



*Laboratoire d'Acoustique Musicale*  
Université Paris 6,  
11, rue de Lourmel, 75015 Paris CEDEX, FRANCE

Mémoire pour le Master Sciences et Technologies de l'UPMC  
Mention Sciences de l'Ingénieur  
Spécialité MIS, Parcours ATIAM

---

# Etude articulatoire de la parole produite en environnement bruyant

**Lucie BAILLY**

Lucie.Bailly@ircam.fr

Collaboration LAM/ICP

06 juillet 2005

---

**MAÎTRE DE STAGE** : Maëva Garnier  
Laboratoire d'Acoustique Musical  
(+33) 1 53 95 43 34  
*garnier@lam.jussieu.fr*

**COLLABORATION** : Hélène Løevenbruck, Pauline Welby & Marion Dohen  
Institut de la Communication Parlée  
46 Avenue Félix Viallet 38031 Grenoble Cedex 01 FRANCE  
*loeven@icp.inpg.fr*  
*welby@icp.inpg.fr*  
*dohen@icp.inpg.fr*

**TUTEUR DE STAGE** : Antoine Chaigne  
École Nationale Supérieure des Techniques Avancées (ENSTA)  
Unité d'Enseignement et de Recherche en Mécanique (UME)  
Chemin de la Hunière 91761 Palaiseau cedex FRANCE  
*Antoine.Chaigne@ensta.fr*

---

# Sommaire

<b>1</b>	<b>Introduction</b>	<b>5</b>
<b>2</b>	<b>Cadre de l'étude</b>	<b>7</b>
2.1	Variabilité de la parole . . . . .	7
2.1.1	La problématique de l'effet Lombard . . . . .	7
2.1.1.1	L'effet Lombard, un réflexe purement physiologique? . . . . .	7
2.1.1.2	L'effet Lombard, un désir d'intelligibilité? . . . . .	8
2.1.1.3	Le double rôle de l'hyper-articulation . . . . .	8
2.1.2	Spécificités de la parole Lombard . . . . .	9
2.1.2.1	Spécificités acoustiques . . . . .	9
2.1.2.2	Spécificités visuelles . . . . .	10
2.1.2.3	Spécificités prosodiques . . . . .	10
2.2	Variations articulatoires de la parole . . . . .	12
2.2.1	Physiologie de la production de la parole . . . . .	12
2.2.1.1	La phonation . . . . .	12
2.2.1.2	L'articulation . . . . .	13
2.2.2	Paramètres de l'étude articulatoire . . . . .	14
<b>3</b>	<b>Expérimentation</b>	<b>17</b>
3.1	Élaboration du corpus . . . . .	17
3.1.1	Contraintes liées aux hypothèses de l'étude . . . . .	17
3.1.1.1	Interaction et communication . . . . .	17
3.1.1.2	Nécessité de définir des syntagmes cibles . . . . .	17
3.1.2	Premières restrictions sur le corpus en conséquence . . . . .	18
3.1.2.1	Les syntagmes cibles et les « phrases porteuses » . . . . .	18
3.1.2.2	La forme syntaxique « SVO » . . . . .	19
3.1.2.3	Les trois positions du syntagme . . . . .	19
3.1.3	Corpus semi-spontanés . . . . .	20
3.1.3.1	Des exemples dans la littérature . . . . .	20
3.1.3.2	Des propositions novatrices . . . . .	21
3.1.4	Les contraintes d'acquisition expérimentale . . . . .	21
3.1.5	Choix définitif du corpus . . . . .	22
3.2	Protocole expérimental . . . . .	22
3.2.1	Stimuli utilisés . . . . .	22
3.2.2	Méthodologie . . . . .	23
3.2.2.1	Sujets et dispositif d'enregistrement . . . . .	23
3.2.2.2	Protocole d'enregistrement . . . . .	25
3.2.2.3	Un compromis réussi . . . . .	25

<b>4 Mesures et Analyses</b>	<b>26</b>
4.1 Traitement des données vidéo . . . . .	26
4.1.1 Le contour des lèvres . . . . .	26
4.1.2 Analyse articulatoire : TRAP . . . . .	27
4.2 Traitement des données audio et Egg . . . . .	28
4.2.1 Synchronisation audio/vidéo et decorrélation bruit/parole . . . . .	28
4.2.2 Etiquetage des segments audio . . . . .	28
4.2.3 Calcul de la $F_0$ via les signaux Egg . . . . .	29
<b>5 Résultats et Discussion</b>	<b>30</b>
5.1 Paramètres de l'étude articulatoire . . . . .	30
5.1.1 Étude de l'aire intero-labiale $S$ et de la protrusion $P1$ . . . . .	30
5.1.1.1 Influence de l'environnement sonore . . . . .	30
5.1.1.2 Influence de la position du syntagme . . . . .	33
5.1.2 Étude de la durée $d$ d'énonciation . . . . .	33
5.1.2.1 Influence de l'environnement sonore . . . . .	33
5.1.2.2 Influence de la position du syntagme . . . . .	34
5.2 Paramètres de l'étude acoustique . . . . .	34
5.2.1 Étude de la fréquence fondamentale $F_0$ . . . . .	34
5.2.1.1 Influence de l'environnement sonore . . . . .	34
5.2.2 Étude de l'intensité $SPL$ . . . . .	36
5.2.2.1 Influence de l'environnement sonore . . . . .	36
5.2.2.2 Influence de la position du syntagme . . . . .	36
<b>6 Conclusion</b>	<b>38</b>
Références . . . . .	40
<b>7 Annexes</b>	<b>42</b>
7.1 Corpus . . . . .	42
7.2 Exemple de dialogue du corpus MapTask (Q3NC8) . . . . .	44
7.2.1 Map instructions for subjects . . . . .	44
7.2.2 To the speaker . . . . .	44
7.2.3 Then to the hearer . . . . .	44
7.3 Graphes de $S$ , $P1$ , $F_0$ et $SPL$ pour les 33 phrases . . . . .	47

# Table des figures

2.1	(a) Segmentation en deux mots (« mes ga <b>LOPS</b> »), (b) Segmentation en un mot (« <b>mégaLO</b> »).	11
2.2	(a) Schéma en coupe du larynx, (b) La glotte en trois positions de phonation.	12
2.3	(a) Articulateurs de la parole, (b) Cavités supra-glottiques.	13
2.4	Voyelle antérieure (a) vs voyelle postérieure (b)	15
2.5	Voyelle fermée (a) vs voyelle ouverte (b)	15
2.6	Articulographe	16
3.1	Spectrogrammes et contours de $F_0$ de (a) « Sens », (b) « Tôt », (c) « Non », (d) « Lila ».	19
3.2	Sujets F1 et F2 de l'expérience 'Lip Tracking'.	21
3.3	Protocole et sujets F1 et F2 de l'expérience 'Lip Tracking'.	23
3.4	Enregistrement vidéo de face et de profil du sujet 2 pour deux instants d'élocution différents.	24
4.1	Paramètres géométriques des lèvres pour les mesures de face.	26
4.2	(a) Analyse du contour des lèvres sous <i>TACLE</i> , (b) Calcul de $S$ ( $cm^2$ ), de sa vitesse ( $cm^2.s^{-1}$ ) et de son accélération ( $cm^2.s^{-2}$ ) en fonction du temps ( $s$ ) sous <i>TRAP</i> .	27
4.3	Segmentations SVO et syllabique sous <i>Praat</i> de la phrase 12 : « <b>La mamie lavait le muret</b> ».	28
4.4	(a) Signal audio, signal Egg et sa dérivée ( <i>de haut en bas</i> ), (b) Cycle d'accolement des cordes vocales sur le signal Egg.	29
5.1	Test de Wilcoxon, avec « Rangs » égal à « + » si $Max_S$ dans le silence > $Max_S$ dans le cocktail; « - » si $Max_S$ dans le silence < $Max_S$ dans le cocktail; et « = » si $Max_S$ dans le silence = $Max_S$ dans le cocktail.	31
5.2	Statistiques descriptives de l'évolution de $S_{max\_moy\_cible}$ (en rapport d'augmentation) dans les deux conditions de bruit.	31
5.3	Statistiques descriptives de la différence moyenne de l'aire sous la courbe de $S$ par unité de temps (en $cm^2$ ) sur la phrase, le syntagme et par syllabes (première, deuxième, troisième pour les syntagmes à trois syllabes, et dernière pour les syntagmes à deux et trois syllabes), dans les deux conditions de bruit.	32
5.4	Variations de l'aire moyenne sous la courbe de $S$ par unité de temps (en $cm^2$ ) au passage de la condition silence à la condition de cocktail à 85 dB, sur le syntagme et les syllabes (première, deuxième, troisième pour les syntagmes à trois syllabes de contenu, et dernière pour les syntagmes à deux et trois syllabes de contenu) en fonction de la position initiale (Init), intermédiaire (Inter) ou finale (Final).	33

5.5	Variations dans le bruit cocktail de la durée moyenne (en <i>s</i> ) du syntagme et des syllabes (première, deuxième, troisième pour les syntagmes à trois syllabes de contenu, et dernière pour les syntagmes à deux et trois syllabes de contenu) en fonction de la position initiale (Init), intermédiaire (Inter) ou finale (Final). . . . .	34
5.6	Variations de la $F_0$ ( <b>a</b> ) dans le silence, ( <b>b</b> ) dans le bruit cocktail, lors de l'élocution par le sujet F2 de la phrase 10 : « <i>Marilou mima Mélanie</i> ». . . . .	35
5.7	Comparaison de ( <b>a</b> ) l'ambitus de $F_0$ , ( <b>b</b> ) la $F_0$ maximale des 33 phrases lues par F2 en condition de silence et de bruit cocktail. . . . .	35
5.8	Comparaison de la $F_0$ moyenne des 33 phrases lues par F2 en condition de silence et de bruit cocktail. . . . .	35
5.9	Comparaison du ( <b>a</b> ) <i>SPL</i> moyen, ( <b>b</b> ) <i>SPL</i> maximal des 33 phrases lues par F2 en condition de silence et de bruit cocktail. . . . .	36
5.10	Augmentation dans le bruit cocktail de l'intensité moyenne de la voix (en <i>dB</i> ) dans le syntagme et par syllabes (première, deuxième, troisième pour les syntagmes à trois syllabes de contenu, et dernière pour les syntagmes à deux et trois syllabes de contenu) en fonction de la position initiale (Init), intermédiaire (Inter) ou finale (Final). . . . .	37
5.11	Paramètres acoustiques et articulatoires des productions vocales de F2 à la prononciation de la phrase 9 : « <i>Le loup et le minet levaient le nez</i> », ( <b>a</b> ) dans le silence, ( <b>b</b> ) dans le bruit cocktail de 85dB. . . . .	37
7.1	Givers Map . . . . .	45
7.2	Followers Map . . . . .	46
7.3	Variations des paramètres acoustiques et articulatoires des productions vocales de F2 à la prononciation de la phrase 31 dans le silence et dans le bruit cocktail de 85 dB. . . . .	48
7.4	Variations des paramètres acoustiques et articulatoires des productions vocales de F2 à la prononciation de la phrase 31 dans le silence et dans le bruit cocktail de 85 dB. . . . .	49

# Remerciements

Ce stage a été réalisé grâce au soutien de personnes avec qui ce fut un plaisir de collaborer et qui, par la passion qui les anime dans leur travail, par leur conseils et leurs encouragements, ont su m'aider à avancer dans mon projet professionnel. Ces quelques lignes de formalité qui leur sont destinées mais qui n'ont pu être rédigées que quelques heures avant la date de rendu de ce mémoire ne seront probablement pas à la hauteur de la reconnaissance que je leur dois ; je tiens néanmoins à m'y essayer.

Une première mention à Maëva Garnier tout d'abord, pour m'avoir fait découvrir le domaine de la parole et avoir pris le temps de m'y initier ; pour son appui dans mon travail au cours de ces quatre mois, son énergie contagieuse qui fut salvatrice dans les dernières lignes droites et pour sa bonne humeur communicative.

Un grand merci également à Marion Dohen, Pauline Welby et Hélène Løevenbruck, pour m'avoir accompagnée dans mes premiers pas en prosodie ; je les remercie pour l'aide précieuse qu'elles ont apportée à mon travail, j'ai beaucoup appris à leur contact et je garde un souvenir très agréable de notre collaboration.

Une pensée toute particulière à Nathalie Henrich, pour l'attention qu'elle m'a portée et pour s'être hasardée dans le grand ménage de mes angoisses d'orientation et de mes questions sans réponses. Je tiens également à remercier Pierre Leveau et Guillaume Pellerin, dont ce fut un plaisir de partager l'antre ; merci pour avoir supporté mes questions diverses et variées en Matlab, en Latex ou encore sur notre petit ventilateur bien impuissant devant les 40° avec qui nous cohabitons depuis deux mois. Je remercie Jean-Dominique Polack pour m'avoir accueillie dans son équipe et tous les membres du LAM pour leur sympathie et disponibilité.

Je tiens à remercier Christophe Savariaux pour son aide précieuse pendant les enregistrements, les sujets Claire Lalevée, Aude Noiray et Coriandre Vilain sans qui ce mémoire n'aurait pas grand sens ; et je remercie plus largement tous les membres de l'ICP qui m'ont accueilli lors de mes séjours de recherche à Grenoble et qui m'ont séduite par leur gentillesse et leur joie de vivre.

# Chapitre 1

## Introduction

Cette étude s'inscrit dans le cadre de recherche des mécanismes d'adaptation en parole. Dans le contexte d'une interaction communicative entre locuteurs parlant sous certaines conditions (relatives au bruit ambiant, à la formalité sociale, à l'attention de l'interlocuteur, à l'urgence de la situation, etc.), les impératifs perceptifs de la parole et les objectifs de compréhension du message impliquent une audibilité et une intelligibilité accrues de la parole. Nous nous intéressons ici aux variations du message vocal lorsque le locuteur est placé en environnement bruyant. Ces phénomènes de régulation des productions vocales dans le bruit sont connus sous le nom d'*effet Lombard*. Leur étude remonte au début du siècle avec Etienne Lombard [Lombard, 1911].

Ce travail vient compléter un projet se proposant d'étudier les adaptations d'un locuteur qui communiquerait dans un environnement bruyant à travers l'analyse acoustique et prosodique (*i.e.* articulatoire et intonative) de ses productions vocales. Des études antérieures ont révélé parmi ces adaptations une modification des paramètres acoustiques (fréquence fondamentale  $F_0$ , régions formantiques, intensité) de la parole (*cf.* [Junqua, 1992]). Force est de constater néanmoins le peu d'études portant sur les variations articulatoires générées en parole forte (*cf.* [Schulman, 1989]), et l'absence à ce jour d'étude articulatoire en parole Lombard. L'objet de ce travail vise à préciser la nature de ces variations articulatoires en relation avec différents facteurs linguistiques et environnementaux. Observe-t-on une hyper-articulation ? Si oui, à quoi sert-elle en définitive ? Est-elle due aux contraintes physiologiques répondant à la simple nécessité de parler plus fort dans le bruit ou bien est-elle davantage motivée par la volonté (consciente ou non) de préserver l'intelligibilité de la communication ? Partant de cette hypothèse, elle ne devrait pas, au cours de l'élocution, se manifester de façon uniforme pour tous les mots d'un énoncé.

L'analyse prosodique de la parole produite en conditions bruitées renseigne sur les stratégies que le locuteur adopte pour préserver son intelligibilité en dépit de la perturbation qui lui est imposée. Nous espérons ici mettre en évidence les mécanismes articulatoires mis en jeu dans l'élaboration de ces stratégies. Ces mécanismes adaptatifs reposent très certainement sur les variations conjointes des mouvements des articulateurs de la parole tels que la mâchoire, la langue, le voile du palais, les lèvres ou le visage. Cependant, nous n'avons encore aucune idée a priori des stratégies articulatoires adoptées en conditions perturbées. C'est pourquoi nous menons une étude pilote qui se limite ici à l'observation des mouvements labiaux et dont les résultats nous permettront d'affiner notre axe de recherche.

Nous nous intéressons donc ici à l'étude de la protrusion labiale (projection des lèvres lors de la production de « ou » par exemple) et des variations de l'aire intero-labiale dans

différents environnements bruyants et espérons ainsi mettre en relation les gestes articulatoires avec les caractéristiques acoustiques de la parole et les variations intonatives et accentuelles de l'élocution. Cette étude articulatoire étant pionnière dans la recherche en parole Lombard, nous pouvons nous attendre à tout phénomène d'adaptation de la part du locuteur. Nous faisons néanmoins l'hypothèse que l'hyper-articulation est un mécanisme régulateur mis en place pour soutenir majoritairement et/ou conjointement l'intensité, la mélodie, la forme syntaxique, la sémantique, et le visuel de la parole dans le bruit. Une analyse prosodique portant sur les contours intonatifs issus des mêmes travaux expérimentaux est menée en parallèle du traitement articulatoire à l'Institut de la Communication Parlée (ICP), pour déterminer en quoi l'hyper-articulation soutient effectivement la segmentation de la phrase.

Nous espérons à l'issue de cette étude acquérir une meilleure compréhension de la production de la parole, dont l'intérêt n'est plus à justifier dans la perspective de nouveaux systèmes de reconnaissance vocale en ambiance bruitée. L'étude de l'adaptation vocale en conditions perturbées est susceptible de nous renseigner sur les mécanismes d'apprentissage de nouvelles stratégies pour la production de la parole. Les retombées de ce travail pourront à ce titre enrichir les connaissances en orthophonie, où la recherche de stratégies de rééducation suit les avancées des théories sur le forçage vocal.

