG Jianwu, 2000. « Some articulatory and acoustic changes associated with emphasis in spoken English », dans les *Actes de la conférence ICSLP 2000, Pékin, Chine,* vol. 3, p. 247-250.

FERNALD Anne, 1989. « Intonation and communicative intent in mothers' speech to infants: Is the melody the message? », *Child Development*, 60(6), p. 1497-1510.

FERNALD Anne, 1991. « Prosody in speech to children: Prelinguistic and linguistic functions », dans VASTA R. (Ed.), *Annals of Child Development*, *volume 8,* Jessica Kingsley, Londres (GB), p. 43-80.

FERNALD Anne, 1993. « Human maternal vocalizations to infants as biologically relevant signals: an evolutionary perspective », dans BLOOM P. (Ed.), *Language acquisition. Core readings*, Harvester Wheatsheaf, p. 51-93.

FUJISAKI Hiroya & SUDO H., 1971. « Synthesis by rule of prosodic features of connected Japanese », dans les *Actes de la conférence.VIIth International Congress of Acoustics, Budapest, Hongrie*, vol.3, p. 133-136.

GALLIGAN R., 1987. « Intonation with single words: Purposive and grammatical use », *Journal of Child Language*, 14, p. 1-21.

GHAZANFAR Asif A., MAIER Joost X., HOFFMAN Kari L. & LOGOTHETIS Nikos, 2005. « Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex », *Journal of Neuroscience,* 25, p. 5004-5012.

GOFFMAN Lisa & SMITH Anne, 1996. « A kinematic analysis of contrastive stress », *The Journal of the Acoustical Society of America*, 99(4), p. 2494.

GOGATE Lakshmi J. & BAHRICK Lorraine E., 1998. « Intersensory Redundancy Facilitates Learning of Arbitrary Relations between Vowel Sounds and Objects in Seven-Month-Old Infants », *Journal of Experimental Child Psychology*, 69, p. 133-149.

GOLDIN-MEADOW Susan J., 1999. « The development of gesture with and without speech in hearing and deaf children », dans MESSING L. & CAMPBELL R. (Eds.), *Gesture, speech and sign,* Oxford University Press, New-York (USA), p. 117-132.

GOLDIN-MEADOW Susan J. & BUTCHER Cynthia, 2003. « Pointing toward two-word speech in young children », dans KITA S. (Ed.), *Pointing : Where language, culture, and cognition meet*, Lawrence Erlbaum Associates, Mahwah, (N. J., USA), p. 85-107.

GRACCO Vincent L., 1988. « Timing factors in the coordination of speech movements », *The Journal of Neuroscience*, 8(12), p. 4628-4636.

GRAF Hans Peter, COSATTO Eric & EZZAT Tony, 2000. « Face analysis for the synthesis of photo-realistic talking heads », dans les *Actes de la IEEE International Conference on Automatic Face and Gesture Recognition, Grenoble, France,* p. 189-194.

GRAF Hans Peter, COSATTO Eric, STROM Volker & HUANG Fu Jie, 2002. « Visual Prosody: Facial Movements Accompanying Speech », dans les *Actes de la 5ième IEEE International Conference on Automatic Face and Gesture Recognition (FGR'02),* p. 381-386.

GRANSTRÖM Björn, HOUSE David & LUNDEBERG Magnus, 1999. « Prosodic cues in multimodal speech perception », dans les *Actes de la conférence ICPhS 1999, San Francisco, USA,* vol.1, p. 655-658.

GRANSTRÖM Björn & HOUSE David, 2004. « Audiovisual Representation of Prosody in Expressive Speech Communication », dans les *Actes de la conférence Speech Prosody 2004, Nara, Japon,* p. 393-400.

GRANSTRÖM Björn & HOUSE David, 2005. « Audiovisual representation of prosody in expressive speech communication », *Speech Communication*, 46, p. 473-484.

GRANT Ken W., ARDELL Lee Ann H., KUHL Patricia K. & SPARKS David W., 1986. « The Transmission of Prosodic Information Via an Electrotactile Speechreading Aid », *Ear and Hearing*, 7(5), p. 328-335.

GRANT Ken W. & BRAIDA Louis D., 1991. « Evaluating the Articulation Index for audiovisual input », *The Journal of the Acoustical Society of America*, 89, p. 2952-2960.

GRANT Ken W. & SEITZ Philip-Franz, 2000. « The use of visible speech cues for improving auditory detection of spoken sentences », *The Journal of the Acoustical Society of America*, 108(3), p. 1197-1208.

GRICE Martine; LADD Robert D. & ARVANITI Amalia, 2000. « On the place of phrase accents in intonational phonology », *Phonology*, 17, p. 143-185.

GUAÏTELLA Isabelle, 1991. « Etude des relations entre geste et prosodie à travers leurs fonctions rythmique et symbolique », dans les *Actes de la XIIièmes conférence ICPhS, Aix-en-Provence, France,* p. 266-269.

GUAÏTELLA Isabelle, 1994. « Interaction entre l'activité gestuelle et l'activité vocale dans la communication : éléments théoriques et méthodologiques », dans les *Actes des XXèmes Journées d'Etude sur la Parole, Trégastel, France,* p. 389-393.

GUAÏTELLA Isabelle, CAVÉ Christian & SANTI Serge, 1993. « Relations entre geste et voix : Le cas des sourcils et de la fréquence fondamentale », dans *Images et Langages : Multimodalité et modélisation cognitive. Actes du colloque interdisciplinaire du Comité National de la Recherche Scientifique*, *Paris, France,* p. 261-268.

GUSSENHOVEN Carlos, 1983. « Testing the Reality of Focus Domains », *Language and Speech*, 26(1), p. 61-80.

HADAR U., STEINER T.J., GRANT E.C. & ROSE F.C., 1983. « Head movement correlates of juncture and stress at sentence level », *Language and Speech*, 26, p. 117–129.

HALLÉ Pierre, BOYSSON-BARDIES DE Bénédicte & VIHMAN Marylin May, 1991. « Beginnings of prosodic organization: Intonation and duration patterns of disyllables produced by Japanese and French infants », *Language and Speech*, 34(4), p. 299-318.

HALLIDAY Michael A. K., 1967. *Intonation and grammar in British English*, Mouton, La Hague.

HALLIDAY Michael A. K., 1975. *Learning how to mean*, Edward Arnold, Londres (GB).

HALLIDAY Michael A. K., 1979. « One's child protolanguage », dans BULLOWA M. (Ed.), *Before Speech: The beginning of interpersonal communication*, Cambridge University Press, Cambridge.

HARDISON Debra M., 2003. « Acquisition of second-language speech: Effects of visual cues, context and talker variability », *Applied Psycholinguistics*, 24, p. 495-522.

HARRINGTON Jonathan, FLETCHER Janet & ROBERTS Corinne, 1995. « Coarticulation and the accented/unaccented distinction: evidence from jaw movement data », *Journal of Phonetics*, 23, p. 305-322.

HAZAN Valérie, SENNEMA Anke & FAULKNER Andrew, 2002. « Audiovisual perception in L2 learners », dans les *Actes de la conférence ICSLP 2002, Denver, USA,* p. 1685-1688.

HEESE G., 1957. « Akzente und Betleitgeärden », *Sprachforum*, 2, p. 274-285.

HIRSH-PASEK Kathryn, KEMLER NELSON Deborah G., JUSCZYK Peter W., WRIGHT CASSIDY Kimberly, DRUSS Benjamin & KENNEDY Lori J., 1987. « Clauses are perceptual units for young infants », *Cognition*, 26, p. 269-286.

HIRST Daniel & DI CRISTO Albert, 1993. « Rythme syllabique, rythme mélodique et représentation hiérarchique de la prosodie du français », *Travaux de l'Institut de Phonétique d'Aix-en-Provence*, 15, p. 9-24.

HOUSE David, BESKOW Jonas & GRANSTRÖM Björn, 2001a. « Timing and Interaction of Visual Cues for Prominence in Audiovisual Speech Perception », dans les *Actes de la conférence Eurospeech 2001, Aalborg, Danemark,* vol. 1, p. 387-390.

HOUSE David, BESKOW Jonas & GRANSTRÖM Björn, 2001b. « Interaction of visual cues for prominence », *Working Papers Lund,* 49, Lund University, Department of Linguistics, p. 62-65.

IACOBONI Marco, 2005. « Understanding others: Imitation, Language, Empathy », dans HURLEY S. & CHATER N. (Eds.), *Perspectives on imitation: From cognitive neuroscience to social science*, MIT Press, Cambridge (USA), vol. 1, chap. 2.

INDEFREY P. & LEVELT W. J. M., 2004. « The spatial and temporal signatures of word production components », *Cognition,* 92, p. 101-144.

JELLEMA Tjeerd & PERRET Dave, 2003. « Perceptual History Influences Neural Responses to Face and Body Postures », *Journal of Cognitive Neuroscience,* 15, p. 961-971.

JIANG Jintao, ALWAN Abeer, BERNSTEIN Lynne E., KEATING Patricia & AUER Ed, 2000. « On the Correlation between Facial Movements, Tongue Movements and Speech Acoustics », dans les *Actes de la conférence ICSLP 2000, Pékin, Chine,* vol. 1, p. 42-45.

JUN Sun-Ah & FOUGERON Cécile, 2000. « A Phonological Model of French Intonation », dans BOTINIS A. (Ed.), *Intonation: Analysis, modelling and technology*, Kluwer Academic Publishers, Dordrecht, p. 209-242.

JUN Sun-Ah & FOUGERON Cécile, 2002. « Realizations of Accentual Phrases in French Intonation », *Probus*, 14, p. 147-172.

KEATING Patricia, BARONI Marco, MATTYS Sven, SCARBOROUGH Rebecca, ALWAN Abeer, AUER Edward T. & BERNSTEIN Lynne E., 2003. « Optical Phonetics and Visual Perception of Lexical and Phrasal Stress in English », dans les *Actes de la conférence ICPhS 2003, Barcelone, Espagne,* p. 2071-2074.

KELLER Eric, 1987. « The variation of absolute and relative measures of speech activity », *Journal of Phonetics*, 15(4), p. 335-347.

KELSO J.A. Scott, VATIKIOTIS-BATESON Eric, SALTZMAN Elliot & KAY Bruce A., 1985. « A qualitative dynamic analysis of reiterant speech production: phase portraits, kinematics, and dynamic modelling », *The Journal of the Acoustical Society of America*, 77(1), p. 266-280.

KENDON Adam, 1978. « Differential perception and attentional frame in face-to-face interaction: Two problems for investigation », *Semiotica*, 24, p. 305-315.

KENT Raymond D. & NETSELL Ronald W., 1971. « Effects of stress contrasts on certain articulatory parameters », *Phonetica*, 24, p. 23-44.

KIM Jeesun & DAVIS Chris, 2003. « Hearing foreign voices: Does knowing what is said affect masked visual speech detection? », *Perception*, 32, p. 111-120.

KITA Sotaro, 2002. *Pointing. Where Language, Culture, and Cognition Meet*, Lawrence Erlbaum Associates, Mahwah (USA).

KLOUDA Gayle V., ROBIN Donald A., GRAFF-RADFORD Neil R. & COOPER William E., 1988. « The role of callosal connections in speech prosody », *Brain Language*, 35(1), 154–171.

KOZHEVNIKOV V.A. & CHISTOVICH Ludmila A., 1965. « Speech : Articulation and Perception », *Joint Publications Research Service,* vol. 30.543, Washington DC (USA).

KRAHMER Emiel, RUTTKAY Zósfia, SWERTS Marc & WESSELINK Wieger, 2002a. « Perceptual Evaluation of Audiovisual cues for prominence », dans les *Actes de la conférence ICSLP 2002, Denver, USA,* p. 1933-1936.

KRAHMER Emiel, RUTTKAY Zósfia, SWERTS Marc & WESSELINK Wieger, 2002b. « Pitch, Eyebrows and the Perception of Focus », dans les *Actes de la conférence Speech Prosody 2002, Aix-en-Provence, France,* p. 443-446.

KRAHMER Emiel & SWERTS Marc, à paraître. « Perceiving Focus », chapitre écrit pour un ouvrage publié suite au *LSA workshop on Topic and Focus*.

KUHL Patricia K. & MELTZOFF Andrew N., 1982. « The bimodal perception of speech in infancy », *Science*, 218, p. 1138-1141.

KUHL Patricia K. & MELTZOFF Andrew N., 1984. « The Intermodal Representation of Speech in Infants », *Infant Behavior and Development*, 7(3), p. 361-381.

KURATATE Takaaki, MUNHALL Kevin G., RUBIN Philip, VATIKIOTIS-BATESON Eric & YEHIA Hani, 1999. « Audio-visual synthesis of talking faces from speech production correlates », dans les *Actes de la conférence Eurospeech 1999, Budapest, Hongrie,* p. 1279-1282.

LADD Robert D., 1996. *Intonational Phonology*, Cambridge University Press.

LALLOUACHE Mohamed-Tahar, 1990. « Un poste 'visage-parole'. Acquisition et traitement de contours labiaux », dans les *Actes des XVIIIèmes Journées d'Etude de la Parole, Montréal, Canada,* p. 282-286.

LALLOUACHE Mohamed-Tahar, 1991. *Un poste Visage-Parole couleur. Acquisition et traitement automatique des contours de lèvres,* thèse de doctorat, Institut National Polytechnique de Grenoble.

LARKEY Leah S., 1983. « Reiterant speech: An acoustic and perceptual validation », *The Journal of the Acoustical Society of America*, 73(4), p. 1337-1345.

LEBIB Riadh, PAPO David, DE BODE Stella & BAUDONNIÈRE Pierre-Marie, 2003. « Evidence of a visual-to-auditory cross-modal sensory gating phenomenon as reflected by the human P50 event-related brain potential modulation », *Neuroscience Letters*, 341, p. 185-188.

LEGERSTEE Maria, 1990. « Infants use multimodal information to imitate speech sounds », *Infant Behavior and Development*, 13(3), p. 343-354.

LEHISTE Ilse, 1970. *Suprasegmentals*, The MIT Press, Cambridge (MA,USA).

LEVITT Andrea G. & WANG Qi, 1991. « Evidence for language-specific rhythmic influences in the reduplicative babbling of French – and English-learning infants », *Language and Speech*, 34(3), p. 235-249.

LEWKOWICZ David J., 1998. « Infants' Response to the Audible and Visible Properties of the Human Face: II. Discrimination of Differences between Singing and Adult-Directed Speech », *Developmental Psychobiology*, 32(4), p. 261-274.

LIBERMAN Mark Y. & PRINCE Alan, 1977. « On Stress and Linguistic Rhythm », *Linguistic Inquiry,* 8(2), p. 249-336.

LIBERMAN Mark Y. & STREETER Lynn A., 1978. « Use of nonsense-syllable mimicry in the study of prosodic phenomena », *The Journal of the Acoustical Society of America*, 63(1), p. 231-233.

LIBERMAN Mark Y. & PIERREHUMBERT Janet, 1984. « Intonational invariance under changes in pitch range and length », dans ARONOFF M. & OEHRLE R. (Eds.), *Language sound to structure: studies in phonology presented to Morris Halle by his teacher and students*, MIT Press, p. 157-233.

LIBERMAN Alvin M. & MATTINGLY Ignatius G., 1985. « The motor theory of speech perception revised », *Cognition*, 21, p. 1-36.

