



**HAL**  
open science

**Relations sensori-motrices lors de communication  
parlée : Application chez les jeunes adultes et séniors  
normo-entendants et les patients sourds implantés  
cochléaire**

Lucie Scarbel

► **To cite this version:**

Lucie Scarbel. Relations sensori-motrices lors de communication parlée : Application chez les jeunes adultes et séniors normo-entendants et les patients sourds implantés cochléaire. Linguistique. Université Grenoble Alpes, 2016. Français. NNT : 2016GREAS007 . tel-01312203

**HAL Id: tel-01312203**

**<https://theses.hal.science/tel-01312203v1>**

Submitted on 4 May 2016

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

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

## THÈSE

Pour obtenir le grade de

### DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

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

Arrêté ministériel : 7 août 2006

Présentée par

**Lucie SCARBEL**

Thèse dirigée par **Jean-Luc SCHWARTZ, Denis BEAUTEMPS**  
et **Marc SATO**

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

## **Relations sensori-motrices lors de la communication parlée :**

Application chez les jeunes adultes et seniors normo-entendants et les patients sourds implantés cochléaires

Thèse soutenue publiquement le « **16 février 2016** »,  
devant le jury composé de :

**Jacqueline LEYBAERT**

Professeur, Université Libre de Bruxelles, Rapporteur

**Lucie MENARD**

Professeur, Université du Québec à Montréal, Rapporteur

**Pascal BARONE**

Directeur de recherche, Université Toulouse, Examineur et Président du jury

**Denis BEAUTEMPS**

Chargé de recherche, Université Grenoble-Alpes, Co-directeur de thèse

**Marc SATO**

Chargé de recherche, Université Aix-Marseille, Co-directeur de thèse

**Jean-Luc SCHWARTZ**

Directeur de recherche, Université Grenoble-Alpes, Directeur de thèse





*« Qu'importe la surdit  de l'oreille quand l'esprit entend ? La seule surdit , la vraie surdit , la surdit  incurable, c'est celle de l'intelligence... »*

Victor Hugo.



## *Remerciements*

*Je tiens tout d'abord à remercier Pascal Barone, Jacqueline Leybaert et Lucie Ménard d'avoir accepté de faire partie de mon jury. Merci pour tous ces compliments qui m'ont fait rougir et merci pour le temps passé à lire ma thèse et à m'écouter et à me questionner lors de la soutenance. Plus précisément, Lucie, merci pour vos remarques judicieuses et précises proposées lors de votre rapport, lors de la soutenance et post-soutenance, Jacqueline, merci pour ce rapport élogieux et merci pour les questions ouvertes qui m'ont permis de réfléchir plus en profondeur sur les applications possible de ma thèse vers d'autres populations, et enfin Pascal, merci d'avoir accepté de prendre le rôle de président du jury en plus de celui d'examinateur et merci pour votre bienveillance qui m'a permis de me détendre quelque peu pendant la soutenance...*

*Un énorme merci global à mes trois directeurs, Jean-Luc Schwartz, Marc Sato et Denis Beautemps. Difficile de dire de manière concise tout ce que je souhaite, je vous ai déjà remercié plusieurs fois et je vous remercie encore de m'avoir fait confiance et de m'avoir proposé de réaliser cette thèse avec vous, et d'avoir été tous les trois présents à votre manière.*

*Marco, je sais que je t'en ai fait voir et que je t'ai souvent exaspéré, mais sache que je suis contente d'avoir été ta tête de pioche ! Merci de m'avoir laissée de l'indépendance quand j'ai fini par oser te le demander, merci pour ta motivation pour la recherche, transmissible, et tes encouragements. Merci pour ton humour et tes blagues linguistiques, pour la découverte des cathédrales et églises de Cologne et d'ailleurs, celle du glaçage de chaussure et pour tes citations diverses et variées !*

*Denis, merci pour ton aide pour le recrutement des patients implantés, pour le lien avec le CHU, merci aussi de m'avoir écouté quand tu as pu, et de m'avoir accompagnée pendant la thèse avec toute la gentillesse et la bienveillance dont tu es capable.*

*Et enfin Jean-Luc, at last but not least (comme tu dis !) je ne saurai te remercier autant que je le voudrai. Tu as su, pendant la thèse, m'accompagner et me guider avec une gentillesse incroyable pour que je finisse par avoir un tant soit peu confiance en moi et en mes capacités, tu as su faire ressortir le meilleur de moi-même en m'écoutant et en me faisant parler, et aussi parfois, quand il y en avait besoin en « m'engueulant » à ta manière douce mais efficace ! Tu as aussi et surtout réussi à me faire réaliser ce que je voulais vraiment faire, en me posant ta question magique dont je me ressers vis-à-vis de mon entourage : « toi, Lucie, sans aucune peur de notre jugement, qu'est ce qui t'exciterait de faire dans la vie ? ». La réponse, pourtant évidente, je la croyais impossible, mais tu l'as entendue, et tu m'as demandé pourquoi je pensais qu'il était impossible que j'aille vers la recherche sur les langues des signes. C'est donc en partie grâce à toi que mes yeux se sont mis à briller quand je parle de mon projet de recherche en particulier et de langue des signes en général. Je te remercie aussi d'avoir essayé (et réussi, je crois) à comprendre mon intérêt pour cette langue, en supportant les lectures et les films que je t'ai infligés ! Donc, un immense merci pour tout ça, et aussi pour toutes les petites attentions dont tu es capable.*

*En résumé, je suis vraiment ravie d'avoir passé ces trois années et demie à vos côtés à tous les trois, on recommence quand ?!*

*Je souhaite remercier sincèrement les participants, entendants et implantés qui ont accepté de prendre le temps de venir participer à mes expériences sans que cette thèse n'aurait pas lieu d'être, ainsi que l'équipe ORL du CHU de Grenoble, qui m'a permis de recruter mes participants.*

*Je tiens à remercier également tous les collègues du Gipsa-lab, tant Ampériens que Stendhaliens. Merci les Ampériens de m'avoir accueillie quand je venais faire un tour par chez vous. Un merci tout particulier à Pascal Perrier, pour ton aide lors de la rédaction de la partie portant sur théories de la production de parole. Les Stendhaliens (passés et présents, je ne vous nommerai pas tou-tes-s mais vous êtes tou-tes-s visé-es-s), merci pour ces cinq années et demies à vos côtés, merci pour les pauses café/thé et plus si affinité, les ateliers nems/gaufres, les repas de Noël et autres pique-niques de fin d'années et merci d'avoir été les testeurs de mes essais culinaires parfois bancals !*

*Plus précisément, un gros merci à Pipou et Cricri (d'Amour !) pour votre aide au début de mes manipes, à l'enregistrement et au traitement des stimuli, et pour vos réponses patientes à mes questions diverses et variées, et pour les moments passés en conf<sup>o</sup> à Lyon, Liège ou Cologne ! Aussi, je n'oublie pas Anne et Hélène, vous m'avez encouragée et formée pendant mon master avec beaucoup de bonté, et vous avez grandement participé à mon envie de continuer en thèse, donc merci à vous deux !*

*Je pense aussi à mes co-bureaux, Avril, Marjorie et Bénédicte, qui ont supporté mes chants de Noël, mes décorations diverses et variées (sans rancune Béné !), et surtout mes craquages surexcités réguliers à propos à propos de notre Dieu à nous ou d'autres choses... Je pense que vous allez vous ennuyer quand je serai partie ! Le bureau sera nettement plus calme !*

*Je remercie également toute l'équipe PCMD pour votre accueil malgré ma non-affiliation officielle à l'équipe ! A vos côtés, j'ai pu découvrir le fonctionnement d'une équipe et plus généralement d'un labo. Merci à ceux qui ont trouvé le temps de venir écouter ma pré-soutenance, Pascal, Amélie, Coriandre, Marjorie, Avril et Diane (cherchez l'intrus-e !) grâce à vous mon ppt est devenu plus lisible et mon discours plus clair !*

*Before going to my friends, I wanted to thank also Robin Thompson: you're the first foreign researcher I met in the field of Sign Languages, and I don't regret it! Thanks for our discussions about my research project; I really hope we can work together some days...*

*Je passe maintenant à mes ami-e-s, les oublis seront évidemment fortuits ! Merci évidemment au Dr. Klox. Ma chère Chloë, tu m'as encouragée et soutenue tout au long de la thèse et avant cela du master, tu m'as servi et me sers de modèle, en me montrant que tout est possible (oui oui), tu as même tenté de m'initier à la course (raté, j'ai toujours dit que c'était dangereux), merci pour cette écoute dont tu es capable, ces conseils avisés quand il y en a besoin, ces moments de franches rigolades, le autour d'un klix...*

*Jeannoue, merci pour le soutien lointain mais efficace, les conseils de conf<sup>o</sup> où j'aurai peut-être dû aller, le soutien pendant l'école d'été, et les délires de ouf-ouf depuis 5 ans, Matilda a bien de la chance d'avoir une mère pareille :-)*

*Je tiens aussi à remercier la fine équipe du Club Piscine (+ Marjo) : Avril (encore elle ?!), Marjo (encore elle ?!), Laurette et Diannoue (oui, j'aime les surnoms). On ne nage pas forcément beaucoup (surtout quand on décrète qu'on a mal à la tête et au ventre et qu'on décide que ce serait mieux de rester au chaud à manger des gâteaux avant d'aller au resto, spéciale dédicace à Laure), mais on rigole bien ! Merci Diane pour les essais pré-cadeaux, c'était judicieux !*

*Je n'oublie pas de remercier chaleureusement Paolo et Fabian : merci pour notre escapade Ecossaïse post-ICPHS, c'était judicieux de passer une semaine à faire autre chose que rédiger la thèse ou le projet post-doc ! Merci Fabian pour tes conseils de résumés de début et fin de chapitre, je m'en suis inspirée, et Paolito, merci pour ta gentillesse et ton amitié, pour ton amour transmissible des randos et des paysages lunaires !*

*Un gros merci à Aurélie pour tous ces moments passés ensemble (je sais que tu ne liras que les remerciements, alors je m'applique un peu), pour l'invitation au voyage en Croatie en deuxième année qui m'a permis d'affronter la troisième année, pour ces services mutuels échangés...*

*Merci aussi à Micka et Cathouille pour les balades, l'amitié et l'initiation à l'équitation (désolée de pas avoir persévéré, mais les manèges c'est moyennement mon truc en fait...)*

*Merci aussi à mes ex-camarades de licence, Sam, Pierto et Gaëlle, merci pour ces bons moments dans nos différents QG ; les garçons, merci de titiller mon esprit avec les discussions politiques et/ou intellectuelles (ou plus ou moins, n'est-ce pas Pierto), on ne s'ennuie pas avec vous !*

*Merci aux copains-« chiens », qu'ils fassent partie des groupes balades, parc ou club canin. Un merci tout particulier à Arthur, pour son intérêt quant à mes recherches et ma thèse : Arthur quand tu liras ceci, ne te sens pas obligé de lire le reste !*

*Enfin, un gros merci en bloc pour les amis Ardéchois ou assimilés (MarieO, MarieL et la p'tite Marie, Vivi, Laureline, Mickey, Gaspary, Clo et les autres) et les amis du Ptit Vélo (impossible de tous vous citer, mais mention spéciale à Titi pour le hug de veille de soutenance, qui m'a donné du courage).*

*En avant dernier lieu, bien qu'ils fassent partie des plus importants, je voudrai remercier les membres de ma famille. Mes petits parents, merci de votre soutien et merci d'être fiers de moi, je sais que ça n'a pas été toujours évident de comprendre mes choix, que la recherche pour vous c'est bien lointain... Je me souviendrai toujours quand, une quinzaine d'année plus tôt, je vous ai dit que j'avais envie de faire des études d'ethnologie, et que vous aviez répondu « mais ça sert à quoi ? ». J'ai finalement bifurqué un peu par hasard vers la phonétique puis vers les sciences cognitives, et j'ai toujours senti, même si vous ne vouliez pas me le montrer, le questionnement « mais ça sert à quoi ? Elle va devenir quoi ma fille ? » J'espère que vous avez pu comprendre un peu mieux ce que j'ai fait pendant trois ans grâce à la soutenance, maintenant, vous pouvez vous préparer à essayer de comprendre mon nouveau projet de recherche !*

*Pour rester dans le thème « famille », merci aussi à mon fabuleux beau-frère, David, pour le soutien apporté, pour tes questions quand tu essayais de comprendre le monde de la recherche et mon sujet de thèse, merci d'avoir dédramatisé mon sujet en en riant ! Merci aussi à ma filleule-nièce Lison et à ma petite nièce Nina, vous comprendrez plus tard pourquoi vous n'avez pas beaucoup vu votre marraine et tante, mais si un jour vous lisez ces mots, soyez bien conscientes de l'amour que je vous porte, et sachez aussi ceci : tout est possible dans la vie quand on le veut, donc n'ayez pas peur de faire des choix mêmes s'ils paraissent compliqués, conseil de Zaza !*

*Enfin, un merci tout particulier à ma sœur, Aline. Merci pour ta confiance indéfectible envers ta petite sœur, tes encouragements, ta compréhension... Tu me comprends souvent bien plus que moi-même, et c'est surtout ça une sœur !*

*Et pour finir, merci à ma petite Miette, qui, même si elle n'en a quasi certainement pas conscience, m'a servi de soupape de sécurité en m'obligeant à prendre des bons bols d'air (pollué), qui a supporté mon stress et mon énervement (et oui, comme dirait Marjorie, les bestioles sont sûrement les victimes collatérales les plus touchées de la fin de thèse...), et qui m'aide actuellement à écrire ces remerciements, qu'elle souhaiterait plus courts pour pouvoir aller profiter du soleil le plus vite possible...*





**Relations sensori-motrices lors de communication parlée : Application chez les jeunes adultes et seniors normo-entendants et les patients sourds implantés cochléaire**

---

La communication parlée peut être vue comme un processus interactif impliquant un couplage fonctionnel entre les systèmes moteur et sensoriel. L'objectif de ce travail de thèse est de tester ces possibles liens perceptivo-moteurs, aussi bien lors de la perception que de la production de la parole, au travers de multiples paradigmes comportementaux et auprès de différentes populations de participants. Le protocole expérimental mis en place est composé de trois paradigmes expérimentaux classiques : un premier paradigme de close-shadowing, visant à explorer le format partiellement moteur de stimuli auditifs et audiovisuels de parole ; un second paradigme permettant de mettre en évidence des corrélations entre la production et la perception de voyelles ; et enfin, un dernier paradigme d'imitation consciente et inconsciente de fréquence fondamentale.

Suite à la validation de notre protocole expérimental auprès d'une population contrôle de jeunes adultes normo-entendants, nous avons étudié une seconde population constituée d'adultes normo-entendants seniors, et ceci afin d'évaluer la conséquence d'un déclin des fonctions cognitives et langagières. Les résultats obtenus ont permis de suggérer une activation fonctionnelle des liens perceptivo-moteurs lors de la perception et de la production de parole chez l'ensemble des participants.

La troisième population testée était constituée de patients sourds post-linguaux puis implantés, afin de déterminer l'impact d'une déprivation sensorielle ainsi que les éventuels réapprentissage liés à leur implantation sur ces liens perceptivo-moteurs. De manière surprenante, les résultats ont mis en évidence des relations sensori-motrices actives chez ces participants, et ce même très peu de temps après l'implantation.

Pris ensemble, les résultats observés avec ces trois paradigmes expérimentaux et auprès de ces trois groupes de participants attestent de la nature perceptivo-motrice de la parole. De manière importante, malgré des performances dégradées, ces interactions entre systèmes sensoriels et moteur lors de la perception et de la production de parole resteraient fonctionnelles auprès des deux populations, celle des adultes normo-entendants seniors et celle des patients sourds post-linguaux porteurs d'un implant cochléaire.

**Mots clefs** : lien perceptivo-moteur, surdité, implant cochléaire, parole, perception-production.

**Sensori-motor relationship in speech communication: Application in young and elderly normal-hearing adults and deaf cochlear implanted patients**

---

Speech communication can be considered as an interactive process involving a functional coupling between sensory and motor systems. The aim of this thesis was to test possible perceptuo-motor linkages during both speech perception and production, using distinct behavioral paradigms and populations. The experimental protocol was made of three classic experiments: a first paradigm of close-shadowing, aiming at exploring the partially motor format of audio and audiovisual stimuli; a second paradigm allowing to correlate production and perception of vowels; and a third paradigm of conscious and unconscious imitation of pitch.

The experimental protocol was validated with a first group of young hearing adults. The second population studied was composed of elderly normal-hearing participants, in order to evaluate the consequences of both cognitive and linguistic declines. Results allowed us to suggest a functional activation of perceptuo-motor linkage during speech production and perception.

The third population we tested comprised post-lingually deaf patients wearing a cochlear implant. Our objective was to determine the impact of the sensorial deprivation and the re-learning processes, associated with their implantation, on perceptuo-motor linkages. Unexpectedly, results showed an active sensori-motor relationship in those participants, even shortly after the cochlear implantation.

Altogether, our results confirmed the perceptuo-motor nature of speech. Importantly, in spite of degraded performances, these interactions between the sensory and the motor systems during speech production and perception remained functional in both the elderly normal-hearing population and the post-lingually deaf patients, wearing a cochlear implant.