# Chapitre 2

## Cadre de l'étude

### 2.1 Variabilité de la parole

Notre parole s'ajuste à l'environnement sonore dans lequel elle est produite grâce au retour auditif que nous en avons. En effet, parler ne consiste pas en la juxtaposition de phones invariants, comme en témoignent les difficultés liées à la reconnaissance automatique de la parole (*cf.* [Junqua, 1992]). Les variations du signal de parole se déclinent suivant l'influence des différents facteurs mis en jeu dans la communication, comme le retour auditif et visuel, le niveau de bruit ambiant, sa nature (brouhaha, bruits de circulations, musique, etc.), le type d'interaction verbale (dialogue en intimité, discours en public, etc.), l'attention portée à l'interlocuteur (ami, collègue, enfant, etc.), la formalité sociale, l'urgence de la situation (*cf.* [Dohalská and Mejvaldová, 2000]), l'état psychologique du locuteur, sa charge de travail, etc. La coordination des gestes articulatoires impliqués dans la réalisation d'une séquence de la parole est elle-même soumise à l'influence de ces divers facteurs contextuels.

#### 2.1.1 La problématique de l'effet Lombard

##### 2.1.1.1 L'effet Lombard, un réflexe purement physiologique ?

L'*effet Lombard* désigne l'effet se traduisant par des modifications de la parole en présence de bruit dans des situations d'atténuation du retour auditif de sa voix (bruit ambiant, port de protections auditives, etc.). Il résulte de l'augmentation de l'intensité phonatoire sous l'influence du bruit. Cet effet a été constaté en 1909 par le docteur E.Lombard, oto-rhino-laryngologiste à l'hôpital Laboisière de Paris, par le rapprochement de deux faits d'observation :

- « Les sourds par lésion de l'appareil labyrinthique élèvent anormalement la voix »
- « Un sujet dont l'audition est intacte élève anormalement la voix au milieu d'un vacarme intense »

Selon lui, en présence d'un milieu bruyé, « le locuteur, qui n'entend pas mieux sa propre voix que les bruits extérieurs, s'efforce de remédier à cette *audition défailante* en *augmentant* l'intensité du son glottique ». Tout se passe comme si « le locuteur se crie en quelque sorte à lui-même comme on crie à l'oreille d'un sourd », et ce d'une manière réflexe.

### 2.1.1.2 L'effet Lombard, un désir d'intelligibilité ?

Certains auteurs (*cf.* [Lane and Tranel, 1971], [Lane and Sisson, 1970]) conçoivent la régulation de la voix comme motivée par la recherche d'intelligibilité envers l'interlocuteur, prenant le dessus sur la volonté de s'entendre soi-même. De ce point de vue, **l'effet Lombard n'a de sens que lorsque le locuteur se trouve dans un contexte de communication**. Ainsi donc, l'effet Lombard est interprété dans la littérature tantôt comme une accommodation aux exigences de la communication intelligible (*cf.* [Lane and Tranel, 1971]), tantôt comme un réflexe physiologique (*cf.* [Garde, 1965]), mais il n'en reste pas moins provoqué par la nécessité que le locuteur a de se faire entendre et/ou de s'entendre soi-même dans le contexte concret d'une interaction locuteur/auditeur.

Dans cette étude, nous concevons l'effet Lombard comme gouverné par ce *désir d'intelligibilité*. Des études expérimentales (*cf.* [Summers et al., 1988]) ont noté que la parole produite dans le bruit est plus intelligible pour un auditeur que la parole produite dans un milieu non bruité à même rapport signal sur bruit<sup>1</sup>. Les études acoustiques de [Junqua, 1992] démontrent également une forte interaction entre le type de bruit environnant et les résultats perceptifs obtenus en terme d'intelligibilité. La nature de cette interaction dépend du type de vocabulaire utilisé. Il apparaît que le bruit blanc Gaussien altère l'intelligibilité des consonnes plus que ne le fait le bruit de conversations, qui, lui, affecte davantage les voyelles.

### 2.1.1.3 Le double rôle de l'hyper-articulation

L'articulation est elle aussi au coeur d'une problématique dans le processus de production vocale et surtout dans l'intelligibilité. Nous verrons au **2.2.1.2** qu'elle sert à moduler l'information acoustique.

Nous pensons que l'hyper-articulation en parole Lombard, si elle a lieu, sert non seulement à « hyper-moduler » certaines informations acoustiques, mais également à transmettre une information visuelle. L'étude de conditions d'intelligibilité réduite, comme en environnement bruyant, conduit donc à une hyper-articulation et permet de tester ces hypothèses. De même que pour les deux conceptions de l'effet Lombard exposées au 2.1.1, il est probable que ces différentes interprétations de l'hyper-articulation dans la parole ne soient pas strictement opposables et qu'une convergence soit largement envisageable. En effet, le modèle théorique de [Lindblom, 1990] que nous développons ci-dessous sur la production de la parole en hyper-articulation nous oriente vers cet axe de pensée.

Rappelons avant tout la correspondance constatée entre les caractéristiques de la perception auditive et la production de la parole. Des études ont déjà suggéré l'existence d'une boucle de rétroaction entre la production et la perception de la parole, présumant une adaptation réciproque entre l'appareil phonatoire et le système auditif (*cf.* [Greenberg, 1996]). Les personnes atteintes de surdité ont ainsi (et paradoxalement) plus de mal à se faire comprendre : du fait de la dégradation de cette boucle de rétroaction, leur parole ne s'adapte plus à l'environnement dans lequel elle est produite. L'oreille humaine est adaptée particulièrement à la perception de la voix humaine. Sur un plan purement physiologique en premier lieu, la production et la reconnaissance de la parole sont ainsi intimement liées. Citons à titre d'illustration les analogies de gammes

<sup>1</sup>[Pickett, 1956] a montré que l'intelligibilité décroît néanmoins lorsque le locuteur entre en régime de voix criée

de fréquences impliquées dans le signal de la parole et dans la zone de sensibilité auditive : l'étendue spectrale du signal de la parole normale se situe entre 80 Hz et 8000Hz, avec une étendue dynamique de 60 à 70 dB, ce qui correspond à la zone fréquentielle de sensibilité maximale de l'oreille.

L'hypothèse de Bjorn Lindblom est que l'évolution de l'appareil articulatoire a été influencé par la perception auditive. Selon lui, l'évolution des sons implique une sélection des sons de la parole sur deux critères :

- *l'économie de la production* (cf. [Lindblom, 1983]). En effet, le système articulatoire génère des mouvements qui tendent par défaut à une économie motrice et à une facilité de production ; il s'agit d'une contrainte interne au processus moteur de production articulatoire ;
- *la distinctivité perceptive*, impliquée par une audibilité et une intelligibilité accrues de la parole dans certaines situations (relatives au bruit ambiant, aux différents facteurs contextuels etc.) ; il s'agit là d'une contrainte externe au processus moteur de production articulatoire et relevant des impératifs perceptifs de la parole.

Le locuteur est ainsi capable, selon cette théorie, de contrôler à court-terme ses variations phoniques entre ces deux pôles, chacun répondant à deux types de contrainte.

Partant de ce constat, la théorie "**H&H**" de Lindblom (1990) propose deux modes de parole : la parole "**Hyper**" qui s'oriente vers une clarté maximale du signal produit et la parole "**Hypo**" qui vise la production la plus économique possible. Selon lui, l'appareil phonatoire produit des sons situés sur ce continuum de variations phonétiques allant de l'hypo-articulation à l'hyper-articulation. La parole est donc de ce point de vue organisée de façon adaptative, exploitant conjointement l'information mise dans le signal par l'effort du locuteur, et l'information contextuelle qui n'est pas dans le signal acoustique et qui est récupérée par l'auditeur par sa compréhension générale de la situation de communication (cf. [Carré et al., 1995]). C'est le jugement qu'il a de cet équilibre entre ces deux sources d'information qui conduit le locuteur tantôt à "hyperarticuler" pour mettre plus d'information dans le signal, tantôt à "hypoarticuler" et exploiter la quantité d'information importante hors signal afin de minimiser ses efforts d'articulation [Beautemps, 2000].

[Carré et al., 1995] pose la question de savoir si le locuteur s'adapte ainsi pour aider l'auditeur dans sa tâche de reconnaissance de données articulatoires, ou bien pour transmettre une information acoustique plus nette. "Is the greater intelligibility of clear speech due to the fact that articulatory information becomes clearer, or to acoustic patterns becoming clearer?"

L'objet de cette étude vise précisément à examiner les stratégies d'adaptation articulatoire en parole "Hyper", que l'on espère obtenir naturellement par effet Lombard. Les résultats articulatoires obtenus confrontés à ceux de l'analyse prosodique menée en parallèle à l'ICP pourront nous orienter dans la perspective de trancher la problématique ainsi soulevée par Lindblom et de retrouver les conclusions de Schulman (cf. 2.1.2) qui voit dans l'effet Lombard une réorganisation des mouvements articulatoires.

## 2.1.2 Spécificités de la parole Lombard

### 2.1.2.1 Spécificités acoustiques

Certains auteurs (*cf.* [Lane and Tranel, 1971], [Junqua, 1992]) ont axé leurs recherches sur l'augmentation de l'intensité phonatoire et les modifications acoustiques du signal vibratoire générées en parole Lombard. Le ralentissement du rythme, l'élévation de la fréquence fondamentale, celle de l'intensité, et les modifications des formants ont été caractérisés parmi les variations acoustiques générées en parole Lombard.

### 2.1.2.2 Spécificités visuelles

Pour mieux se comprendre, les interlocuteurs s'appuient sur les indications visuelles telles que les mouvements des muscles faciaux, des lèvres, des mains et du corps en général. La lecture labiale est utilisée par le mal-entendant pour analyser le langage lu, mais également de façon naturelle et spontanée par toute personne essayant de comprendre son interlocuteur dans un contexte bruyé perturbant la communication. Les travaux de [Benoît et al., 1992] sur l'évaluation quantitative de l'apport de la composante visuelle dans la perception de la parole en milieu bruyé ont permis de contrôler la quantité et la nature de l'information visuelle transmise à l'auditeur. Benoît identifia pour le français une vingtaine de formes labiales caractéristiques, les visèmes du français, équivalents visuels des phonèmes.

On comprend bien que ces formes labiales plus amples et plus étirées en parole Lombard aident en conséquence à rendre la communication intelligible. Il paraît nettement moins intuitif mais tout aussi envisageable cependant de considérer leur production motivée par ce désir d'intelligibilité, et non pas seulement comme la simple conséquence physiologique du forçage vocal, ainsi que le soutiennent les partisans du « réflexe Lombard ».

### 2.1.2.3 Spécificités prosodiques

La prosodie regroupe les caractéristiques de la parole continue qui contribuent à la perception de celle-ci comme constituant un flux auditif cohérent, rythmé et intonné de façon naturelle. Lorsque nous parlons, nous ne sommes en général pas conscients des mouvements complexes des muscles de la phonation, et il en va de même en particulier pour le contrôle de la hauteur et de l'intensité de la voix. Ces deux paramètres, avec les durées successives des segments syllabiques, constituent dans leur évolution la prosodie de la phrase.

---

## VARIATIONS ACCENTUELLES ET INTONATIVES

---

Dans le flux sonore, certaines syllabes apparaissent comme saillantes, elles sont accentuées. Cette saillance peut résulter de plusieurs faits, qui peuvent être en corrélation les uns avec les autres : les syllabes peuvent se distinguer par le fait qu'elles sont prononcées à un plus haut niveau sonore (« accent d'intensité ou accent », *cf.* <son petit a**MI**>), par leur longueur également, à savoir deux distinctions acoustiques de la parole (intensité et durée), que l'articulation en voix forte pourrait renforcer.

Cependant, les syllabes accentuées peuvent aussi se distinguer des autres syllabes parce qu'elles coïncident avec une rupture mélodique. En effet, la réalisation de la parole est marquée en français par des variations dans la hauteur de la voix, ce que les linguistes appellent l'intonation ou la mélodie.

### Mélodie et *syntaxe* :

Les syllabes portant l'accent d'intensité se caractérisent ainsi par la présence d'un contour mélodique montant ou descendant et sont toujours placées à la fin d'un mot isolé ou d'un groupe de mots. L'accentuation a donc une fonction démarcative, marquant la fin de chaque unité de sens dans le discours et permettant ainsi à l'auditeur de structurer (« segmenter ») la phrase pour reconstruire le sens global de l'énoncé. Un des rôles prépondérants de la mélodie est de donner ainsi à l'auditeur une première information sur la structure syntaxique de l'énoncé (*cf.* [Rossi et al., 1981], [Martin, 1981]). L'exemple de la Figure 2.1 (*cf.* [Welby, 2002] et [Welby, 2003]) illustre ce lien entre mélodie et segmentation.

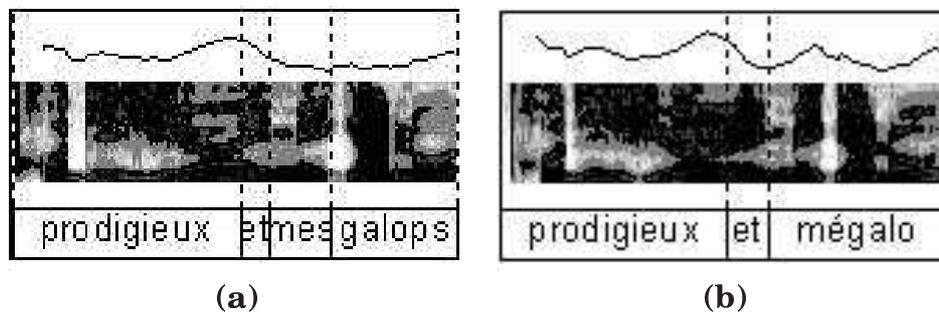


FIG. 2.1 – (a) Segmentation en deux mots (« mes ga**LOPS** »), (b) Segmentation en un mot (« méga**LO** »).

Des courbes mélodiques différentes correspondent à des valeurs énonciatives différentes (interrogation, assertion, phrase non terminée, etc.). Ces mouvements mélodiques sont néanmoins très variables et reflètent nos émotions et caractéristiques individuelles à l'intérieur d'une situation de communication. Dans le bruit, on pourrait donc s'attendre à ce que l'articulation soutienne la segmentation. En effet, il est possible que le locuteur, dans le désir de rendre sa parole intelligible, s'attache principalement à transmettre la forme syntaxique de ses phrases et aider l'auditeur dans sa tâche de segmentation.

### Mélodie et *sémantique* :

La corrélation mélodie-syntaxe fait encore l'objet de nombreuses discussions entre chercheurs phonéticiens et linguistes. Beaucoup pensent que cette corrélation est épisodique (*cf.* [Vaissière, 1977]) et que la mélodie est davantage à relier au sens de la phrase. Selon eux, la mélodie relève d'un domaine de nature sémantique, ou même gestuelle.

Les travaux de [Dohalská and Mejvaldová, 2000] sur l'analyse de la parole professionnelle dans des milieux industriels sujets aux fortes expositions sonores ont montré l'importance de l'intonation dans l'intelligibilité et la compréhension de l'information communiquée, le contenu verbal du message n'étant compréhensible qu'à 44%. L'information intonative se montre plus robuste que l'information purement lexicale, qu'en

est-il de l'information articulatoire ?

## VARIATIONS ARTICULATOIRES

Selon certaines études, (*cf.* [Schulman, 1989]), la parole en voix criée consiste davantage en une réorganisation de l'ensemble des mouvements articulatoires qu'en une amplification des mouvements normaux. Les analyses acoustiques ont montré que les voyelles tendent à être allongées alors que les consonnes sont plus courtes et plus dénaturées. Partant du lien constaté entre la durée des voyelles et l'effort vocal, [Fonagy and Fonagy, 1966] ont vu dans l'allongement des voyelles en parole Lombard la marque que les voyelles portent l'essentiel de l'information perçue par l'auditeur ; les voyelles en effet sont plus audibles que les consonnes à grande distance et ressortent davantage en environnement bruité. A cette explication perceptive, Schulman préfère l'argument selon lequel l'allongement temporel des voyelles découle de l'ouverture articulatoire que nécessite alors le mécanisme de production vocale. [Schulman, 1989] a montré que l'amplification de ces mouvements ne se déduit pas systématiquement d'une relation linéaire avec l'articulation en parole normale, mais implique la réorganisation de mouvements spécifiques motivée par un objectif de communication précis. C'est précisément l'objet de ce mémoire que de préciser la nature de cette réorganisation.

## 2.2 Variations articulatoires de la parole

### 2.2.1 Physiologie de la production de la parole

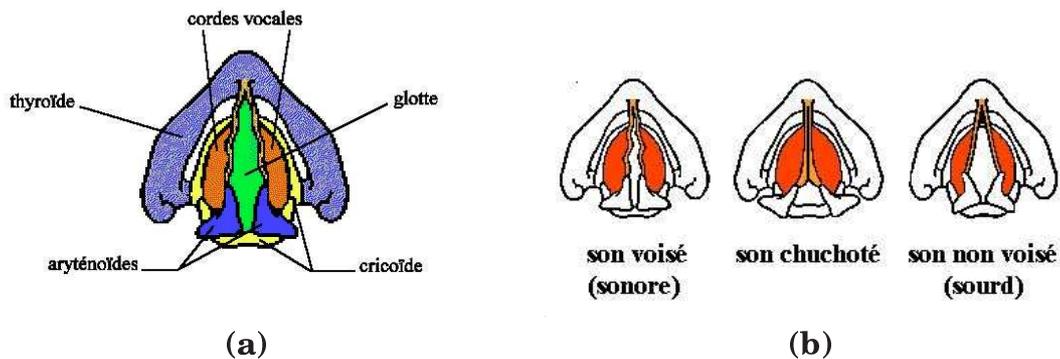


FIG. 2.2 – (a) Schéma en coupe du larynx, (b) La glotte en trois positions de phonation.

La production de la parole est composée de deux fonctions mécaniques de base : la *phonation* et l'*articulation*.

#### 2.2.1.1 La phonation

La phonation est la production du signal acoustique par vibration des cordes vocales. La fréquence fondamentale moyenne  $F_0$  de vibration des cordes vocales est située entre 140Hz et 240Hz pour les femmes, entre 100Hz et 150Hz pour les hommes. La mélodie de la voix résulte de cette vibration et se traduit phonétiquement par l'évolution de la  $F_0$ .