LINDBLOM Björn & RAPP Karin, 1973. « Some temporal regularities of spoken Swedish », *Papers from the Institute of Linguistics of the University of Stockholm*, 21, p. 1-59.

LINDBLOM Björn, 1990. « Explaining phonetic variation: A sketch of the H&H theory », dans HARDCASTLE W.J. & MARCHAL A. (Eds.), *Speech Production and Speech Modelling*, Kluwer Academic Publishers, Pays-Bas, p. 403-439.

LŒVENBRUCK Hélène, 1996. *Pistes pour le contrôle d'un robot parlant capable de réduction vocalique*, thèse de doctorat de sciences cognitives, Institut National Polytechnique de Grenoble, France.

LŒVENBRUCK Hélène, 1999. « An investigation of articulatory correlates of the Accentual Prase in French », dans les *Actes de la 14ième conférence ICPhS 99, San Francisco, USA,* vol. 1, p. 667-670.

LŒVENBRUCK Hélène, 2000. « Effets articulatoires de l'emphase contrastive sur la Phrase Accentuelle en français », dans les *Actes des Journées d'Etude de la Parole, Aussois, France,* p. 165-168.

LŒVENBRUCK Hélène, BACIU Monica, SEGEBARTH Christoph & ABRY Christian, 2005. « The left inferior frontal gyrus under focus: an fMRI study of the production of deixis via syntactic extraction and prosodic focus », *Journal of Neurolinguistics*, 18, p. 237-258.

MACCHI M. J., 1985. *Segmental and suprasegmental features and lip and jaw articulators*, thèse de doctorat, université de New York.

MACLEOD Alison & SUMMERFIELD Arthur Quentin, 1987. « Quantifying the contribution of vision to speech perception in noise », *British Journal of Audiology*, 21, p. 131-141.

MAEDA Shinji, 1976. *A characterization of American English intonation*, thèse de doctorat, MIT, Cambridge.

MAIDMENT John A., 1983. « Language recognition and prosody: further evidence », *Speech, hearing and language: Work in progress, University College London,* 1, p. 133-141.

MANDEL Denise R., JUSCZYK Peter W. & KEMLER NELSON Deborah G., 1994. « Does sentential prosody help infants organize and remember speech information? », *Cognition*, 53(2), p. 155-180.

MASSARO Dominic W., 1987. *Speech Perception by ear and eye: a paradigm for psychological inquiry*, Laurence Erlbaum Associates, Londres (GB).

MASSARO Dominic W., 1989. « Multiple Book Review of Speech Perception by Ear and Eye: A Paradigm for Psychological Inquiry », *Behavioral and Brain Sciences*, 12, p. 741-794.

MASSARO Dominic W., 2002. « Multimodal Speech Perception: A Paradigm for Speech Science », dans GRANSTRÖM B., HOUSE D. & KARLSSON I. (Eds.), *Multimodality in language and speech systems*, Kluwer Academic Publishers, Pays-Bas, p.45-71.

MAYER Jörg, WILDGRUBER Dirk, RIECKER Axel, DOGIL Grzegorz, ACKERMAN Hermann & GRODD Wolfgang, 2002. « Prosody Production and Perceptiom: Converging Evidence from fMRI Studies », dans *les Actes de la conférence Speech Prosody 2002, Aix-en-Provence, France,* pp. 487-490.

MCGRATH Matthew, SUMMERFIELD Quentin & BROOKE Michael, 1984. « Roles of lips and teeth in lipreading vowels », *Proceedings of the Institute of Acoustics*, 6, p. 401-408.

MCGURK Harry & MACDONALD John, 1976. « Hearing lips and seeing voices », *Nature*, 264, p.

746-748.

MEHLER Jacques, JUSCZYK Peter W., LAMBERTZ Ghislaine, HALSTED Nilofar, BERTONCINI Josiane & AMIEL-TISON Claudine, 1988. « A precursor of language acquisition in young infants », *Cognition*, 29(2), p. 143-178.

MERTENS Piet, 1993. « Intonational grouping, boundaries and syntactic structure in French », dans HOUSE D. & TOUATI P. (Eds.), *Proceedings of the ESCA Workshop on Prosody, Lund Working Papers,* 41, p. 155-159.

MERTENS Piet, GOLDMAN Jean-Philippe, WEHRLI Eric & GAUDINAT Arnaud, 2001. « La synthèse de l'intonation à partir de structures syntaxiques riches », *Traitement Automatique des Langues*, 42 (1), p. 142-195.

MEYERS Laura F., 1976. *Aspects of Hausa Tone* (*UCLA Working Papers in Phonetics 32*), Los Angeles: Phonetics Laboratory, University of California at Los Angeles.

MILLER George A., HEISE G.A. & LICHTEN W., 1951. « The intelligibility of speech as a function of the context of the test materials », *Journal of Experimental Psychology*, 41, p. 329-335.

MILLER George A. & NICELY Patricia, 1955. « An Analysis of Perceptual Confusions among some English Consonants », *The Journal of the Acoustical Society of America*, 27(2), p. 338-352.

MONRAD-KROHN Georg Herman, 1947. « Dysprosody or altered "melody of language" », *Brain*, 70, p. 405-415.

MOREL Marie-Annick & DANON-BOILEAU Laurent, 1998. *Grammaire de l'intonation. L'exemple du français oral*, Ophrys, Bibliothèque de Faits de Langues, Paris-Gap (France).

MORFORD Marolyn & GOLDIN-MEADOW Susan J., 1992. « Comprehension and production of gesture in combination with speech in one-word speakers », *Journal of Child Language,* 19(3), p. 559-580.

MORGAN Bayard Quincy, 1953. « Question melodies in American English », *American Speech*, 2, p. 181-191.

MULFORD Randa C., 1988. « First words of the blind child », dans SMITH M.D. & LOCKE J.L. (Eds.), *The emergent lexicon: The child's development of a linguistic vocabulary*, Academic Press, New-York (USA), p. 293-338.

MUNHALL Kevin G., JONES Jeffery A., CALLAN Daniel E., KURATATE Takaaki & VATIKIOTIS-BATESON Eric, 2004. « Visual Prosody and Speech Intelligibility – Head Movement Improves Auditory Speech Perception », *Psychological Science*, 15(2), p. 133-137.

NEELY Keith K., 1956 . « Effects of visual factors on the intelligibility of speech », *The Journal of the Acoustical Society of America*, 28(6), p. 1275-1277.

NELSON W. L., 1983. « Physical principles for economies of skilled movements », *Biological Cybernetics*, 46(2), p. 135-147.

NØLKE Henning, 1994. *Linguistique modulaire : de la forme au sens*, Peeters, Louvain.

OHALA John J. & GILBERT Judy B., 1981. « Listeners' ability to identify languages by their prosody », dans LEON P. & ROSSI M. (Eds.), *Problèmes de prosodie, Vol. II: Experimentations, modèles et functions (Studia Phonetica, 18)*, Didier, Ottawa (Canada), p. 123-131.

OJANEN Ville, MÖTTÖNEN Riikka, PEKKOLA Johanna, JÄÄSKELÄINEN Iiro P., JOENSUU Raimo, AUTTI Taina & SAMS Mikko, 2005. « Processing of audiovisual speech in Broca's area », *NeuroImage*, 25(2), p. 333-338.

O'SHAUGHNESSY Douglas, 1981. « A study of French vowel and consonant durations », *Journal of Phonetics*, 9(4), p. 385-406.

OSTRY David J., KELLER Eric & PARUSH Avraham, 1983. « Similarities in the control of the speech articulators and the limbs: kinematics of tongue dorsum movement in speech », *Journal of Experimental Psychology: Human Perception and Performance*, 9, p. 622-636.

OSTRY David J., GRIBBLE Paul L. & GRACCO Vincent L., 1996. « Coarticulation of Jaw Movements in Speech Production: Is Context Sensitivity in Speech Kinematics Centrally Planned? », *The Journal of Neuroscience*, 16(4), p. 1570-1579.

PARÉ Martin, RICHLER Rebecca, TEN HOVE Martin & MUNHALL Kevin G., 2003. « Gaze behavior in audiovisual speech perception: The influence of ocular fixations on the Mc Gurk effect », *Perception & Psychophysics*, 65(4), p. 553-567.

PASDELOUP Valérie, 1990. *Modèles de règles rythmiques du français appliqué à la synthèse de parole*, thèse de doctorat, Université de Provence.

PENTLAND Alex P. & DARELL Trevor, 1994. « Visual Perception of human bodies and faces for multi-modal interfaces », dans les *Actes de la conférence ICSLP 1994, Yokohama, Japon*, p. 543-546.

PERKELL Joseph S., ZANDIPOUR Madji, MATTHIES Melanie M. & LANE Harlan, 2002. « Economy of effort in different speaking conditions. I. A preliminary study of intersubject differences and modeling issues », *The Journal of the Acoustical Society of America*, 112(4), p. 1627-1641.

PERRET David I., 2005. « Seeing the future. Our brain cells predict what is coming! », *Présentation orale donnée à la conférence "Anatomical and functional organisation of the inferior parietal lobule and its role in the mirror system : evidence from monkey and human studies", Lyon, France,* 15 septembre 2005.

PERRIER Pascal, ABRY Christian & KELLER Eric, 1989. « Vers une modélisation des mouvements du dos de la langue », *Journal d'Acoustique*, 2, p. 69-78.

PIERREHUMBERT Janet, 1980. *The Phonology and Phonetics of English Intonation*, thèse de doctorat, MIT.

PIERREHUMBERT Janet & BECKMAN Mary E., 1988. *Japanese Tone Structure*, MIT Press, Cambridge (USA).

PIERREHUMBERT Janet & HIRSHBERG Julia, 1990. « The meaning of intonational contours in discourse », dans COHEN P., MORGAN J. & POLLACK M. (Eds.), *Intentions in Communication*, Cambridge, (MA, USA), MIT Press, p. 271-311.

PIKE Kenneth L., 1945. *The intonation of American English*, University of Michigan Press, Ann Arbor.

POST Brechtje, 2000. *Tonal and Phrasal Structures in French Intonation*, thèse de doctorat, The Netherlands Graduate School of Linguistics, Thesus.

PURSON Alain, SANTI Serge, BERTRAND Roxanne, GUAÏTELLA Isabelle, BOYER Jacques & CAVÉ Christian, 1999. « The relationships between voice and gesture: Eyebrow movements and questioning », dans les *Actes de Eurospeech 1999, Budapest, Hongrie,* p. 1735-1738.

RAFFLER ENGEL Walburga VAN, 1980. *Aspects of Nonverbal Communication*, Swets and Zeitlinger Lisse (Allemagne).

RAMUS Franck, 2002. « Language discrimination by newborns. Teasing apart phonotactic, rhythmic, and intonational cues », *Annual Review of Language Acquisition*, 2, p. 85-115.

REISBERG Daniel, MCLEAN John & GOLDFIELD Anne, 1987. « Easy to Hear but Hard to Understand: A Lip-reading Advantage with Intact Auditory Stimuli », dans DODD B. & CAMPBELL R. (Eds.), *Hearing by eye: The psychology of lip-reading*, Lawrence Erlbaum Associates, Hillsdale (USA), p. 97-114.

RILLIARD Albert & AUBERGÉ Véronique, 2001. « Prosody evaluation as a diagnostic process: subjective vs. objective measurements », dans *SSW4-2001*, paper 140.

RISBERG Arne & AGELFORS Eva, 1978. « On the identification of intonation contours by hearing impaired listeners », *Speech Transmission Laboratory - Quarterly Progress Report and Status Repor*t, 19(2-3), p. 51-61.

RISBERG Arne & LUBKER J., 1978. « Prosody and speechreading », *Speech Transmission Laboratory - Quarterly Progress Report and Status Repor*t, 19(4), p. 1-16.

ROBERT-RIBES Jordi, 1995. *Modèles d'intégration audiovisuelle de signaux linguistiques : de la perception humaine à la reconnaissance automatique des voyelles*, thèse de Doctorat, Institut National Polytechnique de Grenoble.

ROLLAND Guillaume & LŒVENBRUCK Hélène, 2002. « Characteristics of the Accentual Phrase in French: an Acoustic, Articulatory and Perceptual Study », dans les *Actes de la conférence Speech Prosody 2002, Aix-en-Provence, France*, p. 611-614.

ROMARY Laurent & PIERREL Jean-Marie, 1989. « The use of the Dempster-Shafer rule in the lexical component of a man-machine oral dialogue system », *Speech Communication,* 8(2), p.159-176.

ROSENBLUM Lawrence D. & SALDAÑA Helena M., 1996. « An audiovisual test of kinematic primitives for visual speech perception », *Journal of Experimental Psychology: Human Perception and Performance*, 22(2), p. 318-331.

ROSS E.D., 1981. « The aprosodias: Functional-anatomic organization of the affective components of language in the right hemisphere », *Archives of Neurology*, 38(9), p. 561–569.

ROSSI Mario, 1985. « L'intonation et l'organisation de l'énoncé », *Phonetica,* 42, p. 135-153.

ROSSI Mario, 1999. « La focalisation », dans *L'intonation, le système du français: description et modélisation*, Ophrys, Chap. II-6, p. 116-128.

SCHWARTZ Jean-Luc, 2004. « La parole multisensorielle: Plaidoyer, problèmes, perspective », dans les *Actes des Journées d'Etude de la Parole 2004, Fès, Maroc,* p. xi-xviii.

SCHWARTZ Jean-Luc, ROBERT-RIBES Jordi & ESCUDIER Pierre, 1998. « Ten years after Summerfield ... a taxonomy of models for audiovisual fusion in speech perception » dans CAMPBELL R., DODD B. J. & BURNHAM D. (Eds.), *Hearing by Eye II: Advances in the Psychology of Speechreading and Auditory-visual Speech*, Psychology Press, Hove (GB), p. 85-108.

SCHWARTZ Jean-Luc, BERTHOMMIER Frédéric & SAVARIAUX Christophe, 2004. « Seeing to hear better: evidence for early audio-visual interactions in speech identification », *Cognition*, 93, p. B69-B78.

SEKIYAMA Kaoru & TOHKURA Yoh'ichi, 1993. « Inter-language differences in the influence of visual cues in speech perception », *Journal of Phonetics*, 21, p. 427–444.

SEKIYAMA Kaoru & BURNHAM Denis, 2004. « Issues in the development of auditory-visual speech perception: Adults, infants and children », dans les *Actes de la conférence ICSLP 2004, île de Jeju, Corée,* p. 1137-1140.

SELKIRK Elisabeth, 1984. *Phonology and Syntax: the relation between sound and structure*, MIT Press, Cambridge (USA).

SENNEMA Anke, HAZAN Valérie & FAULKNER Andrew, 2003. « The role of visual cues in L2 consonant perception », dans les *Actes de la conférence ICPhS 2003, Barcelone, Espagne*, p. 135-138.

SLAMA-CAZACU Tatiana, 1976. « Nonverbal components in message sequence : "Mixed syntax" », dans MCCORMACK W.C. & WURM S.A. (Eds.), *Language and man : Anthropological issues*, Mouton, The Hague, p. 217-227.

SNOW David & BALOG Heather L., 2002. « Do children produce the melody before the words? A review of developmental intonation research », *Lingua*, 112(12), p. 1025-1058.

SORIN Christel, 1991. « Synthèse de la parole à partir du texte : état des recherches et des applications », *Journées GRECO/PRC*, *Toulouse, France*.