**Keywords:** perceptuo-motor linkage, deafness, cochlear implant, speech, perception-production.

# Table des matières

<b>TABLE DES MATIERES .....</b>	<b>III</b>
<b>TABLE DES FIGURES.....</b>	<b>IX</b>
<b>TABLE DES TABLEAUX .....</b>	<b>XV</b>
<b>INTRODUCTION GENERALE.....</b>	<b>1</b>
<b>1. OBJECTIFS ET PLAN DE THESE .....</b>	<b>1</b>
1. Problématique de la thèse .....	1
2. Organisation du mémoire de thèse.....	2
<b>2. RELATIONS SENSORI-MOTRICES LORS DE LA PERCEPTION ET DE LA</b>	
<b>PRODUCTION DE PAROLE.....</b>	<b>2</b>
1. Théories auditives de la perception de la parole .....	3
2. Théories motrices de la perception de la parole .....	5
3. Théories perceptivo-motrices de la perception de parole .....	7
4. Théories motrices de la production de parole.....	9
5. Théories auditives de la production de parole.....	10
6. Théories perceptivo-motrices de la production de parole .....	11
7. Arguments en faveur des théories perceptivo-motrices.....	12
<b>PARTIE I : RELATIONS PERCEPTIVO-MOTRICES PAR DES PARTICIPANTS NORMO-</b>	
<b>ENTENDANTS JUNIORS.....</b>	<b>15</b>
<b>1. INTRODUCTION.....</b>	<b>17</b>
<b>2. ORGANISATION GENERALE DE L'EXPERIENCE ET HYPOTHESES GLOBALES ...</b>	<b>17</b>
1. Paradigme expérimental de close-shadowing auditif et audiovisuel .....	18
2. Paradigme d'idiosyncrasies lors de la perception/production de voyelles .....	21
3. Paradigme de convergence et d'imitation.....	24
4. Procédure expérimentale commune aux trois expériences .....	26
1. Test JND.....	29
<b>3. EXPERIENCE DE CLOSE-SHADOWING.....</b>	<b>27</b>
<b>1. Principe et hypothèses.....</b>	<b>27</b>
1. Expérience non-bruitée.....	27
2. Expérience bruitée .....	28
<b>2. Matériel et méthodes.....</b>	<b>29</b>
1. Expérience non-bruitée.....	29
2. Expérience bruitée .....	32
<b>3. Résultats.....</b>	<b>34</b>
1. Expérience non-bruitée.....	34
2. Expérience bruitée vs. Expérience non-bruitée .....	35
<b>4. Discussion.....</b>	<b>38</b>
<b>4. EXPERIENCE D'IDIOSYNCRASIES .....</b>	<b>43</b>
1. Principe et hypothèses.....	43

<b>2. Matériel et méthodes</b> .....	<b>43</b>
1. Participants .....	43
2. Stimuli .....	44
3. Procédure expérimentale.....	46
4. Analyses.....	47
<b>3. Résultats</b> .....	<b>47</b>
1. Tâche de production.....	47
2. Tâche de perception .....	48
3. Corrélation perception/production .....	49
4. Autres analyses.....	51
<b>4. Discussion</b> .....	<b>53</b>
<b>5. EXPERIENCE DE CONVERGENCE PHONETIQUE ET D'IMITATION VOLONTAIRE</b>	<b>54</b>
1. Principe et hypothèses.....	<b>54</b>
2. Matériel et méthodes.....	<b>55</b>
1. Participants .....	55
2. Stimuli .....	55
3. Procédure expérimentale.....	56
4. Analyses.....	57
3. Résultats.....	<b>58</b>
4. Discussion.....	<b>60</b>
6. CONCLUSION .....	<b>61</b>
<b>PARTIE II : RELATIONS PERCEPTIVO-MOTRICES CHEZ LES PERSONNES NORMO-</b>	
<b>ENTENDANTES SENIORS</b> .....	<b>65</b>
1. INTRODUCTION.....	<b>67</b>
2. ORGANISATION GENERALE DE L'EXPERIENCE ET HYPOTHESES GLOBALES ...	<b>67</b>
1. L'enjeu des « séniors ».....	<b>67</b>
2. Perception de la parole chez les personnes normo-entendantes séniors .....	<b>68</b>
1. Perception auditive .....	68
2. Perception audiovisuelle.....	70
3. Production de la parole chez les personnes normo-entendantes séniors.....	<b>72</b>
1. Contrôle sensori-moteur de la parole chez les séniors.....	72
2. Production des voyelles chez les séniors.....	73
3. Contrôle de la fréquence fondamentale chez les séniors .....	74
4. Lien perceptivo-moteur chez les personnes normo-entendantes séniors .....	<b>75</b>
5. Proposition et hypothèses expérimentales .....	<b>76</b>
1. Participants .....	76
2. Procédures expérimentales communes aux trois expériences .....	77
3. EXPERIENCE DE CLOSE-SHADOWING.....	<b>77</b>
1. Principes et hypothèses .....	<b>77</b>
2. Méthodologie.....	<b>78</b>
1. Stimuli .....	78

2.	Procédure expérimentale.....	78
3.	Analyses.....	78
<b>3.</b>	<b>Résultats séniors seuls .....</b>	<b>78</b>
1.	Temps de réaction .....	78
2.	Pourcentage de réponses correctes .....	79
<b>4.</b>	<b>Comparaison avec la population implantée et les populations normo-entendante</b>	
<b>« junior » dans le bruit et sans bruit.....</b>		<b>80</b>
1.	Temps de réaction .....	80
2.	Pourcentages de réponses correctes.....	81
<b>5.</b>	<b>Discussion.....</b>	<b>82</b>
<b>4.</b>	<b>EXPERIENCE D'IDIOSYNCRASIES .....</b>	<b>83</b>
<b>1.</b>	<b>Principes et hypothèses .....</b>	<b>83</b>
<b>2.</b>	<b>Méthodologie.....</b>	<b>84</b>
1.	Stimuli .....	84
2.	Procédure expérimentale.....	84
3.	Analyses.....	84
<b>3.</b>	<b>Résultats population séniors .....</b>	<b>84</b>
1.	Tâche de production.....	84
2.	Tâche de perception .....	85
3.	Corrélation perception/production .....	87
<b>4.</b>	<b>Comparaison entre la population séniors et la population junior.....</b>	<b>89</b>
<b>5.</b>	<b>Discussion.....</b>	<b>91</b>
<b>5.</b>	<b>EXPERIENCE DE CONVERGENCE ET IMITATION.....</b>	<b>92</b>
<b>1.</b>	<b>Principes et hypothèses .....</b>	<b>92</b>
<b>2.</b>	<b>Méthodologie.....</b>	<b>93</b>
1.	Stimuli .....	93
2.	Procédure expérimentale.....	93
3.	Analyses.....	93
<b>3.</b>	<b>Résultats population séniors .....</b>	<b>93</b>
<b>4.</b>	<b>Comparaison avec la population junior .....</b>	<b>96</b>
<b>5.</b>	<b>Discussion.....</b>	<b>97</b>
<b>6.</b>	<b>CONCLUSION .....</b>	<b>97</b>
<b>PARTIE III : RELATIONS PERCEPTIVO-MOTRICES CHEZ LES PATIENTS SOURDS</b>		
<b>POST-LINGUAUX PORTEURS D'UN IMPLANT COCHLEAIRE .....</b>		
		<b>101</b>
<b>1.</b>	<b>INTRODUCTION.....</b>	<b>103</b>
<b>2.</b>	<b>INTRODUCTION : SURDITE ET IMPLANT COCHLEAIRE .....</b>	<b>103</b>
<b>1.</b>	<b>Surdité.....</b>	<b>103</b>
1.	Les différents types de surdité dans le monde et en France.....	103
2.	Les différents modes de communication des personnes atteintes de surdité .....	104
<b>2.</b>	<b>Prothèses auditives et implant cochléaire.....</b>	<b>107</b>
1.	Prothèses auditives .....	107
2.	Principe de l'Implant cochléaire.....	107

<b>3. Production et perception de la parole chez les patients porteurs d'un implant cochléaire</b>	<b>109</b>
1. Perception.....	109
2. Production .....	114
3. Lien perceptivo-moteur.....	116
<b>4. Rôle des interactions perceptivo-motrices et leur réorganisation .....</b>	<b>119</b>
1. Proposition et hypothèses expérimentales .....	119
2. Participants .....	120
3. Procédures expérimentales communes aux trois expériences .....	121
<b>3. EXPERIENCE DE CLOSE-SHADOWING.....</b>	<b>121</b>
<b>1. Principes et hypothèses .....</b>	<b>121</b>
<b>2. Méthodologie.....</b>	<b>122</b>
1. Stimuli .....	122
2. Procédure expérimentale.....	122
3. Analyses.....	123
<b>3. Résultats implantés seuls.....</b>	<b>123</b>
1. Temps de réaction .....	123
2. Pourcentages de réponse correctes .....	124
3. Corrélations des résultats avec l'étiologie.....	125
<b>4. Comparaison avec les populations normo-entendantes .....</b>	<b>125</b>
1. Temps de réaction .....	126
3. Pourcentages de réponses correctes.....	128
<b>5. Discussion.....</b>	<b>129</b>
<b>4. EXPERIENCE D'IDIOSYNCRASIES .....</b>	<b>130</b>
<b>1. Principes et hypothèses .....</b>	<b>130</b>
<b>2. Méthodologie.....</b>	<b>131</b>
1. Stimuli .....	131
2. Procédure expérimentale.....	131
3. Analyses.....	131
<b>3. Résultats population sourde implantée.....</b>	<b>131</b>
1. Production .....	131
2. Perception.....	132
3. Corrélations perception/production .....	133
4. Corrélations facteurs étiologiques .....	137
<b>4. Comparaison avec les populations normo-entendante « junior » et « sénior ».....</b>	<b>137</b>
1. Tâches de production et de perception .....	137
2. Corrélations production-perception .....	140
<b>5. Discussion.....</b>	<b>140</b>
<b>5. EXPERIENCE DE CONVERGENCE ET IMITATION.....</b>	<b>141</b>
<b>1. Principes et hypothèses .....</b>	<b>141</b>
<b>2. Méthodologie.....</b>	<b>142</b>
1. Stimuli .....	142
2. Procédure expérimentale.....	142
3. Analyses.....	142

<b>3. Résultats population sourde implantée.....</b>	<b>142</b>
1. Résultats généraux .....	142
2. Corrélation avec les facteurs étiologiques .....	145
<b>4. Comparaison avec les populations normo-entendantes .....</b>	<b>145</b>
<b>5. Discussion.....</b>	<b>146</b>
<b>6. CONCLUSION .....</b>	<b>147</b>
<b>DISCUSSION GENERALE.....</b>	<b>151</b>
<b>1. Synthèse des principaux résultats .....</b>	<b>153</b>
1. Toutes populations confondues .....	153
2. Relations sensori-motrices chez les séniors .....	158
3. Relations sensori-motrices chez les patients sourds implantés .....	159
<b>2. Conclusion générale et perspectives.....</b>	<b>160</b>
<b>BIBLIOGRAPHIE.....</b>	<b>163</b>



## Table des figures

Figure 1 : Représentation spectrographique des formants F1/F2 des syllabes /di/ et /du/ (Figure empruntée à Galantucci et coll., 2006).....	6
Figure 2 : Illustration de la boucle de procédés perceptivo-moteurs de Skipper et coll. (2007) lors de la perception audiovisuelle (Figure empruntée à Skipper et coll., 2007) .....	8
Figure 3 : Modèle à double voie de perception de la parole et du langage d'Hickok et Poeppel, la voie dorsale étant représentée en bleu, et la voie ventrale en rose (Figure empruntée à Hickok et Poeppel, 2007).....	9
Figure 4 : Modèle de production de parole proposé par la théorie de la phonologie articulatoire (Figure empruntée à Browman et Goldstein, 1990) .....	10
Figure 5 : Modèle DIVA de la production de parole (Figure empruntée à Guenther, 2006) .....	11
Figure 6 : Un des résultats obtenus par Fowler et coll. (2003) : différence de temps de réaction lors des réponses simples selon la compatibilité (colonnes blanches) ou l'incompatibilité (colonnes grises) entre le stimulus perçu et la consonne attribuée au participant (figure empruntée à Fowler et coll. 2003) .....	19
Figure 7 : Ellipses de dispersion des dix voyelles orales du français chez quatre participants adultes de l'étude de Ménard et coll. 2008 (Figure empruntée à Ménard et coll., 2008) .....	22
Figure 8 : Corrélation entre les valeurs normalisées de F1 en production et en perception pour la voyelle [e] chez les participants de l'étude de Ménard et Schwartz, 2014 (Figure empruntée à Ménard et Schwartz, 2014) .....	24
Figure 9 : Une partie des résultats obtenus par Garnier et coll., 2014 : changements imitatifs lors des tâches d'imitation consciente (figure de gauche) et inconsciente (figure de droite) (abscisse : déviance de la F0 du stimulus ; ordonnée : déviance de la production lors des tâches par rapport à la production initiale) (figure adaptée de Garnier et coll. 2014).....	26
Figure 10 : Ordre de passation et durée des différentes expériences.....	26
Figure 11 : Procédure expérimentale de l'expérience de close-shadowing.....	30
Figure 12 : Calcul des temps de réaction. Temps de réaction mesurés entre l'instant de début de la plosion du stimulus et l'instant de début de la réponse pour les réponses orales et entre la plosion du stimulus et l'appui touche pour les réponses manuelles. ....	31
Figure 13 : Exemple d'annotation des données de l'expérience de close-shadowing : .....	32

Figure 14 : Moyenne des temps de réaction (en ms) pour toutes les conditions de l'expérience non-bruitée (les barres d'erreurs correspondent aux écart-types).....	34
Figure 15 : Moyenne des pourcentages de réponses correctes pour toutes les conditions de l'expérience non-bruitée (les barres d'erreurs correspondent aux écart-types).....	35
Figure 16 : Moyenne des temps de réaction des expériences en conditions bruitée et non-bruitée (en ms, les barres d'erreurs correspondent aux écart-types). ....	36
Figure 17 : Moyenne des pourcentages de réponses correctes en conditions bruitée et non-bruitée (les barres d'erreurs correspondent aux écart-types).....	37
Figure 18 : Schéma représentant le modèle de Skipper et collègues (2007) pour la perception d'inputs de parole auditifs.....	39
Figure 19 : Interprétation possible du mode de réponse à partir du modèle de Skipper et collègues (2007).....	40
Figure 20 : Schéma représentant le modèle de Skipper et collègues (2007) pour la perception d'inputs de parole multisensoriels.....	42
Figure 21: Représentation graphique sur l'espace F1/F2' des stimuli utilisés pour la tâche de perception de voyelles de l'expérience d'idiosyncrasies (les stimuli utilisés sont représentés par les points verts sur le graphique – ils sont numérotés dans les Tableaux 3 et 4 par leur numérotation d'origine parmi les 84 stimuli de départ).....	45
Figure 22 : Procédure expérimentale de l'expérience "Idiosyncrasies" .....	47
Figure 23 : Valeurs normalisées de F1 pour /e/ produites par chaque participant. ....	48
Figure 24 : Valeurs normalisées de F1 pour /ɛ/ produite par chaque participant. ....	48
Figure 25 : Valeurs normalisées de F1 lors de la perception de la voyelle /e/, par participant.....	49
Figure 26 : Valeurs normalisées de F1 lors de la perception de la voyelle /ɛ/, par participant.....	49
Figure 27 : Corrélations pour la voyelle /e/.....	50
Figure 28 : Corrélations pour la voyelle /ɛ/.....	50
Figure 29 : Corrélations entre les valeurs produites et les valeurs perçues (les barres d'erreurs correspondent aux écart-types). ....	51
Figure 30 : Pourcentages de stimuli par catégorie, chez chaque participant normo-entendant junior. ....	53
Figure 31 : Rapport entre les pourcentages relatifs de stimuli catégorisés comme les voyelles /e/ et /ɛ/, par participant normo-entendant junior. ....	53

Figure 32 : Procédure expérimentale de l'expérience de Convergence et Imitation .....	57
Figure 33 : Moyenne des pourcentages de changement lors des tâches de convergence et d'imitation (les barres d'erreurs correspondent aux erreurs standards)..	58
Figure 34 : Valeurs de pente par participant, dans les deux tâches. ....	58
Figure 35 : Valeurs de coefficient de corrélation par participant, dans les deux tâches. ....	59
Figure 36 : Résultats obtenus lors du test JND par les participants Juniors .....	60
Figure 37 : Corrélation entre les résultats du test de discrimination de fréquence et les valeurs de pente des tâches d'imitation et de convergence, chez les participants normo-entendants juniors .....	60
Figure 38 : Moyenne des temps de réaction chez les participants séniors (les barres d'erreurs correspondent aux écart-types). ....	79
Figure 39 : Moyenne des pourcentages de réussite chez les participants séniors (les barres d'erreurs correspondent aux écart-types).....	80
Figure 40: Moyennes des temps de réaction chez les participants juniors en condition bruitée (Juniors – B) et non-bruitée (Juniors - NB), ainsi que chez les participants séniors (Séniors) (en ms, les barres d'erreurs correspondent aux écart-types).....	81
Figure 41 : Moyenne des pourcentages de réponses correctes les participants normo-entendants juniors en condition bruitée (Juniors – B) et non-bruitée (Juniors-NB) et les participants séniors (les barres d'erreurs correspondent aux écart-types).....	82
Figure 42 : Valeurs normalisées de F1 (en Barks) pour /e/ produites par chaque participant séniors.....	85
Figure 43 : Valeurs normalisées de F1 (Barks) pour /ε/ produites par chaque participant séniors.....	85
Figure 44 : Valeurs normalisées de F1 lors de la perception de la voyelle /e/, pour chaque participant séniors. ....	86
Figure 45 : Valeurs normalisées de F1 lors de la perception de la voyelle /ε/, pour chaque participant séniors. ....	86
Figure 46 : Corrélation perception-production pour la voyelle /e/ chez les participants séniors. ....	87
Figure 47 : Corrélation perception-production pour la voyelle /ε/ chez les participants séniors. ....	87
Figure 48 : Corrélation entre les valeurs acoustiques produites et perçues par les participants séniors (les barres d'erreurs correspondent aux écart-types). ....	88