Les poumons jouent un rôle de soufflerie qui propulse une colonne d'air ascendante dans la trachée artère. La colonne d'air pulsée traverse le larynx, qui constitue l'organe phonateur. L'espace entre les cordes vocales est appelé la glotte (*cf.* Figure 2.2). La glotte s'ouvre lors de l'inspiration et se referme lors de la phonation, permettant aux cordes vocales de vibrer sous l'effet de la dépression de part et d'autre de l'espace glottique, ce qui génère un flux sonore appelé *voisement*. La production du voisement implique que les cordes vocales soient entièrement accolées et mises en vibration par le flux d'air ventilé par les poumons et véhiculée dans la trachée. Ainsi, la théorie acoustique de production de la parole distingue le mode « voisé », lorsque les cordes vocales vibrent périodiquement, et le mode « non voisé », lorsqu'elles ne vibrent pas. En réalité, le voisement se combine souvent avec une émission de bruit lors de la phonation (bruit d'aspiration, bruit de frication, bruit structurel, etc.).

### 2.2.1.2 L'articulation

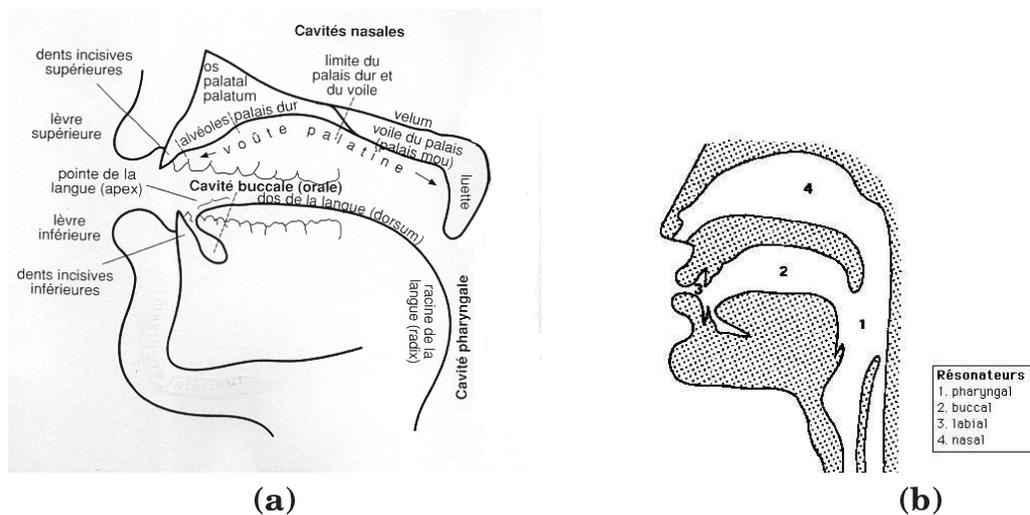


FIG. 2.3 – (a) Articulateurs de la parole, (b) Cavités supra-glottiques.

L'articulation inclut la modulation du signal acoustique par les articulateurs (principalement les lèvres, la langue et le palais mou) et la résonance de ce signal dans les cavités supra-glottiques (le pharynx, la bouche, les fosses nasales et la cavité labiale) (*cf.* Figure 2.3). Les deux premières sont toujours sollicitées pour l'articulation des sons de la parole alors que les cavités nasale et labiale n'interviennent que pour la réalisation de sons spécifiques. Si les lèvres sont projetées vers l'avant et arrondies, il se forme effectivement un résonateur à la sortie du canal buccal, le résonateur labial ; si au contraire, elles sont appliquées contre les dents, le résonateur labial ne se forme pas.

Les phonèmes ainsi produits, *i.e.* les unités distinctives minimales, sont divisés en consonnes sourdes et voisées et en voyelles. On peut exciter le résonateur complexe de l'appareil phonatoire de différentes manières, le mode d'excitation étant fonction du phonème à produire :

#### ARTICULATIONS VOCALIQUES

#### Excitation *glottique* du conduit vocal :

La réalisation des voyelles (et des liquides également) nécessite l'émission de voisement (cf. 2.2.1.1). Le larynx génère alors un signal sonore périodique, modulé par le conduit vocal qui favorise certaines zones fréquentielles (les régions formantiques). C'est ce qui fait qu'un « on » paraît plus grave qu'un « i ».

---

## ARTICULATIONS CONSONNANTIQUES

---

L'air expiré fournit la source d'énergie nécessaire à la production du bruit caractéristique des consonnes, voisées ou non.

### **Excitation du conduit vocal par une *impulsion acoustique* :**

L'articulation consonnantique s'effectue dans ce cas par de petites explosions créées par le relâchement brusque d'une occlusion réalisée le long du chenal expiratoire. Cette occlusion, qui rétrécit ou bloque le passage de l'air, entraîne une accumulation de la pression à l'intérieur de la bouche et c'est le bruit de l'air que nous entendons lorsque le relâchement soudain de l'occlusion se produit. Le bruit d'explosion est ainsi caractéristique des consonnes *occlusives* (exemple : [t], [k])

### **Excitation du conduit vocal par un *bruit de friction* :**

L'articulation naît ici du bruit du passage de l'air pulmonaire au travers des organes phonateurs, c'est-à-dire qu'il se produit un bruit d'écoulement (« friction ») là où le chenal expiratoire est suffisamment rétréci, au « point de constriction ». Ce resserrement donne naissance aux articulations consonnantiques *constrictives* (exemple : [s], [z]).

Trois phénomènes articulatoires peuvent accompagner la production du bruit consonnantique : le voisement (les consonnes voisées comme [g] se doublent alors d'une composante vocale), la nasalité et la labialité. Le voisement a pour conséquence une atténuation des bruits d'explosion et de friction, et la disparition du bruit d'écoulement glottal.

En résumé, il y a vibration des cordes vocales, donc présence de  $F_0$ , au cours de l'émission des voyelles et des consonnes voisées ([b], [d], [g], [v], [z], [m], [n], [l], etc.). Les consonnes sourdes ([p], [t], [k], [f], etc.) introduisent une interruption dans la ligne mélodique décrivant l'évolution de  $F_0$ , de la même manière que les pauses ou les silences.

Parmi les consonnes voisées, certaines font apparaître des chutes de  $F_0$  significatives, de l'ordre de 10 à 20Hz, par rapport aux voyelles adjacentes et à l'évolution globale de la courbe mélodique. Ces chutes sont dues essentiellement à une diminution de la pression de la colonne d'air expiratoire (pression sous-glottique) pendant la réalisation de ces consonnes. C'est le cas pour [b], [d], [g], [v], [z]. Les consonnes [l], [m] et [n] n'entraînent pas de perturbation locale de la courbe mélodique. Nous expliquerons par la suite (cf. 3.1.3) l'importance de ces considérations pour le choix des phonèmes dans l'élaboration de notre corpus.

## 2.2.2 Paramètres de l'étude articulatoire

La figure ci-dessus met en évidence le mouvement de certains des articulateurs de la parole (de la langue en particulier) selon le type de la voyelle prononcée (antérieure ou postérieure, ouverte ou fermée, etc.). On observe de même lors de la production d'une

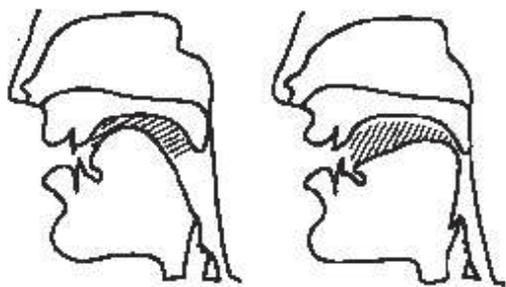


FIG. 2.4 – Voyelle antérieure (a) vs voyelle postérieure (b)

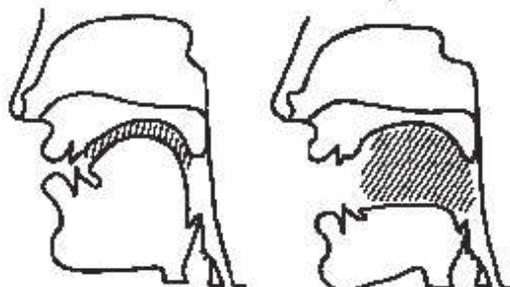


FIG. 2.5 – Voyelle fermée (a) vs voyelle ouverte (b)

consonne un mouvement très singulier de la langue et de la mâchoire selon les traits phonétiques de la consonne (mode d'articulation, lieu d'articulation, voisement). Nous décrivons ici les mouvements à étudier pour une analyse articulatoire complète :

### ***Rotation de la mâchoire***

La mâchoire étant l'articulateur qui supporte en grande partie la langue, son mouvement constitue le premier mouvement de base. Le mouvement de rotation de la mâchoire entraîne un mouvement de bascule de la langue.

### ***Mouvement lingual***

Les mouvements importants liés à la masse globale de la langue sont :

- le mouvement du corps de la langue (arrière-bas/avant-haut) qui correspond approximativement à un mouvement [a]/[i] ;
- Le mouvement du dos de la langue qui permet de se ramasser en boule vers l'arrière, comme pour le son "ou" dans "loup" ou au contraire de se détendre.

Les mouvements concernant plus particulièrement la pointe de la langue, si important pour les consonnes telles que [s], [t] ou [l] sont :

- Le mouvement vertical de la pointe de la langue qui permet le contact, partiel ou total, avec le palais dur et les incisives supérieures ;
- Le mouvement d'avancée/rétraction de la pointe de la langue qui gère les mouvements avant/arrière, en particulier pour la distinction entre [s] et [sh].

### ***Aire intero-labiale***

Le mouvement d'ouverture/fermeture des lèvres constitue un mouvement important du français, puisque c'est essentiellement celui qui permet toutes les consonnes plosives labiales ([p], [b], [m], etc.).

### ***Protrusion labiale***

Un autre mouvement caractéristique est le mouvement de protrusion. Il s'agit d'un mouvement de rétraction des lèvres, qui permet de passer de la voyelle [i] à la voyelle [y].

La langue et la mâchoire sont donc deux articulateurs fondamentaux dont les posi-

tions sont déterminantes dans la forme du conduit vocal. Cependant, nous nous limitons dans le cadre de ce mémoire à l'étude des mouvements labiaux. Seules les évolutions de l'aire intero-labiale et de la protrusion seront donc analysées à l'issue de notre expérience pilote. Un raffinement du traitement articulatoire est envisagé à la suite de cette étude via un appareil de mesure basé sur l'enregistrement électromagnétique (EMA) des mouvements linguaux et des contacts palataux, l'articulographe<sup>2</sup> (cf. Figure 2.6). Le système est composé d'un casque sur lequel sont fixées trois bobines émettrices positionnées respectivement au niveau du front, du menton et de la nuque du sujet. Des courants alternatifs à 3 fréquences différentes sont utilisés pour chacune de ces 3 bobines, générant ainsi 3 champs magnétiques différents à l'intérieur du casque. Des bobines réceptrices sont fixées au niveau des incisives supérieures du locuteur (point référentiel) et des incisives inférieures pour localiser la mâchoire. Deux autres capteurs sont positionnés sur la langue (pointe, partie laminaire) et un cinquième capteur est placé sur le velum. La tension induite dans chacun de ces 5 récepteurs est convertie utilisant une loi de la physique selon laquelle la puissance d'un champ magnétique d'une bobine réceptrice est inversement proportionnelle au cube de sa distance à la bobine émettrice. La position et donc le déplacement de chacun des 5 récepteurs par rapport au bobines émettrices peut ainsi être calculée, ce qui permet l'étude des mouvements de la langue et du velum.

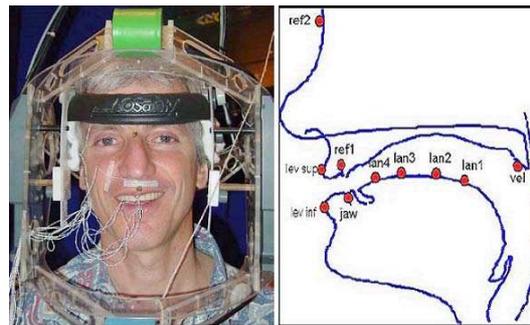


FIG. 2.6 – Articulographe

<sup>2</sup>L'articulographe AG100 dont dispose l'ICP a été développé par la société Carstens (Allemagne)

# Chapitre 3

## Expérimentation

Une première expérimentation pilote a été menée en collaboration avec l'ICP. Nous présentons ici sa mise au point et le traitement des données articulatoires.

### 3.1 Élaboration du corpus

#### 3.1.1 Contraintes liées aux hypothèses de l'étude

##### 3.1.1.1 Interaction et communication

Selon nos hypothèses, l'effet Lombard n'est pas pur réflexe physiologique mais répond au désir du locuteur d'être plus audible et plus intelligible. En effet, nous pensons que l'articulation en parole Lombard sert à augmenter l'intelligibilité de la parole, et n'en est pas seulement une conséquence. C'est pourquoi nous nous attendons à une hyper-articulation qui appuierait la thèse d'une articulation orientée au service des impératifs perceptifs et du désir d'intelligibilité.

Pour pouvoir tester de telles hypothèses, nous sommes contraints de placer les locuteurs à enregistrer dans des situations de communication et d'interaction. Le locuteur doit éprouver dans notre corpus le besoin de se faire entendre et/ou de s'entendre soi-même comme il le ferait naturellement au cours d'une conversation perturbée par le bruit ambiant. Par ailleurs, d'une façon plus générale, nous ne croyons pas à la parole « pour soi ». On parle toujours à quelqu'un. D'un point de vue méthodologique, nous sommes donc assez opposés à de la lecture. D'où la première nécessité à nos yeux de placer le locuteur en interaction avec un auditeur.

##### 3.1.1.2 Nécessité de définir des syntagmes cibles

À ce stade, nous pourrions a priori nous contenter d'enregistrer une conversation telle quelle, sans soumettre les locuteurs à aucune autre contrainte que de parler dans un environnement sonore imposé. Une conversation spontanée enregistrée dans un environnement bruyant pourrait très bien convenir à l'étude de l'intensité globale de la voix ou de l'élévation de la  $F_0$ .

Cependant, dans notre étude où nous cherchons à observer l'évolution des mouvements articulatoires, il nous faut :

- pouvoir comparer des locuteurs entre eux et comparer les productions vocales d'un même locuteur dans le silence et dans le bruit ;

- pouvoir comparer l’articulation produite en début, fin et milieu de phrase, pour tester l’hypothèse que l’hyper-articulation soutient l’intonation et donc la segmentation (*cf.* 2.1.2.3 et 3.1.2.3) de l’énoncé. En effet, l’étude prosodique d’un même mot, placé en trois positions différentes dans un énoncé mais gardant a fortiori la même valeur sémantique, permet de tester l’influence de la structure syntaxique dans l’hyper-articulation.

De plus, l’outil statistique est le seul moyen à notre disposition dans l’étude de la parole pour traiter les données expérimentales et en dégager un comportement modélisable. Or la parole spontanée débitée sans aucun paramètre contrôlable présente dans l’analyse statistique des difficultés qui nous contraignent à opérer différemment. En effet, la parole spontanée n’est pas comparable. Elle diffère d’un locuteur à l’autre, et pour un même locuteur, elle est caractérisée par une trop grande variabilité dans les formes syntaxiques et le lexique utilisés. Pour ces deux raisons nous sommes obligés de définir :

- certains mots cibles ( ou « syntagmes » cibles) ;
- des patrons intonatifs fixés, de façon à ce que les mots cibles se retrouvent à des positions comparables ;
- des formes syntaxiques non ambiguës telles qu’il n’y ait pas plusieurs interprétations possibles de la phrase, et par conséquent, pas plusieurs formes intonatives différentes.

### 3.1.2 Premières restrictions sur le corpus en conséquence

#### 3.1.2.1 Les syntagmes cibles et les « phrases porteuses »

##### *Une analyse contraignante*

- En vue de permettre le bon suivi de la fréquence fondamentale  $F_0$ , seuls des phonèmes sonores ont été utilisés dans la définition des syntagmes et des phrases porteuses, *i.e.* celles contenant le mot cible. Nous nous sommes donc limités à l’utilisation des voyelles et des consonnes voisées : m, n, l, r ( et v éventuellement). Nous nous sommes ainsi restreints aux consonnes voisées qui ne perturbent pas la ligne mélodique (*cf.* 2.2.1). De plus, les spectrogrammes de ces phonèmes présentent de nombreuses similitudes avec ceux des voyelles ; on y découvre, le plus souvent, un fort premier formant de basse fréquence et l’analyse de segmentation des phrases au spectrogramme s’en trouve facilitée (*cf.* Figure 3.1).
- L’enchaînement alternant consonnes (C) et voyelles (V) selon le schéma syllabique /CV/ aide au bon suivi des mouvements articulatoires, *i.e.* à la détection des début et fin de syllabe, des pics de vitesse et d’accélération dans le tracé des données articulatoires. En effet, dans le cas des voyelles qui se suivent ou des consonnes qui se suivent, il est très difficile de repérer les frontières. Les frontières entre une consonne et une voyelle sont plus faciles à repérer. Cette condition est importante pour le degré de précision dans la mesure des durées des syllabes.
- Enfin, l’équilibre de la structure des syntagmes nominaux doit être respecté pour donner sens à l’analyse comparative. Les premiers résultats pourraient éventuellement aboutir à des tendances articulatoires distinctes selon le type de syntagme étudié : syntagmes à deux syllabes ou trois syllabes de contenu.