STEIN Barry E. & MEREDITH M. Alex, 1993. *The Merging of the Senses*, MIT Press, Cambridge (USA).

STERN Daniel N., SPIEKER Susan & MACKAIN Kristine, 1982. « Intonation contours as signals in maternal speech to prelinguistic infants », *Developmental Psychology*, 18, p. 727-735.

STETSON Raymond H., 1905. « A motor theory of rhythm and discrete succession », *Psychological Review*, 12, p. 293-350.

STONE Maureen, 1981. « Evidence for a rhythm pattern in speech production: observations of jaw movement », *Journal of Phonetics*, 9, p. 109-120.

SUMBY W.H. & POLLACK Irwin, 1954. « Visual contribution to Speech Intelligibility in Noise », *The Journal of the Acoustical Society of America*, 26(2), p. 212-215.

SUMMERFIELD Arthur Quentin, 1979. « Use of visual information for phonetic perception », *Phonetica*, 36, p. 314-331.

SUMMERFIELD Arthur Quentin, 1987. « Some preliminaries to a comprehensive account of audio-visual speech perception », dans DODD B. & CAMPBELL R. (Eds.), *Hearing by eye: the psychology of lipreading*, Lawrence Erlbaum Associates, Londres (GB), p. 3-51.

SUMMERS W. VAN, 1987. « Effects of stress and final-consonant voicing on vowel production: Articulatory and acoustic analyses », *The Journal of the Acoustical Society of America*, 82(3), p. 847-863.

SWERTS Marc, KRAHMER Emiel & AVESANI Cinzia, 2002. « Prosodic marking of information status in Dutch and Italian: a comparative analysis », *Journal of Phonetics*, 30, p. 629-654.

SWERTS Marc & KRAHMER Emiel, 2004. « Congruent and Incongruent Audiovisual Cues to Prominence », dans les *Actes de la conférence Speech Prosody 2004, Nara, Japon,* p. 69-72.

SWERTS Marc & KRAHMER Emiel, 2005. « Cognitive processing of audiovisual cues to prominence », dans les *Actes de la conférence AVSP 2005, île de Vancouver, Canada,* p. 29-30.

T'HART Johan, COLLIER René & COHEN Antonie, 1990. *A perceptual study of intonation*, Cambrige University Press, Cambridge.

THOMPSON Dorothy M., 1934. « On the detection of emphasis in spoken sentences by means of visual, tactual, and visual-tactual cues », *Journal of General Psychology*, 11, p. 160-172.

THORSEN Nina, 1980. « A study of the perception of sentence intonation – evidence from Danish », *The Journal of the Acoustical Society of America*, 67(3), p. 1014-1030.

TOMASELLO Michael, 2003. *Constructing a Language. A Usage-Based Theory of Language Acquisition*, Harvard University Press, Cambridge (USA), Londres (GB).

TOUATI Paul, 1987. « Structures prosodiques du suédois et du français », *Lund Working Papers, 21*, Lund University Press.

TOUATI Paul, 1989. « De la prosodie française du dialogue. Rapport du projet KIPROS », *Working Papers, Lund University*, 35, p. 203-214.

TOURATIER Christian, 2000. *La sémantique*, Armand Collin, Paris (France).

TWIST D., SQUIRES N., SPIELHOLZ Neil I. & SILVERGLIDE R., 1991. « Event-related potentials in disorders of prosodic and semantic linguistic processing », *Neuropsychiatry, Neurosurgery, and Behavioral Neurology*, 4, p. 281–304.

TZOURIO N., ELI MASSIOUI F., CRIVELLO F., JOLIOT M., RENAULT B., & MAZOYER B., 1997. « Functional anatomy of human auditory attention studied with PET », *NeuroImage*, 5, p. 63–77.

VAISSIÈRE Jacqueline, 1971. *Contribution à la synthèse par règle du français*, thèse de doctorat, université de Grenoble.

VAISSIÈRE Jacqueline, 1997. « Langues, prosodies et syntaxe », *A.T.A.L.A.*, 38(1), p. 53-82.

VAN WASSENHOVE Virginie, GRANT Ken W. & POEPPEL David, 2005. « Visual speech speeds up the neural processing of auditory speech », *Proceedings of the National Academy of Sciences of the United States of America*, 102, p. 1181-1186.

VATIKIOTIS-BATESON Eric & KELSO J. A. Scott, 1993. « Rhythm type and articulatory dynamics in English, French and Japanese », *Journal of Phonetics*, 21, p. 231-265.

VATIKIOTIS-BATESON Eric, EIGSTI Inge-Marie, YANO Sumio & MUNHALL Kevin G., 1998. « Eye Movement of Perceivers During Audiovisual Speech Perception », *Perception & Psychophysics*, 60(6), p. 926-940.

VIHMAN Marilyn May, 1996. *Phonological Development – The Origins of Language in the Child*, Blackwell Publishers.

VOLTERRA Virginia & CASELLI Cristina M., 1986. « First stages of language acquisition through two modalities in deaf and hearing children », *Journal of Neurological Sciences*, 5, p. 109-115.

VROOMEN Jean H.M., 1992. *Hearing voices and seeing lips: Investigations in the psychology of lipreading*, thèse de Doctorat, Katolieke Universiteit, Brabant.

WATKINS Kate E., STRAFELLA Antonio P. & PAUS Tomáš, 2003. « Seeing and hearing speech excites the motor system involved in speech production », *Neuropsychologia*, 41(8), p. 989-994.

WEINTRAUB Sandra, MESULAM Marek-Marsel & KRAMER L., 1981. « Disturbances in Prosody: A Right-hemisphere Contribution to Language », *Archives of Neurology*, 38(12), p. 742-744.

WELBY Pauline, 2002. « The realization of early and late rises in French intonation: A production study », dans les *Actes de la conférence Speech Prosody 2002, Aix-en-Provence, France,* p. 695-698.

WELBY Pauline, 2003. *The slaying of Lady Mondegreen, being a study of French tonal association and alignment and their role in speech segmentation*, thèse de doctorat, Ohio State University.

WESTBURY John R. & FUJIMURA Osamu, 1989. « An articulatory characterization of contrastive emphasis in correcting answers », *The Journal of the Acoustical Society of America*, 85(Suppl. 1), S98.

YEHIA Hani C., RUBIN Philip E. & VATIKIOTIS-BATESON Eric, 1998. « Quantitative association of vocal-tract and facial behavior », *Speech Communication*, 26(1-2), p. 23-43.

ZATORRE Robert J., EVANS Allan C., MEYER Ernst & GJEDDE Albert, 1992. « Lateralization of phonetic and pitch discrimination in speech processing », *Science*, p. 256, 846–849.

– Annexes –

A. Annexe 1 - Corpus AV1

(1) [Jean]$_{S1}$ [veut ménager]$_{V4}$ [nos jolis nouveaux navets]$_{O7}$.

(2) [Romain]$_{S2}$ [ranima]$_{V3}$ [la jolie maman]$_{O5}$.

(3) [Mélanie]$_{S3}$ [vit]$_{V1}$ [les mauvais loups malheureux]$_{O7}$.

(4) [Véroniqua]$_{S4}$ [mangeait]$_{V2}$ [les mauvais melons]$_{O5}$.

(5) [Les mauvais loups]$_{S4}$ [mangeront]$_{V3}$ [Jean]$_{O1}$.

(6) [Mon mari]$_{S3}$ [veut ranimer]$_{V4}$ [Romain]$_{O2}$.

(7) [Les loups]$_{S2}$ [suivaient]$_{V2}$ [Marilou]$_{O3}$.

(8) [Le beau marin]$_{S4}$ [vit]$_{V1}$ [Véroniqua]$_{O4}$.

B. Annexe 2 - Corpus AV2

(1) [Romain]$_{S2}$ [ranima]$_{V3}$ [la jolie maman]$_{O5}$.

(2) [Véroniqua]$_{S4}$ [mangeait]$_{V2}$ [les mauvais melons]$_{O5}$.

(3) [Mon mari]$_{S3}$ [veut ranimer]$_{V4}$ [Romain]$_{O2}$.

(4) [Les loups]$_{S2}$ [suivaient]$_{V2}$ [Marilou]$_{O3}$.

(5) [La nounou]$_{S3}$ [mariera]$_{V3}$ [Li]$_{O1}$.

(6) [Le lama lent]$_{S4}$ [lut]$_{V1}$ [Marinella]$_{O4}$.

(7) [Marinella]$_{S4}$ [va laminer]$_{V4}$ [Numu]$_{O2}$.

(8) [Lou]$_{S1}$ [mima]$_{V2}$ [la lama]$_{O3}$.

(9) [Le nominé]$_{S4}$ [lut]$_{V1}$ [les longs mots]$_{O3}$.

(10) [La nounou]$_{S3}$ [vit]$_{V1}$ [Lou]$_{O1}$.

(11) [Les loups]$_{S2}$ [mimaient]$_{V2}$ [Marilou]$_{O3}$.

(12) [Lou]$_{S1}$ [ramena]$_{V3}$ [Manu]$_{O2}$.

(13) [Li]$_{S1}$ [ralluma]$_{V3}$ [les moulinets]$_{O4}$.

C. Annexe 3 - Données Optotrak : analyse détaillée des résultats pour chaque paramètre articulatoire

Une analyse détaillée des résultats pour chaque paramètre sera d'abord effectuée. Compte-tenue de la quantité de données, cette analyse sera inévitablement assez longue. Dans un but de clarification et de lecture aisée, elle sera ainsi rédigée entièrement à chaque fois pour le locuteur B et rédigée de façon synthétique pour les autres locuteurs. Afin de clarifier les choses, les résultats seront présentées rigoureusement de la même façon pour chaque paramètre et notamment au niveau des graphiques et tableaux. Des comparaisons aisées et pratiques pourront ainsi être faites. Un bilan et une tentative de généralisation seront ensuite décrits.

C.1.1.1. Analyse de la durée

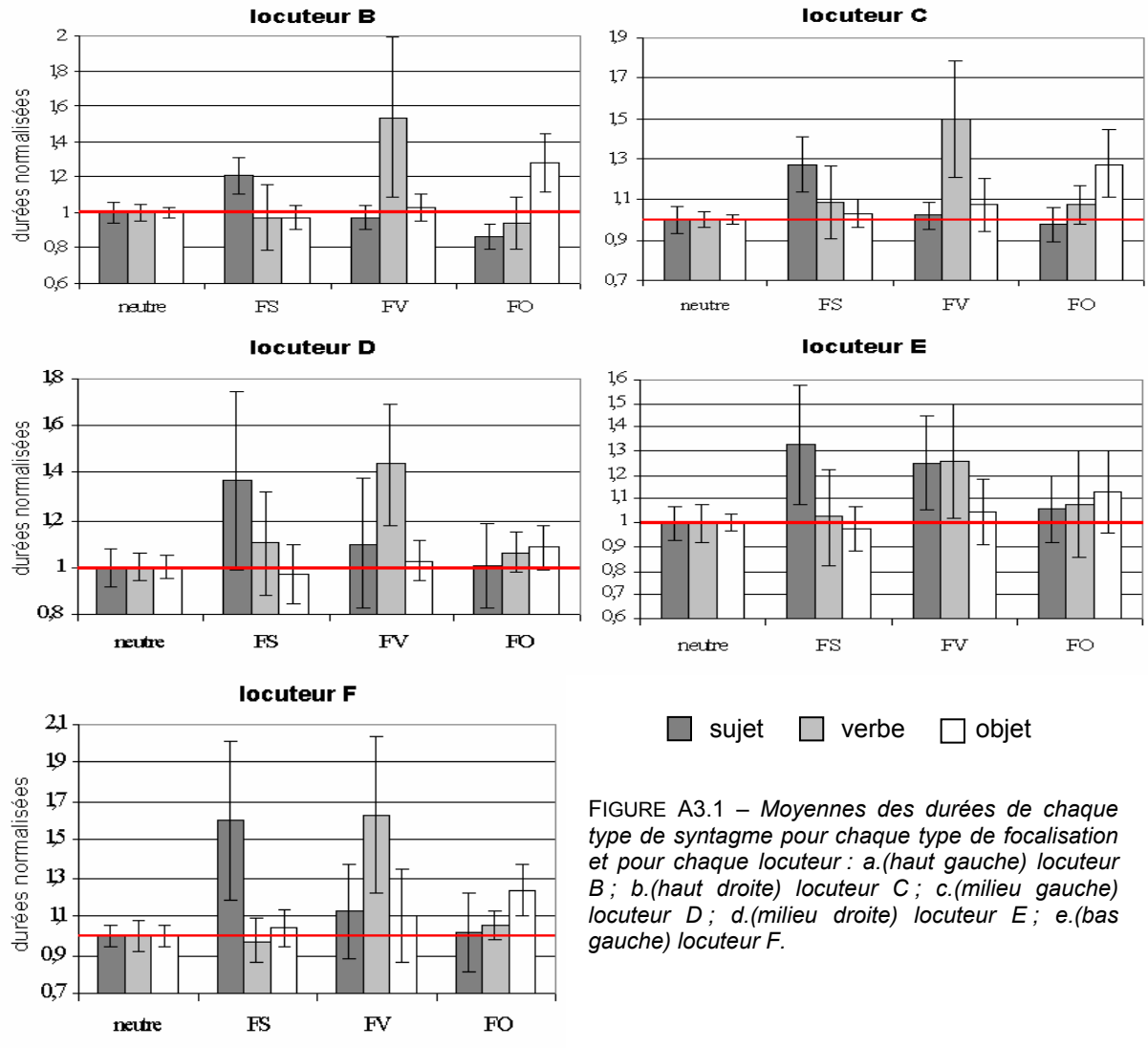


FIGURE A3.1 – *Moyennes des durées de chaque type de syntagme pour chaque type de focalisation et pour chaque locuteur : a.(haut gauche) locuteur B ; b.(haut droite) locuteur C ; c.(milieu gauche) locuteur D ; d.(milieu droite) locuteur E ; e.(bas gauche) locuteur F.*

locuteur	B	C	D	E	F
effet congruence	F(1,25)=198,734 p<0,001	F(1,25)=323,851 p<0,001	F(1,25)=109,381 p<0,001	F(1,25)=50,017 p<0,001	F(1,25)=239,567 p<0,001
effet type de focalisation	F(1,315,50)=15,553 p<0,001	F(2,50)=13,895 p<0,001	F(2,50)=15,465 p<0,001	F(2,50)=5,812 p=0,005	F(1,126,50)=10,672 p=0,001
interaction	F(1,304,50)=5,621 p=0,006	F(2,50)=6,636 p=0,003	F(2,50)=10,232 p<0,001	F(2,50)=7,301 p=0,002	F(1,527,50)= 7,923 p=0,003
test t congruence	t=9,792 p<0,001	t=13,937 p<0,001	t=9,547 p<0,001	t=9,017 p<0,001	t=11,706 p<0,001
test t post-foc	t=-0,842 p=0,402	t=4,311 p<0,001	t=2,171 p=0,033	t=0,786 p=0,434	t=2,077 p=0,041
test t pré-foc	t=-2,810 p=0,007	t=4,026 p<0,001	t=3,155 p=0,003	t=5,296 p<0,001	t=3,76 p<0,001

TABLE A3.1 – *Résultats des tests statistiques menés sur les données de durées selon la méthode décrite à la section A.2.1.2.d du chapitre III et pour chaque locuteur.*

C.1.1.1.a. Locuteur B

Contrastes intra-énoncés

Le graphique a. de la figure A3.1 montre clairement qu'il y a un contraste entre ce qui est focalisé et ce qui ne l'est pas au sein d'un même énoncé. Ce contraste moyen est de 38,7%. L'analyse statistique (facteur congruence[139] de l'ANOVA[140]) permet de montrer que ce contraste intra-énoncé est bien significatif : F(1,25)=198,734 (p<0,001).