Figure 49 : Pourcentages de stimuli par catégorie, chez chaque participant normo-entendant sénior .....	89
Figure 50 : Rapport entre les pourcentages relatifs de stimuli catégorisés comme les voyelles /e/ et /ɛ/ par participant normo-entendant sénior. ....	89
Figure 51 : Moyenne de pourcentage de changement lors des tâches de convergence et d'imitation chez les participants séniors (les barres d'erreurs correspondent aux écart-types). ....	94
Figure 52 : Valeurs de pente par participant sénior, dans les deux tâches. ....	94
Figure 53 : Valeur de coefficient de corrélation par participant sénior, dans les deux tâches.....	95
Figure 54 : Résultats obtenus lors du test JND par les participants séniors .....	95
Figure 55 : Corrélation entre les résultats du test de discrimination de fréquence et les valeurs de pente lors des tâches de convergence et d'imitation, chez les participants normo-entendants séniors .....	96
Figure 56 : Coefficients de pente lors des tâches de convergence et d'imitation chez les participants juniors et les participants séniors (les barres d'erreurs correspondent aux écart-types). ....	97
Figure 57 : Clés manuelles de la LPC pour le Français (Bayard, 2014). ....	106
Figure 58 : Partie externe (A) et interne (B) d'un implant cochléaire de marque Cochlear®.....	107
Figure 59 : Fonctionnement d'un implant cochléaire.....	108
Figure 60 : Schéma des interactions perceptivo-motrices chez les normo-entendants (1) et schéma hypothétique des interactions perceptivo-motrices chez les patients sourds post-linguaux porteurs d'un implant cochléaire (2) .....	120
Figure 61 : Moyenne des temps de réaction pour les participants implantés (en ms, les barres d'erreurs correspondent aux écart-types). ....	123
Figure 62 : Moyenne des pourcentages de réponses correctes pour les participants implantés (les barres d'erreurs correspondent aux écart-types). ....	125
Figure 63 : Moyenne des temps de réaction chez les participants implantés (CI), chez les participants normo-entendants juniors en condition bruitée (Juniors – bruit) et non-bruitée (Juniors – non bruit), ainsi que chez les participants normo-entendants séniors (Séniors) (en ms, les barres d'erreurs correspondent aux écart-types).....	127
Figure 64 : Moyenne des pourcentages de réponses correctes chez les participants implantés (CI) et les participants normo-entendants juniors en condition bruitée (Juniors – B) et non-bruitée (Juniors-NB) et les participants séniors (les barres d'erreurs correspondent aux écart-types) .....	129

Figure 65 : Valeurs normalisées de F1 (en Barks) pour /e/ produite par chaque participant implanté.....	132
Figure 66 : Valeurs normalisées de F1 (en Barks) pour /ɛ/ produite par chaque participant implanté.....	132
Figure 67 : Valeurs normalisées de F1 lors de la perception de la voyelle /e/, pour chaque participant implanté .....	133
Figure 68 : Valeurs normalisées de F1 lors de la perception de la voyelle /ɛ/, pour chaque participant implanté .....	133
Figure 69 : Corrélacion perception-production pour la voyelle /e/ chez les participants implantés.....	134
Figure 70 : Corrélacion perception-production pour la voyelle /ɛ/ chez les participants implantés.....	134
Figure 71 : Corrélacion entre les distances /ɛ/ - /e/ lors des tâches de perception et de production.....	135
Figure 72 : Corrélacion entre les distances /ɛ/ - /e/ lors des tâches de perception et de production sans les deux participants ne présentant pas de contrastes entre les deux voyelles lors de la tâche de production.....	135
Figure 73 : Corrélacion entre les valeurs acoustiques produites et perçues par les participants implantés (les barres d'erreurs correspondent aux écart-types).....	136
Figure 74 : Pourcentages de stimuli par catégorie, chez chaque participant implanté. ....	136
Figure 75 : Rapport entre les pourcentages relatifs de stimuli catégorisés comme les voyelles /e/ et /ɛ/, par participant implanté.....	137
Figure 76 : Corrélacion entre les valeurs acoustiques globales produites et perçues chez les trois groupes de participants (les barres d'erreurs correspondent aux écart-types).....	139
Figure 77 : Résultats obtenus lors du test JND chez les participants implantés	143
Figure 78 : Corrélacion entre les valeurs de pente des deux tâches et les résultats du test de JND. ....	143
Figure 79 : Moyenne de pourcentage de changement lors des tâches de convergence et d'imitation chez les participants implantés (les barres d'erreurs correspondent aux écart-types). ....	144
Figure 80 : Valeurs de pente par participant implanté, dans les deux tâches. ...	144
Figure 81 : Valeur de coefficient de corrélacion par participant implanté, dans les deux tâches.....	144

Figure 82 : Coefficients de pente lors des tâches de convergence et d'imitation chez les participants implantés, les participants normo-entendants juniors, et les participants séniors (les barres d'erreurs correspondent aux écart-types). .....146

Figure 83 : Moyennes des temps de réaction (A) et des pourcentages de réponses correctes (B) par groupe et par condition. ....154

Figure 84 : Corrélations globales entre les valeurs produites et perçues chez les trois populations. ....156

Figure 85 : Coefficients de pente des tâches de convergence et d'imitation chez les trois groupes de participants. ....157

# Table des tableaux

Tableau 1 : Durée des /a/ initiaux (en ms) pour chaque stimulus.....	29
Tableau 2 : Effets significatifs et interactions pour les temps de réaction et les pourcentages de réponses correctes.....	36
Tableau 3 : Moyenne des temps de réaction et des pourcentages de réponses correctes pour toutes les conditions des expériences en conditions bruitée et non-bruitée.....	36
Tableau 4 : Valeurs des formants pour les stimuli utilisés dans la tâche de perception (en Barks et en Hertz) .....	46
Tableau 5 : Valeurs des bandes passantes pour les stimuli utilisés dans la tâche de perception (en Hertz) .....	46
Tableau 6 : Tests de significativité par voyelle selon l'ordre de passation des tâches. ....	52
Tableau 7 : Valeurs de F0 pour chacune des neuf voyelles présentées au participant, selon son sexe et sa moyenne de F0 .....	56
Tableau 8 : Valeurs moyennes de F1 pour chaque tâche et chaque groupe (en Barks).....	89
Tableau 9 : Etiologie des participants implantés (le terme "prothèse contrôle" réfère à une prothèse auditive classique, portée dans l'oreille non-implantée). L'expérience de l'implant correspond à la durée depuis l'activation de l'implant, alors que la durée de surdit� correspond à la période entre le début de la surdit� et l'activation de l'implant, la passation des différents participants ayant été réalisée sur plus d'un an.) .....	121
Tableau 10 : Effets significatifs et interactions pour les temps de réaction et les pourcentages de réponses correctes chez les participants implantés.....	124
Tableau 11 : Moyenne des temps de réaction (RTs, en ms) et des pourcentages de réponses correctes chez les participants implantés, par condition.....	124
Tableau 12 : Tableaux des effets significatifs et des interactions sur les temps de réaction et les pourcentages de réponses correctes lors de l'ANOVA comparant le groupe de participants implantés et les trois groupes de participants normo-entendants (Juniors bruités et non-bruités, Séniors) .....	127
Tableau 13 : Valeurs moyennes de F1 pour chaque tâche et chaque groupe (en Barks).....	139
Tableau 14 : Corrélation entre les paradigmes de close-shadowing et de convergence et d'imitation, chez les trois populations.....	158



# Introduction générale

---



## 1. Objectifs et plan de thèse

### 1. Problématique de la thèse

Dans le domaine de la communication orale, un débat fondamental et ancien concerne la nature des traitements et représentations mis en œuvre lors de la perception de la parole. Différentes propositions ont été émises ces cinquante dernières années, opposant des approches purement auditives, qui proposent que percevoir de la parole mobiliserait uniquement les aires auditives de notre cerveau, avec des théories purement motrices qui supposent que percevoir de la parole c'est percevoir les gestes articulatoires de notre interlocuteur. Nous nous situons entre ces deux grandes familles de théories et approches, dans un cadre perceptivo-moteur, et supposons que la perception de parole est à la fois sensorielle et motrice, c'est-à-dire que les aires auditives, visuelles et motrices entrent en collaboration étroite pour permettre à l'auditeur de décoder le signal de parole. Plus précisément, un auditeur, lorsqu'il perçoit un signal de parole, va le comparer avec ses propres représentations motrices internes, construites et instanciées dans le cerveau depuis le développement de la parole, afin d'aboutir à un meilleur décodage et donc une meilleure compréhension du message. Réciproquement, le locuteur va planifier des buts sensoriels (c'est-à-dire auditifs et somatosensoriels) à atteindre, et lors de la production il va comparer à ses objectifs initiaux les retours auditif et somatosensoriel qu'il reçoit afin de contrôler au mieux sa production, et de l'adapter si nécessaire, en cas de bruit ambiant par exemple.

La question centrale de cette thèse était de tester ces possibles liens perceptivo-moteurs aussi bien lors de la perception que de la production de la parole au travers de divers paradigmes comportementaux et différentes populations de participants. En premier lieu, nous avons donc élaboré un protocole expérimental composé de trois paradigmes comportementaux classiques, et nous l'avons testé et validé sur une population contrôle constituée de jeunes adultes normo-entendants.

Un premier objectif consistait en l'examen de patients sourds post-linguaux puis implantés afin de déterminer l'impact d'une déprivation sensorielle et les réapprentissage liés à leur implantation sur ces liens perceptivo-moteurs. Notre raisonnement était que les liens perceptivo-moteurs, fonctionnels avant la surdité chez ces patients, se sont trouvés inactifs pendant la déprivation sensorielle puis ont été confrontés au nouvel input apporté par l'implant. Notre questionnement principal était de comprendre comment sont réorganisées les relations sensori-motrices après l'implantation, vis-à-vis des anciennes et des nouvelles représentations sensorielles reçues par le patient.

Un second objectif a été d'évaluer le lien perceptivo-moteur chez une population normo-entendante âgée, afin, d'une part, d'évaluer la conséquence d'un déclin des fonctions cognitives et langagières chez ces personnes, et d'autre part, de déterminer si les résultats obtenus chez la population implantée, majoritairement âgée, étaient influencés par l'âge des participants.

## **2. Organisation du mémoire de thèse**

Ce mémoire de thèse sera organisé en trois grandes parties. Préalablement, en guise d'introduction, nous effectuerons une revue de la littérature portant sur les relations sensori-motrices lors de la production et de la perception de parole. Cette revue sera relativement brève, car un certain nombre de points en seront repris et développés dans les parties expérimentales qui suivront. Elle visera essentiellement à rappeler les grands enjeux des débats entre théories auditives, motrices et perceptivo-motrices de la communication parlée, et les types de paradigmes ayant permis de mettre en évidence les liens perceptivo-moteurs dans les domaines de la perception de la parole d'une part, de la production de la parole d'autre part.

Dans une première partie nous exposerons alors notre protocole expérimental général en présentant les différentes expériences le composant, la méthodologie utilisée et, enfin, les résultats observés auprès d'une population de jeunes adultes normo-entendants (dits « juniors »), correspondant à notre population contrôle par rapport aux deux autres populations testées (normo-entendants âgés, dits « séniors », et patients sourds porteurs d'un implant cochléaire). Dans une seconde partie, nous proposerons une brève revue de questions portant sur la production et la perception de parole chez les personnes séniors, avant de présenter les résultats obtenus chez notre groupe de personnes normo-entendantes séniors, en comparaison avec ceux de la population junior. Puis, lors de la troisième partie de ce manuscrit, nous effectuerons une présentation générale de la surdité et de l'implant cochléaire, nous introduirons différentes études précédemment menées sur la perception, la production de parole et les liens perceptivo-moteurs chez les personnes sourdes post-linguales puis implantées, avant d'exposer les résultats obtenus sur notre groupe de patients sourds post-linguaux porteurs d'un implant cochléaire. Enfin, nous discuterons de ces résultats, en cherchant à la fois à en tirer des enseignements généraux sur le rôle des relations sensori-motrices dans la communication parlée, et des enseignements plus spécifiques sur les possibles modifications des mécanismes de production et de perception associées à l'âge d'une part, à la mise en œuvre de l'implant cochléaire chez les sujets sourds post-linguaux d'autre part.

### **2. Relations sensori-motrices lors de la perception et de la production de parole**

Un débat classique dans le domaine de la perception de la parole concerne l'implication du système moteur et du lien fonctionnel entre représentations sensorielles, notamment auditives mais aussi visuelles et somatosensorielles, et motrices. Ainsi, les théories auditives de la perception de parole réfutent l'implication du système moteur et proposent l'engagement de processus auditifs de décodage phonétique à partir du signal acoustique (Diehl et coll., 2004). A contrario, Liberman et Mattingly (1985) supposent que pour percevoir la parole, nous utilisons des représentations procédurales motrices basées sur notre expérience de locuteur. Enfin, les théories perceptivo-motrices postulent que les représentations motrices soient utilisées en lien avec les représentations

sensorielles dans le traitement et décodage des informations phonétiques (Skipper et coll., 2007, Schwartz et coll. 2012).

De la même manière que pour la perception de parole, les différentes interprétations des processus de production de parole peuvent être classifiées en trois catégories, selon l'espace dans lequel sont spécifiés les buts que le locuteur cherche à atteindre pour garantir une bonne perception par des auditeurs. Certains modèles proposent que les buts soient définis dans le domaine articulaire. C'est le cas de la Task Dynamics (Saltzman, 1986), et de la Phonologie Articulaire qui lui est associée (Browman et Goldstein, 1992), ou du modèle de Öhman (Öhman, 1967). Dans ces modèles les buts sont spécifiés en termes de position et de degré de fermeture des constriction dans le conduit vocal (les « Vocal Tract Variables » de la Task Dynamics). D'autres modèles considèrent que les buts sont spécifiés dans l'espace des caractéristiques spectrales du signal, essentiellement sous la forme des zones spectrales de forte énergie (les « formants » pour les voyelles, ou les zones d'énergie maximales pour les consonnes). C'est ce qu'a proposé en particulier Stevens dans sa Théorie Quantique (Stevens, 1972, Stevens, 1988) et Perkell dans son modèle du contrôle de la production de la parole (Perkell et coll., 1995, Perkell et coll., 2000). La prédominance de la composante auditive dans la spécification des buts de la parole est aussi au cœur du modèle DIVA proposé par Guenther et collègues (Guenther et coll., 1998). Selon ce modèle, l'acquisition de la production de la parole chez l'enfant passe par la volonté d'atteindre des buts auditifs spécifiques définis comme des régions de l'espace des formants. La répétition de ces tâches tendues vers des objectifs auditifs permet l'émergence de buts somatosensoriels dans l'espace du conduit vocal. La production d'un son chez l'adulte n'est alors plus seulement associée au but auditif initial, mais aussi aux buts somatosensoriels ainsi élaborés, qui se traduisent pas des objectifs en termes de sensations tactiles ou proprioceptives. Le contrôle de la production de la parole met alors en jeu à la fois le feedback auditif, le feedback proprioceptif et le feedback tactile (Guenther & Vladusich, 2012). Les récents travaux d'Ostry et collègues (Nasir & Ostry, 2006 ; Lametti et coll., 2012), ou de Max et collègues (Max et coll., 2008 ; Feng et coll., 2011) ont confirmé l'importance de ces différents feedback dans le contrôle de la production de la parole. On peut donc aujourd'hui affirmer que les buts de la parole chez l'adulte sont multimodaux. L'expérience de Feng et coll. (2011) appliquant simultanément des perturbations somatosensorielles et auditives de la production de la parole a cependant suggéré que parmi ces composantes multimodales la composante auditive serait dominante.

### **1. Théories auditives de la perception de la parole**

D'une manière générale, les théories auditives de la perception de parole, comme dit précédemment, considèrent l'objet de la perception de parole comme uniquement acoustique. Ainsi, la théorie de l'invariance acoustique, proposée par Stevens et Blumstein (Stevens et Blumstein, 1978 ; Blumstein et Stevens, 1979) en accord avec la théorie quantique de Stevens (Stevens, 1972, 1989), revendique deux affirmations

principales : tout d'abord, des patrons acoustiques – ou, dans des versions progressives de la théorie, des patrons auditifs – invariants seraient présents dans le signal de parole ; ensuite, les auditeurs humains utiliseraient ces patrons invariants pour accéder aux structures phonologiques de leur langue à partir des sons de parole qu'ils perçoivent. Cependant dans cette théorie, le lien entre les gestes articulatoires et les informations issues du signal acoustique n'est pas complètement rejeté. En effet, c'est dans la relation non-linéaire – dénommée « quantale » par Stevens (1972, 1989) – entre gestes et sons que s'élaborent les invariants acoustiques, les non-linéarités délimitant naturellement des zones gestuelles associées à de faibles variations acoustiques, dites stables, séparées par des zones de forte instabilité. Ainsi, les invariants acoustiques référerait à ces zones articulatoires et porteraient donc intrinsèquement des informations sur le type de gestes qui les produisent.

La théorie de la variabilité adaptative (Lindblom, 1988, 1990) suppose également que la nature des traitements liés à la perception de parole soit acoustique et non articulatoire. En revanche, elle rejette l'idée d'invariants acoustiques ou articulatoires, du fait de productions articulatoires des locuteurs variables selon la rapidité de parole, les caractéristiques inhérentes au locuteur (âge, genre, région d'origine etc...) et les conditions environnementales (bruit environnant par exemple). Le locuteur ne contrôlerait pas sa production selon des invariants acoustiques ou articulatoires, mais selon des espaces de variation acoustique acceptables, et dépendant des capacités de traitement et de compréhension de l'auditeur. En effet, la théorie de la variabilité adaptative postule que l'auditeur, de son côté, décode le message de son interlocuteur sur la base des informations acoustiques disponibles mais aussi d'un ensemble complexe d'informations contextuelles qui lui sont spécifiques. Ainsi se met en place une véritable « négociation communicative » entre locuteur et auditeur, conférant aux phonèmes des propriétés acoustiques non pas invariantes mais « adaptativement variables » selon un « pouvoir discriminatif efficace » (Lindblom, 1990) qui permettent aux phonèmes d'être interprétés correctement en fonction du contexte global de communication.

Enfin, plus récemment, les approches générales auditives (Diehl et coll., 2003), spécifiées principalement par opposition aux théories motrices de la parole, affirment que les sons de parole sont traités par de purs mécanismes perceptifs – sans référence aux processus de production qui leur ont donné naissance – selon des processus d'apprentissage perceptif et d'adaptabilité aux sons similaires aux processus auditifs utilisés pour les sons environnementaux non-langagiers. Selon cette approche, les auditeurs catégoriseraient les sons de parole de la même manière que les sons non-langagiers au travers de mécanismes généraux de l'audition communs aux humains mais également aux animaux (ayant été montré que les oiseaux ou singes, à force d'apprentissage, peuvent différencier différentes syllabes).

## 2. Théories motrices de la perception de la parole

Une seconde proposition d'interprétation des mécanismes mis en œuvre lors de la perception de parole est portée par la théorie motrice et la théorie directe-réaliste. Ces théories supposent que la parole est perçue en référence à la façon dont elle est produite.