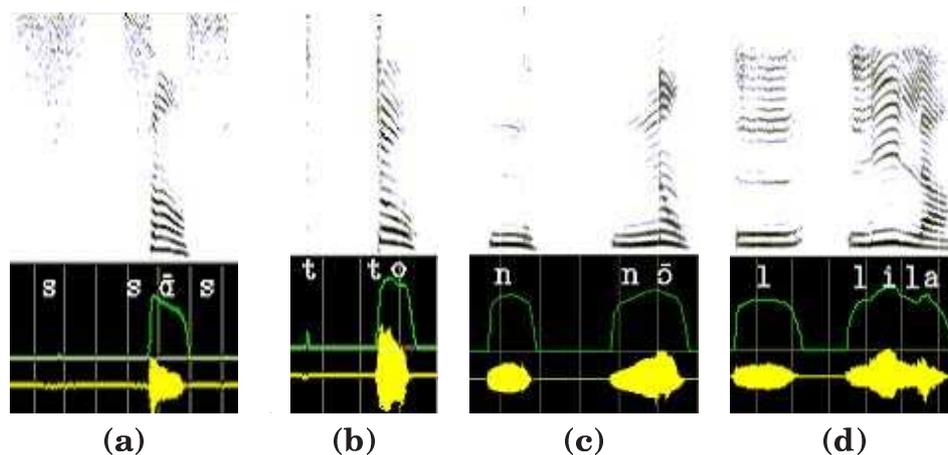


FIG. 3.1 – Spectrogrammes et contours de  $F_0$  de (a) « Sens », (b) « Tôt », (c) « Non », (d) « Lila ».

### *Choix des syntagmes*

Nous avons dans notre recherche accordé la plus grande importance au *sens des phrases et des syntagmes* susceptibles d'être énoncés au cours de l'expérience. En effet, nous tentons de reproduire le naturel d'une conversation, et les corpus se limitant à l'énonciation de syllabes isolées, ou de mots et de phrases dénués de sens ne peuvent répondre à de telles exigences. Nous partons du principe selon lequel l'hyper-articulation facilite et dirige non seulement la perception de la parole mais aussi la compréhension du message. Elle doit donc notamment aider à la reconnaissance et à l'accès lexical.

Les syntagmes ciblés choisis en définitive sont classés en deux catégories renfermant chacune deux sous-catégories :

1. Syntagmes à 2 syllabes de contenu
  - (a) avec un article : La mamie, le minet, les moulins, les lamas ;
  - (b) sans article : Loulou, Maman, Manon, Ninon ;
2. Syntagmes à 3 syllabes de contenu
  - (a) avec un article : Les monuments, les moulinets, la nominée, le minaret ;
  - (b) sans article : Mélanie, Marilou, Mélina, Marina.

#### **3.1.2.2 La forme syntaxique « SVO »**

Nous nous sommes interrogés sur la tournure syntaxique à adopter pour les phrases du corpus. L'étude de Marion Dohen [Dohen et al., 2003] sur le schéma « Sujet, Verbe, Complément » (SVO) montre de meilleurs résultats articulatoires. En effet, les patrons intonatifs ne présentent dans ce cas aucune ambiguïté, ce qui garantit l'unicité de l'interprétation que le locuteur peut faire de la phrase. Nous avons donc gardé ce type syntaxique dans notre corpus.

#### **3.1.2.3 Les trois positions du syntagme**

Hélène Løevenbruck [Løevenbruck, 2000] montre une différence de patron mélodique selon la position (initiale, intermédiaire ou finale) qu'occupe le syntagme dans la phrase

en parole normale. Notre étude visant à tester l'hypothèse selon laquelle l'hyper-articulation en parole Lombard soutient éventuellement le comportement intonatif, nous avons conservé la position du syntagme comme paramètre d'étude, de sorte que chaque mot-cible est étudié dans les trois positions dans les phrases du corpus.

À ce stade, aux vues de ces contraintes liées aux hypothèses de notre étude, nous avons réfléchi à un moyen qui nous permettrait de faire dire aux sujets les syntagmes cibles suivant des patrons intonatifs et des tournures syntaxiques imposés, tout en préservant le plus de *spontanéité* possible et surtout de sorte qu'ils gardent le désir d'*intelligibilité*.

### 3.1.3 Corpus semi-spontanés

Nous avons réfléchi à différentes alternatives pour contourner l'impasse d'une étude statistique en parole spontanée et satisfaire ces exigences d'interaction, de contenu sémantique, et de cibles imposées. En premier lieu, nous avons donc axé nos recherches vers l'élaboration d'un corpus semi-spontané, qui impliquerait les locuteurs au sein d'un dialogue interactif et dont ils auraient seuls l'initiative. Ils seraient orientés pour prononcer un certain nombre de mots cibles à étudier, tout en gardant la liberté de les utiliser à leur guise afin de préserver la spontanéité d'une interaction communicative.

#### 3.1.3.1 Des exemples dans la littérature

##### *Les Questions/Réponses*

Une première idée, la plus intuitive a priori, consisterait à poser des questions orientées au locuteur, dans l'attente de mots cibles spécifiques en guise de réponse.

Cependant, compte tenu des restrictions imposées dans le cadre de notre étude, on ne peut raisonnablement laisser au locuteur l'initiative des réponses, si l'on espère en même temps lui faire prononcer les syntagmes attendus dans des énoncés ne contenant que les consonnes m, n, l, r et v. De plus, les réponses risquent de varier de manière significative d'un locuteur à l'autre ce qui rendrait les données expérimentales inexploitable par méthode statistique.

##### *Le « MapTask »*

Le « MapTask » est l'alternative déjà utilisée par [Brown et al., 1983] pour impliquer le locuteur dans le naturel d'un dialogue communicatif. Le corpus initial est composé de 128 dialogues dont une partie est accessible en ligne. Le schéma et le protocole du MapTask sont simples et correspondent assez bien à nos objectifs. Il s'agit d'un « jeu de piste virtuel » au cours duquel le locuteur doit suivre les indications de son partenaire, que nous nommerons le meneur, pour arriver à destination et remporter la partie. Chacun possède une grille de jeu différente (*cf.* exemple annexe 7.2), la grille du meneur présentant le chemin à suivre. Les indications d'orientation portent le nom des syntagmes à étudier. Le locuteur s'oriente sur sa grille en interrogeant son guide. Il lui est simplement imposé de prononcer les mots-cibles dans une position définie dans la phrase, de sorte que le locuteur vienne à les répéter de manière quasi-naturelle.

Une seule position est ainsi testée par mot-cible. Etant données ces règles du jeu, les phrases porteuses peuvent être très différentes d'un syntagme à l'autre.

### 3.1.3.2 Des propositions novatrices

#### *Le puzzle*

Nous avons réfléchi à un moyen qui permettrait de réduire les variations des phrases porteuses sur le corpus.

Les syntagmes sont répartis sur les pièces d'un puzzle. Le locuteur déplace donc les mots-cibles pour reconstituer l'image du puzzle. Tout déplacement doit être formulé à voix haute selon un schéma syntaxique imposé :

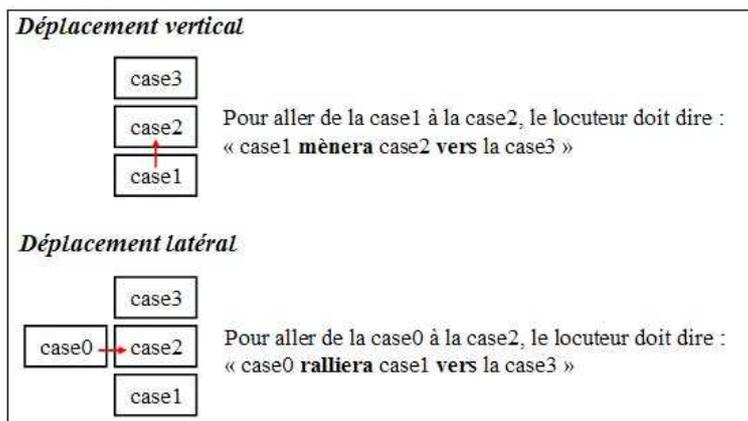


FIG. 3.2 – Sujets F1 et F2 de l'expérience 'Lip Tracking'.

Ce schéma permet ainsi d'énoncer le syntagme dans les trois positions possibles. Si le puzzle est bien construit, les trois positions sont testées pour chaque mot-cible. Cependant, le jeu dans cette première version n'instaure aucune interaction communicative. Il est donc nécessaire de faire intervenir un "maître du jeu". Certains déplacements sont maintenant interdits, ils sont schématisés par des cases noires. Seul le maître du jeu possède la version du puzzle avec les cases noires. Avant tout déplacement, le locuteur doit en l'annonçant obtenir l'accord du maître du jeu.

Non seulement la mise au point de tels puzzles relève du défi, mais alors leur résolution demande trop de temps et de réflexion, même dans des conditions de silence. L'excès de charge cognitive se fait alors ressentir sur le flux de parole et la spontanéité du joueur.

#### *Le jeu interactif*

Nous nous sommes alors attelés à l'élaboration d'un jeu plus simple dans sa résolution. Son principe de base est similaire à celui du puzzle : déplacer un pion sur une grille présentant soit des cases noires soit des cases où figurent les syntagmes. Le locuteur annonce ses déplacements au maître du jeu (comme précédemment) en suivant le trajet le plus direct possible. La grille est construite de sorte que tous les mots-cibles soient cités au cours d'une même partie, et qu'ils le soient dans les trois positions. Pour se faire, trois phrases porteuses sont nécessaires (*cf.* annexe \*\*\*\*\*).

### 3.1.4 Les contraintes d'acquisition expérimentale

Les méthodes d'enregistrement articulatoire et acoustique imposent de nouvelles contraintes au corpus. Le corpus élaboré au final résulte à ce titre d'un compromis

qui doit permettre l'acquisition et l'analyse des données expérimentales, ce qui inévitablement limite le champ d'interaction auditeur/locuteur et le contenu sémantique des phrases.

Il existe deux méthodes d'acquisition des données articulatoires développées à l'ICP :

- le '*Lip Tracking*' (cf. [Lallouache, 1991]) basé sur l'enregistrement audiovisuel des mouvements labiaux du locuteur, qui nécessite le port de lunettes opaques à des fins de protection, compte tenu des conditions d'éclairage.
- l'*articulographe* (cf. 2.2.2).

Le dispositif du Lip Tracking est beaucoup moins pénible pour le locuteur que ne l'est l'articulographe. Plaçant le confort du sujet comme la première de nos priorités, l'expérience étant déjà par elle-même assez éprouvante (en effet, le locuteur est plongé dans des environnements sonores désagréables), nous avons opté pour l'enregistrement vidéo par Lip Tracking pour mener notre première expérience pilote. Un raffinement des données est possible par la suite via l'articulographe. Nous devons en contrepartie renoncer à la modalité visuelle au cours de l'expérience. Les corpus semi-spontanés type Map-Task, puzzle ou jeu interactif présentés au paragraphe 3.1.3 deviennent inapplicables.

### 3.1.5 Choix définitif du corpus

Ainsi, les contraintes expérimentales ont fortement restreint la latitude dont nous disposons a priori pour l'élaboration décisive du corpus. Le corpus élaboré pour notre expérience pilote consiste en 33 phrases type SVO contenant un ou deux mots-cibles à analyser en 3 positions (initiale, intermédiaire, finale). Dans le cadre de l'étude pilote, chaque phrase du corpus est donc lue par l'un des expérimentateurs dans un environnement non bruyant, en parole chuchotée (sans  $F_0$ ), de sorte à ne pas influencer le locuteur dans la mélodie à utiliser. Le corpus dans sa dernière version figure en annexe 7.1. Les trois phrases porteuses sont différentes pour un même syntagme et réparties aléatoirement dans le corpus, afin d'éviter l'effet de focalisation (marquage inatontif et accentuel) sur le mot-cible par habitude de la phrase porteuse (cf. [Dohen et al., 2003]). Certaines phrases sont inspirées des premières phrases porteuses élaborées pour les corpus semi-spontanés du jeu interactif ou du puzzle ; elles ne présentent donc pas beaucoup de sens énoncées hors jeu mais leur étude confrontée aux résultats d'autres expériences qui permettraient une réelle mise en situation de jeu avec la modalité visuelle peut se révéler intéressante.

Nous avons donc finalement opté pour un corpus lu par le locuteur, dans l'optique d'une prochaine expérience qui tiendrait compte de la modalité visuelle et qui serait menée sur la base du jeu interactif.

## 3.2 Protocole expérimental

### 3.2.1 Stimuli utilisés

L'effet Lombard est dans la réalité expérimentale combiné à d'autres sources de variabilité (stress, fatigue, etc.), de sorte qu'il est difficile de le séparer de ces autres effets. D'où l'importance pour une telle étude d'enregistrer la parole dans des conditions de bruit et de communication bien contrôlées. Pour répondre à de tels objectifs, nous avons

choisi les deux bruits suivants, les plus étudiés dans la littérature :

- un bruit blanc. Il s'agit du bruit le plus « normalisé » dans la mesure où il couvre toute la largeur de la bande des fréquences audibles (de 20Hz à 20kHz);
- un bruit d'ambiance type « cocktail-party ». Ce bruit plonge le locuteur dans un environnement réaliste et intéressant car il s'agit d'un des bruits les plus communément rencontrés dans notre quotidien (bar, restaurant, discothèque, etc.) et il couvre la gamme des fréquences de la parole en particulier (de 80Hz à 8000Hz). Ce bruit est issu d'une base de données **BD\_BRUIT**, spécialement conçue pour permettre l'étude des perturbations de la production de parole par le bruit environnant.

Ces bruits ont été diffusés par hauts-parleurs avec un niveau de 85dBC.

### 3.2.2 Méthodologie

#### 3.2.2.1 Sujets et dispositif d'enregistrement

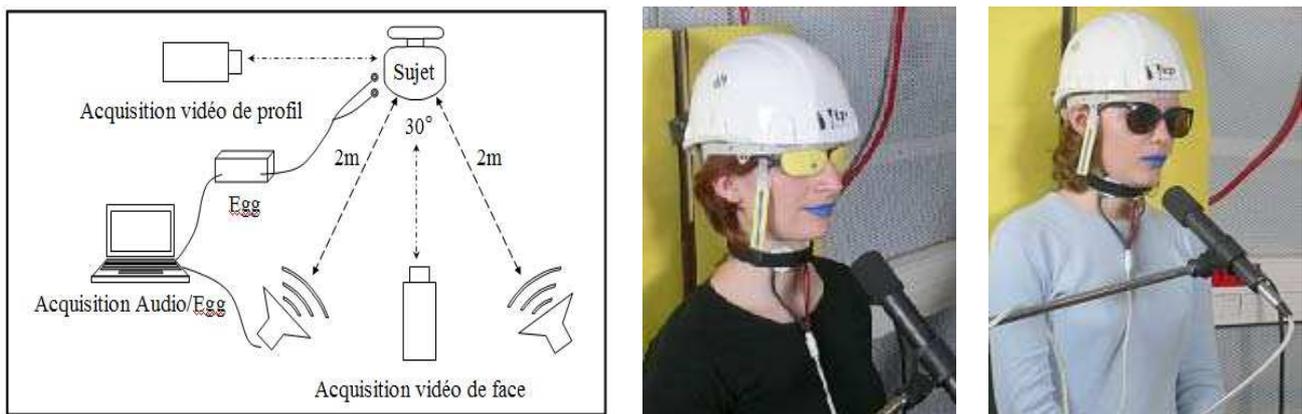


FIG. 3.3 – Protocole et sujets F1 et F2 de l'expérience 'Lip Tracking'.

Deux locutrices (F1 et F2) ont été filmées et enregistrées avec le dispositif labiométrique du 'Lip Tracking' (cf. [Lallouache, 1991]). Les enregistrements audio-visuels des deux locutrices ont été effectués dans la chambre anéchoïque de l'ICP. Pendant une séance d'enregistrement, le signal acoustique (constitué de la parole du locuteur et du bruit diffusé par hauts-parleurs) ainsi que le signal électroglottographique (Egg) du sujet ont été enregistrés simultanément.

#### *Acquisition du signal audio*

La chaîne de mesure du signal acoustique est constituée d'un microphone cardioïde dynamique (shure) placé à 50cm du sujet, et d'un amplificateur. Le signal est numérisé directement sur une carte son Sound Blaster Extigy à une fréquence d'échantillonnage de 44.1 kHz sur 16 bits.

#### *Acquisition du signal Egg*

L'électroglottographe est un dispositif qui permet d'obtenir une image de l'accolement et de l'ouverture des cordes vocales au moyen de la mesure de l'impédance électrique entre deux électrodes placées de part et d'autre du larynx (*cf.* Figure 3.3). Bien qu'il ne donne pas une information réelle de la surface d'ouverture de la glotte, il permet par contre une très bonne représentation du cycle d'oscillation des cordes vocales, et donc une bonne mesure de sa fréquence ( $F_0$ ) car il est dénué de bruits extérieurs à la parole (bruits diffusés par hauts-parleurs et bruits aérodynamiques).

Lors d'une calibration préalable, le sujet émet un son tenu, dont l'intensité acoustique est mesurée à l'aide d'un sonomètre analogique, placé au niveau du microphone. Le signal Egg est enregistré en synchronie avec le signal acoustique, à l'aide d'un électroglottographe à deux voies.

### *Acquisition des images vidéo*



FIG. 3.4 – Enregistrement vidéo de face et de profil du sujet 2 pour deux instants d'élocution différents.

Le 'Lip Tracking' est basé sur une détection des contours labiaux sur des images vidéo de face et de profil du visage du locuteur dont les lèvres ont été maquillées en bleu (*cf.* Figure 3.3). L'ICP possède un banc d'acquisition et de stockage d'images vidéo. L'acquisition des images est réalisée par l'intermédiaire de 2 caméras 3CCD<sup>1</sup>(*Ref.* JVC KY 15E), de face et de profil (*cf.* 3.3). Leur résolution est de l'ordre de 638x582 pixels et le rapport signal sur bruit de l'ordre de 55 dB. Le stockage des images est réalisée via des magnétoscopes Betacam de marque Sony (*Ref.* UVW 1400, UVW 1600 et UVW 1800). Une table de mixage permet de mixer la sortie des 2 caméras et obtenir ainsi sur une même bande des images de face et profil. Un exemple d'image obtenue est représentée Figure 3.4.