L'ANOVA[141] permet également de mettre en relief un effet significatif du type de focalisation : F(1,315,50)=15,553 (p<0,001). Ceci est dû au fait que de manière générale, lorsqu'il y a focalisation sur le verbe, la durée moyenne de toutes les syllabes de l'énoncé focalisé sur le verbe est plus importante que pour les autres types de focalisation.

L'effet d'interaction[142] est lui aussi significatif : F(1,304,50)=5,621 (p=0,006). Car lorsque la focalisation porte sur le verbe, le contraste entre ce qui est focalisé dans l'énoncé et ce qui ne l'est pas est plus important que pour les autres types de focalisation.

Comparaison des énoncés dans les cas neutre et focalisé

Allongement focal - Le graphique a. de la figure A3.1 permet de voir que lorsqu'un syntagme est focalisé, la durée moyenne des syllabes qui le composent est, de façon claire, supérieure à sa valeur dans le cas neutre (barres nettement au-dessus de 1). Cette augmentation est en moyenne de 20,9% pour le sujet, de 53,7% pour le verbe et de 28,3% sur l'objet. On observe donc un net allongement

[139] Congruence : facteur intra-sujet à deux niveaux : cas congruents (S&FS, V&FV et O&FO) et cas incongruents (V&FS et O&FS, S&FV et O&FV et S&FO et V&FO).
[140] ANOVA à deux facteurs intra-sujets : congruence (deux niveaux) et type de focalisation (trois niveaux) ; pour une description complète et détaillée du test voir la section A.2.1.2.d du chapitre III.
[141] ANOVA à deux facteurs intra-sujets : congruence (deux niveaux) et type de focalisation (trois niveaux) ; pour une description complète et détaillée du test voir la section A.2.1.2.d du chapitre III.
[142] Interaction congruence×type de focalisation.

des syllabes focales chez ce locuteur et celle-ci est statistiquement significative[143] (cf. table A3.1 : t=9,792 p<0,001).

Invariance post-focale - Le graphique a. de la figure A3.1 permet aussi de constater qu'il n'y a pas de variation de la durée des syllabes post-focales. Ceci est confirmé par l'étude statistique qui montre que les durées des syllabes post-focales ne sont pas significativement supérieures à 1 (cf. table A3.1 : t=-0,842 p=0,402).

Allongement pré-focal - Pour ce locuteur, le sujet pré-focal (focalisation verbe) et le verbe pré-focal (focalisation objet) ne sont pas allongés et leur durée diminue même par rapport au cas neutre (-3,1% pour le sujet pré-focal et -6% pour le verbe pré-focal). Les barres S&FV et V&FO de la figure A3.1 sont ainsi en-dessous de 1. Ceci est confirmé par un test statistique qui montre que la durée des syllabes des éléments pré-focaux est significativement plus petite que 1 (cf. table A3.1 : t=-2,81 p=0,007). Pour ce locuteur, il n'y a donc pas d'allongement pré-focal et on observe même une diminution significative de la durée des syllabes pré-focales.

C.1.1.1.b. Locuteur C

Contrastes intra-énoncés (voir graphique b de la figure A3.1)

- contraste moyen : 30,5% (21,3% pour FS, 45% pour FV et 25,3% pour FO) significatif (F(1,25)=323,851 p<0,001);
- effet significatif du type de focalisation : F(2,50)=13,895 (p<0,001). Lorsqu'il y a focalisation sur le verbe, les durées moyennes de tout l'énoncé sont plus importantes.
- effet d'interaction significatif : F(2,50)=6,636 (p=0,003), contraste plus important lorsque la focalisation porte sur le verbe.

Comparaison des énoncés dans les cas neutre et focalisé (voir graphique b de la figure A3.1)

Allongement focal
- allongement moyen : 34,8% (27,1% pour S&FS, 49,6% pour V&FV, 27,6% pour O&FO) ;
- significatif : t=13,937 (p<0,001).

Invariance post-focale
- pas d'invariance post-focale ;
- allongement faible (3,7% pour FS et 2,9% pour FV) mais significatif : t=4,311 (p<0,001).

Allongement pré-focal
- allongement pré-focal : 4,7% (2% pour S&FV et 7,3% pour V&FO) ;
- significatif : t=4,026 (p<0,001) ;
- Quand le verbe est focalisé le sujet n'est pas allongé, c'est donc bien uniquement l'élément qui précède directement la focalisation qui est allongé.

C.1.1.1.c. Locuteur D

Contrastes intra-énoncés (voir graphique c de la figure A3.1)

- contraste moyen : 25,3% (33,6% pour FS, 37,4% pour FV et 5% pour FO) significatif (F(1,25)=109,381 p<0,001).

[143] Test t de comparaison à 1.

- effet significatif du type de focalisation : F(2,50)=15,465 (p<0,001). Lorsqu'il y a focalisation sur le verbe, les durées moyennes des syllabes de tout l'énoncé sont plus importantes.
- effet d'interaction significatif : F(2,50)=10,232 (p<0,001), contraste plus important pour les focalisations sur le sujet et sur le verbe.

Comparaison des énoncés dans les cas neutre et focalisé (voir graphique c. de la figure A3.1)

Allongement focal

- allongement moyen : 29,8% (37,2% pour S&FS, 43,7% pour V&FV, 8,4% pour O&FO) ;
- significatif : t=9,547 (p<0,001).

Invariance post-focale

- pas d'invariance post-focale ;
- allongement faible (3,7% pour FS et 2,9% pour FV) mais significatif : t=2,171 (p=0,033).

Allongement pré-focal

- allongement pré-focal : 8,1% (9,8% pour S&FV et 6,3% pour V&FO) ;
- significatif : t=3,155 (p=0,003) ;
- Cet allongement n'est visible que sur le verbe directement pré-focal lors de la focalisation sur l'objet donc il s'agit très probablement d'une anticipation.

C.1.1.1.d. Locuteur E

Contrastes intra-énoncés (voir graphique d. de la figure A3.1)

- contraste moyen : 16,8% (32,8% pour FS, 11,3% pour FV et 6,3% pour FO) significatif (F(1,25)=50,017 p<0,001).
- effet significatif du type de focalisation : F(2,50)=5,812 (p=0,005). Lorsqu'il y a focalisation sur l'objet les durées moyennes de toutes les syllabes de l'énoncé sont plus importantes.
- effet d'interaction significatif : F(2,50)=7,301 (p=0,002), contraste plus important lorsque la focalisation porte sur le sujet.

Comparaison des énoncés dans les cas neutre et focalisé (voir graphique d. de la figure A3.1)

Allongement focal

- allongement moyen : 23,9% (32,6% pour S&FS, 25,9% pour V&FV, 13,1% pour O&FO) ;
- significatif : t=9,017 (p<0,001).

Invariance post-focale

- invariance post-focale (t=0,786 p=0,434).

Allongement pré-focal

- allongement pré-focal : 16,4% (25% pour S&FV et 7,7% pour V&FO) ;
- significatif : t=5,296 (p<0,001) ;
- On note également un allongement du sujet dans le cas de la focalisation objet de 6%. Il n'est donc pas certain que l'allongement pré-focal soit ici la conséquence de la mise en place d'une stratégie d'anticipation. Cet allongement général pourrait correspondre à un effort de la part du locuteur pour rendre l'énoncé tout entier plus compréhensible par l'interlocuteur.

C.1.1.1.e. Locuteur F

Contrastes intra-énoncés (voir graphique e. de la figure A3.1)

- contraste moyen : 43,8% (59,4% pour FS, 51,5% pour FV et 20,6% pour FO) significatif (F(1,25)=239,567 p<0,001) ;
- effet significatif du type de focalisation : F(1,126,50)=10,672 (p=0,001). Lorsqu'il y a focalisation sur l'objet les durées moyennes sont plus importantes.
- effet d'interaction congruence×type de focalisation significatif : F(1,527,50)=7,923 (p=0,003), contraste plus important lorsque la focalisation porte sur le sujet et sur le verbe

Comparaison des énoncés dans les cas neutre et focalisé (voir graphique e. de la figure A3.1)

Allongement focal

- allongement moyen : 49% (60,2% pour S&FS, 63% pour V&FV, 23,9% pour O&FO) ;
- significatif : t=11,706 (p<0,001).

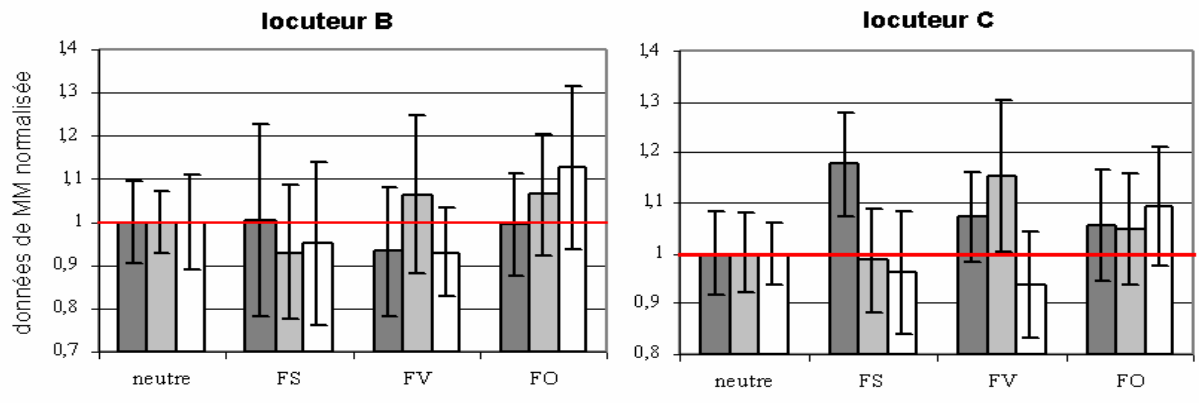
Invariance post-focale

- pas d'invariance post-focale globale ;
- allongement post-focal (FS : 0,9% et FV : 10,2%) significatif : t=2,077 (p=0,041).
- On peut conclure qu'il y a bien invariance pour FS. C'est parce qu'il y a un fort allongement après FV qu'on obtient un allongement significatif.

Allongement pré-focal

- allongement pré-focal : 9% (12,9% pour S&FV et 5,1% pour V&FO) ;
- significatif : t=3,76 (p<0,001).
- On note que la durée moyenne des syllabes du sujet dans le cas de focalisation sur le verbe ne varie quasiment pas (+1,5%). Ceci nous permet de penser qu'il s'agit ici d'une stratégie d'anticipation de la focalisation puisqu'elle n'est mise en place que sur l'élément précédant directement la focalisation.

C.1.1.2. Analyse des mouvements de la mandibule



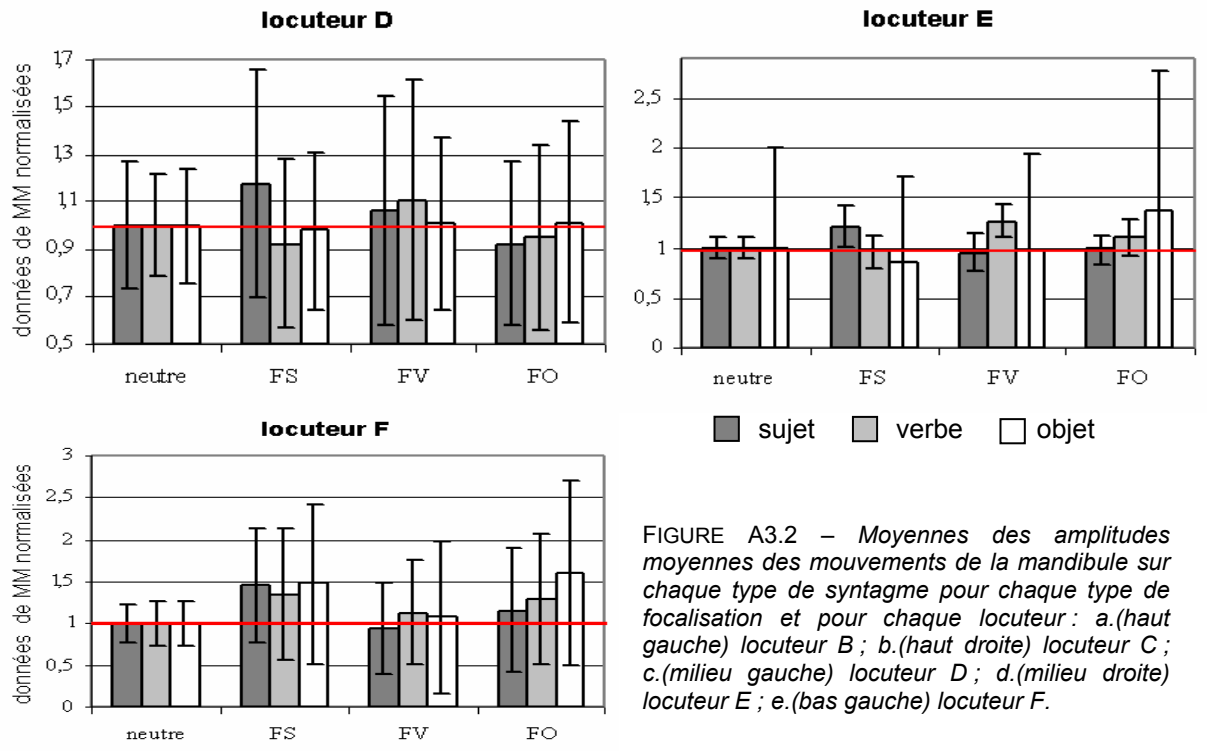


FIGURE A3.2 – *Moyennes des amplitudes moyennes des mouvements de la mandibule sur chaque type de syntagme pour chaque type de focalisation et pour chaque locuteur : a.(haut gauche) locuteur B ; b.(haut droite) locuteur C ; c.(milieu gauche) locuteur D ; d.(milieu droite) locuteur E ; e.(bas gauche) locuteur F.*

locuteur	B	C	D	E	F
effet congruence	F(1,25)=29,189 p<0,001	F(1,25)=309,813 p<0,001	F(1,25)=27,429 p<0,001	F(1,25)=188,256 p<0,001	F(1,25)=30,499 p<0,001
effet type de focalisation	F(2,50)=5,77 p=0,006	F(1,719,50)=0,061 p=0,941	F(2,50)=0,873 p=0,424	F(2,50)=7,201 p=0,002	F(2,50)=3,105 p=0,054
interaction	F(2,50)=0,644 p=0,529	-	-	F(2,50)=0.350 p=0.706	-
test t congruence	t=3,011 p=0,004	t=10,285 p<0,001	t=2,147 p=0,035	t=11,547 p<0,001	t=4,261 p<0,001
test t post-foc	t=-3,731 p<0,001	t=-3,204 p=0,002	t=-0,881 p=0,381	t=-3,846 p<0,001	t=2,822 p=0,006
test t pré-foc	t=-0,095 p=0,925	t=4,570 p<0,001	t=0,117 p=0,907	t=1,026 p=0,31	t=1,212 p=0,232

TABLE A3.2 – *Résultats des tests statistiques menées sur les données de mouvements de la mâchoire selon la méthode décrite à la section A.2.1.2.d du chapitre III pour chaque locuteur.*

C.1.1.2.a. Locuteur B

Contrastes intra-énoncés

Le graphique a. de la figure A3.2 permet de constater qu'il existe un contraste entre l'amplitude moyenne des mouvements de la mandibule sur l'élément focalisé par rapport à cette même amplitude sur le reste de l'énoncé. Ce contraste moyen est de 9,9% (6,5% pour FS, 13,4% pour FV et 9,8% pour

FO). L'analyse statistique (facteur congruence[144] de l'ANOVA[145], cf. table A3.2) montre que ce contraste est nettement significatif : F(1,25)=29,189 (p<0,001).