La théorie motrice de la perception de la parole, proposée initialement par Liberman et coll. (1952, 1967, 1985), est ancrée sur l'hypothèse que la perception de la parole implique la récupération perceptive des gestes du locuteur. Cette récupération se ferait par des mécanismes de perception spécifiques (« speech is special »), différents des mécanismes auditifs généraux, et reposant sur le fait que l'auditeur est aussi un locuteur et dispose donc des « clés de production » qui lui permettraient d'inférer les gestes à partir des sons – dans une version ultérieure de la théorie, c'est même en réalité l'intention motrice qui serait récupérée par l'auditeur, et non les gestes eux-mêmes. Plus spécifiquement, il est supposé que les auditeurs ont recours à un invariant présent au niveau des commandes motrices à l'origine du signal perçu, et non dans le signal de parole. Autrement dit, le signal de parole contiendrait des informations portant sur les gestes articulatoires prévus par le locuteur, et chez l'auditeur, un « module phonétique » serait présent dans le cerveau faisant appel aux connaissances articulatoires de l'auditeur communes à celles du locuteur. Cette explication permet d'éviter la complexité de la variabilité de production causée par la coarticulation, puisque les commandes (pré)motrices seraient elles invariables.

La théorie réaliste directe (Fowler, 1986, 1996), inspirée de la théorie motrice, suggère également que l'objet de la perception de la parole est moteur, mais considère que l'accès aux causes de la stimulation sensorielle est un processus général et non spécifique à la parole. Plus précisément, la perception de parole impliquerait une récupération directe des mouvements du conduit vocal et non inférentielle (comme proposé par la théorie motrice). Ainsi, le signal acoustique contiendrait des informations invariantes sur les gestes articulatoires produits par le locuteur, et lors de la perception, l'auditeur récupérerait directement les gestes articulatoires produits et présents dans le signal.

Très simplement résumé, ces deux théories supposent donc que l'objet de la perception de parole est moteur et non sensoriel, par opposition aux théories ou aux approches générales auditives, et sont opposées essentiellement par le fait que pour la théorie réaliste directe, le signal de parole reflète la conséquence directe des gestes articulatoires sur l'environnement sonore, et permet ainsi d'accéder directement aux gestes articulatoires, alors que selon la théorie motrice, le signal reflète l'intention, c'est-à-dire les commandes (pré)motrices, auxquelles l'auditeur accède par un système spécifique ancré dans ses capacités de production.

En 2006, dans une revue sur la théorie motrice, Galantucci et collègues organisent les données en faveur des théories motrices autour de trois arguments principaux. Le premier argument provient de la coarticulation et du fait que les

caractéristiques acoustiques des sons sont dépendantes du contexte et donc non invariantes. Ainsi un même geste articulatoire peut produire des sons différents. Un exemple classique est qu'une consonne plosive telle que /d/ produite devant une voyelle antérieure comme /i/ n'aura pas les mêmes transitions formantiques que si elle était produite devant une voyelle postérieure comme /u/ (voir Figure 1). Cependant, malgré cette absence de correspondance directe entre les sons et les phonèmes, les gestes moteurs porteraient, eux, une caractéristique constante, donc un invariant : dans notre exemple, un même geste de closure du conduit vocal dans la région alvéo-dentale au moyen d'un geste de l'apex de la langue. Cet invariant articulatoire fournirait le corrélât phonétique de la catégorie phonologique et constituerait l'objet ultime de la perception de parole.

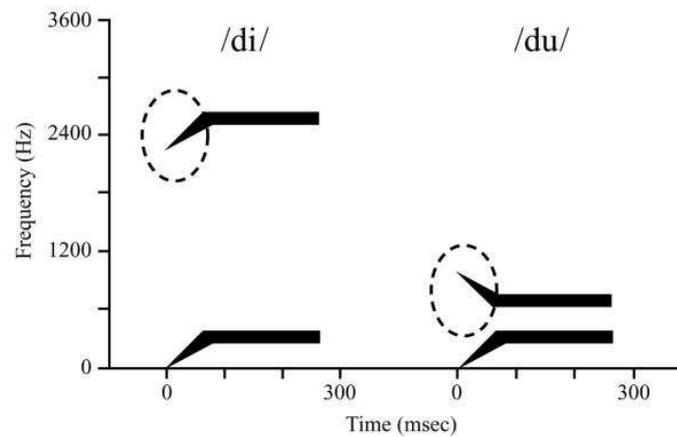


Figure 1 : Représentation spectrographique des formants F1/F2 des syllabes /di/ et /du/ (Figure empruntée à Galantucci et coll., 2006)

Le second argument apporté par Galantucci et coll. est celui de la multisensorialité lors de la perception de parole. En effet, l'apport d'une information supplémentaire au signal auditif, comme la vision du visage de son interlocuteur, va avoir un impact sur la perception de parole. Par exemple, l'effet McGurk (McGurk et McDonald, 1976) apporte une illustration classique quant à l'influence de la vision sur la perception : lorsqu'on présente un signal audiovisuel incohérent, c'est-à-dire lorsque l'information auditive et l'information visuelle sont différentes (par exemple un signal composé de la syllabe auditive /ba/ accompagnée de la syllabe visuelle /ga/) on peut observer la perception d'un stimulus qui ne soit conforme ni à l'entrée visuelle ni à l'entrée auditive, par exemple une troisième syllabe fournissant un compromis perceptif entre les deux informations (la syllabe /da/ dans l'exemple précédent). Également, l'addition de la modalité visuelle à un signal auditif bruité entraîne un meilleur décodage et donc une meilleure compréhension du message perçu, résultat montré Sumbly et Pollack (1954) puis confirmé par de nombreux travaux depuis. La modalité haptique, impliquée lorsque l'auditeur touche le visage de son interlocuteur, semble aussi interagir avec la modalité auditive, ainsi que l'ont mis en évidence Fowler et Deckle (1991) sur l'identification de signaux acoustiquement ambigus, et qui a été confirmé en électrophysiologie par Treille et coll. (2014). Ces différentes études menées sur la

perception multi-sensorielle de la parole apportent donc un argument important aux théories motrices puisque si le signal acoustique est décodé en le traduisant en gestes moteurs, alors la vision, ou le toucher, des gestes articulatoires devrait fournir une aide pour la perception et la compréhension du message.

Enfin, le dernier argument de Galantucci et coll. est porté par les expériences de close-shadowing (ou répétition rapide de syllabes). Dans ce paradigme, on observe des capacités de répétition de syllabes très rapide, ce qui a été interprété par le fait que, si les objets de la perception de parole sont les gestes articulatoires, alors les stimuli perçus sont directement dans le « bon format » pour être reproduits extrêmement rapidement. Nous développerons cet argument de manière plus détaillée lors de la partie I. 2. a., puisqu'il sera utilisé comme l'un de nos paradigmes de base dans la suite de cette thèse.

### **3. Théories perceptivo-motrices de la perception de parole**

Au-delà des théories uniquement motrices ou auditives, un certain nombre de théories perceptivo-motrices ont proposé que les connaissances motrices implicites des auditeurs et les représentations motrices associées soient utilisées en relation avec les représentations et les processus auditifs dans le traitement et le décodage des informations phonétiques.

Tout d'abord, Stevens et coll. (Halle et Stevens, 1959 ; Stevens, 1960 ; Halle et Stevens, 1962 ; Stevens et Halle, 1967 ; Stevens 1972) ont proposé un premier modèle d'analyse par synthèse de la perception de la parole, composé de deux branches, correspondant l'une à la production et l'autre à la perception. Plus précisément, lors de la perception de parole, l'input auditif est préalablement analysé par le système auditif. Ensuite, cette première analyse génère une représentation motrice, qui sera comparée avec l'input auditif (stocké de manière temporaire). Si aucune correspondance n'est observée entre l'input auditif et la représentation motrice, la procédure est répétée jusqu'à ce qu'une correspondance valide soit obtenue.

S'appuyant sur les travaux précédents de Stevens, Skipper et coll. (2007) ont proposé une version plus élaborée du modèle d'analyse par synthèse dans le cadre de la perception audiovisuelle de la parole, au travers d'une boucle de procédés perceptivo-moteurs (voir Figure 2). Ici, lors de la perception d'un signal audiovisuel, un premier stade de traitements multisensoriels va générer des hypothèses phonétiques sur les phonèmes produits par le locuteur. Dans le cas d'hypothèses ambiguës, non strictement déterminées, les commandes motrices permettant de produire ces phonèmes vont être stimulées, afin de produire des copies d'efférence qui seront finalement comparées avec l'input multisensoriel et permettront de contraindre l'interprétation phonétique finale.

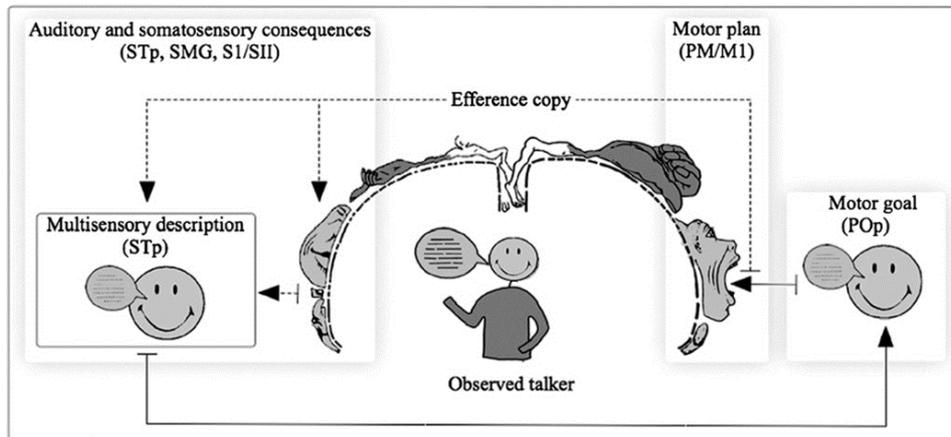


Figure 2 : Illustration de la boucle de procédés perceptivo-moteurs de Skipper et coll. (2007) lors de la perception audiovisuelle (Figure empruntée à Skipper et coll., 2007)

Proposant également une implication conjointe de processus multisensoriels et moteurs, la Théorie de la Perception pour le Contrôle de l'Action (PACT : Perception for Action Control Theory ; Schwartz et coll., 2002, 2012) suppose que les processus de décodage acoustico-phonétiques seraient en partie gérés par le système auditif, mais avec une possible implication importante du système moteur. L'appel à des représentations motrices dans le décodage du message serait basé sur des connaissances acquises au cours du développement. Ainsi, l'enfant, en s'accordant sur les productions de parole auxquelles il est confronté, et également en testant sa voix dans différents contextes environnementaux, va construire des représentations motrices qui lui sont propres, et plus tard, devenu adulte il continuera à utiliser ces représentations. Également, au-delà des connaissances motrices de chaque individu, les propositions de la PACT se basent sur les caractéristiques des phonèmes des langues du monde ; ceux-ci respecteraient à la fois des principes de production efficace (économie articulatoire), mais aussi des principes de distance acoustique et visuelle optimale, favorisant la perception de ces phonèmes. Plus généralement, ces différents arguments illustrent l'interaction et l'influence constante entre la production et la perception de parole.

Enfin, il est important de présenter un dernier modèle, neurobiologique, de la perception de parole, celui d'Hickok et Poeppel (2000, 2004, 2007), appelé modèle de double voie ventrale et dorsale (Figure 3), qui contraste deux voies de traitement impliquées dans la perception de parole et du langage. La voie ventrale partant des régions temporales supérieures de traitement dans le cortex auditif et allant vers le gyrus temporal moyen, permettrait l'accès au lexique et serait la voie essentielle de la compréhension. La voie dorsale, correspondant aux régions temporales postérieures, pariétales et frontales, analyserait la façon dont les sons sont produits, c'est-à-dire les gestes moteurs, et serait essentiellement impliquée dans les mécanismes d'apprentissage sensori-moteur pour la production de la parole.

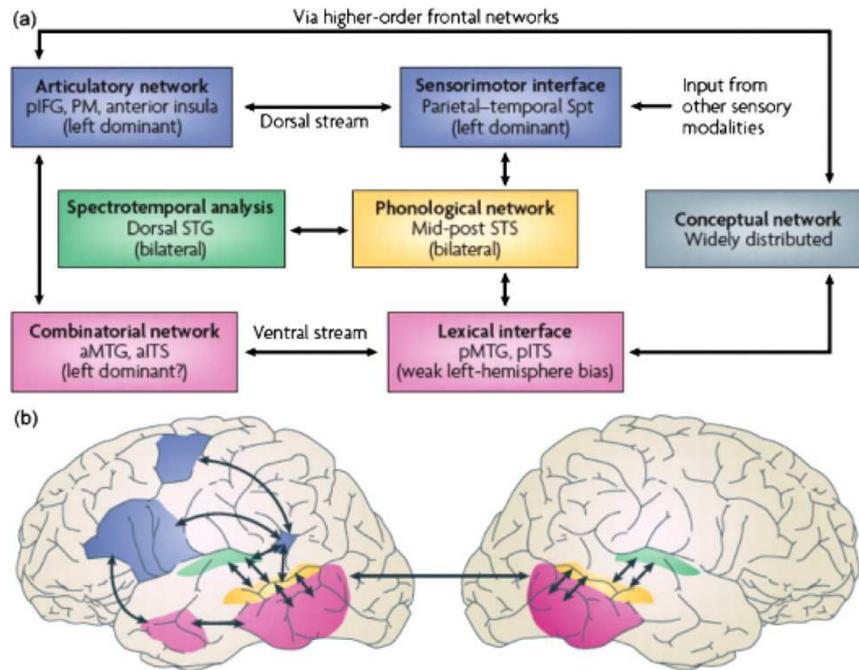


Figure 3 : Modèle à double voie de perception de la parole et du langage d'Hickok et Poeppel, la voie dorsale étant représentée en bleu, et la voie ventrale en rose (Figure empruntée à Hickok et Poeppel, 2007)

La voie dorsale d'apprentissage sensori-moteur serait plus particulièrement active lors du développement de la parole et de l'acquisition du langage chez les enfants. Chez l'adulte, la voie dorsale serait principalement sollicitée lors de tâches complexes de perception de parole, par exemple lors de tâches phonologiques ou de mémoire de travail verbale ; la boucle complète, quant à elle, servirait à adapter la production de parole lors de processus de correction ou de régulation articulatoire, comme on le verra plus loin (pour une revue, consulter Guenther, 2006).

#### 4. Théories motrices de la production de parole

La théorie motrice est principalement représentée dans la théorie de la phonologie articulatoire, développée par Browman et Goldstein (1989, 1986, 1992 ; et Goldstein et Fowler, 2003), qui proposent que l'unité de base de la production de parole soit le geste articulatoire ; ce geste étant défini comme une action articulatoire ayant un but distinctif. La définition du but de la production en tant que geste articulatoire permettrait tout d'abord une représentation de l'acte articulatoire pour le contrôle du processus de production, autrement dit l'adaptation des commandes motrices en fonction d'un retour sensoriel articulatoire ; mais également une représentation de contrastes phonologiques, c'est-à-dire que le geste catégoriserait l'unité phonologique en étant à la fois invariant au sein d'une catégorie phonologique, et spécifique pour permettre un contraste vis-à-vis des autres catégories.

Plus précisément la théorie de la phonologie articulatoire organise la production de parole sous forme de trajectoires articulatoires (voir Figure 4). Dans un premier temps, la séquence phonologique associée à l'énoncé souhaité génère une « partition

gestuelle» définissant la succession temporelle de positions du conduit vocal correspondant aux différents phonèmes le composant. Ces positions sont définies en termes de lieux d'articulation et de degré de constriction du conduit vocal. Ensuite, un modèle de tâche dynamique met en lien les positions attendues du conduit vocal avec les trajectoires permettant de les réaliser. Enfin, lors d'un dernier stade, un modèle de conduit vocal va générer le signal acoustique à partir des trajectoires articulatoires.

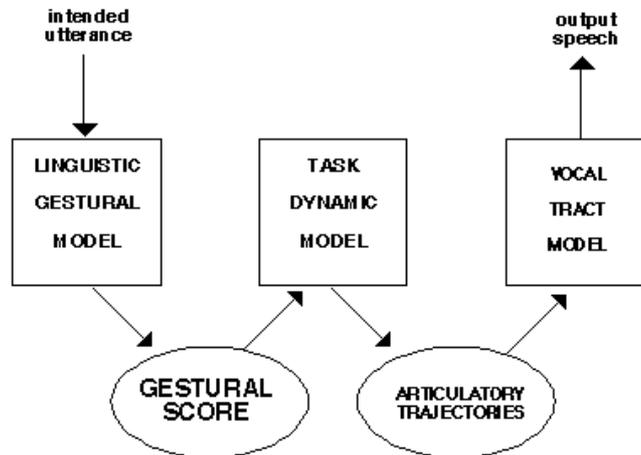


Figure 4 : Modèle de production de parole proposé par la théorie de la phonologie articulatoire (Figure empruntée à Browman et Goldstein, 1990)

## 5. Théories auditives de la production de parole

Les théories auditives de la production de parole (Guenther et coll., 1998 ; Perkell et coll., 1995 ; Perkell et coll. 2000 ; pour une revue, voir Perrier, 2005a) défendent l'idée que le but ultime de la production de parole est uniquement acoustique/auditif, les mouvements de parole étant utilisés pour parvenir à ce but et plusieurs arguments sont proposés dans ce sens. Tout d'abord, la difficulté de perception sensorielle du degré de constriction du conduit vocal est soulignée ; en effet, le feedback sensoriel ou tactile ne suffit souvent pas à percevoir le lieu et le mode de constriction, particulièrement pour les voyelles. Egalement, un autre argument essentiel en faveur des théories auditives de la production de parole est apporté par la variabilité des mouvements articulatoires, puisqu'ils s'adaptent lors de la coarticulation (un même phonème ne sera pas produit de manière identique selon les phonèmes qui le précèdent ou le suivent), lorsque la vitesse de production varie (lors de production rapide, les phonèmes seront plus influencés les uns par les autres) et selon l'accentuation présente dans la phrase ou dans le mot produits. Ainsi, plus précisément, l'exemple du phonème /r/ en anglais américain peut être exploré : ce phonème pouvant être produit avec deux configurations particulièrement contrastées (soit avec le dos de la langue positionné au niveau du palais, soit avec la pointe de la langue positionnée sur les alvéoles), ces deux configurations, aboutissant au même son, vont être choisies alternativement par le locuteur en fonction du contexte phonémique avoisinant le /r/.

## 6. Théories perceptivo-motrices de la production de parole

En partant en général des théories auditives, les théories perceptivo-motrices de la production de parole proposent une intégration entre systèmes moteur et auditif pour le contrôle orofacial lors de la production. Ainsi, si la production de parole impliquerait bien des cibles auditives permettant un contrôle « feedforward », un retour sensoriel fournirait une boucle de rétroaction (système feedback), permettant de corriger la production lorsqu'elle est erronée ou perturbée.