La règle fixée sur le casque permet de définir une position de référence immobile pendant toute la durée de l'enregistrement. Cette règle de référence est utilisée dans le traitement des données vidéo (*cf.* 4.1) pour le calcul des variations de la protrusion des lèvres. En effet, la protrusion de la lèvre supérieure (ou inférieure) est évaluée par la distance séparant l'extrémité de la lèvre supérieure (ou inférieure) et la règle de référence (*cf.* Figure 4.2 (a)). Cette distance en pixels a priori est convertie en centimètres

<sup>1</sup>Les CCD (Charge Coupled Devide) sont des dispositifs d'analyse photosensibles de type matriciel dont la densité des photo-éléments détermine leur résolution. Les CCD présentent une grande dynamique et d'excellentes précisions géométriques (pas de déformation d'image).

par les enregistrements vidéo, au commencement de l'expérience, d'un quadrillage placé au niveau de la tête du sujet, de face et de profil.

### 3.2.2.2 Protocole d'enregistrement

Le premier sujet F1 a dû porter des lunettes opaques au cours de l'expérience, afin de respecter les mesures préventives décidées à l'ICP depuis la création de ce dispositif. L'ICP utilise un shutter de 1/250e sur les caméras, ce qui nécessite un grand apport de lumière en compensation (néons fluorescents de 5 500 Watts). Une bonne acquisition de données nécessite un éclairage maximal pour favoriser la détection automatique des contours labiaux, et par conséquent le port des lunettes opaques à des fins de protection. Pour le sujet F1, chaque phrase du corpus a donc été lue par l'un des expérimentateurs dans un environnement non bruyant en parole chuchotée, ainsi que nous l'avions décidé dans notre protocole. Il était demandé au sujet de répéter en parole voisée la phrase entendue, d'abord dans le silence puis dans le bruit cocktail et enfin dans le bruit blanc à 85dBC, calibrés ux oreilles du locuteur avec un sonomètre.

Cependant, dans certaines manipulations où l'on ne peut faire autrement que de faire lire le locuteur, l'ICP adapte son dispositif en remplaçant les lunettes opaques par les lunettes de soleil de base et en redirigeant la lumière indirectement : les projecteurs ne sont alors pas dirigés directement sur la face du locuteur mais pointent vers le plafond, ce qui diffuse la lumière mais aussi l'atténue. Jusqu'à présent, l'ICP a eu très rarement besoin de ce type de modifications dans son dispositif et évite d'y avoir recours. Néanmoins, compte tenu de l'importance de la modalité visuelle dans notre étude où notre priorité consiste à impliquer le locuteur dans une situation d'interaction et de communication, nous avons eu l'accord de l'ICP pour modifier notre protocole et utiliser les lunettes de soleil. Limités par le temps, nous n'avons pu cependant mettre en place le jeu interactif pour cette étude pilote. Nous avons donc décidé de garder le même corpus de phrases à faire lire au sujet F2. Par la suite, nous pourrions donc tester le jeu interactif sur le même locuteur et comparer ses productions vocales dans le contexte du jeu et dans celui d'une lecture de corpus.

### 3.2.2.3 Un compromis réussi

Pour le sujet F1, le bruit ambiant a été interrompu avant chaque production du locuteur, afin de diffuser les phrases chuchotées dans le silence et lui permettre de les répéter dans le bruit. Cependant les interruptions régulières du bruit ambiant dans le premier protocole empêchaient la locutrice de se sentir réellement plongée en environnement bruyant. C'est pourquoi nous avons demandé au sujet F2 de lire les phrases du corpus, sans interruption du bruit ambiant. Ainsi, bien que nous ne soyons pas convaincus par la pertinence d'un corpus lu, le deuxième protocole s'est avéré mieux adapté à notre étude. Nous nous sommes donc concentrés sur l'analyse des productions vocales du sujet F2 dans ce mémoire.



Pour les mesures de face, l'aire peut être calculée à partir de l'aperture et de l'étirement :

$$S = k.(A \times B), \quad (4.1)$$

la constante  $k$  dépendant du locuteur et se situant autour de 0.75.

Pour les mesures de profil, l'aire intero-labiale peut se calculer à partir de l'étirement et de la commissure :

$$S = k' \times B.(Ref. - C), \quad (4.2)$$

avec  $k'$  constante et Ref. la mesure de la protrusion en position préphonatoire.

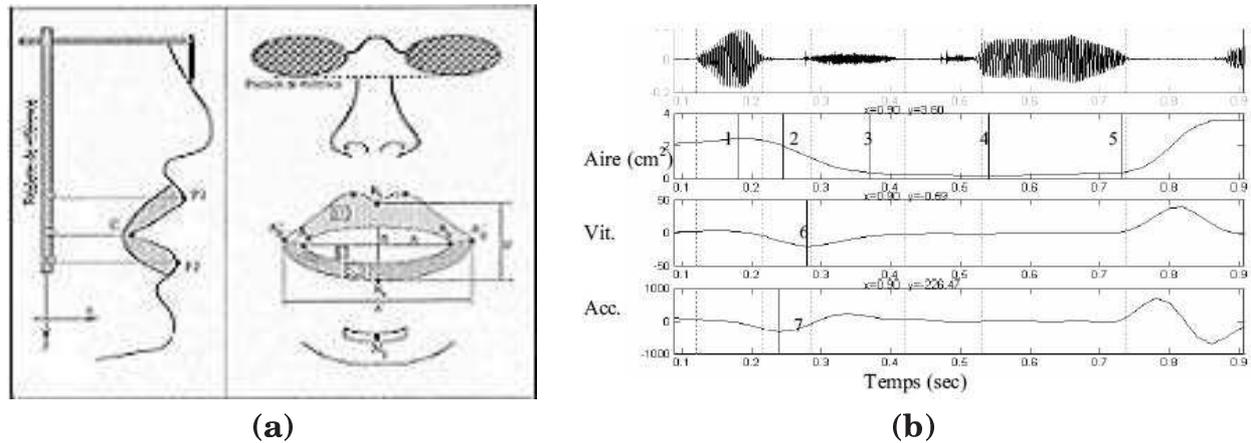


FIG. 4.2 – (a) Analyse du contour des lèvres sous *TACLE*, (b) Calcul de  $S$  ( $cm^2$ ), de sa vitesse ( $cm^2.s^{-1}$ ) et de son accélération ( $cm^2.s^{-2}$ ) en fonction du temps ( $s$ ) sous *TRAP*.

Le principe de traitement consiste à appliquer un chroma-key<sup>2</sup> sur l'image, de manière à seuiliser en noir les zones teintées en bleues, à filtrer les images via un filtre médian, puis à effectuer une détection de contours sur ces zones noires et enfin extraire les paramètres demandés à partir de ces contours. Ce système est très robuste et permet de détecter des aires intero-labiales extrêmement faibles pouvant atteindre  $0.5mm^2$ .

#### 4.1.2 Analyse articulatoire : TRAP

« TRAP<sup>3</sup> » est un logiciel intégré pour le TRAitement des signaux de Parole, développé à l'ICP (Løevenbruck, Savariaux & Lefebvre) et écrit sous environnement Matlab. Il permet d'effectuer des traitements et des analyses sur différents types de signaux : signaux acoustiques au format binaire (EMA, Matlab ou WAV) et signaux de mesures labiales de type TACLE (TAC). Il permet de visualiser simultanément les signaux de la parole et l'évolution temporelle des paramètres de l'étude articulatoire ( $S$  et  $P1$ ), de calculer les vitesses et accélérations de ces paramètres (cf. Figure 4.2 (b)) et de détecter les extrema (locaux et absolus) de chaque signal.

À partir des fichiers TAC répertoriant les données des segments d'images et des fichiers WAV correspondant aux segments audio (cf. 4.2), nous avons donc procédé pour

<sup>2</sup>« Incrustation en chrominance ». Il s'agit d'un code de chromaticité, pour modifier les couleurs. Il permet l'incrustation vidéo : un objet est filmé sur un fond uni de couleur pure ; après numérisation, on enlève ce fond par sélection de couleurs et on le remplace par un autre.

<sup>3</sup>[http://www.icp.inpg.fr/~savario/\\_activite/zip/trap\\_v4.zip](http://www.icp.inpg.fr/~savario/_activite/zip/trap_v4.zip)

chaque phrase du corpus à la détection des maxima et des minima de  $S$ ,  $P1$  et de leurs vitesses. Limités par le temps, nous avons préféré nous concentrer sur l'analyse du déroulement de  $S$ . Ce paramètre est en effet très pertinent sur le plan articulatoire et sur le plan acoustique pour l'arrondissement (cf. [Abry et al., 1980]). Une voyelle arrondie peut se caractériser par une faible protrusion, mais surtout par une aire latéro-labiale élevée. De plus,  $S$  est fortement corrélé à l'aperture des lèvres et à l'étirement labial, comme le montre la relation 4.1 au 4.1.1.

## 4.2 Traitement des données audio et Egg

### 4.2.1 Synchronisation audio/vidéo et decorrélation bruit/parole

La synchronisation des signaux audio et Egg avec les signaux vidéo segmentés a été effectuée à l'aide de l'intercorrélacion entre les signaux audio et Egg acquis sur l'ordinateur et le signal audio acquis avec la vidéo.

Un algorithme de débruitage (cf. [Ternström et al., 2002]) a permis de décorréler le bruit ambiant (bruit blanc et bruit cocktail) de la parole dans l'enregistrement audio. L'algorithme s'est avéré très efficace pour débruiter l'enregistrement effectué dans le bruit cocktail, moins pour le bruit blanc.

Nous nous sommes concentrés sur le traitement des données audio dans les conditions de silence et de cocktail.

### 4.2.2 Etiquetage des segments audio

Cette étape est un préalable à l'étude de paramètres prosodiques, phonologiques et acoustiques sur les unités segmentales : syllabes et mot. Nous avons procédé à l'étiquetage syntaxique « Sujet-Verbe-Objet » (SVO) des 33 phrases du corpus et à l'étiquetage en syllabes des 48 mots cibles, dans les conditions de silence et de bruit cocktail. L'étiquetage a été effectué à partir de la représentation temporelle et spectrale de la parole et de son écoute au moyen de l'éditeur de signal PRAAT<sup>4</sup> [Boersma and Weenink, 1996] (cf. Figure 4.3).

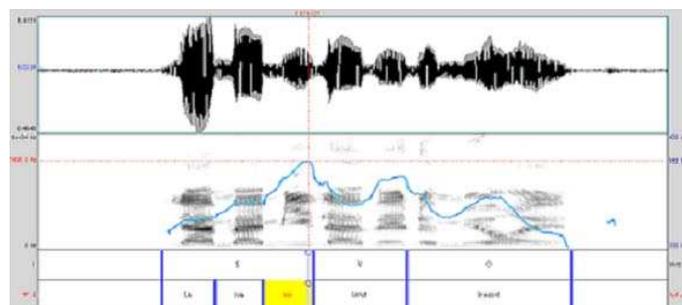


FIG. 4.3 – Segmentations SVO et syllabique sous *Praat* de la phrase 12 : « **La mamie lavait le muret** ».

Les frontières posées manuellement se situent dans une marge d'erreur au dessous de 10ms et la durée des segments n'est statistiquement affectée que de façon négligable compte tenu des prévisions d'allongement des phonèmes et des mots dans le bruit.

<sup>4</sup><http://www.praat.org>

### 4.2.3 Calcul de la $F_0$ via les signaux Egg

Le signal Egg donne une information sur la source laryngée et sur la périodicité du signal acoustique. Il se présente sous la forme d'une onde caractérisée par une modulation rapide, synchrone avec le cycle d'ouverture et de fermeture des cordes vocales, ainsi qu'une modulation basse fréquence, révélatrice des mouvements lents du larynx (déglutition, etc.) ou des mouvements de rotation du cou.

Nous avons déjà évoqué au 3.2.2.1 le lien entre la  $F_0$  et le signal Egg. Plus précisément la  $F_0$  est égale à l'inverse de l'intervalle de temps entre deux fermetures de la glotte, détectées sur la dérivée du signal Egg (cf. Figure 4.4).

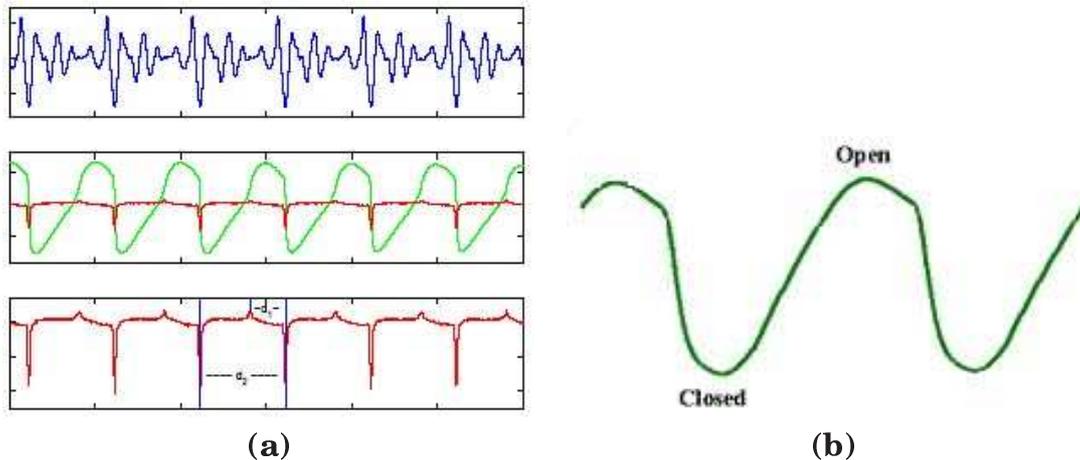


FIG. 4.4 – (a) Signal audio, signal Egg et sa dérivée ( de haut en bas), (b) Cycle d'accolement des cordes vocales sur le signal Egg.

Cette évaluation fidèle de la  $F_0$  à partir du signal Egg permettra dans l'analyse des résultats (cf. 5.2.1) d'observer l'évolution de l'intonation, en particulier les hausses et descentes mélodiques et de les confronter aux extrema de l'aire intero-labiale détectés sous TRAP.

Les signaux Egg donnent aussi de façon fiable une indication sur la durée d'ouverture de la glotte relative à la durée du cycle glottique, appelée le quotient ouvert ( $Oq$ , avec  $Oq = d_1/d_2$  selon les notations de la Figure 4.4). Ce paramètre  $Oq$  apporte des informations sur la variation de la qualité vocale. Nous nous sommes limités dans ce mémoire à l'étude de l'évolution de la  $F_0$  pour tester l'hypothèse d'une hyper-articulation soutenant la segmentation mélodique de la phrase. Il peut être intéressant dans la continuité de cette étude d'étudier l'évolution du quotient ouvert pour observer l'influence de l'hyper-articulation sur la qualité vocale.

# Chapitre 5

## Résultats et Discussion

Nous avons considéré la condition de bruit imposée (silence ou cocktail) et la position du syntagme cible dans la phrase (initiale, intermédiaire, finale) comme facteurs intéressants pour cette étude articulatoire. Une analyse statistique prenant en compte l'influence du nombre de syllabes dans la cible (2 ou 3) pourrait venir compléter cette étude pilote.

Nous nous proposons au 5.1 d'observer les variations statistiques de l'aire intero-labiale  $S$ , de la protrusion  $P1$  et de la durée  $d$  d'énonciation en fonction de la position des syntagmes dans la phrase et de l'environnement bruyant auquel a été soumis le locuteur. Nous complétons l'étude de ces paramètres articulatoires par l'étude acoustique des variations pendant l'élocution de la fréquence fondamentale  $F_0$  et de l'intensité sonore  $SPL$  de la voix de F2 (cf. 5.2). La confrontation de ces deux études articulatoire et acoustique nous permettra de comparer les extrema de  $S$ ,  $F_0$  et  $SPL$ , en amplitude et dans le temps, dans l'idée de tester nos hypothèses quant au rôle joué par l'hyper-articulation en parole Lombard.

### 5.1 Paramètres de l'étude articulatoire

#### 5.1.1 Étude de l'aire intero-labiale $S$ et de la protrusion $P1$

##### 5.1.1.1 Influence de l'environnement sonore

Le traitement des données vidéo sous TRAP a permis de déceler, au sein d'un même syntagme et dans les deux conditions sonores (silence et bruit cocktail), 0, 1, 2, 3 ou 4 maxima locaux de  $S$  ( $Max_S$ ), à raison d'un maximum au plus par syllabe. En effet, les mots cibles de notre corpus ont au plus 3 syllabes mais la détection des maxima a été effectuée sur les syntagmes, à savoir sur l'ensemble  $\{\text{déterminants et mots cibles}\}$  pour les 8 syntagmes correspondant à des noms communs. Pour un même syntagme, l'étude montre en moyenne un nombre identique de maxima locaux dans les deux environnements ; ils diffèrent par une amplitude plus élevée dans le cocktail, quelle que soit la position de la cible dans la phrase (cf. Tableau 5.1).

$S$  et suit une évolution monotone dans le silence comme dans le bruit ; il s'agit du syntagme « Loulou », pour lequel l'étude de l'aire aux lèvres n'est pas significative (cf. Tableau 5.1). On distingue dans les mots cibles les voyelles ouvertes [a], [i], [e], [ɛ], [ã], [ĩ] et les voyelles fermées [õ], [o] et [y]<sup>1</sup>. L'étude de l'aire aux lèvres pour les voyelles fermées

---

<sup>1</sup>cf. Notations de l'Alphabet Phonétique International (API)

n'est pas pertinente mais l'étude de la protrusion pour ces voyelles montre en revanche de nettes augmentations dans la condition bruyante. Pour l'ensemble des syntagmes, les maxima de  $S$  sont en grande majorité calés sur les voyelles ouvertes. Il semble ainsi que  $S$  soit un facteur pertinent pour l'étude articulatoire de ces phonèmes puisque l'aire intéro-labiale rend compte à la fois de l'aperture et de l'étirement des lèvres (significative respectivement pour la prononciation de [a], [ã], [ε] et de [i], [e], etc.).