L'ANOVA permet également de mettre en valeur un effet significatif du type de focalisation[146] : F(2,50)=309,813 (p=0,006). En effet, chez ce locuteur l'amplitude des mouvements mandibulaire est toujours plus importante lorsque la focalisation porte sur l'objet aussi bien sur l'élément focalisé que sur les autres éléments de l'énoncé.

Enfin bien que l'effet d'interaction[147] ne soit pas significatif : F(2,50)=0,664 (p=0,529). Il apparaît que c'est pour le cas de focalisation sur le verbe que le contraste entre ce qui est focalisé et ce qui ne l'est pas dans un énoncé est le plus important.

Comparaison des énoncés dans le cas neutre et focalisé

Hyper-articulation focale - Le graphique a. de la figure A3.2 permet de voir que pour les cas de focalisation sur le verbe et sur l'objet, l'amplitude moyenne des mouvements mandibulaires de l'élément focalisé est plus importante que dans le cas neutre (barres au-dessus de 1). L'augmentation moyenne par rapport au cas neutre est de 6,4% pour FV et de 12,8% pour FO. Pour le cas de la focalisation sur le sujet, cette différence est assez faible (0,6%). On observe globalement chez ce locuteur une augmentation de l'amplitude moyenne des mouvements mandibulaires par rapport au cas neutre lors de la focalisation et celle-ci est statistiquement significative : t=3,011 (p=0,004).

Hypo-articulation post-focale - Le graphique a. de la figure A3.2 permet de constater qu'il y a une baisse de l'amplitude des mouvements mandibulaires après focalisation. Cette baisse est de 5,9% après le sujet focalisé et de 7% après le verbe focalisé. On observe donc chez ce locuteur une hypo-articulation post-focale significative : t=-3,731 (p<0,001).

Hyper-articulation pré-focale - De façon générale, on n'observe pas chez ce locuteur d'hyper-articulation pré-focale significative : t=-0,095 (p=0,925). Le graphique a. de la figure A3.2 permet d'ailleurs de constater qu'il y a même, chez ce locuteur, une baisse assez nette de l'amplitude des mouvements mandibulaires sur le sujet pré-focal (6,9%) pour la focalisation sur le verbe. Par contre, il y a bien hyper-articulation du verbe pré-focal (focalisation sur l'objet) d'en moyenne 6,5%. Statistiquement cette hyper-articulation ne ressort pas à cause de l'hypo-articulation lors de la focalisation sur le verbe. De plus, lors de la focalisation sur l'objet,on n'observe pas d'hyper-articulation du sujet pré-focal (focalisation sur l'objet). Par conséquent, on peut penser que lorsque la focalisation porte sur l'objet, ce locuteur met en place une stratégie d'anticipation de la focalisation.

C.1.1.2.b. Locuteur C

Contrastes intra-énoncés (cf graphique b. de la figure A3.2)

- contraste moyen : 14,9% (23% pour FS, 14,8% pour FV et 4% pour FO) significatif (F(1,25)=309,813 p<0,001) ;
- effet non significatif du type de focalisation : F(1,719,50)=0,061 (p=0,941) ;
- C'est pour FS que le contraste entre l'élément focalisé et le reste de l'énoncé est le plus important.

[144] Congruence : facteur intra-sujet à deux niveaux : cas congruents (S&FS, V&V et O&FO) et cas incongruents (V&FS et O&FS, S&FV et O&FV et S&FO et V&FO).
[145] ANOVA à deux facteurs intra-sujets : congruence (deux niveaux) et type de focalisation (trois niveaux) ; pour une explication complète et détaillée voir la section A.2.1.2.d du chapitre III.
[146] Type de focalisation : facteur intra-sujet à trois niveaux : FS, FV et FO.
[147] Interaction congruence×type de focalisation.

Comparaison des énoncés dans le cas neutre et focalisé (cf graphique b de la figure A3.2)

Hyper-articulation focale

- hyper-articulation significative : t=10,285 (p<0,001) ;
- +14% (+17,6% pour S&FS, +15,3% pour V&FV et +9,1% pour O&FO).

Hypo-articulation post-focale

- hypo-articulation post-focale significative : t=-3,204 (p=0,002) ;
- -2,6% pour FS et -6,1% pour FV.

Hyper-articulation pré-focale

- hyper-articulation pré-focale significative : t=4,57 (p<0,001) ;
- +7,2% pour S&FV et +4,8% pour V&FO ;
- +5,5% pour S&FO donc sans doute pas de stratégie d'anticipation.

C.1.1.2.c. Locuteur D

Contrastes intra-énoncés (cf graphique c. de la figure A3.2)

- contraste moyen : 12,5% (22,3% pour FS, 7,2% pour FV et 8% pour FO) significatif (F(1,25)=27,429 p<0,001) ;
- effet non significatif du type de focalisation : F(2,50)=0,873 (p=0,424) ;
- C'est pour FS que le contraste entre l'élément focalisé et le reste de l'énoncé est le plus important.

Comparaison des énoncés dans le cas neutre et focalisé (cf graphique c. de la figure A3.2)

Hyper-articulation focale

- hyper-articulation significative : t=2,147 (p=0,035) ;
- 9,9% (+17,4% pour S&FS, +10,8% pour V&FV et +1,5% pour O&FO) ;

Hypo-articulation post-focale

- pas d'hypo-articulation post-focale significative (t=-0,881 p=0,381).

Hyper-articulation pré-focale

- pas d'hyper-articulation pré-focale significative (t=0,117 p=0,907).

C.1.1.2.d. Locuteur E

Contrastes intra-énoncés (cf graphique d. de la figure A3.2)

- contraste moyen : 31,5% (29,8% pour FS, 30,8% pour FV et 33,8% pour FO) significatif (F(1,25)=188,256 p<0,001) ;
- effet significatif du type de focalisation : F(2,50)=7,201 (p=0,002) ; Lorsqu'il y a focalisation sur l'objet, l'amplitude des mouvements mandibulaires est plus importante que pour les autres types de focalisation aussi bien sur l'élément focalisé que sur les autres éléments de l'énoncé ;
- effet d'interaction non significatif (F(2,50)=0,350 p=0,706), on constate que c'est pour FO que le contraste entre l'élément focalisé et le reste de l'énoncé est le plus important.

Comparaison des énoncés dans le cas neutre et focalisé (cf graphique d. de la figure A3.2)

Hyper-articulation focale

- hyper-articulation significative : t=11,547 (p<0,001) ;
- 28,5% (+20,6% pour S&FS, +26,8% pour V&FV et +38% pour O&FO) ;

Hypo-articulation post-focale

- hypo-articulation post-focale significative : t=-3,846 (p<0,001) ;
- -9,2% pour FS et -3,2% pour FV.

Hyper-articulation pré-focale

- pas d'hyper-articulation pré-focale significative (t=1,026 p=0,31).

C.1.1.2.e. Locuteur F

Contrastes intra-énoncés (cf graphique e. de la figure A3.2)

- contraste moyen : 18,5% (5,3% pour FS, 12,7% pour FV et 37,5% pour FO) significatif (F(1,25)=30,499 p<0,001) ;
- pas d'effet significatif du type de focalisation : F(2,50)=3,105 (p=0,054) ;
- On constate que c'est pour FO que le contraste entre l'élément focalisé et le reste de l'énoncé est le plus important.

Comparaison des énoncés dans le cas neutre et focalisé (cf graphique d de la figure A3.2)

Hyper-articulation focale

- hyper-articulation significative : t=4,261 (p<0,001) ;
- 39,5% (+45,8% pour S&FS, +13,1% pour V&FV et +59,6% pour O&FO).

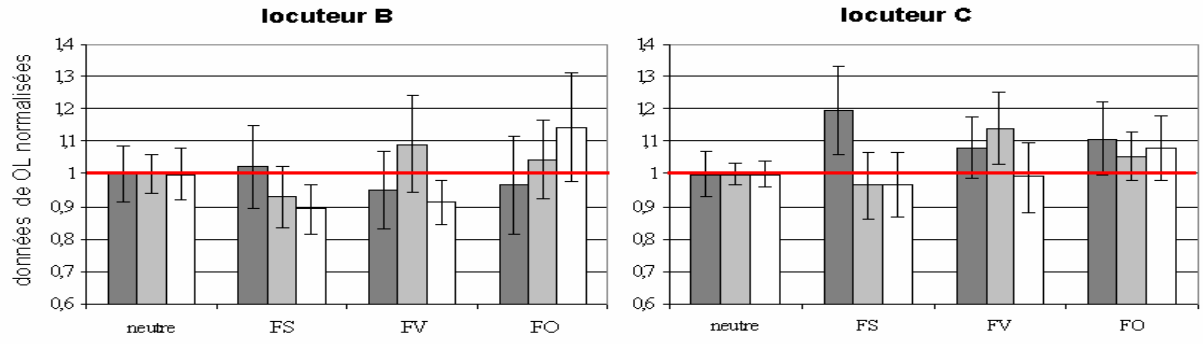
Hypo-articulation post-focale

- pas d'hypo-articulation post-focale ;
- MAIS hyper-articulation post-focale significative : t=2,822 (p=0,006) ;
- +40,6% pour FS et +7,1% pour FV.

Hyper-articulation pré-focale

- pas d'hyper-articulation pré-focale significative (t=1,212 p=0,232).

C.1.1.3. Analyse de l'ouverture des lèvres



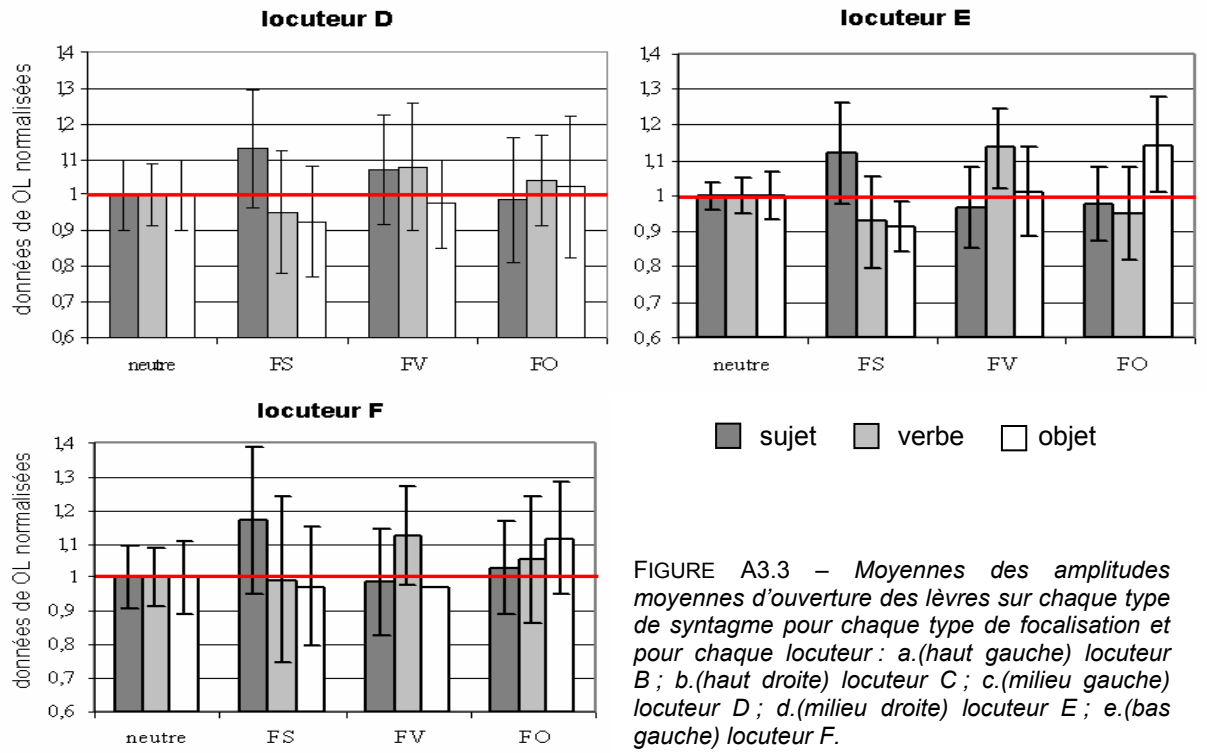


FIGURE A3.3 – *Moyennes des amplitudes moyennes d'ouverture des lèvres sur chaque type de syntagme pour chaque type de focalisation et pour chaque locuteur : a.(haut gauche) locuteur B ; b.(haut droite) locuteur C ; c.(milieu gauche) locuteur D ; d.(milieu droite) locuteur E ; e.(bas gauche) locuteur F.*

locuteur	B	C	D	E	F
effet congruence	F(1,25)=11,106 p<0,001	F(1,25)=149,141 p<0,001	F(1,25)=49,207 p<0,001	F(1,25)=111,371 p<0,001	F(1,25)=97,902 p<0,001
effet type de focalisation	F(2,50)=11,725 p<0,001	F(2,50)=0,129 p=0,880	F(2,50)=0,592 p=0,557	F(2,50)=2,071 p=0,137	F(2,50)=0,582 p=0,562
interaction	F(2,50)=0,907 p=0,41	-	-	-	-
test t congruence	t=5,278 p<0,001	t=10,036 p<0,001	t=4,234 p<0,001	t=9,131 p<0,001	t=7,261 p<0,001
test t post-foc	t=-10,005 p<0,001	t=-2,371 p=0,02	t=-3,453 p=0,001	t=-3,665 p<0,001	t=-0,976 p=0,332
test t pré-foc	t=-0,139 p=0,89	t=6,012 p<0,001	t=3,286 p=0,002	t=-2,443 p=0,018	t=0,974 p=0,335

TABLE A3.3 – *Résultats des tests statistiques menées sur les données d'ouverture des lèvres selon la méthode décrite dans la section A.2.1.2.d du chapitre III pour chaque locuteur.*

C.1.1.3.a. Locuteur B

Contrastes intra-énoncés

Le graphique a. de la figure A3.3 montre clairement qu'il existe un contraste entre l'ouverture des lèvres de ce qui est focalisé et celle de ce qui ne l'est pas au sein d'un même énoncé. Ce contraste moyen est de 13,8% (11,3% pour FS, 16,2% pour FV et 13,8% pour FO). L'analyse statistique (effet

congruence[148] de l'ANOVA[149], cf. table A3.3) permet de montrer que ce contraste intra-énoncé est bien significatif : F(1,25)=11,106 (p<0,001).