Ce feedback sensoriel est supposé être à la fois auditif, mais aussi somato-sensoriel, résultant des sensations apportées par les mouvements des articulateurs dans le conduit vocal. Gracco et Abbs (1987) sont parmi les premiers à pointer que la parole implique un feedback continu c'est-à-dire que l'exécution continue d'un programme moteur nécessite un flot également continu d'informations sensorielles apportées par les muscles et les sens cutanés à travers les régions respiratoires, laryngales et orofaciales.

Dans cette dernière catégorie, on peut mentionner le modèle de contrôle optimal GEPPETO (Perrier et coll., 2005b) et le modèle neuronal DIVA (Directions Into Velocities of Articulators) (Guenther, 2006), présenté en Figure 5. Ce modèle part d'une carte des sons de parole, qui définit des instructions motrices pour les articulateurs. Pendant la production, le système feedforward est mis en place, et les feedback sensoriels (c'est-à-dire auditif et somato-sensoriels) sont activés ; les retours sensoriels sont envoyés aux cartes d'erreurs auditives et somato-sensorielles où ils sont comparés avec les buts de production initiaux. En cas d'erreur de production, la carte de contrôle feedback va redéfinir et réadapter les commandes motrices afin d'aboutir à une production correspondant le mieux possible aux buts initiaux.

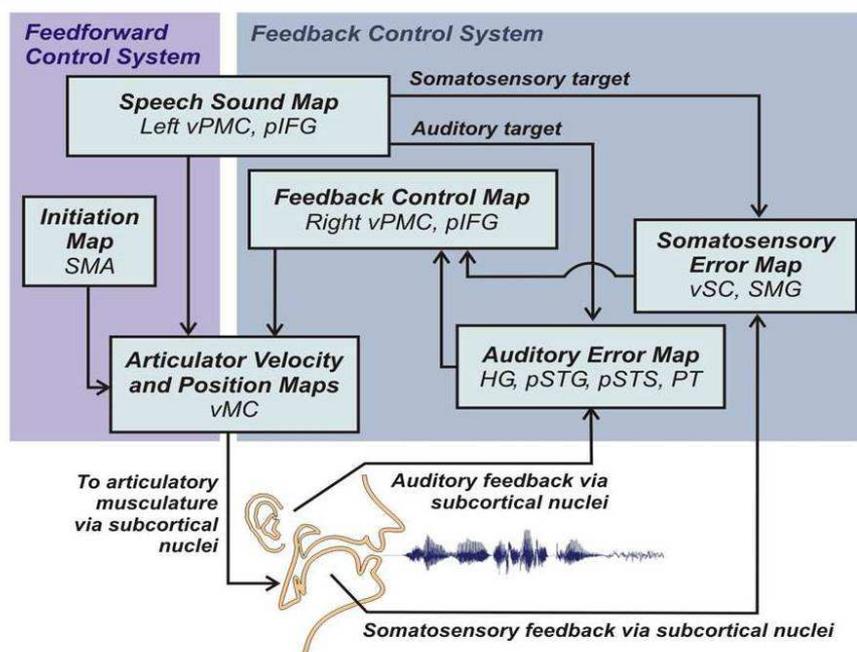


Figure 5 : Modèle DIVA de la production de parole (Figure empruntée à Guenther, 2006)

## 7. Arguments en faveur des théories perceptivo-motrices

Différents paradigmes expérimentaux comportementaux ont été utilisés pour illustrer les relations sensori-motrices, permettant donc de démontrer l'interaction entre les processus perceptifs et les connaissances articulatoires.

Précédemment, nous avons énoncé les arguments en faveur de la théorie motrice, avancés par Galantucci et coll. (2006), c'est-à-dire la coarticulation, la perception multisensorielle et le close-shadowing ; ces trois arguments peuvent être également associés aux théories perceptivo-motrices de la perception de parole. Par exemple, les interactions entre modalités auditive et visuelle lors de la perception de signaux audiovisuels peuvent représenter un argument pour une théorie perceptivo-motrice. En effet, le fait que le signal acoustique soit décodé en association avec des connaissances inférées sur les gestes moteurs fournit une explication toute naturelle à l'influence de la vision des gestes articulatoires sur les capacités de perception (voir aussi Skipper et coll. 2007).

Au-delà des études comportementales, un nombre croissant d'études neuro-anatomiques et neuro-physiologiques confirment l'existence d'un lien actif entre les systèmes auditif et moteur, à la fois lors de la production et la perception de parole (pour une revue, consulter la thèse de Grabski, 2012).

En effet, des études d'imagerie cérébrale (en Imagerie par Résonance Magnétique fonctionnelle – IRMf – ou en Magnétoencéphalographie – MEG) ont démontré à de nombreuses reprises l'implication d'aires corticales typiquement activées dans les processus de production de parole (c'est-à-dire le gyrus frontal inférieur gauche, le cortex prémoteur ventral, le cortex moteur primaire et/ou le cortex somato-sensoriel) pendant des tâches variées de perception de parole (Binder et coll., 2004 ; Wilson et Iacoboni, 2006 ; Grabski et coll., 2013a), particulièrement lors de conditions adverses, par exemple dans le bruit (Zekveld et coll., 2006), ou lors de la perception d'un accent étranger (Callan et coll., 2004).

D'un autre côté, les expériences de Stimulation Magnétique Transcranienne (Transcranial Magnetic Stimulation, TMS) confirment l'implication du système moteur dans la perception de parole, à la fois lors de la perception auditive et visuelle (Fadiga et coll., 2002 ; Möttönen et coll., 2004). Également, différentes études neurophysiologiques basées sur l'utilisation de la TMS montrent que la perturbation du système moteur pourrait entraîner de faibles modifications significatives sur les processus de perception (Meister et coll., 2007 ; Sato et coll., 2009 ; d'Ausilio et coll., 2009 ; Möttönen et Watkins, 2011 ; Grabski et coll., 2013b), même si les perturbations restent faibles et sont parfois difficiles à interpréter.

Concernant la production de parole, plusieurs expériences comportementales représentent des arguments importants en faveur d'interactions sensori-motrices. Tout d'abord nous pouvons citer les paradigmes de perturbation de la production, comme les expériences de lip-tube (Savariaux et coll., 1995, 1999). Dans ce paradigme, les

participants doivent produire la voyelle [u] de manière isolée, une première fois sans perturbation, une seconde fois avec perturbation (insertion d'un tube entre les lèvres) puis une dernière fois post-perturbation. Les résultats observés montrent une adaptation de la production lors de la perturbation, et même post-perturbation, démontrant des stratégies compensatoires afin d'adapter la production et ainsi de tenter d'atteindre les mêmes buts sensoriels.

La perturbation du feedback auditif entraîne là aussi des modifications dans la production. Par exemple, au cours de l'effet Lombard, si un bruit important est ajouté pendant la production, les locuteurs vont automatiquement adapter leur articulation, en sur-articulant ou en augmentant leur intensité (Lombard, 1911). Également, plusieurs études ont montré que lors de l'ajout d'un délai entre la production et le feedback auditif, les locuteurs ont tendance à ralentir leur production allant même jusqu'à entraîner des erreurs de parole (Lee, 1990 ; Stuart et coll., 2002 ; Jones et Striemer, 2007 ; Chen et coll., 2013). Le fait que des perturbations puissent entraîner des réadaptations claires du système de production montre le rôle important des feedbacks auditifs et somato-sensoriels et fournit des arguments en faveur du rôle de boucles sensori-motrices.

En lien avec la modification du feedback auditif, la surdité est également un autre argument important pour montrer le rôle des mécanismes d'intégration sensori-motrice dans la production de parole. En effet, alors qu'une personne née sourde présente une production perturbée et parfois peu intelligible (Cowie et Douglas-Cowie, 1983), une personne devenue sourde plus tardivement conserve une parole peu perturbée et intelligible (Matthies et coll., 2008), suggérant une acquisition puis un maintien des représentations motrices, avant de se dégrader au fur et à mesure des années. Dans les deux cas (surdité de naissance ou tardive), la production perturbée aura tendance à s'améliorer après la pose d'un implant cochléaire et donc la récupération d'une certaine sorte de retour auditif, cette amélioration pouvant être interprétée comme une preuve de l'importance du retour auditif lors des processus feedback-feedforward (Perkell et coll., 2012).

Enfin deux autres types de paradigmes sont également intéressants pour illustrer les relations sensori-motrices : les phénomènes de compatibilité motrice, où certaines études ont démontré une corrélation très forte entre la façon dont les phonèmes, et plus particulièrement les voyelles, sont distribuées dans le répertoire perceptif et moteur chez un même locuteur (Bell-Berti et coll., 1979 ; Fox, 1982 ; Ménard et Schwartz, 2014) ; et les phénomènes d'imitation consciente et inconsciente lors de situations non-interactives ou interactives, démontrant une nouvelle fois que les commandes motrices s'adaptent à l'environnement à partir de l'input auditif reçu. Nous développerons ces deux paradigmes plus en détail dans la partie suivante.



# Partie I : Relations perceptivo- motrices par des participants normo- entendants Juniors

---



## 1. Introduction

L'objectif principal de cette première partie est de présenter le protocole expérimental mis en place pour évaluer les liens perceptivo-moteurs lors de la perception et de la production de parole. Ainsi, dans un premier temps, nous présenterons une revue de la littérature exhaustive portant sur les trois paradigmes expérimentaux comportementaux composant le protocole expérimental général, et inscrirons ces paradigmes au sein des théories perceptivo-motrices de la perception et de la production de parole. Ensuite, pour chaque paradigme, nous exposerons la méthodologie utilisée, puis nous présenterons les résultats obtenus pour la population contrôle de jeunes adultes normo-entendants.

## 2. Organisation générale de l'expérience et hypothèses globales

Le protocole expérimental mis en place avait pour but de tester l'existence d'interactions entre systèmes sensoriels et articulatoire lors de la perception et de la production de la parole au travers de trois expériences comportementales menées sur des participants normo-entendants « juniors » (c'est-à-dire âgés de moins de 40 ans), « séniors » (âgés de plus de 60 ans) et, enfin, chez des participants sourds post-linguaux porteurs d'un implant cochléaire (âgés de 27 à 72 ans).

En regard des résultats observés chez les participants normo-entendants juniors, nous souhaitons déterminer dans quelles mesures les interactions possibles entre systèmes sensoriels et moteur pourraient s'établir chez des patients sourds post-linguaux après une implantation cochléaire, quels seraient les dysfonctionnements éventuels du couplage perceptivo-moteur chez ces patients, et enfin quel serait le lien entre la ré-activation du couplage perceptivo-moteur – sous une forme différente liée aux nouvelles stimulations sensorielles prodiguées par l'implant cochléaire – et la probable récupération des capacités de perception et de production de parole après plusieurs mois d'implantation. Nous avons également décidé d'étendre notre étude aux participants seniors, afin de tester d'éventuelles modifications de l'efficacité et du rôle du couplage perceptivo-moteur sur cette population et de distinguer pour les résultats des patients implantés, quelle est la part relative à leur implantation par rapport à leur âge.

Nous avons choisi de nous centrer sur un protocole exclusivement comportemental, considérant les difficultés inhérentes à l'usage de certaines techniques de neuroimagerie et neurophysiologie (imagerie par résonance magnétique fonctionnelle, électroencéphalographie, magnéto-encéphalographie) sur les participants porteurs d'implant cochléaire. Plusieurs designs expérimentaux sont susceptibles de tester et d'étudier les relations perceptivo-motrices lors de la perception et de la production de parole, cependant nous avons choisi trois paradigmes expérimentaux bien précis : un premier paradigme de répétition syllabique rapide réalisé selon différentes modalités de présentation auditive et audiovisuelle (paradigme dit de « close-shadowing »), un second paradigme comparant l'organisation perceptivo-motrice de voyelles du Français lors de leur production et perception (paradigme dit d'« idiosyncrasies ») et enfin un troisième

paradigme testant les capacités d'imitation automatique et volontaire des participants (paradigme dit de « convergence et imitation »).

### 1. Paradigme expérimental de close-shadowing auditif et audiovisuel

Le premier paradigme expérimental que nous souhaitons utiliser est donc le close-shadowing, consistant pour les participants à produire des stimuli de parole immédiatement après les avoir entendus et utilisant de fait la compétence naturelle de répétition de parole pour tester les liens perceptivo-moteurs en parole.

Dans leur étude pionnière, Porter et Castellanos (1980) ont comparé les temps de répétition lors de deux tâches de perception de parole impliquant des séquences voyelle-consonne-voyelle (VCV) (/aba/, /apa/, /ama/, /aka/, /aga/). Lors de la première tâche, les participants devaient répéter le plus rapidement possible la séquence VCV entendue. Pour ce faire, ils devaient d'abord produire la voyelle initiale /a/ puis produire la consonne dès qu'ils l'avaient perçue et identifiée, suivie de la voyelle finale /a/. Lors de la seconde tâche, à choix simple cette fois-ci, les participants devaient répéter la voyelle initiale et, lors de la transition consonantique, ils devaient alors prononcer aussi vite que possible la syllabe /ba/ ce, quel que soit le stimulus perçu. Comme attendu, les auteurs ont observé des temps de répétition plus rapides lors de la tâche de répétition à choix simple par rapport à la tâche de close-shadowing. Cependant, de manière intéressante, cette différence était extrêmement faible : de l'ordre de 30 à 60 ms de différence entre les deux tâches. Galantucci et collègues (2006) (pour plus de détails sur cet article, se reporter à l'introduction générale, partie 2) ont par la suite comparé ces résultats avec des temps de réaction obtenus par Luce en 1986. Celui-ci a utilisé le même type de paradigme consistant en une tâche à choix simple et une tâche à choix multiple, avec des stimuli comparables, mais en demandant aux participants de répondre en appuyant sur une touche plutôt que de produire la réponse oralement. Ainsi, lors de la tâche à choix multiple, les participants devaient appuyer sur la touche correspondant à la syllabe qu'ils percevaient, et lors de la tâche à choix simple, ils devaient appuyer sur la même touche quelle que soit la syllabe entendue. Dans l'étude de Luce (1986), les différences de temps de réaction entre les deux tâches étaient bien plus importantes que chez Porter et Castellanos (1980) lors des tâches de répétition orale : de 100 à 150 ms de différence chez Luce, contre 30 à 60 ms chez Porter et Castellanos. Pour Galantucci et collègues, ces différences appuient l'hypothèse selon laquelle percevoir la parole c'est percevoir les gestes articulatoires qui la sous-tendent et que, lors de cette tâche, la perception des gestes à l'origine de la syllabe perçue va dès lors directement contrôler la réponse orale et la rendre ainsi plus rapide. En soutien de cette hypothèse, Fowler et collègues (2003) ont publié une étude de close-shadowing basée sur celle de Porter et Castellanos (1980) dans laquelle les participants devaient répéter des syllabes lors d'une tâche à choix simple et une tâche à choix multiple, en réponse à trois types de stimuli : /apa/, /ata/ et /aka/. Lors de la tâche à choix simple, ils ont attribué une des trois syllabes à chacun des participants, qui devaient répéter le /a/ initial et ensuite produire leur syllabe CV dès que la consonne du stimulus était présentée, ceci indépendamment du type de consonne

perçu. Lors de la tâche à choix multiple, les participants devaient simplement répéter le plus vite possible toutes les syllabes VCV. Comme chez Porter et Castellanos (1980), Fowler et collègues ont obtenu de faibles différences de temps de réaction lors de la tâche à choix simple par rapport à la tâche à choix multiple. Dans la tâche à choix simple, ils ont également comparé les temps de réaction entre les trois groupes de participants (un par syllabe attribuée) et ont observé des temps de réaction plus courts pour les stimuli correspondant à la syllabe attribuée (voir Figure 6). Ces résultats ont été interprétés par Fowler et collègues (2003), tout comme par Galantucci et collègues (2006), comme suggérant que les stimuli acoustiques de parole sont perçus comme des gestes articulatoires fournissant un but de réponse très précoce qui viendrait moduler le temps de réaction en fonction de la compatibilité entre le stimulus et la réponse attendue.

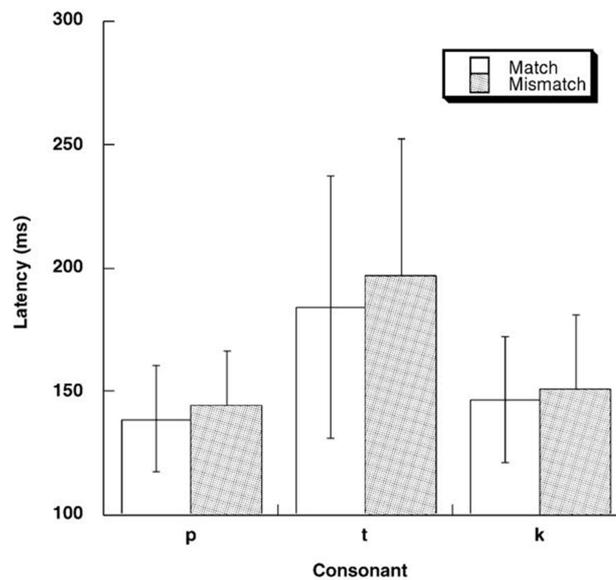


Figure 6 : Un des résultats obtenus par Fowler et coll. (2003) : différence de temps de réaction lors des réponses simples selon la compatibilité (colonnes blanches) ou l'incompatibilité (colonnes grises) entre le stimulus perçu et la consonne attribuée au participant (figure empruntée à Fowler et coll. 2003)

Dans le cadre de cette thèse, nous avons choisi d'utiliser le paradigme de close-shadowing dans un contexte de multisensorialité. Il n'est plus à prouver que la lecture labiale est une aide pour la compréhension de la parole. Si l'importance de la lecture labiale chez les participants malentendants est connue depuis longtemps, les participants normo-entendants sont également capables de lire sur les lèvres (Cotton, 1935) et nous savons depuis Sumby et Pollack (1954) que l'addition de la modalité visuelle à un signal auditif bruité entraîne un meilleur décodage et donc une meilleure compréhension du message perçu. Ces différentes études menées sur la perception audiovisuelle démontrant la complémentarité entre modalité visuelle et auditive constituent un argument possible en faveur des théories perceptivo-motrices de la perception de la parole. En effet, si nous décodons le signal acoustique de parole en le traduisant en gestes moteurs, même partiellement, alors il semble naturel que la vision

des gestes articulatoires participe également à la perception et la compréhension du message linguistique.