	Rangs	Nombre de syntagmes
$S$ dans le cocktail – $S$ dans le silence	–	0
	+	45
	=	3

FIG. 5.1 – Test de Wilcoxon, avec « Rangs » égal à « + » si  $Max_S$  dans le silence  $>$   $Max_S$  dans le cocktail; « – » si  $Max_S$  dans le silence  $<$   $Max_S$  dans le cocktail; et « = » si  $Max_S$  dans le silence =  $Max_S$  dans le cocktail.

On observe donc effectivement une hyper-articulation labiale en parole Lombard, qui se réalise par une aperture, un écartement et/ou une protrusion plus grandes selon le phonème à produire. L'étude de la dérivée temporelle de  $S$  pourra dans une prochaine étude raffiner nos résultats en nous renseignant sur la vitesse d'élocution et la manière dont la locutrice hyper-articule.

Les résultats des maxima de  $S$  dans la condition cocktail ont été normalisés pour permettre l'analyse comparative des différentes cibles. Ainsi, pour chaque syntagme, la valeur des maxima de  $S$  dans la condition bruitée a été divisée par la valeur des maxima de  $S$  prise dans la condition de silence. Nous avons ainsi défini le pourcentage d'augmentation des maxima de  $S$  au passage de la condition de silence à la condition cocktail. Nous avons ensuite calculé la moyenne  $S_{max\_moy\_cible}$  sur toutes les cibles de ces pourcentages d'augmentation. Cette première approximation définit ainsi une unique valeur maximale de  $S$  par syntagme et  $S_{max\_moy\_corpus}$  définit la moyenne des  $S_{max\_moy\_cible}$  sur les 33 phrases du corpus. Le tableau 5.2 montre que l'aire intéro-labiale augmente de 40,5% en moyenne dans le bruit cocktail. Le minimum détecté, égal au minimum détecté dans le silence, est atteint pour le phonème [u] de la cible « Loulou ». Le maximum d'augmentation de 115%, est atteint pour la cible « Manon » en position intermédiaire.

$S_{max\_moy\_corpus}$	Ecart-type	Minimum	Maximum
1,405	0,259	1	2,152

FIG. 5.2 – Statistiques descriptives de l'évolution de  $S_{max\_moy\_cible}$  (en rapport d'augmentation) dans les deux conditions de bruit.

Nous avons envisagé de raffiner l'analyse en étudiant les syntagmes par syllabes. En effet, le moyennage des pourcentages d'augmentation de  $S$  sur le syntagme « minet » par exemple atténue le pic d'étirement labial du [i] hyper-articulé (plus élevé que le pic d'ouverture des lèvres dû à la prononciation du [ε]). Cette approximation suffit a priori à comparer la prononciation de ce même syntagme dans les deux conditions bruitées et à observer une hyper-articulation labiale significative dans le bruit; les maxima locaux de  $S$  étant en nombre identique et portant sur les mêmes phonèmes dans le silence et dans

le cocktail, l'approximation sur le syntagme a effectivement la même précision pour les deux environnements. Ce moyennage n'a plus de sens en revanche si l'on veut pouvoir comparer les syntagmes entre eux.

Nous avons donc décidé de segmenter l'étude articulatoire par syllabes. Cependant, la plupart des syntagmes de notre corpus ne présente pas de maximum de  $S$  à chaque syllabe ; notre analyse étant basée jusqu'à présent sur la détection des maxima locaux de  $S$  au sein de chaque syntagme, nous ne pouvons rendre compte du comportement articulatoire adopté sur chaque syllabe. La détection des maxima de  $S$  est révélatrice pour l'étude du contrôle moteur du sujet, i.e. pour connaître l'amplitude du geste articulatoire qu'il estime adapté à la situation. Cependant, l'hyper-articulation ne se manifeste pas nécessairement par un maximum articulatoire plus grand en bruit cocktail que dans le silence mais surtout par le fait que globalement sur tout l'intervalle de temps observé, la bouche a été plus ouverte dans le bruit ; d'où l'idée d'observer un autre descripteur.

Nous avons ainsi complété l'étude de l'augmentation des maxima articulatoires par celle des variations d'un nouveau paramètre qui nous permet de mieux différencier les cibles entre elles : l'aire sous la courbe de  $S$ , pour un intervalle de temps correspondant à la phrase, au mot cible ou à une syllabe en particulier. Ce paramètre permet également de mieux rendre compte des phénomènes de « coarticulation » car ne se limite pas à la détection de pics isolés : les articulations se succèdent très rapidement, une première articulation peut ne pas être achevée au commencement de la seconde et la réalisation de chaque phénomène peut varier considérablement en fonction de son voisinage.

Le tableau 5.3 présente les résultats de tests ANOVA portant sur l'aire sous la courbe et prenant en compte la segmentation par syllabes. Les valeurs indiquées ont été normalisées et représentent l'aire sous la courbe par unité de temps observé.

		Silence	Cocktail	Degré de signifiante p
$Aire_S$ ( $cm^2$ )	Phrase	1,333	1,899	$5,3 \cdot 10^{-3}$
	Cibles indifférenciées	1,256	1,758	$1,8 \cdot 10^{-5}$
	Syl. 1	1,278	1,785	0,0023
	Syl. 2	1,067	1,494	0,0243
	Syl. 3	1,173	1,686	0,0034
	Syl. de fin	1,127	1,664	0,0006

FIG. 5.3 – Statistiques descriptives de la différence moyenne de l'aire sous la courbe de  $S$  par unité de temps (en  $cm^2$ ) sur la phrase, le syntagme et par syllabes (première, deuxième, troisième pour les syntagmes à trois syllabes, et dernière pour les syntagmes à deux et trois syllabes), dans les deux conditions de bruit.

On retrouve une hyper-articulation moyenne significative sur la globalité de la phrase et sur les syntagmes cibles indifférenciés. On observe également une hyper-articulation significative ( $p = 0,0002$ ) sur la première syllabe de chaque syntagme.

La segmentation des phrases en voyelles ouvertes et fermées permet de mesurer la pertinence des paramètres articulatoires pour ces deux classes de voyelles. Des tests ANOVA portant sur l'aire sous la courbe et utilisant la segmentation voyelles/consonnes montrent que :

- Pour les voyelles ouvertes :

L'aire sous la courbe de  $S$  augmente significativement ( $p = 3,4 \cdot 10^{-5}$ ) de 50,1% ;

$P1$  n'augmente pas de manière significative ( $p = 0,1072$ ).

– Pour les voyelles fermées :

L'aire sous la courbe de  $S$  n'augmente pas de manière significative ( $p = 3,4 \cdot 10^{-5}$ ) ;

$P1$  n'augmente pas de manière significative ( $p = 0,4496$ ).

Il semble qu'on ne joue pas sur la protrusion dans l'hyper-articulation. Ce résultat pourrait être un indice du fait que l'hyper-articulation est seulement reliée à l'ouverture de la mâchoire pour parler plus fort ; ou bien que  $P1$  n'est pas un descripteur très significatif de l'hyper-articulation.

### 5.1.1.2 Influence de la position du syntagme

Nous avons mené une analyse statistique de l'influence de la position des syntagmes dans la phrase sur  $S_{max\_moy\_corpus}$  (cf. 5.4) mais les résultats ne sont pas significatifs.

		Init	Inter	Final	$P_{init/inter}$	$P_{inter/final}$	$P_{init/final}$
$Aire_S$ ( $cm^2$ )	Cible	+0,544	+0,506	+0,457	0,7106	0,6777	0,5438
	Syl. 1	+0,569	+0,588	+0,364	0,9144	0,2352	0,1948
	Syl. 2	+0,459	+0,414	+0,408	0,7247	0,9636	0,7516
	Syl. 3	+0,490	+0,531	+0,589	0,688	0,69	0,4957
	Syl. de fin	+0,557	+0,488	+0,566	0,5279	0,549	0,947

FIG. 5.4 – Variations de l'aire moyenne sous la courbe de  $S$  par unité de temps (en  $cm^2$ ) au passage de la condition silence à la condition de cocktail à 85 dB, sur le syntagme et les syllabes (première, deuxième, troisième pour les syntagmes à trois syllabes de contenu, et dernière pour les syntagmes à deux et trois syllabes de contenu) en fonction de la position initiale (Init), intermédiaire (Inter) ou finale (Final).

## 5.1.2 Étude de la durée $d$ d'énonciation

L'étiquetage effectué manuellement sous PRAAT a permis de détecter les débuts et fins de phrases, ainsi que les frontières inter-syllabiques de chaque mot cible. Ce travail a permis de relever la durée des phrases, caractéristique du débit de phonation, celle des mots cibles et de leurs syllabes pour le silence et le bruit cocktail.

### 5.1.2.1 Influence de l'environnement sonore

On constate l'allongement significatif des durées dans le bruit sur la phrase et sur les 16 cibles indifféremment de leur position. Le débit de phonation est donc ralenti en parole Lombard. Les résultats des tests ANOVA menés sur la durée montrent un allongement de la phrase de 0.23s en moyenne dans le bruit ( $p = 0.0038$ ). De plus, la locutrice prenait parfois le temps de marquer une pause entre le sujet et le verbe, ou à l'intérieur même du sujet (**ex** : « Momo [pause] et Marilou lavaient les moulins ») dans la condition de silence. L'étiquetage dans le bruit cocktail (et dans le bruit blanc également) montre que les groupes de mots sont davantage enchaînés et que le flux de parole n'est jamais interrompu.

Le ralentissement du rythme montre ainsi qu'en présence de bruit, le locuteur tend à parler plus lentement, malgré la perturbation qui lui est imposée et qui l'incite à fluidifier son débit de parole, *i.e.* sans interruption. Cela confirme les résultats d'études antérieures sur la durée d'énonciation en parole Lombard.

### 5.1.2.2 Influence de la position du syntagme

		Init	Inter	Final	$p_{init/inter}$	$p_{inter/final}$	$p_{init/final}$
$d (s)$	Cible	+0,030	-0,006	+0,095	0,0008	0,0950	0,0094
	Syl. 1	+0,020	+0,010	+0,005	0,08	0,5	0,07
	Syl. 2	+0,010	-0,005	+0,060	0,07	0,0059	0,03
	Syl. 3	+0,001	-0,017	+0,058	0,0419	0,002	0,0027
	Syl. de fin	+0,008	-0,018	+0,086	0,0085	$1,5 \cdot 10^{-5}$	0,006

FIG. 5.5 – Variations dans le bruit cocktail de la durée moyenne (en  $s$ ) du syntagme et des syllabes (première, deuxième, troisième pour les syntagmes à trois syllabes de contenu, et dernière pour les syntagmes à deux et trois syllabes de contenu) en fonction de la position initiale (Init), intermédiaire (Inter) ou finale (Final).

Nous avons ainsi noté des durées de phrases et de syntagmes plus longues en parole Lombard. Ces durées sont nettement plus allongées quand la cible est située en position « Objet » en comparaison des deux autres positions syntaxiques (« Sujet » et « Verbe »). En effet, le Sujet de la phrase s'allonge en moyenne de  $0.08s$  ( $p = 0.2557$ ) et le Verbe de  $0.03s$  ( $p = 0.0909$ ); l'Objet en revanche s'allonge de  $0.1s$  de manière significative ( $p = 1.10^{-5}$ ).

Nous avons étudié la corrélation entre  $S$  et la durée d'énonciation, pour voir si les maxima articulatoires étaient atteints dans les périodes d'allongement significatifs. Il semble qu'il n'y ait pas de corrélation entre ces deux paramètres sur les syllabes ( $R_{syl} = -0.0554$ , où  $R_{syl}$  désigne l'indice de corrélation sur les syllabes).

## 5.2 Paramètres de l'étude acoustique

### 5.2.1 Étude de la fréquence fondamentale $F_0$

#### 5.2.1.1 Influence de l'environnement sonore

Des tests ANOVA portant sur la  $F_0$  ont été réalisés sur les phrases entières. Les résultats acoustiques confirment la littérature :

- la  $F_0$  augmente significativement ( $p = 0$ ) avec le bruit (en moyenne de 229 à 290 Hz donc de 61 Hz);
- la valeur maximale de la  $F_0$  passe en moyenne sur une phrase de 290Hz dans le silence à 373 Hz dans le bruit donc de 83 Hz, et ce, toujours significativement ( $p = 0$ );
- l'ambitus des phrases est en moyenne de 276 Hz dans le silence et de 354 Hz dans le bruit soit un élargissement significatif ( $p = 0$ ) de l'intonation.

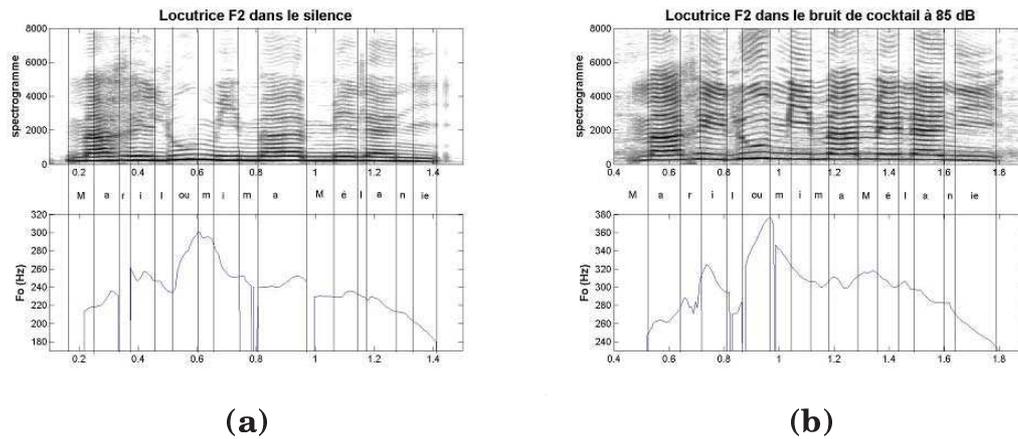


FIG. 5.6 – Variations de la  $F_0$  (a) dans le silence, (b) dans le bruit cocktail, lors de l'élocution par le sujet F2 de la phrase 10 : « *Marilou mima Mélanie* ».

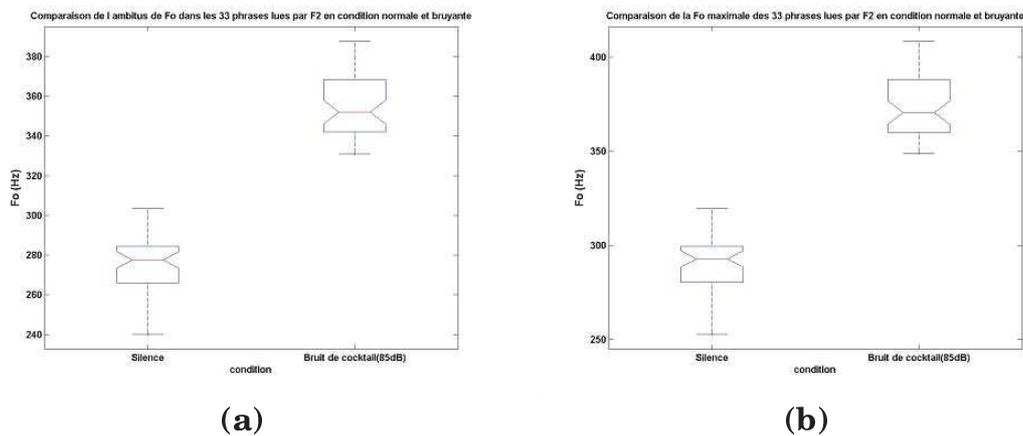


FIG. 5.7 – Comparaison de (a) l'ambitus de  $F_0$ , (b) la  $F_0$  maximale des 33 phrases lues par F2 en condition de silence et de bruit cocktail.

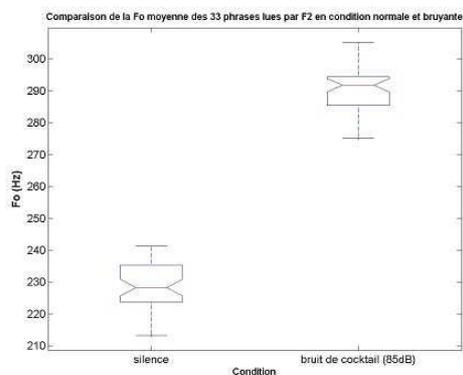


FIG. 5.8 – Comparaison de la  $F_0$  moyenne des 33 phrases lues par F2 en condition de silence et de bruit cocktail.

Une étude plus détaillée est menée en parallèle à l'ICP. Elle porte sur les variations intonatives du locuteur et pourrait nous renseigner sur leur rôle dans la segmentation et la façon dont l'hyper-articulation soutient la syntaxe de l'énoncé.

## 5.2.2 Étude de l'intensité $SPL$

### 5.2.2.1 Influence de l'environnement sonore

Des tests ANOVA menés sur le  $SPL$  en fonction de l'environnement bruité imposé au sujet F2 montrent :

- l'augmentation significative ( $p = 0$ ) de 8 dB de l'intensité moyenne de la voix dans un bruit de cocktail de 85 dB (cf. Figure 5.9 (a));
- l'augmentation significative ( $p = 0$ ) de 10 dB de l'intensité maximale moyenne de 10 dB entre la condition normale et bruyante (cf. Figure 5.9 (b));
- l'augmentation non-significative du registre d'amplitudes utilisées par la locutrice ( $\Delta = Max_{SPL} - Min_{SPL}$ ).