L'ANOVA permet également de mettre en relief un effet significatif du type de focalisation[150] : F(2,50)=11,725 (p<0,001). Ceci est dû au fait que de manière générale, lorsqu'il y a focalisation sur l'objet, l'amplitude moyenne de l'ouverture des lèvres est plus importante sur tout l'énoncé.

L'effet d'interaction[151] est faiblement significatif (F(2,50)=0,907 p=0,41), il apparaît en effet que c'est dans le cas où la focalisation porte sur verbe que le contraste entre ce qui est focalisé et ce qui ne l'est pas dans un énoncé est le plus important.

Comparaison des énoncés dans les cas neutre et focalisé (voir graphique a. de la figure A3.3)

Hyper-articulation focale - Le graphique a. de la figure A3.3 permet de voir que lorsqu'un syntagme est focalisé, l'amplitude moyenne de l'ouverture des lèvres lui correspondant est supérieure de façon claire par rapport à sa valeur dans le cas neutre (barres nettement au-dessus de 1). Cette augmentation est en moyenne de 8,7% (2,3% pour S&FS, 9,4% pour V&FV et 14,4% pour O&FO). On observe donc bien une hyper-articulation focale chez ce locuteur et bien que celle-ci soit assez faible pour le sujet, elle est significative : t=5,278 (p<0,001).

Hypo-articulation post-focale - Le graphique a. de la figure A3.3 permet également de constater qu'il y a une baisse de l'amplitude de l'ouverture des lèvres assez nette après la focalisation (barres inférieures à 1). Cette baisse est de 9% en moyenne après focalisation du sujet et de 8,7% en moyenne après focalisation du verbe. On a donc bien une hypo-articulation post-focale significative chez ce locuteur : t=-10,005 (p<0,001).

Hyper-articulation pré-focale - Le graphique a. de la figure A3.3 permet de constater que lorsque le verbe est focalisé, le sujet n'est pas du tout hyper-articulé du point de vue de l'ouverture des lèvres puisqu'on observe même une baisse de l'amplitude moyenne de l'ouverture des lèvres de 4,9%. Par contre, le verbe pré-focal est hyper-articulé puisque son amplitude augmente de 4,5% en moyenne. Cependant cette hyper-articulation n'est globalement pas significative (t=-0,139 p=0,89). Ce locuteur ne semble donc pas hyper-articuler avant de focaliser.

C.1.1.3.b. Locuteur C

Contrastes intra-énoncés (cf graphique b. de la figure A3.3)

- ce contraste existe bien pour FS et FV, par contre pour FO on constate qu'il existe bien un contraste entre le verbe et l'objet mais l'amplitude du sujet est quant à elle plus importante que celle de l'objet focalisé.
- contraste moyen : 16,8% (23% pour FS, 10,5% pour FV et -0,2% pour FO) significatif (F(1,25)=149,141 p<0,001) ;
- pas d'effet significatif du type de focalisation (F(2,50)=0,129 p=0,88) ;
- On constate que c'est pour FS que le contraste entre ce qui est focalisé et ce qui ne l'est pas dans un énoncé est le plus important.

[148] Congruence : facteur intra-sujet à deux niveaux : cas congruents (S&FS, V&FV et O&FO) et cas incongruents (V&FS et O&FS, S&FV et O&FV et S&FO et V&FO).
[149] ANOVA à deux facteurs intra-sujets : congruence (deux niveaux) et type de focalisation (trois niveaux) ; pour une explication complète et détaillée du test voir la section A.2.1.2.d du chapitre III.
[150] Type de focalisation : facteur intra-sujet à trois niveaux : FS, FV et FO.
[151] Interaction congruence×type de focalisation.

Comparaison des énoncés dans les cas neutre et focalisé (voir graphique b. de la figure A3.3)

Hyper-articulation focale

- hyper-articulation significative : t=10,036 (p<0,001) ;
- +13,9% (+19,4% pour S&FS, +14,1% pour V&FV et +8,1% pour O&FO).

Hypo-articulation post-focale

- hypo-articulation post-focale légère mais significative : t=-2,371 (p=0,02) ;
- -3,5% pour FS et -0,9% pour FV.

Hyper-articulation pré-focale

- hyper-articulation pré-focale significative : t=6,012 (p<0,001) ;
- +8% pour S&FV et +4,5% pour V&FO ;
- +10.9% pour S&FO donc que chez ce locuteur, l'hyper-articulation pré-focale ne relève pas forcément d'une stratégie d'anticipation locale, mais plutôt d'une production plus soignée de l'énoncé tout entier.

C.1.1.3.c. Locuteur D

Contrastes intra-énoncés (cf graphique c. de la figure A3.3)

- ce contraste existe bien pour FS et FV, par contre pour FO on constate qu'il existe bien un contraste entre le sujet et l'objet mais l'amplitude du verbe pré-focal est quant à elle plus importante que celle de l'objet focalisé.
- contraste moyen : 5,1% (19,1% pour FS, 5,4% pour FV et 0,8% pour FO) significatif (F(1,25)=49,207 p<0,001) ;
- pas d'effet significatif du type de focalisation (F(2,50)=0,592 p=0,557) ;
- On constate que c'est lorsque la focalisation porte sur le sujet que le contraste entre ce qui est focalisé et ce qui ne l'est pas dans un énoncé est le plus important.

Comparaison des énoncés dans les cas neutre et focalisé (voir graphique c. de la figure A3.3)

Hyper-articulation focale

- hyper-articulation significative : t=4,234 (p<0,001) ;
- +7,6% (+12,8% pour S&FS, +7,7% % pour V&FV et +2,2% % pour O&FO).

Hypo-articulation post-focale

- hypo-articulation post-focale significative : t=-3,453 (p=0,001) ;
- -6,3% % pour FS et -2,6% pour FV.

Hyper-articulation pré-focale

- hyper-articulation pré-focale significative : t=3,286 (p=0,002) ;
- +7,2% pour S&FV (or +7,7% seulement pour V&FV)
- +4,1% pour V&FO et seulement +2,2% pour O&FO donc l'hyper-articulation pré-focale est plus importante que l'hyper-articulation focale ;
- Cette hyper-articulation vient certainement de la mise en place d'une stratégie d'anticipation. En effet, elle n'est visible que sur le verbe directement pré-focal lors de la focalisation sur l'objet, par exemple.

C.1.1.3.d. Locuteur E

Contrastes intra-énoncés (cf graphique d. de la figure A3.3)

- contraste moyen : 17,4% (19,9% pour FS, 14,5% pour FV et 17,9% pour FO) significatif (F(1,25)=111,371 p<0,001) ;
- pas d'effet significatif du type de focalisation : F(2,50)=2,071 (p=0,137) ;
- On constate que c'est lorsque la focalisation porte sur le sujet que le contraste entre ce qui est focalisé et ce qui ne l'est pas dans un énoncé est le plus important.

Comparaison des énoncés dans les cas neutre et focalisé (voir graphique d. de la figure A3.3)

Hyper-articulation focale

- hyper-articulation significative : t=9,131 (p<0,001) ;
- +13,3% (+12% pour S&FS, +13,5% % pour V&FV et +14,4% % pour O&FO).

Hypo-articulation post-focale

- hypo-articulation post-focale globalement significative : t=-3,665 (p<0,001) ;
- -8% pour FS ;
- pas d'hypo-articulation post-focale sur FV : +1,2%.

Hyper-articulation pré-focale

- pas d'hyper-articulation pré-focale significative ;
- hypo-articulation pré-focale significative : t=-2,443 (p=0,018). ;
- -3,3% pour S&FV et –4,9% pour V&FO.

C.1.1.3.e. Locuteur F

Contrastes intra-énoncés (cf graphique e. de la figure A3.3)

- contraste moyen : 13,5% (18,6% pour FS, 14,4% pour FV et 7,4% pour FO) significatif (F(1,25)=97,902 p<0,001) ;
- pas d'effet significatif du type de focalisation : F(2,50)=0,582 (p=0,562) ;
- On constate que c'est lorsque la focalisation porte sur le sujet que le contraste entre ce qui est focalisé et ce qui ne l'est pas dans un énoncé est le plus important.

Comparaison des énoncés dans les cas neutre et focalisé (voir graphique e. de la figure A3.3)

Hyper-articulation focale

- hyper-articulation significative : t=7,261 (p<0,001) ;
- +13,6% (+17% pour S&FS, +12,3% pour V&FV et +11,5% pour O&FO).

Hypo-articulation post-focale

- hypo-articulation post-focale très légère et non significative : t=-0,976 (p=0,332) ;
- -1,7% pour FS et –3% pour FV.

Hyper-articulation pré-focale

- pas d'hyper-articulation pré-focale significative : t=0,974 (p=0,335) ;
- hyper-articulation pré-focale pour FO : +5,4%.

C.1.1.4. Analyse de l'étirement des lèvres

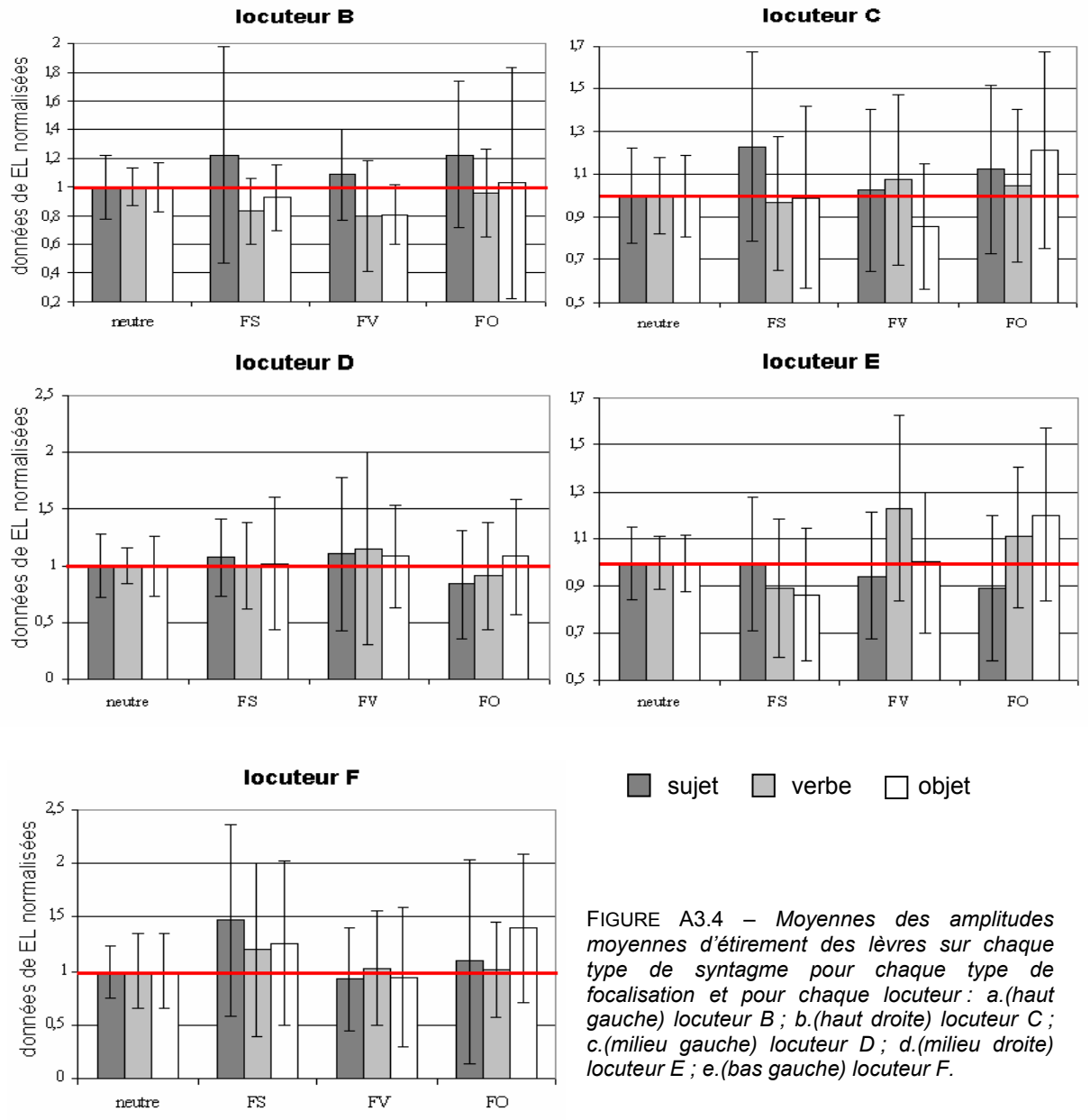


FIGURE A3.4 – *Moyennes des amplitudes moyennes d'étirement des lèvres sur chaque type de syntagme pour chaque type de focalisation et pour chaque locuteur : a.(haut gauche) locuteur B ; b.(haut droite) locuteur C ; c.(milieu gauche) locuteur D ; d.(milieu droite) locuteur E ; e.(bas gauche) locuteur F.*

locuteur	B	C	D	E	F
effet congruence	F(1,25)=0,558 p=0,462	F(1,25)=11,836 p=0,002	F(1,25)=3,774 p=0,063	F(1,25)=47,12 p<0,001	F(1,25)=11,581 p=0,002
effet type de focalisation	F(1,725,50)=2,158 p=0,134	F(2,50)=3,66 p=0,033	F(1,421,50)=1,240 p=0,298	F(2,50)=4,046 p=0,024	F(2,50)=3,687 p=0,033
interaction	-	F(2,50)=0,534 p=0,59	-	F(2,50)=1,424 P=0,250	F(2,50)=1,573 p=0,218
test t congruence	t=0,214 p=0,831	t=3,571 p=0,001	t=1,705 p=0,092	t=3,524 p=0,001	t=3,586 p=0,001
test t post-foc	t=-5,653 p<0,001	t=-1,532 p=0,130	t=0,783 p=0,436	t=-2,365 p=0,021	t=1,497 p=0,139
test t pré-foc	t=0,577 p=0,567	t=0,746 p=0,459	t=0,071 p=0,943	t=0,657 p=0,514	t=-0,549 p=0,585

TABLE A3.4 – *Résultats des tests statistiques menées sur les données d'étirement des lèvres selon la méthode décrite à la section A.2.1.2.d du chapitre III pour chaque locuteur.*

C.1.1.4.a. Locuteur B

Contrastes intra-énoncés

Le graphique a. de la figure A3.4 montre que les variations de l'étirement des lèvres sont assez inattendues chez ce locuteur. Il semblerait que ce soit toujours le syntagme sujet qui corresponde à un étirement des lèvres plus important, quelle que soit la condition de focalisation considérée (les barres S&FS, S&FV et S&FO sont les plus hautes). Les autres syntagmes ne ressortent jamais. D'ailleurs l'analyse statistique (effet congruence[152] de l'ANOVA[153], cf. table A3.4) montre que les contrastes intra-énoncés ne sont pas significatifs : F(1,25)=0,558 (p=0,462).

Il n'y a pas non plus d'effet significatif du type de focalisation[154] (F(1,725,50)=2,158 p=0,134).

Comparaison des énoncés dans les cas neutre et focalisé

Hyper-articulation focale

Le même constat que précédemment peut être fait ici, seul le syntagme sujet est hyper-articulé et ce pour toutes les conditions de focalisation.

Hypo-articulation post-focale

Le graphique a. de la figure A3.4 permet par contre de constater qu'il y a une nette baisse de l'amplitude de l'ouverture des lèvres après la focalisation (barres inférieures à 1). Cette baisse est de 12,1% en moyenne après la focalisation du sujet et de 19% en moyenne après focalisation du verbe. On a donc bien une hypo-articulation post-focale significative chez ce locuteur : t=-5,653 (p<0,001).