L'apport de la modalité visuelle a été également démontré dans des études de répétition sans contrainte de temps (paradigme dit de "shadowing"). Ainsi, les expériences menées par Reisberg et coll. (1987) ont permis de mettre en évidence une meilleure compréhension, par des participants, de matériaux linguistiques complexes ou produits dans une langue étrangère ou avec un accent étranger. Reisberg et coll. ont testé deux groupes de participants anglophones et mesuré la précision de leur production lors de la répétition de phrases en français ou en allemand, dans des conditions de présentation auditive et audiovisuelle. Les participants ont obtenu de meilleurs scores en termes de précision de répétition dans la condition audiovisuelle en comparaison à la condition uniquement auditive. Reisberg et coll. ont ensuite testé un groupe de participants anglophones qui devaient répéter des stimuli en anglais produits avec un accent belge, dans des conditions auditives et audiovisuelles, lors de trois expériences : une première avec des phrases simples, une seconde avec des phrases plus complexes et enfin une troisième avec des mots rares. Encore une fois, les participants ont obtenu des meilleurs scores dans la condition audiovisuelle. Par la suite, Davis et Kim (2001) ont testé des scores de précision de répétition de phrases coréennes, par des locuteurs anglophones naïfs, dans une expérience de répétition différée, lors de conditions auditive et audiovisuelle. Suite à la tâche de répétition, les participants devaient écouter des stimuli puis devaient déterminer s'ils avaient déjà entendu le stimulus ou non. Lors des deux tâches, de répétition et de catégorisation auditive, les performances étaient meilleures pour la condition audiovisuelle de présentation.

Cependant, ces tâches de répétition se sont faites sans pression de temps. Or la tâche de close-shadowing fournit une fenêtre riche sur la dynamique temporelle des processus de décision. A l'inverse, les expériences de close-shadowing ont rarement incorporé la modalité visuelle, se privant d'une connaissance sur le rôle des interactions audiovisuelles en lien avec les relations perceptivo-motrices. La seule étude à notre connaissance est celle de Beautemps et coll. (2003) qui ont enregistré quatre participants devant réaliser une tâche de close-shadowing en condition auditive seule et en condition audiovisuelle, en réponse à des stimuli VCV impliquant les consonnes /p/, /t/ et /k/. Les auteurs ont obtenu des temps de réaction plus longs (entre 227 et 314 ms) que ceux obtenus par Porter et Castellanos (1980). Cette différence de temps de réaction est expliquée par le manque d'entraînement des participants à la tâche. Elle est probablement également due à la précision des réponses plus élevée (6,8% contre 23% chez Porter et Castellanos). De manière intéressante, Beautemps et collègues ont observé un avantage perceptif lors de la condition audiovisuelle par rapport à la condition auditive seule chez deux des quatre participants (avec un gain du visuel de 13,4 à 14,7%). Chez un des participants, ce gain a été obtenu uniquement pour la consonne /p/, en raison de la haute visibilité de cette consonne bilabiale. Cependant, cette étude ne comportant que quatre participants, il nous a semblé nécessaire de la répliquer avec un plus grand nombre de participants.

Nous proposons donc, à travers notre premier paradigme expérimental, d'étudier quel est l'apport de la modalité visuelle dans une tâche de close-shadowing. De plus, pour parfaire l'étude d'un tel apport possible de la modalité visuelle, nous avons également utilisé des stimuli acoustiques bruités, en plus de stimuli de parole non-bruités, de façon à dégrader leur perception auditive mais à accroître le rôle éventuel de la lecture labiale.

## **2. Paradigme d'idiosyncrasies lors de la perception/production de voyelles**

Le second paradigme que nous souhaitons mettre en œuvre consiste en la production et la perception de voyelles du Français et vise à déterminer une possible corrélation entre les caractéristiques perceptivo-motrices des voyelles produites et perçues.

Des études précédentes ont de fait observé des corrélations possibles entre la façon dont les phonèmes sont répartis dans le répertoire perceptif et moteur d'un même locuteur. Bell-Berti et collègues (1979) ont ainsi étudié les caractéristiques perceptives et articulatoires du contraste /i/ vs. /I/ en anglais américain. Ils ont comparé deux groupes de participants lors d'une tâche de catégorisation des voyelles /i/ et /I/, à partir d'un continuum acoustique ; les deux groupes ayant été sélectionnés à partir de données électromyographiques (EMG) observées lors de la production de ces mêmes voyelles et sur la mise en évidence de deux stratégies articulatoires de différenciation de cette paire de voyelles antérieures hautes (production basée sur la hauteur ou sur la tension de la langue). Des idiosyncrasies (c'est-à-dire une variabilité inter-locuteurs et ici inter-groupes) ont été observées lors de la tâche de catégorisation et, de manière importante, ces idiosyncrasies ont été corrélées avec les idiosyncrasies articulatoires dans la tâche de production.

De manière similaire, Fox (1982) a analysé les corrélations acoustiques au sein du triangle vocalique lors de la production et de la perception de voyelles de l'anglais américain. Suite à la production des voyelles et à la mesure des caractéristiques acoustiques des voyelles produites pour chaque participant, les participants ont ensuite dû effectuer une tâche de classification auditive de ces mêmes voyelles. Pour ce faire, les voyelles étaient présentées par paires et les participants devaient choisir si ces voyelles étaient proches ou lointaines sur une échelle de neuf points, ceci afin de déterminer la répartition des voyelles dans l'espace vocalique de chaque participant. Fox (1982) a observé des corrélations entre les caractéristiques acoustiques des voyelles produites par les participants et la façon dont les participants semblent répartir les voyelles perçues dans leur espace vocalique.

En 2008, Ménard et collègues ont étudié la variabilité de production des voyelles du Français, en étudiant plus particulièrement l'organisation des caractéristiques de hauteur de ces voyelles. Ils ont enregistré 27 locuteurs de deux régions francophones, répartis en trois groupes : un groupe d'enfants de 4 ans, un groupe d'enfants de 8 ans et un groupe d'adultes, et leur ont demandé de produire les dix voyelles du français. Leurs productions ont fait l'objet d'une analyse spécifique afin de quantifier la distribution des

voyelles hautes, mi-hautes, mi-basses et basses sur l'axe F1 (premier formant) : tout d'abord la moyenne des valeurs de F1 pour chaque voyelle a été mesurée en Barks, une échelle perceptive de hauteur, utilisée fréquemment lors des mesures de formants ; ensuite, les valeurs normalisées des voyelles mi-hautes et mi-basses ont été calculées comme suit (où  $M_1$  = Moyenne des valeurs de F1 des voyelles 3 voyelles hautes /i y u/ ;  $M_4$  = Moyenne des valeurs de F1 de la voyelle basse /a/ ;  $Y_V$  = Valeur normalisée d'une voyelle donnée v, mi-haute ou mi-basse) :

$$Y_v = (\text{Moyenne}_v - M_1) / (M_4 - M_1)$$

Ménard et collègues ont obtenu une variabilité inter-locuteurs forte en ce qui concerne les positions des différentes voyelles au sein de l'espace vocalique (voir Figure 7). Le point crucial de cette étude est que, au niveau des contrastes de hauteur, l'implémentation semble être différente selon les participants. Ainsi, certains participants ont produit les voyelles hautes et mi-hautes (comme /i/ et /e/) très proches en terme de F1, d'autres ont produit des voyelles mi-basses et basses (comme /e/ et /a/) relativement proches et enfin, un dernier groupe de locuteur a produit les voyelles mi-hautes et mi-basses avec des valeurs proches pour F1. Plus intéressant encore, les locuteurs semblent avoir produit les voyelles appartenant à un même degré de hauteur (soit, /e ø o/ et /ε œ ɔ/) selon des valeurs stables de F1, indépendamment du lieu d'articulation et de l'arrondissement des voyelles.

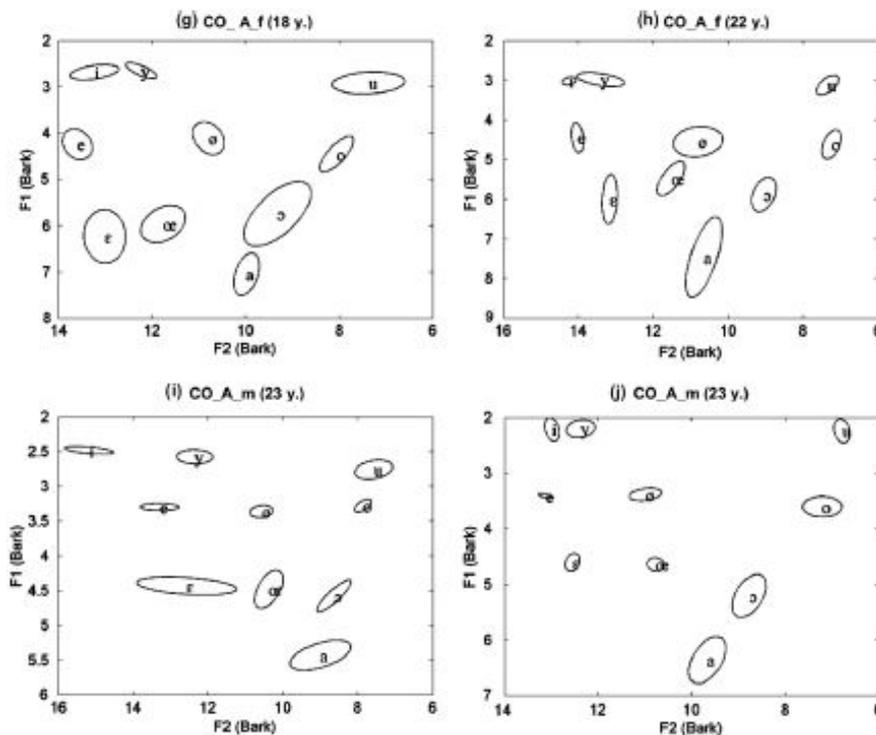


Figure 7 : Ellipses de dispersion des dix voyelles orales du français chez quatre participants adultes de l'étude de Ménard et coll. 2008 (Figure empruntée à Ménard et coll., 2008)

Suite à cette étude sur la production de voyelles, en 2014, Ménard et Schwartz ont étudié la répartition des contrastes de hauteurs de voyelles perçues. Ils ont repris les données d'une tâche perceptive ayant été exécutée à la même période que l'expérience de production, dans laquelle il avait été demandé à douze participants ayant réalisé la tâche de production décrite plus haut d'effectuer une tâche de catégorisation de voyelles. Les stimuli utilisés ont été synthétisés à partir du modèle VLAM, un modèle articulatoire du conduit vocal permettant de générer un espace maximal acoustique pour les voyelles susceptibles d'être produites par des locuteurs d'âges différents. Ménard et Schwartz ont sélectionné trois groupes de stimuli : un groupe de voyelles susceptibles d'être produites par un adulte, et trois groupes de voyelles susceptibles d'être produites par un enfant à trois âges donnés (nouveau-né, 4 ans et 8 ans). Les deux groupes de participants enfants ont dû catégoriser 76 stimuli « enfant » correspondant à leur âge et 38 stimuli adultes tandis que le groupe de participants adultes a dû catégoriser les 38 stimuli « adultes » et 76 stimuli « nouveau-né ». Pour les analyses, les stimuli ont été classés selon la façon dont les participants les avaient catégorisés, c'est-à-dire que « chaque catégorie de voyelle a été associée à une série de stimuli perçus comme des membres de cette catégorie » ; ensuite, pour chaque voyelle, la moyenne des valeurs de F1 et de F2 a été calculée en Barks et les valeurs normalisées pour les voyelles mi-hautes et mi-basses ont été calculées de la même manière que dans la tâche de production de l'expérience précédente.

Les résultats obtenus ont révélé des idiosyncrasies lors de la perception des voyelles, et, pour la plupart des participants, les cibles de voyelles appartenant à des degrés de hauteur similaires ont été alignées sur des valeurs de F1 communes, et ce indépendamment du lieu d'articulation et de l'arrondissement, comme lors de la production de voyelles. Mais surtout, en comparant les caractéristiques des voyelles produites et perçues, les auteurs ont observé que les valeurs formantiques normalisées associées à chaque degré de hauteur pour un participant donné lors de la tâche de perception étaient corrélées avec celles observées chez le même participant lors de la tâche de production de voyelles (voir Figure 8). Ainsi, par exemple, un locuteur, qui produit des voyelles hautes et mi-hautes proches, va également catégoriser à proximité les voyelles hautes et mi-hautes. Ces résultats démontrent l'existence d'une organisation phonémique commune au système perceptif et moteur et constituent donc un argument supplémentaire aux théories perceptivo-motrices.

Nous proposons donc, pour notre second paradigme, de reprendre le protocole expérimental des deux études de Ménard et collègues dans le but de tester le lien possible entre les systèmes sensoriels et moteur lors de la production et de la perception de parole.

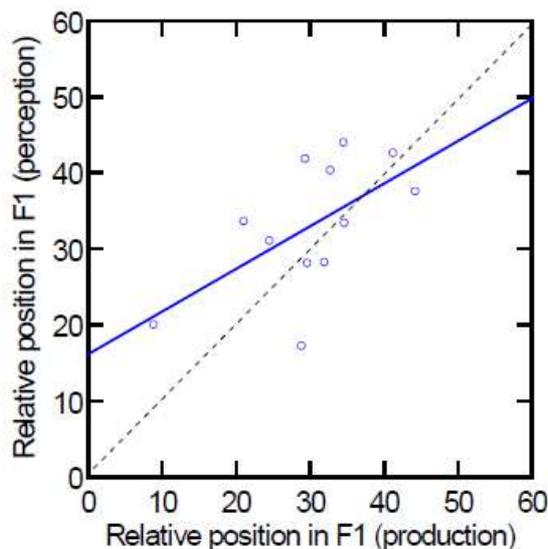


Figure 8 : Corrélation entre les valeurs normalisées de F1 en production et en perception pour la voyelle [e] chez les participants de l'étude de Ménard et Schwartz, 2014 (Figure empruntée à Ménard et Schwartz, 2014)

### 3. Paradigme de convergence et d'imitation

Le troisième paradigme expérimental que nous avons décidé d'utiliser porte sur la convergence phonétique et l'imitation volontaire. Ce type d'expérience représente un bon moyen de tester l'intégration sensorimotrice lors de la production de parole.

Dans des situations de communication interactive, les locuteurs s'adaptent perpétuellement à leur interlocuteur, ceci probablement afin de faciliter les échanges entre les acteurs d'une telle situation communicative. Ces mécanismes d'adaptation, dits de convergence, peuvent être volontaire, c'est-à-dire que le locuteur imite consciemment son interlocuteur, mais aussi involontaire (voir le récent numéro de *Frontiers in Psychology* consacré à ce thème, « Speech imitation: the cognitive underpinnings of adaptive vocal behaviour »). L'imitation inconsciente, présente lorsque le locuteur converge de manière apparemment spontanée vers son interlocuteur, a été explorée lors de nombreuses études qui ont observé une convergence à la fois sur des paramètres non-langagiers comme les gestes mais surtout sur des paramètres acoustiques de la parole, comme l'intensité, la fréquence fondamentale (F0) ou les formants (pour une revue récente, voir Lelong, 2012). Ces effets de convergence phonétique peuvent être expliqués par l'aspect social de la communication : la convergence des productions vers une zone commune de paramètres acoustiques et articulatoires pourrait de fait faciliter l'échange, et de manière plus générale, pourrait être associée à un désir d'appartenance à un groupe social.

Si la plupart des études menées sur la convergence phonétique reposent sur des échanges conversationnels en condition naturelle, quelques études ont montré que la convergence phonétique pouvait également avoir lieu dans le cadre de situation non-conversationnelle et contrôlée en laboratoire. Ainsi, Delvaux et Soquet (2007) ont étudié

lors d'une tâche de dénomination les capacités de convergence phonétique de participants exposés à des stimuli de parole provenant de locuteurs ayant un accent régional différent du leur. Ils ont observé que les productions des participants convergeaient vers celles des locuteurs perçus ce de manière automatique et involontaire. De manière remarquable, un after-effect a également été observé : plus de dix minutes après la fin de l'exposition aux stimuli, les participants continuaient à présenter dans leurs productions une certaine influence des caractéristiques acoustiques préalablement perçues. Les chercheurs ont supposé que ce phénomène de convergence phonétique pourrait être expliqué non pas uniquement par des mécanismes sociaux-psychologiques mais aussi par des mécanismes automatiques d'accommodation de plus bas niveau, et pourraient être à la base des variations phonétiques observées au sein de communautés en contact.

Par la suite, Sato et coll. (2014) et Garnier et coll. (2014) ont réalisé des expériences d'imitation volontaire et involontaire de voyelles isolées, lors de situations de production de parole non-conversationnelle. Ils ont enregistré trois groupes de participants. Le premier groupe a dû produire des voyelles selon une cible visuelle orthographique, puis selon une cible auditive et, enfin de nouveau selon une cible orthographique. Cette procédure de production de voyelles orthographiques et auditives, donnée sans aucune consigne d'imitation, avait pour but de tester de possibles mécanismes de convergence phonétique lors de la présentation des cibles auditives par rapport à la première présentation de cibles orthographiques (soit une modification des productions des participants fonction des voyelles perçues auditivement) et également de possibles after-effects lors de la présentation finale des cibles orthographiques (soit la persistance de effets de convergence observés). Le second groupe a réalisé exactement la même procédure expérimentale, avec cette fois la consigne d'imiter les voyelles et la voix perçue. Un sous-groupe de participants a réalisé les deux expériences afin que les chercheurs puissent comparer la convergence phonétique et l'imitation volontaire. Enfin, une troisième expérience a été réalisée dans le but de tester si les possibles after-effects observés lors de ces tâches étaient liés non pas à la production des locuteurs mais à la simple exposition auditive aux voyelles présentées. La procédure expérimentale était similaire en tous points aux précédentes, excepté que lors de la présentation des cibles auditives, il était demandé aux participants de réaliser une tâche de catégorisation manuelle de ces cibles vocaliques. Lors de ces trois expériences, les cibles acoustiques ont été manipulées en termes de F0 et de F1. Les analyses acoustiques réalisées sur les productions des participants ont montré des changements imitatifs très forts lors de la tâche d'imitation volontaire, et des changements imitatifs plus faibles mais cependant significatifs lors de la tâche de convergence (voir Figure 9). L'analyse des productions du sous-groupe de participants ayant réalisé les deux tâches n'a démontré aucune corrélation entre la tâche de convergence et celle d'imitation. Enfin, de manière intéressante, des after-effects ont été observés lors des trois expériences, c'est-à-dire que les participants ont convergé involontairement vers les cibles acoustiques qu'ils avaient précédemment produites, imitées ou catégorisées. Pris ensemble, ces résultats suggèrent

que la convergence phonétique s'appuie sur des mécanismes sensori-moteurs de bas niveau, en dehors de tout cadre conversationnel, et que la production de parole s'adapte continuellement sur l'environnement de parole externe.