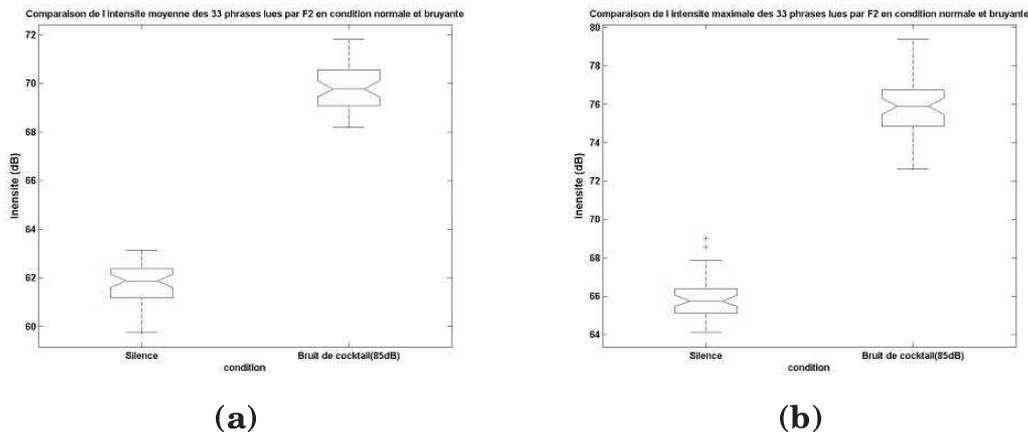


FIG. 5.9 – Comparaison du (a)  $SPL$  moyen, (b)  $SPL$  maximal des 33 phrases lues par F2 en condition de silence et de bruit cocktail.

On observe également dans le bruit que les maxima d'intensité suivent parfois l'évolution des maxima articulatoires (cf. annexes 7.3). Cette corrélation reste assez faible néanmoins ( $R_{syl} = 0,37$ ) sur les syllabes. On ne note aucune corrélation entre la durée et le  $SPL$  ( $R - syl = -0,054$ ). Le type de voyelle n'a pas d'influence sur ces corrélations. Effet, on trouve des indices de corrélation sur les voyelles ouvertes égaux à 0,06 pour l'étude corrélée du  $SPL$  et de  $d$ , à 0,37 pour le  $SPL$  et  $S$  et à  $4 \cdot 10^{-4}$  pour  $d$  et  $S$ . Ces indices valent pour l'étude des voyelles fermées, respectivement -0,44, 0,40 et -0,24.

### 5.2.2.2 Influence de la position du syntagme

Des tests ANOVA ont été menés sur les syntagmes en fonction de leur position. Les résultats de ces tests sont répertoriés dans le tableau 5.10 mais dépassent largement le seuil de signifiante.

Aux vues du signal  $SPL$  (cf. Figure 5.11), il semble que le phrasé de la parole soit pensé au niveau du groupe syntaxique (S,V,O) dans le silence et davantage au niveau de la voyelle dans le bruit cocktail. Une segmentation du corpus en voyelles/consonnes serait à ce titre certainement plus révélatrice qu'une segmentation syllabique.

$SPL_{moyen}$ (dB)		Init	Inter	Final
	Cible	+7,26	+7,75	+8,5
	Syl. 1	+8,2	+8,5	+9
	Syl. 2	+6,99	+7,06	+7,99
	Syl. 3	+6,5	+7,3	+7,8
Syl. de fin	+6,9	+7,7	+7,9	

FIG. 5.10 – Augmentation dans le bruit cocktail de l'intensité moyenne de la voix (en dB) dans le syntagme et par syllabes (première, deuxième, troisième pour les syntagmes à trois syllabes de contenu, et dernière pour les syntagmes à deux et trois syllabes de contenu) en fonction de la position initiale (Init), intermédiaire (Inter) ou finale (Final).

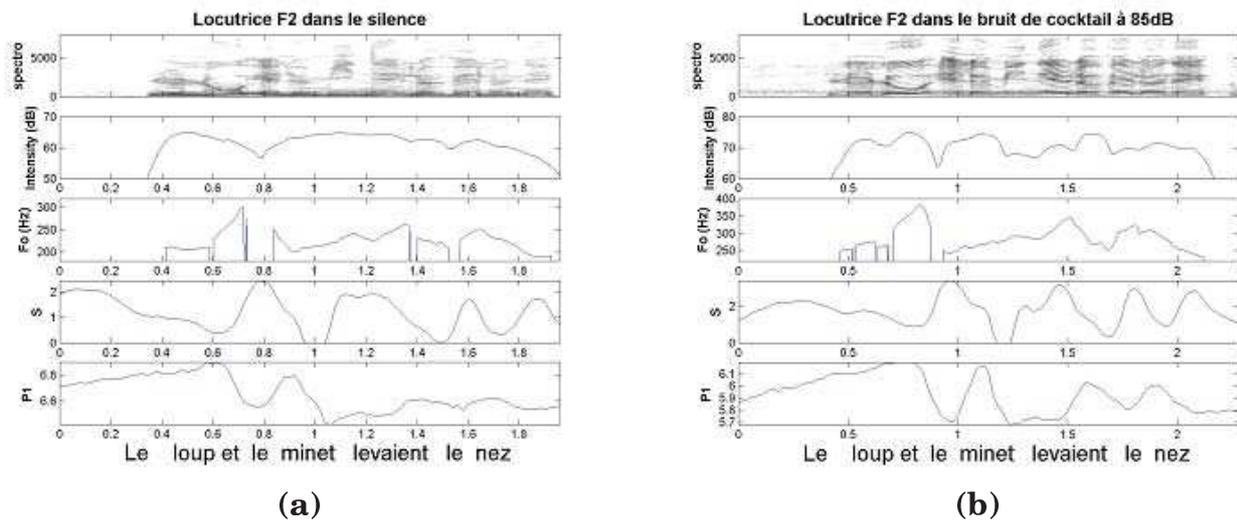


FIG. 5.11 – Paramètres acoustiques et articulatoires des productions vocales de F2 à la prononciation de la phrase 9 : « *Le loup et le minet levaient le nez* », (a) dans le silence, (b) dans le bruit cocktail de 85dB.

# Chapitre 6

## Conclusion

Nous nous proposons dans le cadre de ce mémoire d'étudier les variations articulatoires (de l'aire aux lèvres  $S$  et de la protrusion  $P1$ ) engendrées en parole Lombard et les corrélations possibles avec les variations acoustiques de l'élocution (durée, fréquence fondamentale  $F_0$ , intensité  $SPL$ ). Plus largement, ce travail vise à déterminer le rôle joué par l'hyper-articulation dans la parole Lombard : est-elle orientée vers une clarté maximale de la parole et motivée par la volonté consciente ou non du locuteur de rendre la communication intelligible ? Si oui, comment se manifeste ce désir d'intelligibilité dans ses productions vocales ?

Nous avons mené une expérience pilote pour étudier le mouvement des lèvres, dans l'idée de cerner les premières tendances significatives de l'adaptation du locuteur au bruit ambiant.

D'après les résultats de cette étude, l'effet Lombard entraîne des changements acoustiques significatifs tels que l'allongement de la durée, l'élévation de la fréquence fondamentale, et l'augmentation de l'intensité de la voix, conformément à la littérature. L'expérience menée dans cette étude montre que la parole Lombard se manifeste et/ou se réalise aussi par une hyper-articulation labiale significative. L'augmentation des paramètres acoustiques s'accompagne d'une **hyper-articulation globale** de  $S$ , *i.e.* sur tout l'énoncé et sur tous les mot cibles. Ces variations ont été évaluées à partir d'une segmentation de la phrase par groupes syntaxiques, par syntagmes puis par syllabes. Au niveau syntaxique, nous avons observé aucun effet de la position du syntagme dans la phrase sur les paramètres articulatoires ni sur le  $SPL$ . Au niveau syllabique, nous n'avons pas observé d'effet significatif de la syllabe sur l'ensemble des paramètres acoustiques. Il apparaît également que la protrusion n'augmente pas de façon significative. Ces observations nous orientent a priori dans l'idée d'une hyper-articulation reliée seulement à l'ouverture de la mâchoire consécutive de la volonté de parler plus fort.

Cependant, nous avons remarqué que les voyelles de l'énoncé tendent à être particulièrement accentuées en intensité dans le bruit. Nous avons donc regroupé les syllabes contenant des voyelles ouvertes ([a], [i], [e], [ɛ], [ã], [ĩ]) et celles contenant des voyelles fermées ([ō], [o] et [y]) pour raffiner cette observation. Nous n'avons observé qu'une légère corrélation entre l'aire intéro-labiale et le niveau d'intensité  $SPL$ . Ce résultat vient à l'encontre de nos premières observations qui suggéraient une hyper-articulation seulement liée à la nécessité de parler plus fort dans le bruit. De plus, il semble que la première syllabe de chaque syntagme soit le siège d'une hyper-articulation labiale et que la durée de l'« Objet » soit significativement élevée dans le bruit.

Au vu des résultats de la segmentation par voyelles, d'un point de vue méthodologique et dans l'objectif de l'étude pilote, nous avons pu également déterminer les descripteurs pertinents pour l'étude articulatoire. L'aire aux lèvres constitue un descripteur significatif des voyelles ouvertes puisqu'il est relié à la fois à l'aperture et à l'étirement des lèvres. L'étude conjointe des maxima locaux de  $S$  et de l'aire sous la courbe de  $S$  permet de rendre compte du contrôle moteur du locuteur et de son ouverture globale. Il pourrait être intéressant d'introduire d'autres facteurs influants dans les mouvements labiaux :

- le pincement des lèvres. En effet, l'étude de la fermeture des lèvres peut au même titre que celle de l'ouverture être riche en informations articulatoires. L'hyper-articulation labiale peut se manifester par un resserrement anticipé et/ou plus fort des lèvres à la prononciation de certains phonèmes tels [m] ou [p], etc.
- les mouvements anticipatoires et terminaux du locuteur. Ces mouvements peuvent contenir des informations pertinentes quant aux stratégies d'adaptation articulatoire.

L'étude pilote a ainsi permis de mettre en évidence une hyper-articulation globale de  $S$ , l'absence d'effet de la position du mot dans la phrase ou de la syllabe dans le mot, et la non-pertinence de la protrusion dans l'hyper-articulation dans le bruit. Ces premiers résultats sembleraient plutôt pencher en faveur d'une hyper-articulation au service du renforcement de l'intensité. Cependant, la corrélation entre les mouvements articulatoires et le niveau  $SPL$  n'est pas très importante. De plus, nous avons observé que la phrase semble davantage pensée au niveau des voyelles que des groupes syntaxiques en parole Lombard. Nous envisageons donc de raffiner l'étude articulatoire en examinant de plus près le rôle des voyelles. Cette réflexion pourrait rejoindre l'idée d'un patron vocalique en environnement bruité, déjà suggéré par [Dohalská and Mejvaldová, 2000].

Il reste à tester si l'hyper-articulation ne serait pas à destination visuelle, ce qui pourrait également expliquer nos résultats. La communication place les locuteurs en interaction auditive et visuelle. L'importance de la multimodalité de la parole n'est en effet plus à souligner. Nous envisageons dans la suite du travail de comparer différentes situations de communication en faisant varier le niveau de bruit, le feedback auditif et visuel de l'interlocuteur.

De plus, l'étude intonative sur la fréquence fondamentale menée en parallèle à l'ICP avec P. Welby permettra d'avoir d'autres informations sur le rôle de la prosodie en milieu bruyant. L'étude pilote que nous avons menée permet d'orienter nos recherches sur la position Objet, la voyelle et la corrélation entre l'intensité et la  $F_0$ . Les résultats de l'étude intonative, confrontés à l'analyse articulatoire traitée dans ce mémoire, pourraient révéler dans l'hyper-articulation labiale une accentuation de la syntaxe et de la segmentation.

Enfin, il reste à réaliser des expériences plus écologiques, avec plus de locuteurs et différentes techniques de mesures articulatoires (comme l'articulographe). Pour approcher le naturel et la spontanéité d'une communication, nous pensons aussi adapter notre protocole pour intégrer la modalité visuelle. Nous envisageons dans la continuité de cette étude de constituer un nouveau protocole basé sur le jeu interactif exposé dans ce mémoire et qui nous permettrait de rétablir une situation d'interaction et de communication.

# Bibliographie

- [Abry et al., 1980] Abry, C., Boë, L.-J., Corsi, P., Descout, R., Gentil, M., and Graillot, P. (1980). Labialité et phonétique. données fondamentales et études expérimentales sur la géométrie et la motricité labiales. *Publications de l'Université des Langues et Lettres de Grenoble*.
- [Audouy, 2000] Audouy, M. (2000). *Traitement d'images vidéo pour la capture de mouvements labiaux*. PhD thesis, Institut National Polytechnique de Grenoble.
- [Beautemps, 2000] Beautemps, D. (2000). Parole hyper-articulée : données et analyses acoustiques pour des plosives en français. *Actes des XXIIIèmes Journées d'Etude sur la Parole*, pages 437–440.
- [Benoît et al., 1992] Benoît, C., Lallouache, M.-T., Mohamadi, T., and Abry, C. (1992). A set of french visemes for visual speech synthesis. *Talking Machines : Theories, Models, and Designs*, pages 485–504.
- [Boersma and Weenink, 1996] Boersma, P. and Weenink, D. (1996). Praat, a system for doing phonetics by computer. *version 3.4. IFA rapport. 132*, page 182.
- [Brown et al., 1983] Brown, G., Anderson, A., Yule, G., and Shillcock, R. (1983). *Teaching Talk*. Cambridge : Cambridge University Press.
- [Carré et al., 1995] Carré, R., Lindblom, B., and MacNeilage, P. (1995). Rôle de l'acoustique dans l'évolution du conduit vocal humain. *Comptes Rendus de l'Académie des Sciences*, 30 :471–476.
- [Dohalská and Mejvaldová, 2000] Dohalská, M. and Mejvaldová, J. (2000). Rôle de la prosodie dans le communication en milieu bruité. *XXIIIèmes Journées d'Etude sur la Parole*, pages 265–268.
- [Dohen et al., 2003] Dohen, M., Løevenbruck, H., Cathiard, M.-A., and Schwartz, J.-L. (2003). Audiovisual perception of contrastive focus in french. *Proceedings of the AVSP'03 Conference*, pages 245–250.
- [Fonagy and Fonagy, 1966] Fonagy, I. and Fonagy, J. (1966). Sound pressure level and duration. *Phonetica*, 15 :14–21.
- [Garde, 1965] Garde, E.-J. (1965). Le réflexe cochléo-phonatoire : état actuel de la question après quelques faits expérimentaux et pathologiques. *Revue Lar. Otol.Rhinol.*, 86 :631–638.
- [Greenberg, 1996] Greenberg, S. (1996). Auditory processing of speech. *Principles of Experimental Phonetics*, pages 362–407.
- [Junqua, 1992] Junqua, J.-C. (1992). The variability of speech produced in noise. *Proceedings of ESCA on Speech Processing in Adverse Conditions*, pages 43–51.
- [Lallouache, 1991] Lallouache, M.-T. (1991). *Acquisition et traitement automatique des contours de lèvres*. PhD thesis, Institut National Polytechnique de Grenoble.

- [Lane and Sisson, 1970] Lane, H. and Sisson, C. (1970). Regulation of voice communication by sensory dynamics. *J. Acoust. Soc. Am.*, 47(2) :618–624.
- [Lane and Tranel, 1971] Lane, H. and Tranel, B. (1971). The lombard sign and the role of hearing in speech. *Speech and Hearing*, 14 :677–709.
- [Lindblom, 1983] Lindblom, B. (1983). Economy of speech gestures. *The Production of Speech*, pages 217–245.
- [Lindblom, 1990] Lindblom, B. (1990). Explaining phonetic variation : A sketch of the h&h theory. *Speech Production and Speech Modeling*, pages 403–439.
- [Lœvenbruck, 2000] Lœvenbruck, H. (2000). Effets articulatoires de l'emphase contrastive sur la phrase accentuelle en français. *XXIIIèmes Journées d'Etude sur la Parole*, pages 165–168.
- [Lombard, 1911] Lombard, E. (1911). Le signe de l'élevation de la voix. *Annales des Maladies de l'oreille et du larynx*, 37(2) :101–119.
- [Martin, 1981] Martin, P. (1981). Pour une théorie de l'intonation. *L'intonation : de l'acoustique à la sémantique*, pages 234–271.
- [Pickett, 1956] Pickett, J. (1956). Effects of vocal force on the intelligibility of speech sound. *J. Acous. Soc. Am.*, 28(5) :902–905.
- [Rossi et al., 1981] Rossi, M., Cristo, A. D., Martin, P., and Nishinuma, Y. (1981). L'intonation : de l'acoustique à la sémantique. *Collection études Linguistiques*, 25.
- [Schulman, 1989] Schulman, R. (1989). Articulatory dynamics of loud and normal speech. *J. Acoust. Soc. Am.*, 85(1) :295–312.
- [Summers et al., 1988] Summers, W., Pisoni, D., Bernacki, R., Pedlow, R., and Stockes, M. (1988). Effects of noise on speech production : Acoustic and perceptual analyses. *J. Acoust. Soc. Am.*, 84(3) :917–928.
- [Ternström et al., 2002] Ternström, Södersten, and Bohman (2002). Cancellation of simulated environmental noise as a tool. *Journal of voice*, 16(2) :195–206.
- [Tubach, 1989] Tubach, J.-P. (1989). *La parole et son traitement automatique*. Masson.
- [Vaissière, 1977] Vaissière, J. (1977). Quelques expériences d'analyse perceptuelle en français. *Actes des VIIIèmes Journées d'Etude sur la Parole*, pages 183–189.
- [Welby, 2002] Welby, P. (2002). The realization of early and late rises in french intonation : A production study. *Proceedings of Speech Prosody 2002*, pages 695–698.
- [Welby, 2003] Welby, P. (2003). *The Association and Alignment of French Intonational Rises and their Role in Speech Segmentation*. PhD thesis, The Ohio State University.