Hyper-articulation pré-focale

Le graphique a. de la figure A3.4 permet de constater qu'il n'y a pas d'hyper-articulation pré-focale.

[152] Congruence : facteur intra-sujet à deux niveaux : cas congruents (S&FS, V&FV et O&FO) et cas incongruents (V&FS et O&FS, S&FV et O&FV et S&FO et V&FO).
[153] ANOVA à deux facteurs intra-sujets : congruence (deux niveaux) et type de focalisation (trois niveaux) ; pour une explication complète et détaillée voir la section A.2.1.2.d du chapitre II.
[154] Type de focalisation : facteur intra-sujet à trois niveaux : FS, FV et FO.

C.1.1.4.b. Locuteur C

Contrastes intra-énoncés (voir graphique b. de la figure A3.4)

- contraste moyen : 17% (24,9% pour FS, 13,4% pour FV et 12,7% pour FO) significatif (F(1,25)=11,836 p=0,002) ;
- effet significatif du type de focalisation : F(2,50)=3,66 (p=0,033). On constate en effet que l'amplitude des mouvements d'étirement des lèvres est globalement plus grande sur tous les syntagmes lors des focalisations sujet et objet.
- Pas d'effet d'interaction (F(2,50)=0,534 p=0,59), mais on constate tout de même que c'est dans le cas de la focalisation sur le sujet que le contraste entre ce qui est focalisé et ce qui ne l'est pas dans un énoncé est le plus important.

Comparaison des énoncés dans les cas neutre et focalisé (voir graphique b. de la figure A3.4)

Hyper-articulation focale

- hyper-articulation significative : t=3,571 (p=0,001) ;
- +17,3% (+23,1% pour S&FS, +7,5% pour V&FV et +21,3% pour O&FO).

Hypo-articulation post-focale

- pas d'hypo-articulation post-focale significative : t=-1,532 (p=0,13) ;
- -1,8% pour FS et –14,4% pour FV.

Hyper-articulation pré-focale

- pas d'hyper-articulation pré-focale significative : t=0,746 (p=0,459) ;
- +2,7% pour S&FV et +4,7% pour V&FO.

C.1.1.4.c. Locuteur D

Contrastes intra-énoncés (voir graphique c. de la figure A3.4)

- contraste moyen : 11% (6% pour FS, 5,9% pour FV et 21,2% pour FO) mais non significatif (F(1,25)=3,774 p=0,063) ;
- pas d'effet significatif du type de focalisation (F(1,421,50)=1,24 p=0,298).
- On constate tout de même que c'est lorsque la focalisation porte sur l'objet que le contraste entre ce qui est focalisé et ce qui ne l'est pas dans un énoncé est le plus important.

Comparaison des énoncés dans les cas neutre et focalisé (voir graphique c. de la figure A3.4)

Hyper-articulation focale

- Pas d'hyper-articulation significative : t=1,705 (p=0,092) ;
- +10,2% (+7,1% pour S&FS, +15,3% pour V&FV et +8,2% pour O&FO).

Hypo-articulation post-focale

- pas d'hypo-articulation post-focale significative : t=0,783 (p=0,436) ;
- +1,1% pour FS et +8,6% pour FV.

Hyper-articulation pré-focale

- pas d'hyper-articulation pré-focale significative : t=0,071 (p=0,943) ;
- +10,2% pour S&FV et –9,2% pour V& FO.

C.1.1.4.d. Locuteur E

Contrastes intra-énoncés (voir graphique d. de la figure A3.4)

- contraste moyen : 19,4% (11,7% pour FS, 25,9% pour FV et 20,5% pour FO) significatif (F(1,25)=47,12 p<0,001) ;
- effet significatif du type de focalisation : F(2,50)=4,046 (p=0,024). On constate en effet que l'amplitude des mouvements d'étirement des lèvres est plus importante dans les cas de focalisation sur le verbe et sur l'objet quel que soit le syntagme considéré.
- effet d'interaction non significatif (F(2,50)=1,424 p=0,25), mais on constate tout de même que c'est lorsque la focalisation porte sur l'objet que le contraste entre ce qui est focalisé et ce qui ne l'est pas dans un énoncé est le plus important.

Comparaison des énoncés dans les cas neutre et focalisé (voir graphique d. de la figure A3.4)

Hyper-articulation focale

- hyper-articulation focale significative : t=3,524 (p=0,001) ;
- +14,5% (-0,4% pour S&FS, +23,3% pour V&FV et +20,5% pour O&FO).

Hypo-articulation post-focale

- hypo-articulation post-focale significative : t=-2,365 (p=0,021) ;
- -12,1% pour FS ;
- mais +0,5% pour FV.

Hyper-articulation pré-focale

- pas d'hyper-articulation pré-focale significative : t=0,657 (p=0,514) ;
- +5,7% pour S&FV et +11% pour V& FO.

C.1.1.4.e. Locuteur F

Contrastes intra-énoncés (voir graphique e. de la figure A3.4)

- contraste moyen : 23,2% (25,6% pour FS, 8,9% pour FV et 35,1% pour FO) significatif (F(1,25)=11,581 p=0,002) ;
- effet significatif du type de focalisation : F(2,50)=3,687 (p=0,033). On constate en effet que l'amplitude des mouvements d'étirement des lèvres est plus importante lorsque la focalisation porte sur le sujet et sur l'objet quel que soit le syntagme considéré.
- effet d'interaction non significatif (F(2,50)=1,573 p=0,218), mais on constate tout de même que c'est dans le cas de la focalisation sur l'objet que le contraste entre ce qui est focalisé et ce qui ne l'est pas dans un énoncé est le plus important.

Comparaison des énoncés dans les cas neutre et focalisé (voir graphique d. de la figure A3.4)

Hyper-articulation focale

- hyper-articulation focale significative : t=3,586 (p=0,001) ;
- +30% (+47,8% pour S&FS +2% pour V&FV et +40,3% pour O&FO).

Hypo-articulation post-focale

- pas d'hypo-articulation post-focale significative : t=1,497 (p=0,139) ;
- +22,7% pour FS ;
- mais +0,5% pour FV.

Hyper-articulation pré-focale

- pas d'hyper-articulation pré-focale significative : t=0,657 (p=0,514) ;
- –5,9% pour V&FO ;
- mais +5,7% pour S&FV.

C.1.1.5. Analyse des gestes de protrusion de la lèvre supérieure

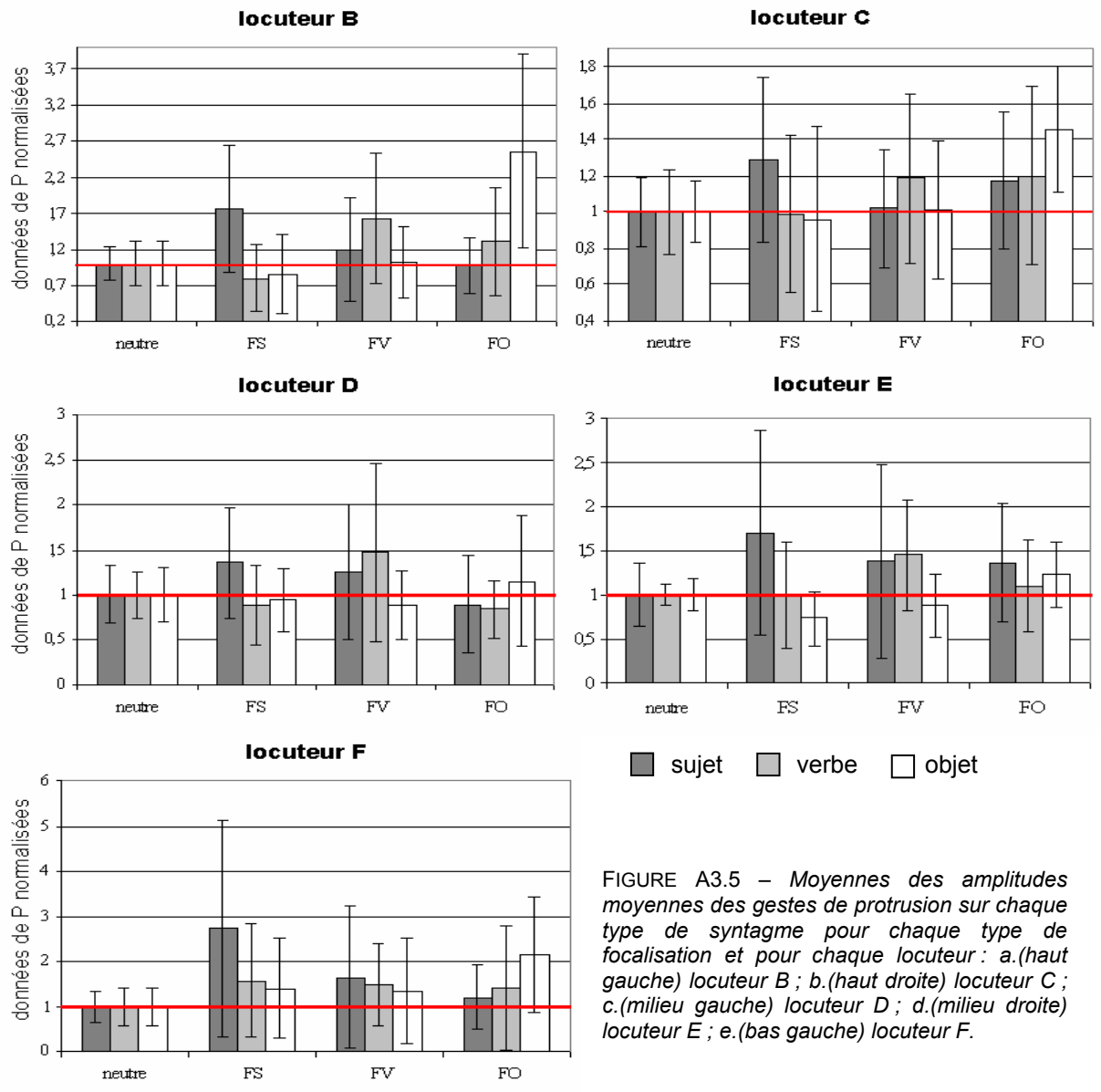


FIGURE A3.5 – *Moyennes des amplitudes moyennes des gestes de protrusion sur chaque type de syntagme pour chaque type de focalisation et pour chaque locuteur : a.(haut gauche) locuteur B ; b.(haut droite) locuteur C ; c.(milieu gauche) locuteur D ; d.(milieu droite) locuteur E ; e.(bas gauche) locuteur F.*

locuteur	B	C	D	E	F
effet congruence	F(1,25)=71,972 p<0,001	F(1,25)=38,174 p<0,001	F(1,25)=5,463 p<0,001	F(1,25)=5,716 p<0,001	F(1,25)=17,08 p<0,001
effet type de focalisation	F(2,50)=10,537 p<0,001	F(2,50)=7,433 p<0,001	F(2,50)=3,187 p=0,05	F(2,50)=0,151 p=0,860	F(1,545,50)=6,081 p=0,099
interaction	F(2,50)=7,818 p<0,001	F(2,50)=0,469 p=0,628	F(2,50)=0,53 p=0,592	-	-
test t congruence	t=8,285 p<0,001	t=6,485 p<0,001	t=4,028 p<0,001	t=5,068 p<0,001	t=5,695 p<0,001
test t post-foc	t=-2,012 p=0,048	t=-0,250 p=0,803	t=-2,522 p=0,014	t=-2,544 p=0,002	t=3,212 p=0,002
test t pré-foc	t=2,676 p=0,01	t=1,889 p=0,065	t=0,566 p=0,574	t=2,043 p=0,046	t=2,644 p=0,011

TABLE A3.5 – *Résultats des tests statistiques menés sur les données de protrusion selon la méthode décrite dans la section A.2.1.2.d du chapitre III pour chaque locuteur.*

C.1.1.5.a. Locuteur B

Contrastes intra-énoncés

Le graphique a. de la figure A3.5 montre clairement qu'il y existe un contraste entre ce qui est focalisé et ce qui ne l'est pas au sein d'un même énoncé. Ce contraste moyen est de 62,_% (93,6% pour FS, 52,6% pour FV et 42,2% pour FO). L'analyse statistique (effet congruence[155] de l'ANOVA[156], cf. table A3.5) permet de montrer que ces contrastes intra-énoncés sont bien significatifs : F(1,25)=71,972 (p<0,001).

L'ANOVA permet également de mettre en relief un effet significatif du type de focalisation[157] : F(2,50)=10,537 (p<0,001). Ceci est dû au fait que de manière générale, lorsqu'il y a focalisation sur l'objet, l'amplitude moyenne des gestes de protrusion est plus importante sur tout l'énoncé.

L'effet d'interaction[158] est aussi significatif : F(2,50)=7,818 (p<0,001). Ceci est dû au fait que le contraste entre ce qui est focalisé et ce qui ne l'est pas au sein d'un énoncé est plus important lorsque la focalisation porte sur l'objet.

Comparaison des énoncés dans les cas neutre et focalisé (voir graphique a. de la figure A3.5)

Hyper-articulation focale - Le graphique a. de la figure A3.5 permet de voir que lorsqu'un syntagme est focalisé l'amplitude moyenne des gestes de protrusion lui correspondant est supérieure de façon claire par rapport à sa valeur dans le cas neutre (les barres S&FS, V&FV et O&FO sont nettement au-dessus de 1). Cette augmentation est en moyenne de 98,4% (76% pour S&FS, 62,9% pour V&FV et 156,2% pour O&FO). On observe donc une hyper-articulation focale importante chez ce locuteur et celle-ci est significative : t=8,285 (p<0,001).

[155] Congruence : facteur intra-sujet à deux niveaux : cas congruents (S&FS, V&FV et O&FO) et cas incongruents (V&FS et O&FS, S&FV et O&FV et S&FO et V&FO).
[156] ANOVA à deux facteurs intra-sujets : congruence (deux niveaux) et type de focalisation (trois niveaux) ; pour une explication complète et détaillée du test voir la section A.2.1.2.d du chapitre III.
[157] Type de focalisation : facteur intra-sujet à trois niveaux : FS, FV et FO.
[158] Interaction congruence×type de focalisation.

Hypo-articulation post-focale - Le graphique a. de la figure A3.5 permet également de constater qu'il y a une baisse de l'amplitude des gestes de protrusion après la focalisation (les barres V&FS, O&FS et O&FV sont en -dessous de 1). Cette baisse est de 17,6% en moyenne après la focalisation du sujet. Cependant on n'observe pas d'hypo-articulation post-focale pour la focalisation verbe (+1,4% pour O&FV). Globalement, on obtient quand même une hypo-articulation tout juste significative chez ce locuteur : t=-2,012 (p=0,048).

Hyper-articulation pré-focale - Le graphique a. de la figure A3.5 permet de constater que lorsque le verbe est focalisé, le sujet est hyper-articulé (+19,3% pour S&FV). Il en est de même pour le verbe pré-focal (focalisation sur l'objet) dont l'hyper-articulation est de 30,9%. Cette hyper-articulation est significative : t=2,676 (p=0,01) et est certainement due à la mise en place d'une stratégie d'anticipation de la focalisation puisque lorsque l'objet est focalisé le sujet pré-focal n'est pas hyper-articulé (-2,8%). Seul l'élément directement pré-focal est donc hyper-articulé.