A partir de ces expériences, nous proposons, dans notre troisième paradigme, de tester les capacités d'imitation volontaire et involontaire, et par là même, de tester l'intégration perceptivo-motrice mise en œuvre lors de la production de parole.

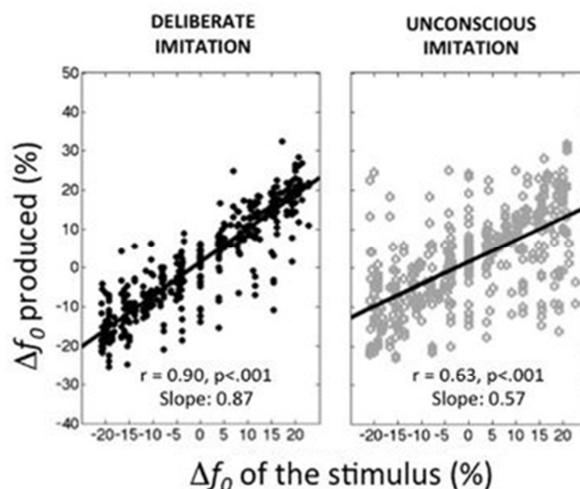


Figure 9 : Une partie des résultats obtenus par Garnier et coll., 2014 : changements imitatifs lors des tâches d'imitation consciente (figure de gauche) et inconsciente (figure de droite) (abscisse : déviance de la F0 du stimulus ; ordonnée : déviance de la production lors des tâches par rapport à la production initiale) (figure adaptée de Garnier et coll. 2014).

#### 4. Procédure expérimentale commune aux trois expériences

L'ordre de passation des trois expériences était le même pour tous les participants : ils devaient tout d'abord réaliser l'expérience de close-shadowing, puis l'expérience d'idiosyncrasies et enfin l'expérience de convergence et imitation (voir Figure 10). Préalablement aux trois expériences, les participants devaient passer un audiogramme afin de vérifier leurs capacités auditives, sur des fréquences de 125Hz, 250Hz, 500Hz, 750Hz, 1000Hz, 1500Hz, 2000Hz, 3000Hz, 4000Hz et 8000 Hz. puis à la fin des expériences, ils devaient réaliser un court test de discrimination de fréquences afin de contrôler leur capacité de discrimination de fréquences (nommé « Test JND », voir ci-dessous).

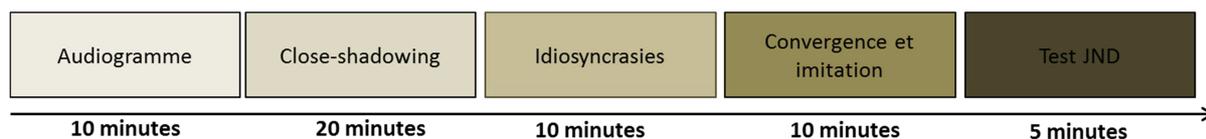


Figure 10 : Ordre de passation et durée des différentes expériences

Les trois expériences ont été réalisées dans une chambre sourde. Les participants étaient assis en face d'un ordinateur à une distance d'approximativement 50 cm. Les

stimuli acoustiques étaient présentés à un niveau sonore confortable via une enceinte, sauf pour l'expérience de close-shadowing impliquant des stimuli bruités, pour laquelle nous avons utilisé des écouteurs. Le niveau sonore était le même pour tous les participants. Le logiciel Presentation (Neurobehavioral Systems, Albany, CA) a été utilisé pour contrôler la présentation des stimuli et pour enregistrer les réponses manuelles. Toutes les productions des participants ont été enregistrées grâce à un microphone AKG 1000S pour les analyses offline, avec un système assurant la synchronisation entre les stimuli présentés et la réponse des participants. Une courte session d'entraînement précédait chaque tâche. Ce protocole expérimental a été validé par le CERNI (Comité d'éthique pour les recherches non interventionnelles) et tous les participants ont donné par écrit leur accord en signant, préalablement aux expériences, un formulaire de consentement.

### **1. Test JND**

Le test JND, permettant de tester les capacités de discrimination de fréquences des participants, a été mis en œuvre sur le modèle du paradigme expérimental proposé par Moore & Vinay (2010). Deux groupes de quatre sons étaient présentés aux participants, composés soit de sons identiques (AAAA), soit de deux sons différents en terme de fréquence (ABAB), et ce dans un ordre aléatoire, le participant devant déterminer dans quel groupe se trouvaient les sons différents entre eux (ABAB).

Les stimuli utilisés étaient des stimuli synthétiques obtenus par addition d'harmoniques d'une fréquence fondamentale  $F_0$  donnée, le spectre respectant celui d'une voyelle /e/ produite soit par un homme soit par une femme – selon le sujet testé – extraite des stimuli utilisés pour la tâche de convergence et d'imitation. Pour le son A la valeur de fréquence fondamentale  $F_0$  choisie était égale à la valeur moyenne produite par le participant, et le son B était systématiquement plus haut que le son A, en commençant par une augmentation de 5% au début du test, cet écart diminuant au cours du test. De plus des variations aléatoires d'intensité étaient appliquées sur l'ensemble AAAA et ABAB d'une écoute à la suivante, afin d'obliger les sujets à se concentrer sur la perception de hauteur et non sur des fluctuations de perception de timbre ou d'intensité.

Lorsque le participant répondait correctement deux fois de suite, la différence de fréquence entre les sons A et B diminuait, alors qu'une erreur entraînait une augmentation de la différence de fréquence entre les deux sons. Le seuil différentiel estimé (« jnd » pour « just noticeable difference ») correspondait à la moyenne sur les 8 derniers essais de l'entre  $F_0$  pour les stimuli A et B.

### **3. Expérience de close-shadowing**

#### **1. Principe et hypothèses**

#### **2. Expérience non-bruitée**

Comme nous l'avons vu précédemment, deux arguments importants en faveur du rôle possible du système moteur dans la perception de la parole sont la rapidité des

mécanismes de close-shadowing et les effets d'interaction et d'intégration multisensorielle. Nous avons souhaité combiner ces deux arguments lors d'une expérience de close-shadowing audiovisuel.

L'expérience consistait en deux tâches de catégorisation : une tâche de close-shadowing où les réponses étaient données oralement, en répétant le plus vite possible la séquence présentée, et une tâche de décision manuelle, où les réponses étaient données manuellement, en appuyant le plus vite possible sur la touche appropriée. Les stimuli à catégoriser correspondaient aux séquences /apa/, /ata/ et /aka/, comme dans l'étude de Fowler et collègues (2003). Les participants étaient informés qu'on allait leur présenter ces séquences, soit de manière auditive soit de manière audiovisuelle.

Lors de la tâche de close-shadowing, on leur demandait de catégoriser et répéter chaque séquence le plus vite possible. Pour ce faire, ils devaient produire la voyelle initiale /a/ puis répéter immédiatement la syllabe CV perçue (/pa/, /ta/ ou /ka/). Lors de la tâche de décision manuelle, les participants devaient catégoriser chaque énoncé en appuyant le plus vite possible avec leur main dominante sur une des trois touches correspondant respectivement à /apa/, /ata/ ou /aka/, l'ordre des touches étant contrebalancé entre les participants.

Nos hypothèses concernant cette première expérience étaient les suivantes :

- Concernant les temps de réaction, nous supposons que les réponses orales seraient plus rapides que les réponses manuelles, conformément aux résultats de Luce (1986).
- Concernant les pourcentages de réponses correctes, nous supposons que les taux d'erreur seraient similaires dans la tâche de close-shadowing et dans la tâche de catégorisation manuelle

### **3. Expérience bruitée**

En considérant les résultats d'une étude « pilote » montrant des scores très élevés et des différences extrêmement faibles de performances (tant du point de vue des pourcentages de réponses correctes que des temps de réaction) entre modalités auditive et audiovisuelle, nous avons décidé de réaliser une expérience de close-shadowing avec des stimuli bruités. L'objectif principal de cette seconde étude était non pas tant de produire une réduction de performances en condition auditive mais de mettre en évidence un avantage produit par l'entrée visuelle en modalité audiovisuelle du fait du signal acoustique détérioré.

Nous avons pour cette seconde expérience portant sur des stimuli bruités trois hypothèses principales :

- Les temps de réactions devraient être globalement plus longs et les pourcentages de réponses correctes plus faibles par rapport à ceux obtenus lors de l'expérience portant sur des stimuli non-bruités.
- L'avantage de la modalité audiovisuelle sur la modalité audio seule devrait être plus marqué dans l'expérience bruitée que dans l'expérience non-bruitée,

tant en termes de pourcentages de réponses correctes que de temps de réaction.

- Le même portrait de différences devrait être observé entre les deux tâches, avec des réponses plus rapides et des scores de réponses correctes équivalents pour la tâche impliquant des réponses orales par rapport à celle impliquant des réponses manuelles.

## 2. Matériel et méthodes

### 1. Expérience non-bruitée

#### 1. Participants

La première expérience de close-shadowing réalisée avec des stimuli non bruités a impliqué 16 participants (11 femmes et 5 hommes), âgés de 20 à 40 ans (moyenne d'âge : 29 ans). Tous les participants étaient normo-entendants, de langue maternelle française, sans déficits auditifs ni troubles de la vision non corrigés et sans troubles du langage.

#### 2. Stimuli

Pour créer les stimuli audio et audiovisuels, nous avons enregistré un locuteur masculin, âgé de 48 ans, de langue maternelle française (ne participant pas aux expériences), dans une chambre sourde. Nous lui avons demandé de produire plusieurs fois les séquences VCV (Voyelle Consonne Voyelle) /apa/, /ata/ et /aka/. Les productions ont été enregistrées grâce à une caméra de haute qualité digitale, placée en face du locuteur afin de filmer son visage, puis numérisées à 44100 Hz pour la partie audio et à 50 Hz sur 572 x 520 pixels pour la partie vidéo au format PAL/SECAM.

Durée initiale attendue	Syllabe	Durée réelle du /a/ initial	Durée réelle du /a/ final
500	aka	444	242
500	apa	439	233
500	ata	438	234
1000	aka	967	243
1000	apa	950	242
1000	ata	961	243
1500	aka	1441	237
1500	apa	1440	246
1500	ata	1438	243
2000	aka	1955	244
2000	apa	1960	243
2000	ata	1963	235

Tableau 1 : Durée des /a/ initiaux (en ms) pour chaque stimulus

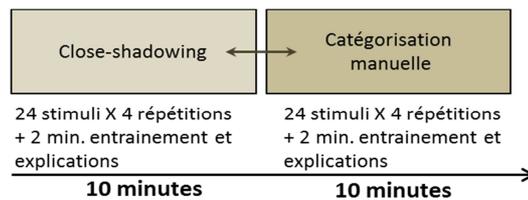
Le corpus utilisé pour l'enregistrement était constitué de séquences VCV /apa/, /ata/ et /aka/, avec quatre durées différentes du /a/ initial (500, 1000, 1500 et 2000 ms). Pour ce faire, nous avons demandé au locuteur de maintenir la production de la voyelle initiale sur les différentes durées. Le logiciel Presentation a été utilisé pour présenter les stimuli à produire sur un écran d'ordinateur placé en face du locuteur, ainsi que pour monitorer les différentes consignes de durées des voyelles initiales. Nous avons ainsi créé

48 stimuli (4 durées initiales x 3 séquences x 4 répétitions). Pour chaque séquence et chaque durée initiale, une répétition a été sélectionnée parmi les quatre (sur la base de la qualité de la vidéo et de la production), afin d'obtenir 12 stimuli. Les signaux audiovisuels ont été ensuite redécoupés avec le logiciel Virtual Dub pour égaliser les durées des stimuli /a/ initiaux (mesurées sur le signal acoustique entre le début et la fin du voisement) et finaux (mesurées sur le signal acoustique entre l'instant de plosion de la consonne et la fin du voisement de la voyelle) entre les 3 stimuli /apa/, /ata/ et /aka/, et ce en prenant en compte la période (20 ms) des échantillons d'images (voir Tableau 1). Pour obtenir les stimuli audio, nous avons simplement extrait la partie audio des stimuli audiovisuels grâce au logiciel VirtualDub.

### 3. Procédure expérimentale

Pour chaque tâche (avec réponses manuelles ou orales), 4 répétitions de chacune des 24 séquences audio et audiovisuelles /apa/, /ata/ et /aka/ étaient présentées de manière randomisée. L'ordre de présentation des tâches était contrebalancé entre les participants. Chaque tâche orale et manuelle durant 8 minutes, la durée totale de l'expérience était d'environ 20 minutes, explications, session d'entraînement aux deux tâches et pauses comprises (voir Figure 11).

Figure 11 : Procédure expérimentale de l'expérience de close-shadowing



## 5. Analyses

Les temps de réaction et les pourcentages de réponses correctes ont été déterminés en combinant les données récupérées par le logiciel Presentation dans le cas de réponses manuelles (données portant sur la réponse fournie, et sur l'instant d'appui de la touche correspondante), avec les enregistrements numérisés des stimuli et des réponses orales (voir Figure 12). Pour ce faire, en nous appuyant sur le logiciel Praat, nous avons d'abord segmenté les productions des participants et les stimuli, selon une procédure semi-automatique. Cette procédure implique une segmentation de chaque production, basée sur un algorithme de détection de l'intensité et de la durée.

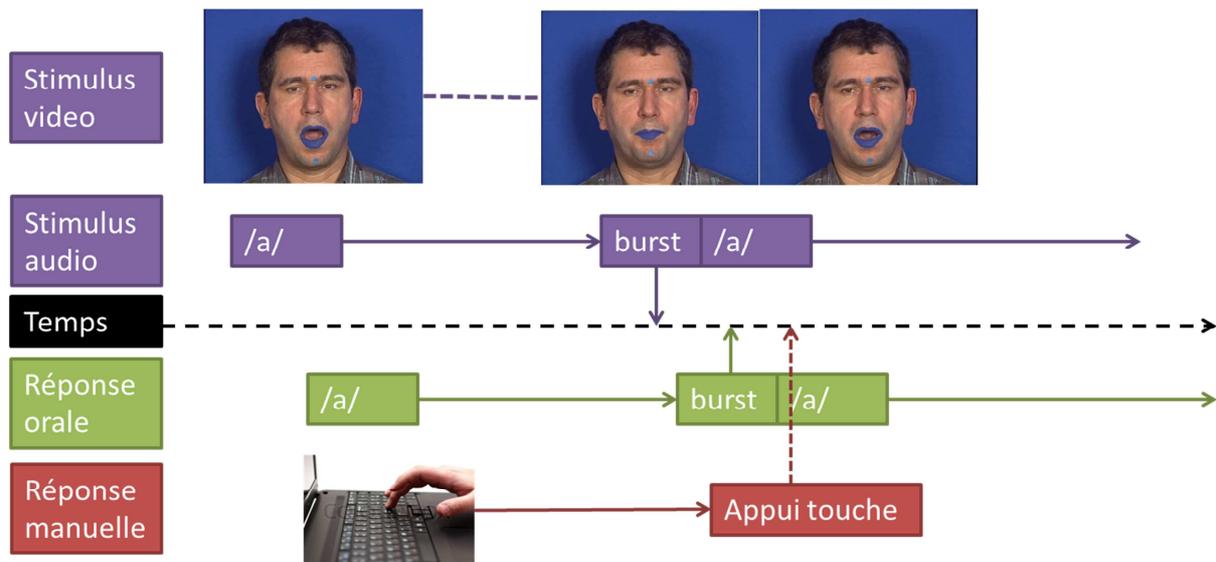


Figure 12 : Calcul des temps de réaction. Temps de réaction mesurés entre l'instant de début de la plosion du stimulus et l'instant de début de la réponse pour les réponses orales et entre la plosion du stimulus et l'appui touche pour les réponses manuelles.

Ensuite pour chaque stimulus présenté, nous avons noté, pour les réponses orales, quelle était la réponse des participants. Nous avons également, sous Praat, détecté manuellement les instants précis de début de plosion de la consonne du stimulus présenté (S) et, dans le cas de réponses orales, de plosion de la consonne de la réponse (R) (voir Figure 13).

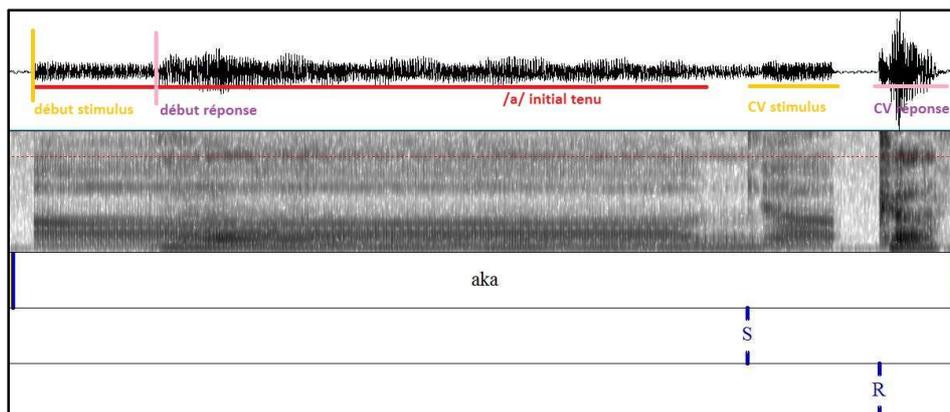


Figure 13 : Exemple d'annotation des données de l'expérience de close-shadowing : la première et la seconde lignes représentent le signal et son spectrogramme, stimulus et réponse étant mêlés ; la troisième ligne indique la syllabe produite ; les quatrième et cinquième ligne indiquent l'emplacement du burst du stimulus (S) et du burst de la réponse (R), respectivement.