# Chapitre 7

## Annexes

### 7.1 Corpus

1. Maman ramena Loulou.
2. Lili et Maman louent les monuments.
3. Mélina lavera Manon.
4. Le minaret rallie la mairie.
5. Les lamas ramenaient Maman.
6. Mélanie ramena les moulinets.
7. Nina et Loulou mimaient les lamas.
8. Ninon lira le roman.
9. Le loup et le minet levaient le nez.
10. Marilou mima Mélanie.
11. Mélanie et Ninon loueront le minaret.
12. La mamie lavait le muret.
13. Loulou nourrit le muret.
14. Les mulets et les lamas rallieront la vallée.
15. Momo et Marilou lavaient les moulins.
16. Marina nourrit les mulets.
17. Manu et la nominée menaient la mamie.
18. Le monument et le minaret rallient la vallée.
19. Loulou et Marina liront le roman.
20. Nina et Mélanie ramenaient Ninon.
21. Manon leva le nez.
22. Le monument et les moulins rallient la vallée.
23. Les monuments rallient le muret.
24. Ninon et la mamie mimaient Marilou.
25. Lili et Mélina ramenaient Marina.
26. Le minaret et les monuments rallient la vallée.

27. La nominée mena Mélina.
28. Lola lavera la mamie.
29. Mélanie et Manon menaient la nominée.
30. Les moulinets ramenaient les morues.
31. Le minet leva le nez.
32. Le minaret et les moulinets rallient la vallée.
33. Les moulins rallient la mairie.

## 7.2 Exemple de dialogue du corpus MapTask (Q3NC8)

### 7.2.1 Map instructions for subjects

(once they're sitting down and have been given their copy of the map)

### 7.2.2 To the speaker

You and your partner have both got a map of the same place (*cf.* Figure 7.1).

Your map has got a route on it; your partner's map does not.

Your job is to describe the route to your partner so that s/he can draw it on her/his map.

Your path is known to be the only reliable route through and around all the various obstacles.

You must try to describe your route carefully so that your partner can avoid the obstacles and hazards on the way.

It is important to avoid these obstacles, rather than to make the routes identical to the last millimeter!

As you do this, keep in mind that the maps have been drawn by different explorers and might not be quite the same.

### 7.2.3 Then to the hearer

You and your partner have both got a map of the same place (*cf.* Figure 7.2).

Your partner's map has got a route on it, which s/he's doing to describe to you.

Your job is to draw the route on your map.

You must draw the route with care, because it's the only route known to avoid the various obstacles you may encounter.

Listen carefully to what your partner says, and ask questions if there's anything you're not sure about.

As you do this, keep in mind that the maps have been drawn by different explorers and might not be quite the same.

Do you understand what you're supposed to do?

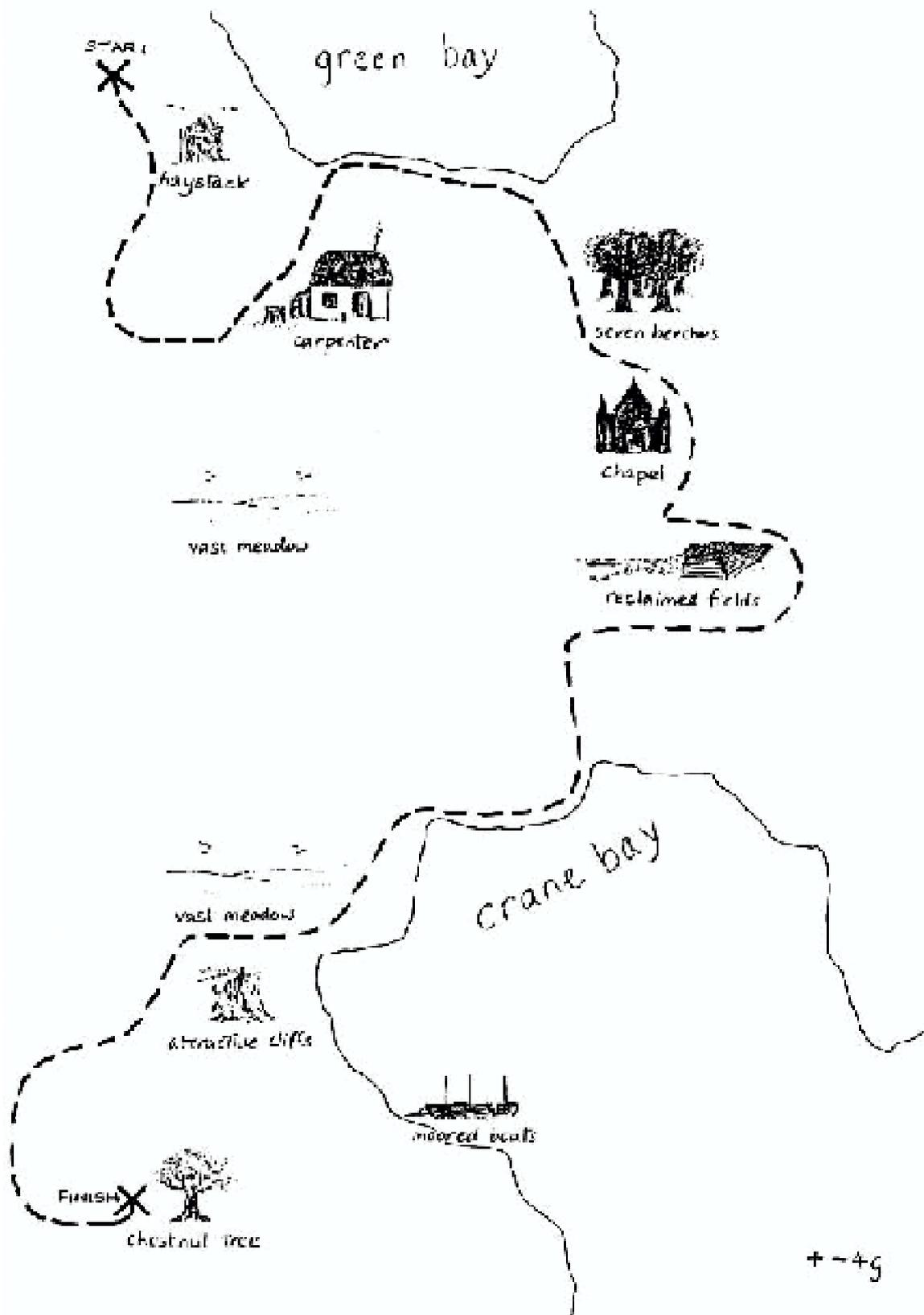


FIG. 7.1 – Givers Map

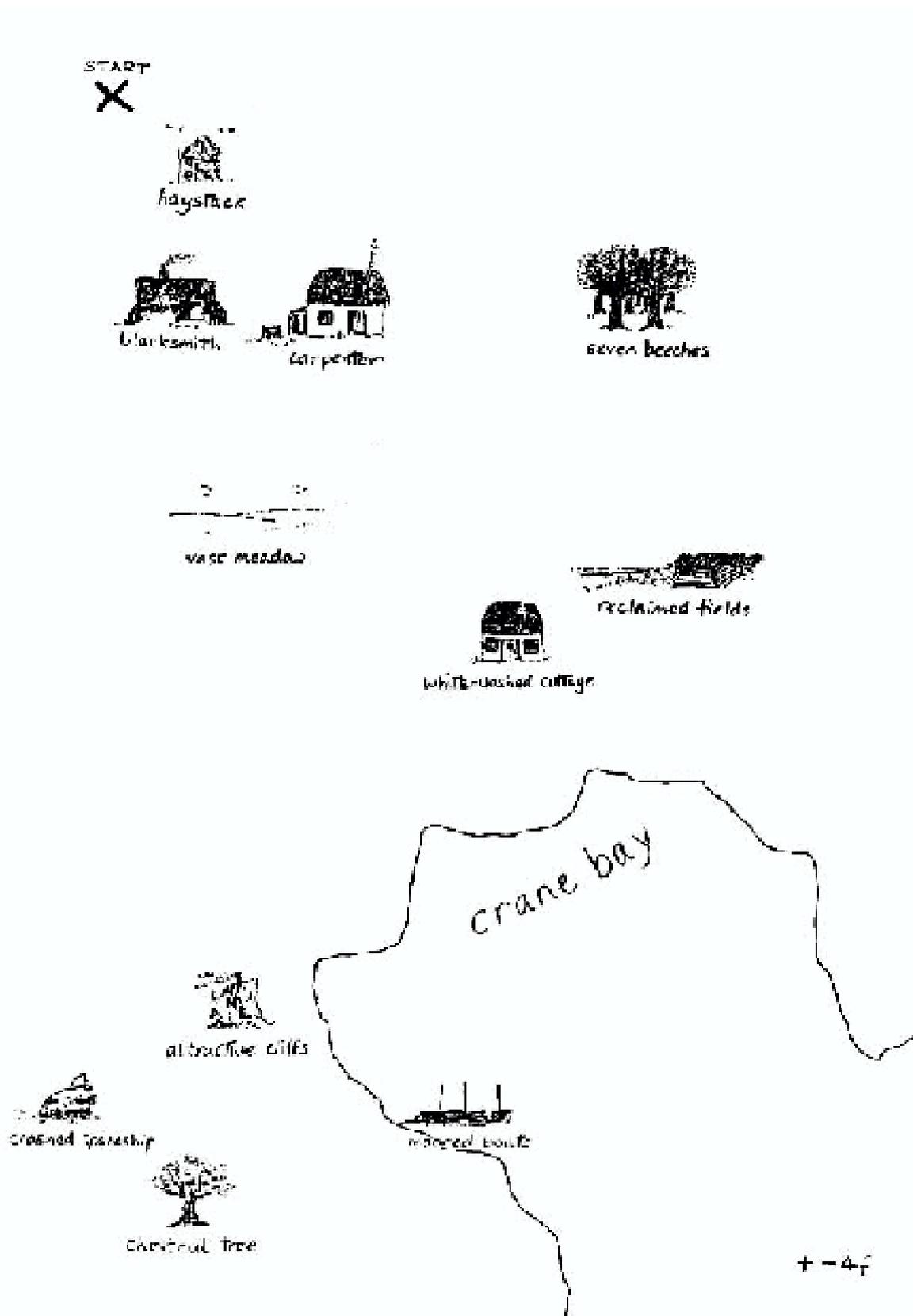


FIG. 7.2 – Followers Map

## 7.3 Graphes de $S$ , $P_1$ , $F_0$ et $SPL$ pour les 33 phrases

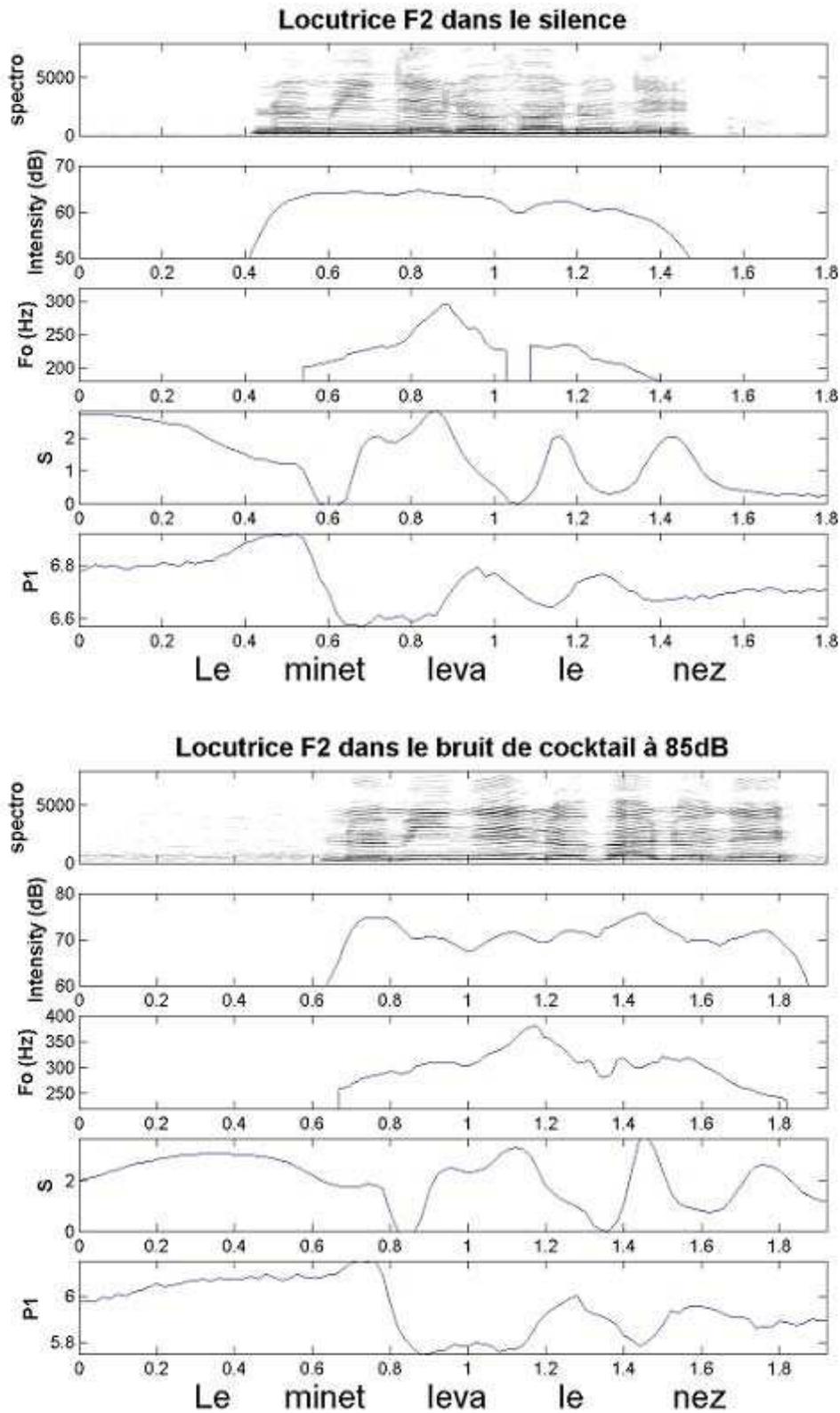


FIG. 7.3 – Variations des paramètres acoustiques et articulatoires des productions vocales de F2 à la prononciation de la phrase 31 dans le silence et dans le bruit cocktail de 85 dB.

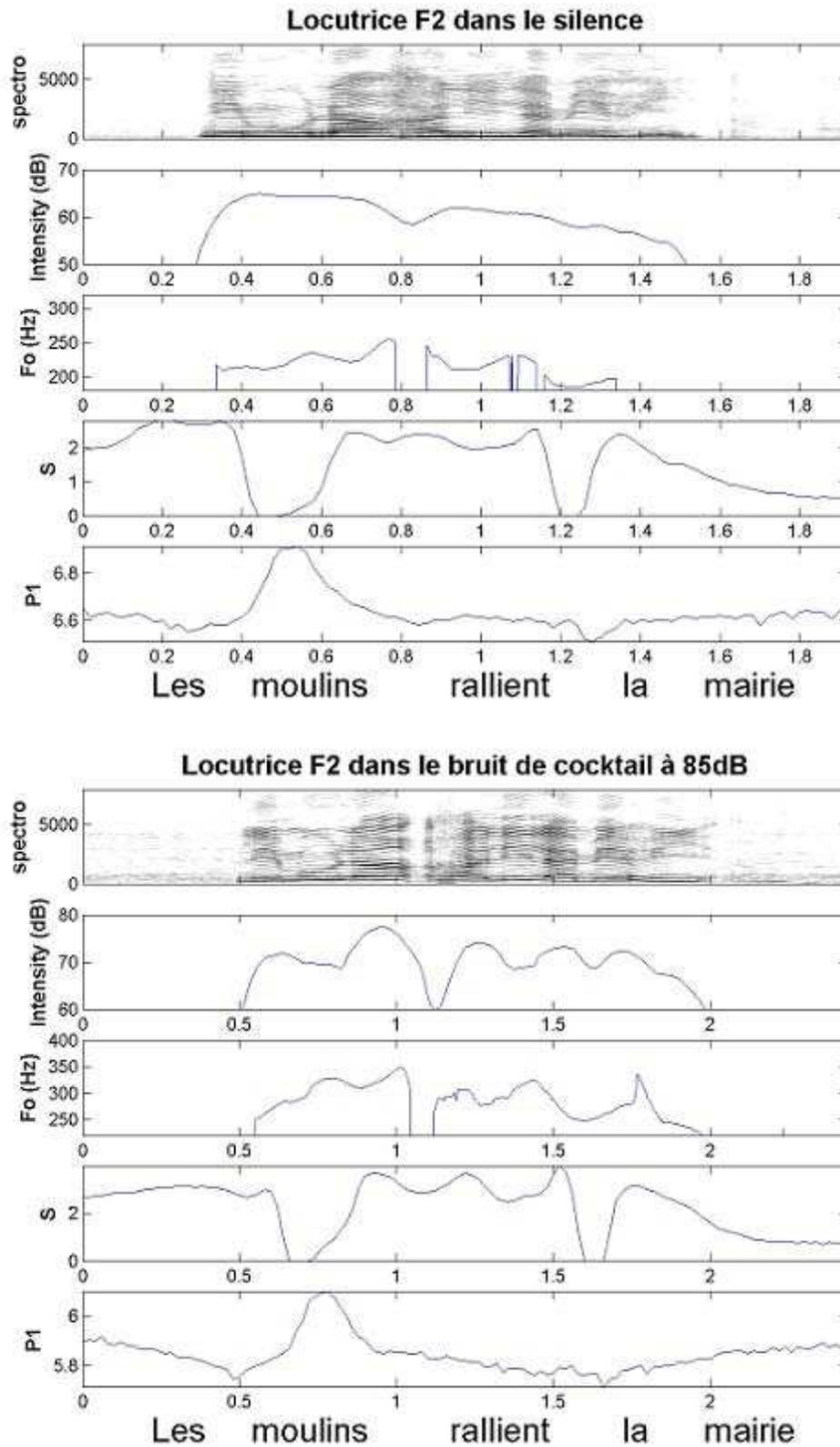


FIG. 7.4 – Variations des paramètres acoustiques et articulatoires des productions vocales de F2 à la prononciation de la phrase 31 dans le silence et dans le bruit cocktail de 85 dB.