C.1.1.5.b. Locuteur C

Contrastes intra-énoncés (voir graphique b. de la figure A3.5)

- contraste moyen : 25% (30,8% pour FS, 17,3% pour FV et 26,9% pour FO) significatif (F(1,25)=38,174 p<0,001) ;
- effet significatif du type de focalisation : F(2,50)=7,433 (p<0,001). On constate en effet que l'amplitude des mouvements de protrusion est plus importante dans le cas de focalisation objet quelque soit le syntagme considéré.
- effet d'interaction non significatif (F(2,50)=0,469 p=0,628), mais on constate tout de même que c'est dans le cas où la focalisation porte sur le sujet que le contraste entre ce qui est focalisé et ce qui ne l'est pas dans un énoncé est le plus important.

Comparaison des énoncés dans les cas neutre et focalisé (voir graphique b. de la figure A3.5)

Hyper-articulation focale

- hyper-articulation focale significative : t=6,485 (p<0,001) ;
- +30,9% (+28,4% pour S&FS +18,7% pour V&FV et +45,7% pour O&FO).

Hypo-articulation post-focale

- pas d'hypo-articulation post-focale significative : t=-0,25 (p=0,803) ;
- -2,4% pour FS ;
- mais +1,2% pour FV.

Hyper-articulation pré-focale

- pas d'hyper-articulation pré-focale significative : t=1,889 (p=0,065) ;
- +1,7% pour S&FV et +20,1% pour V&FO.

C.1.1.5.c. Locuteur D

Contrastes intra-énoncés (voir graphique c. de la figure A3.5)

- contraste moyen : 37,4% (44,4% pour FS, 40,5% pour FV et 27,4% pour FO) significatif (F(1,25)=5,463 p<0,001) ;
- effet tout juste significatif du type de focalisation : F(2,50)=3,187 (p=0,05). On constate en effet que l'amplitude des mouvements de protrusion est plus importante dans le cas de focalisation verbe quelque soit le syntagme considéré

- effet d'interaction non significatif (F(2,50)=0,53 p=0,592), mais on constate tout de même que c'est lorsque la focalisation porte sur le sujet que le contraste entre ce qui est focalisé et ce qui ne l'est pas dans un énoncé est le plus important.

Comparaison des énoncés dans les cas neutre et focalisé (voir graphique c de la figure A3.5)

Hyper-articulation focale

- hyper-articulation focale significative : t=4,028 (p<0,001) ;
- +32% (+35,6% pour S&FS, +47,5% pour V&FV et +12,8% pour O&FO).

Hypo-articulation post-focale

- hypo-articulation post-focale significative : t=-2,522 (p=0,014) ;
- -8,9% pour FS et –11,1% pour FV.

Hyper-articulation pré-focale

- pas d'hyper-articulation pré-focale significative : t=0,566 (p=0,574) ;
- +25,2% pour S&FV ;
- mais –16,4% V&FO.

C.1.1.5.d. Locuteur E

Contrastes intra-énoncés (voir graphique d. de la figure A3.5)

- contraste moyen : 39,4% (83,1% pour FS, 35,8% pour FV et -0,6% pour FO) significatif (F(1,25)=5,716 p<0,001) ;
- mais pas de contraste pour le cas de focalisation objet : -0,6% ;
- pas d'effet du type de focalisation (F(2,50)=0,151 p=0,86) ;
- On constatera que c'est lorsque la focalisation porte sur le sujet que le contraste entre ce qui est focalisé et ce qui ne l'est pas dans un énoncé est le plus important.

Comparaison des énoncés dans les cas neutre et focalisé (voir graphique d. de la figure A3.5)

Hyper-articulation focale

- hyper-articulation focale significative : t=5,068 (p<0,001) ;
- +46% (+69,9% pour S&FS +45,1% pour V&FV et +23,1% pour O&FO).

Hypo-articulation post-focale

- hypo-articulation post-focale significative : t=-2,544 (p=0,002) ;
- -13,2% pour FS et –12,7% pour FV.

Hyper-articulation pré-focale

- hyper-articulation pré-focale significative : t=2,043 (p=0,046) ;
- +38,2% pour S&FV et +10,5% pour V&FO ;
- mais +39,8% S&FO donc sans doute pas de stratégie d'anticipation focale.

C.1.1.5.e. Locuteur F

Contrastes intra-énoncés (voir graphique e. de la figure A3.5)

- contraste moyen : 69,1% (125% pour FS, -0,4% pour FV et 82,6% pour FO) significatif (F(1,25)=17,08 p<0,001) ;
- mais pas de contraste pour le cas de focalisation verbe : -0,4% ;
- pas d'effet du type de focalisation (F(1,545,50)=6,081 p=0,099) ;

- On constate tout de même que c'est lorsque la focalisation porte sur le sujet que le contraste entre ce qui est focalisé et ce qui ne l'est pas dans un énoncé est le plus important.

Comparaison des énoncés dans les cas neutre et focalisé (voir graphique e. de la figure A3.5)

Hyper-articulation focale

- hyper-articulation focale significative : t=5,695 (p<0,001) ;
- +112,4% (+273,9% pour S&FS, +49,2% pour V&FV et +214,2% pour O&FO).

Hypo-articulation post-focale

- pas d'hypo-articulation post-focale ;
- mais hyper-articulation post-focale significative : t=3,212 (p=0,002) ;
- +48,8% pour FS et +34,3% pour FV.

Hyper-articulation pré-focale

- hyper-articulation pré-focale significative : t=2,644 (p=0,011) ;
- +64,7% pour S&FV et +42,6% pour V&FO ;
- mais +20,7% S&FO donc sans doute pas d'une stratégie d'anticipation focale.

D. Annexe 4 – Test de perception visuelle A (parole délexicalisée, locuteur A) : résultats détaillés pour chaque stimulus

phrase	type de focalisation	version	pourcentage de réponses correctes
2	FO	1	91,2
2	FO	1	88
2	FS	2	94,4
2	FS	2	96
2	FV	1	97,6
2	FV	1	97,6
2	neutre	2	92
2	neutre	2	81,6
4	FO	1	86,4
4	FO	1	81,6
4	FS	2	94,4
4	FS	2	92
4	FV	1	36,8
4	FV	1	98,4
4	neutre	2	93,6
4	neutre	2	88,8
6	FO	1	88
6	FO	1	96,8
6	FS	2	83,2
6	FS	2	90,4
6	FV	1	99,2
6	FV	1	96,8
6	neutre	2	84,8
6	neutre	2	88,8
7	FO	1	33,6
7	FO	1	88
7	FS	2	88
7	FS	2	91,2
7	FV	1	36
7	FV	1	92
7	neutre	2	91,2
7	neutre	2	96
	moyenne		86

E. Annexe 5 – Test de perception visuelle B (parole lexicalisée, locuteur A) : résultats détaillés pour chaque stimulus

phrase	type de focalisation	version	pourcentages de bonnes réponses
2	FO	1	68,5
2	FO	2	71,5
2	FS	1	66,7
2	FS	2	66,7
2	FV	1	78,2
2	FV	2	61,2
2	neutre	1	62,4
2	neutre	2	74,5
4	FO	1	27,3
4	FO	2	20,6
4	FS	1	80,0
4	FS	2	86,1
4	FV	1	84,2
4	FV	2	96,4
4	neutre	1	43,6
4	neutre	2	88,5
6	FO	1	64,2
6	FO	2	79,4
6	FS	1	86,7
6	FS	2	86,7
6	FV	1	98,2
6	FV	2	98,2
6	neutre	1	66,1
6	neutre	2	84,8
7	FO	1	89,7
7	FO	2	98,8
7	FS	1	79,4
7	FS	2	89,1
7	FV	1	2,4
7	FV	2	6,7
7	neutre	1	80,0
7	neutre	2	99,4
	moyenne		71,4

F. Annexe 6 – Test de perception visuelle C (parole lexicalisée, locuteur B) : résultats détaillés pour chaque stimulus

phrase	type de focalisation	version	pourcentage de réponses correctes
5	neutre	1	77,9
5	neutre	2	62,6
5	FS	1	56,0
5	FS	2	13,0
5	FV	1	76,1
5	FV	2	94,4
5	FO	1	16,8
5	FO	2	16,5
6	neutre	1	51,5
6	neutre	2	42,4
6	FS	1	38,9
6	FS	2	48,2
6	FV	1	79,5
6	FV	2	87,0
6	FO	1	35,4
6	FO	2	89,0
7	neutre	1	5,6
7	neutre	2	11,0
7	FS	1	54,0
7	FS	2	35,6
7	FV	1	59,6
7	FV	2	31,7
7	FO	1	16,3
7	FO	2	16,9
8	neutre	1	56,9
8	neutre	2	54,9
8	FS	1	44,8
8	FS	2	22,4
8	FV	1	77,6
8	FV	2	44,8
8	FO	1	22,1
8	FO	2	51,9
9	neutre	1	3,6
9	neutre	2	14,6
9	FS	1	56,5
9	FS	2	76,1
9	FV	1	21,7
9	FV	2	30,8
9	FO	1	12,6
9	FO	2	34,8
10	neutre	1	94,6

10	neutre	2	79,8
10	FS	1	72,1
10	FS	2	31,7
10	FV	1	55,4
10	FV	2	27,9
10	FO	1	7,4
10	FO	2	1,8
11	neutre	1	40,5
11	neutre	2	27,7
11	FS	1	11,1
11	FS	2	18,3
11	FV	1	78,0
11	FV	2	96,2
11	FO	1	18,5
11	FO	2	31,9
12	neutre	1	29,7
12	neutre	2	29,3
12	FS	1	77,7
12	FS	2	84,9
12	FV	1	5,8
12	FV	2	9,2
12	FO	1	27,7
12	FO	2	55,6
13	neutre	1	46,3
13	neutre	2	52,2
13	FS	1	60,6
13	FS	2	49,6
13	FV	1	7,3
13	FV	2	9,1
13	FO	1	74,3
13	FO	2	59,5
	moyenne		43,2

G. Annexe 7 – Test de perception audiovisuelle : résultats détaillés pour chaque stimulus

phrase	locuteur	type de focalisation	modalité		
			AV	A	V
2	B	neutre	42,3	42,3	76,9
2	A	neutre	76,9	65,4	19,2
2	B	FO	92,3	80,8	34,6
2	A	FO	69,2	88,5	61,5
2	B	FS	65,4	53,8	80,8
2	A	FS	80,8	80,8	96,2
2	B	FV	65,4	46,2	46,2
2	A	FV	100,0	100,0	65,4
4	B	neutre	76,9	73,1	84,6
4	A	neutre	88,5	61,5	88,5
4	B	FO	38,5	15,4	11,5
4	A	FO	30,8	53,8	26,9
4	B	FS	50,0	15,4	26,9
4	A	FS	100,0	100,0	73,1
4	B	FV	69,2	61,5	88,5
4	A	FV	96,2	80,8	100,0
6	B	neutre	80,8	65,4	69,2
6	A	neutre	73,1	53,8	65,4
6	B	FO	76,9	11,5	50,0
6	A	FO	100,0	96,2	92,3
6	B	FS	96,2	69,2	84,6
6	A	FS	100,0	88,5	92,3
6	B	FV	38,5	38,5	11,5
6	A	FV	84,6	84,6	80,8
7	B	neutre	65,4	69,2	92,3
7	A	neutre	84,6	92,3	34,6
7	B	FO	69,2	38,5	65,4
7	A	FO	100,0	88,5	88,5
7	B	FS	57,7	61,5	53,8
7	A	FS	76,9	80,8	84,6
7	B	FV	30,8	42,3	7,7
7	A	FV	92,3	76,9	23,1
moyennes			74,0	64,9	61,8

Résumé

Le travail présenté dans ce mémoire est sous-tendu par trois observations majeures. D'abord, de nombreux travaux ont mis en évidence que la parole n'était pas uniquement de nature auditive mais aussi visuelle. D'autre part, la prosodie, domaine de l'intonation, du rythme et du phrasé joue un rôle crucial en parole. Enfin, la deixis ou monstration est un phénomène au cœur de la communication parlée et de son acquisition par les jeunes enfants. Or celle-ci peut, entre autre, s'exprimer uniquement avec la parole : il est possible de « montrer de la voix » par la focalisation prosodique par exemple. Ces observations et constatations permettent d'émettre l'hypothèse que la focalisation contrastive prosodique se manifesterait non seulement par la modalité auditive ,comme il a déjà été largement exploré, mais aussi par la modalité visuelle. C'est la piste que les travaux de ce mémoire visent à explorer pour le cas particulier du français. Plusieurs analyses en production de la parole ont ainsi permis, grâce aux enregistrements de six locuteurs avec deux systèmes de mesure différents et complémentaires, de mettre en évidence les stratégies de signalisation visuelle de la focalisation. Il semble que les locuteurs produisent des indices articulatoires visibles selon deux stratégies principales : la stratégie de signalisation absolue et la stratégie de signalisation différentielle. Les analyses ont également permis de montrer que d'autres gestes faciaux non articulatoires (mouvements des sourcils et de la tête) pourraient être liés à la production de la focalisation mais de façon très variable non seulement d'un locuteur à l'autre mais aussi pour un même locuteur. Par ailleurs, des analyses parallèles en perception, ont permis de montrer que les indices visuels mis en évidence en production, étaient effectivement utilisés en perception et qu'ils permettent d'extraire l'information de focalisation quand la modalité auditive est indisponible ou dégradée. Il a été mis en évidence que les indices visuels identifiés en production correspondent au moins en partie à ceux utilisés en perception audiovisuelle. Ces travaux montrent ainsi que la focalisation contrastive en français est « visible » et est « vue ». Ces résultats permettent d'esquisser un modèle cognitif de la production et de la perception audiovisuelles de la focalisation contrastive en français.

Mots clés : multisensorialité, focalisation, deixis, prosodie, production et perception de la parole, français, articulation, gestes faciaux, contrôle multisensoriel.

Abstract

The work described in this dissertation is grounded by three major findings. Firstly, numerous researchers have shown that speech is not only auditory but also visual. Secondly, prosody *i.e.* intonation, rhythm and phrasing, plays a key role in speech. Thirdly, deixis is a core phenomenon in spoken communication and its acquisition by infants. Deixis can be achieved using speech: it is indeed possible to "show with the voice" using prosodic focus for example. These observations enable us to assume that prosodic contrastive focus is rooted not only in audition, as has already been widely explored, but also in vision. The various works presented in this dissertation explore this hypothesis for French. Several production studies analyzing the recordings of six speakers using two different and complementary measurement techniques have shown that focus is signaled visually. Speakers use two different strategies regarding the visible articulatory movements: an absolute strategy and a differential one. The measurements have also shown that other non-articulatory facial gestures may be linked to the production of contrastive focus such as eyebrow and head movements. The link is however widely inter and intra speaker dependent. In parallel, perceptual experiments have enabled us to show that the visual correlates of focus are used for focus information extraction when the auditory modality is absent or degraded. It was also shown that the visual correlates identified in the production studies correspond at least in part to those used in audiovisual perception. These studies have thus shown that prosodic contrastive focus is "visible" and "seen". The findings allow us to sketch a cognitive model of the audiovisual production and perception of contrastive focus in French.

Key words: multisensoriality, focus, deixis, prosody, speech production and perception, French, articulation, facial gestures, multisensory control.