Les temps de réaction ont été calculés entre le début de la plosion des consonnes de S et R pour les réponses orales, et entre le début de la plosion de S et l'appui touche pour les réponses manuelles (voir Figure 13). Les temps de réaction ont été calculés uniquement sur les réponses correctes : les omissions ou tous autres types d'erreurs, comme les substitutions de consonne, la production de deux consonnes ou de deux syllabes dans la tâche à réponses orales ou le double appui touche dans la tâche à réponses manuelles, ont été exclus de cette analyse.

A partir de ces données de réponse et de temps de réaction pour chaque stimulus présenté, la proportion des réponses correctes et la médiane des temps de réaction ont été déterminées pour chaque tâche (réponse orale vs. réponse manuelle), chaque modalité (audio vs. audiovisuelle) et chaque séquence (/apa/ vs. /ata/ vs. /aka/).

## 6. Analyses statistiques

Pour tester statistiquement les différentes hypothèses explicitées précédemment, nous avons effectué deux ANOVA à mesures répétées sur ces données (une ANOVA pour les temps de réaction et la deuxième pour les pourcentages de réponses correctes), avec les variables tâche, modalité et séquence comme variables intra-participants. Nous avons également réalisé des tests post-hoc avec correction de Bonferroni et fixé le seuil de significativité à  $p=.05$  ; également, nous avons vérifié la sphéricité des données grâce au test de Mauchly, et lorsque nécessaire, nous avons effectué des corrections de Greenhouse-Geiser (Greenhouse et Geiser, 1959).

### 2. Expérience bruitée

#### 1. Participants

La seconde expérience de close-shadowing réalisée avec des stimuli bruités a impliqué 14 participants (11 femmes et 3 hommes), âgés de 19 à 34 ans (moyenne d'âge :

24 ans). Tous les participants étaient normo-entendants, de langue maternelle française, sans déficits auditifs ni déficits de la vision corrigée et sans troubles du langage.

## 2. Stimuli

Pour créer les stimuli bruités, tous les stimuli de la base de données utilisée lors de l'expérience non-bruitée ont été mixés avec un bruit blanc filtré passe bas à -6db/octave selon un rapport signal sur bruit de -3dB (l'énergie du signal étant définie à partir du début de la plosion jusqu'à la fin de la voyelle).

## 3. Procédure expérimentale

La procédure expérimentale était la même que pour l'expérience avec des stimuli non bruités (voir partie 3), excepté le fait que les stimuli étaient présentés aux participants via un casque audio et non via une enceinte. Cette nouvelle configuration nous a en effet permis d'enregistrer les stimuli et les réponses des participants sur deux pistes différentes synchronisées d'un enregistreur numérique PMD, afin de faciliter le traitement des données, opérant sur des réponses non contaminées par le bruit présent sur les stimuli. La durée totale de l'expérience portant sur les stimuli bruités était la même que celle de l'expérience portant sur les stimuli non bruités, c'est-à-dire 20 minutes environ.

## 4. Analyses

Pour le traitement des productions de la tâche de close-shadowing en condition bruitée, étant donné que le bruit masquait la plosion du stimulus, nous n'avons pas pu annoter celui-ci directement. Nous avons donc détecté sous Praat le début du stimulus et, en connaissant la durée du /a/ initial, nous avons pu calculer où se situait le début de la plosion. La suite du traitement des données était la même que pour la tâche de close-shadowing en condition non-bruitée (voir partie 0).

Comme précédemment, la proportion des réponses correctes et la médiane des temps de réaction ont été déterminées pour chaque tâche (réponse orale vs. réponse manuelle), chaque modalité (audio vs. audiovisuelle) et chaque séquence (/apa/ vs. /ata/ vs. /aka/).

## 5. Analyses statistiques

Pour tester statistiquement les différentes hypothèses explicitées précédemment, nous avons analysé conjointement les résultats des deux tâches de close-shadowing avec stimuli bruités et non-bruités. Pour ce faire, nous avons effectué deux ANOVA à mesures répétées sur ces données (une ANOVA pour les temps de réaction et la deuxième pour les pourcentages de réponses correctes), avec la variable groupe (bruit vs. non bruit) comme variable inter-participant<sup>1</sup>, et les variables tâche, modalité et séquence comme variables intra-participants. Nous avons également réalisé des tests post-hoc avec correction de

---

<sup>1</sup> Tout au long du manuscrit, comme la différence entre conditions bruitée et non-bruitée correspond à l'emploi de deux groupes de sujets différents, nous emploierons le terme de groupe pour référer à la différence entre conditions bruitée et non-bruitée.

Bonferroni et du fait des comparaisons multiples lors de cette seconde série d'analyses, nous avons fixé le seuil de significativité à  $p = .025$  (sur ce point, il est de plus à noter que tous les effets et interactions observés avaient un seuil de significativité inférieur à  $p < .005$ ). Egalement, nous avons vérifié la sphéricité des données grâce au test de Mauchly, et lorsque nécessaire, nous avons effectué des corrections de Greenhouse-Geiser (Greenhouse et Geiser, 1959).

### 3. Résultats

#### 1. Expérience non-bruitée

##### 1. Temps de réaction

Aucune différence significative n'a été observée entre les stimuli audio et audiovisuels. Par contre, les temps de réaction ont été plus courts lors des réponses orales par rapport aux réponses manuelles (240 ms vs. 462 ms) ( $F(1,14)=81.8$  ;  $p<.001$ ).

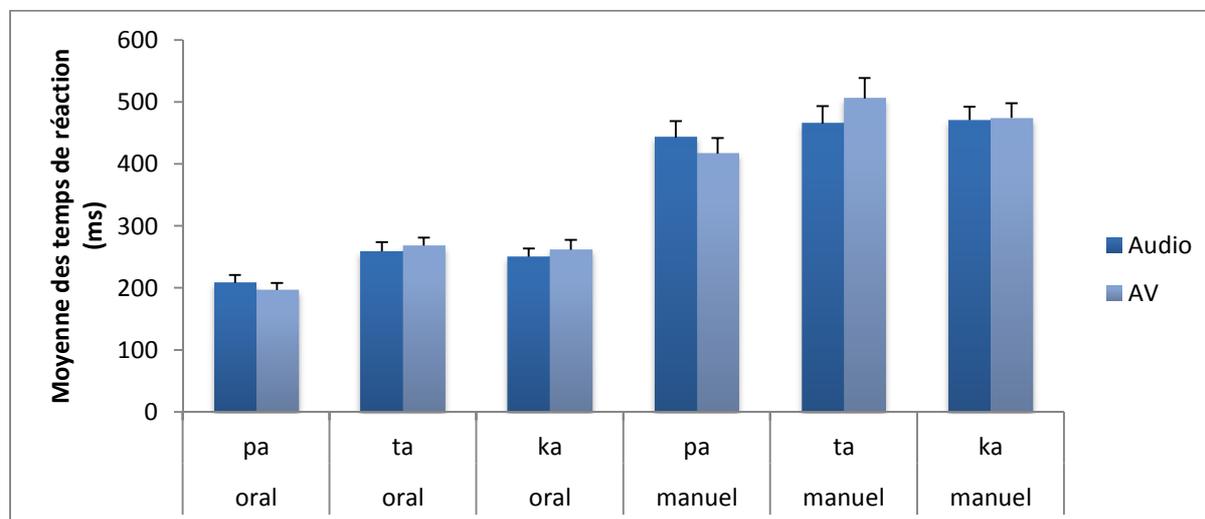


Figure 14 : Moyenne des temps de réaction (en ms) pour toutes les conditions de l'expérience non-bruitée (les barres d'erreurs correspondent aux écart-types).

L'interaction entre les trois facteurs (modalité, réponse et syllabes) s'est avérée significative ( $F(2,28)=4.6$  ;  $p<.01$ ). En effet, en mode de réponse orale, les temps de réaction pour la syllabe /pa/ ne sont pas différents entre les modalités audio (196 ms) et audiovisuelle (208 ms), alors qu'en mode de réponse manuel, on observe un avantage de la modalité audiovisuelle sur la modalité audio, les stimuli audiovisuels étant plus rapidement reconnus (415 ms) que les stimuli audio (442 ms). De manière surprenante, pour la syllabe /ta/, les temps de réaction manuels sont plus longs en réponse aux stimuli audiovisuels (506 ms) par rapport aux stimuli audio (465 ms). Enfin, pour la syllabe /ka/, aucune différence n'a été observée entre les modalités et entre les tâches (voir Figure 14).

##### 2. Pourcentages de réponses correctes

Comme pour les temps de réaction, aucune différence n'a été observée entre les stimuli audios et audiovisuels (95% vs. 96%). Par contre, et de manière non attendue, les

participants ont fait moins d'erreurs lors des réponses manuelles par rapport aux réponses orales (3% vs. 7%)( $F(1,14)=13$  ;  $p<.002$ ). Les participants ont également fait moins d'erreurs pour la syllabe /pa/ (2%) par rapport aux syllabes /ta/ et /ka/ (7%)( $F(1,14)=6.8$  ;  $p<.004$ ). De plus, une interaction significative a été trouvée entre les modes de réponse et les syllabes. En effet, pour les syllabes /ta/ et /ka/, plus de réponses correctes ont été observées pour les réponses manuelles (97% et 97%) par rapport aux réponses orales (90% et 89%). Par contre, pour la syllabe /pa/, aucune différence n'a été observée (voir Figure 15) ( $F(2,28)=5.6$  ;  $p<.01$ ).

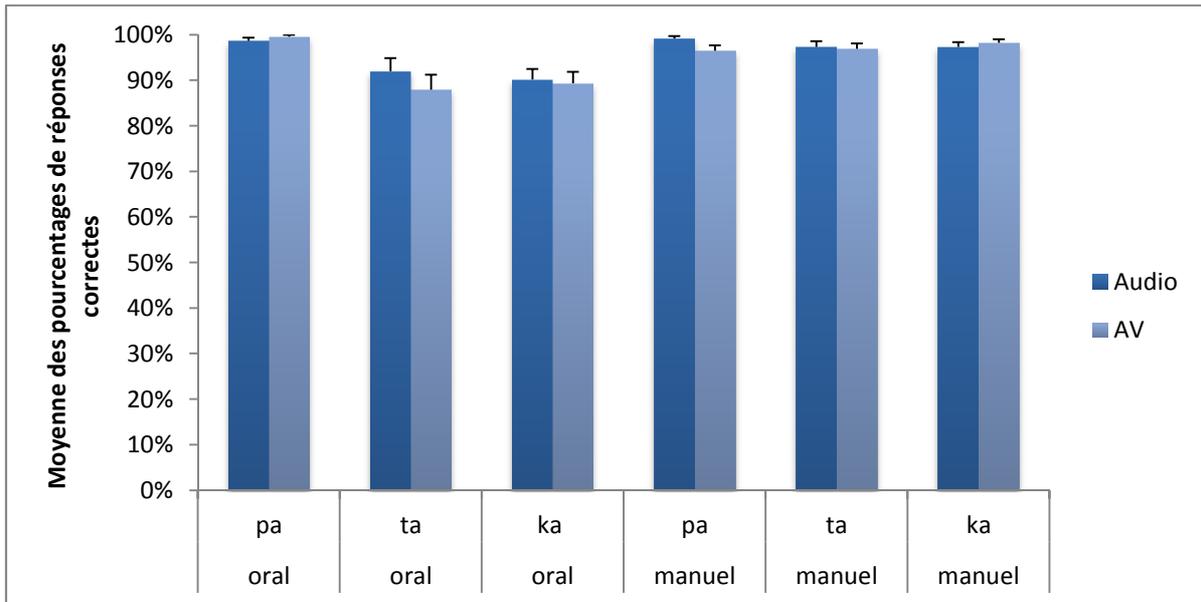


Figure 15 : Moyenne des pourcentages de réponses correctes pour toutes les conditions de l'expérience non-bruitée (les barres d'erreurs correspondent aux écart-types).

## 2. Expérience bruitée vs. Expérience non-bruitée

### 1. Temps de réaction

Comme attendu, l'effet principal du groupe est significatif ( $F(1,27)=24,38$ ;  $p<0.001$ ), avec des temps de réaction pour les stimuli non bruités plus courts que ceux des stimuli bruités (351ms vs 484ms). Les effets principaux de la tâche ( $F(1,27)=151,70$ ;  $p<0.001$ ) et de la modalité ( $F(1,27)=14,79$ ;  $p<0.001$ ) sont aussi significatifs. Pour la tâche, les réponses orales étaient plus rapides que les réponses manuelles (286ms vs 545ms). Par rapport à la modalité, les temps de réaction étaient plus courts dans la modalité audiovisuelle par rapport à la modalité auditive (405ms vs. 425ms). Une interaction significative entre groupe et modalité ( $F(1,27)=21,74$ ;  $p<0.001$ ) montre que l'effet bénéfique de la présentation audiovisuelle est présent avec les stimuli bruités (461ms vs. 507ms) mais pas avec les stimuli non-bruités (354ms vs. 349ms). Par contre, l'interaction entre modalité et réponse n'est pas significative (voir Figure 16, Tableau 2 et Tableau 3).

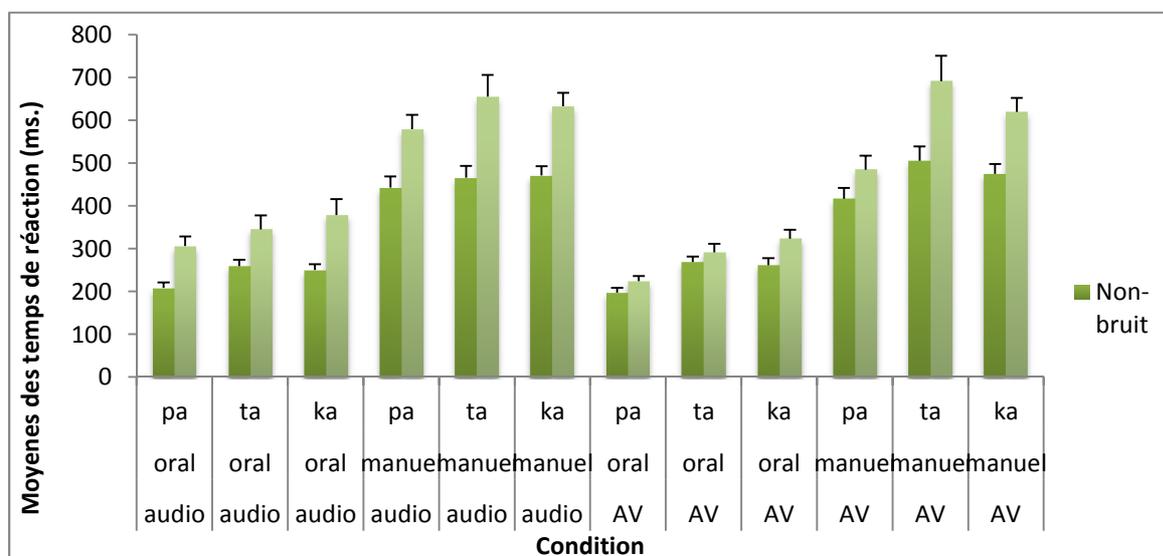


Figure 16 : Moyenne des temps de réaction des expériences en conditions bruitée et non-bruitée (en ms, les barres d'erreurs correspondent aux écart-types).

	Temps de réaction	Proportion de réponses correctes
Groupe	p<.001	p<.001
Modalité	p<.001	p<.001
Réponse	p<.001	p<.001
Syllabe	p<.001	p<.001
Groupe x Modalité	p<.001	p<.001
Groupe x Réponse	p<.001	p<.001
Groupe x Syllabe	n.s.	p<.001
Modalité x Syllabe	p<.001	n.s.
Réponse x Syllabe	n.s.	p<.001
Réponse x Modalité x syllabe	p<.005	n.s.
Groupe x Modalité x Réponse	n.s.	n.s.

Tableau 2 : Effets significatifs et interactions pour les temps de réaction et les pourcentages de réponses correctes.

Modalité	A	A	A	A	A	A	AV	AV	AV	AV	AV	AV
Mode	oral	oral	oral	manuel	manuel	manuel	oral	oral	oral	manuel	manuel	manuel
Syllabe	ka	pa	ta	ka	pa	ta	ka	pa	ta	ka	pa	ta
RTs sans bruit	250	208	259	471	442	465	261	197	268	474	416	506
RTs avec bruit	348	335	373	614	666	632	277	296	318	531	653	601
% sans bruit	90%	99%	93%	98%	99%	98%	89%	100%	88%	98%	96%	97%
% avec bruit	56%	36%	40%	67%	42%	79%	89%	44%	38%	96%	58%	85%

Tableau 3 : Moyenne des temps de réaction et des pourcentages de réponses correctes pour toutes les conditions des expériences en conditions bruitée et non-bruitée.

Ces effets semblent être dépendants des syllabes perçues. Notamment, une interaction à trois facteurs ‘tâche x modalité x syllabe’ a été trouvée ( $F(2,54)=6,49$ ;  $p<0.005$ ). Dans la modalité auditive, aucune différence significative des temps de réaction n’a été observée entre les syllabes à la fois pour les réponses orales et pour les réponses manuelles. Par contre, dans la modalité audiovisuelle, les temps de réaction pour les réponses orales étaient plus rapides pour la syllabe /pa/ par rapport aux syllabes /ta/ et /ka/, alors que les temps de réaction pour les réponses manuelles étaient plus rapides pour /pa/ par rapport à /ka/ et pour /ka/ par rapport à /ta/.

Ainsi, globalement, on obtient un patron de résultats attendus : des temps de réponse plus courts en réponse orale, un effet du bruit ralentissant les temps de réponse, une accélération de la réponse en modalité audiovisuelle par rapport à la modalité auditive en présence de bruit. Il n’apparaît cependant pas d’interaction entre modalité et type de réponse.

## 2. Pourcentage de réponses correctes

L’effet principal du groupe est significatif ( $F(1,27)=266,28$ ;  $p<0.001$ ) avec une proportion de réponses correctes plus élevée pour les stimuli non-bruités (95%) par rapport aux stimuli bruités (61%). D’autres effets principaux ont été significatifs, à la fois pour la tâche ( $F(1,27)=69,40$ ;  $p<0.001$ ) et pour la modalité ( $F(1,27)=52,39$ ;  $p<0.001$ ).

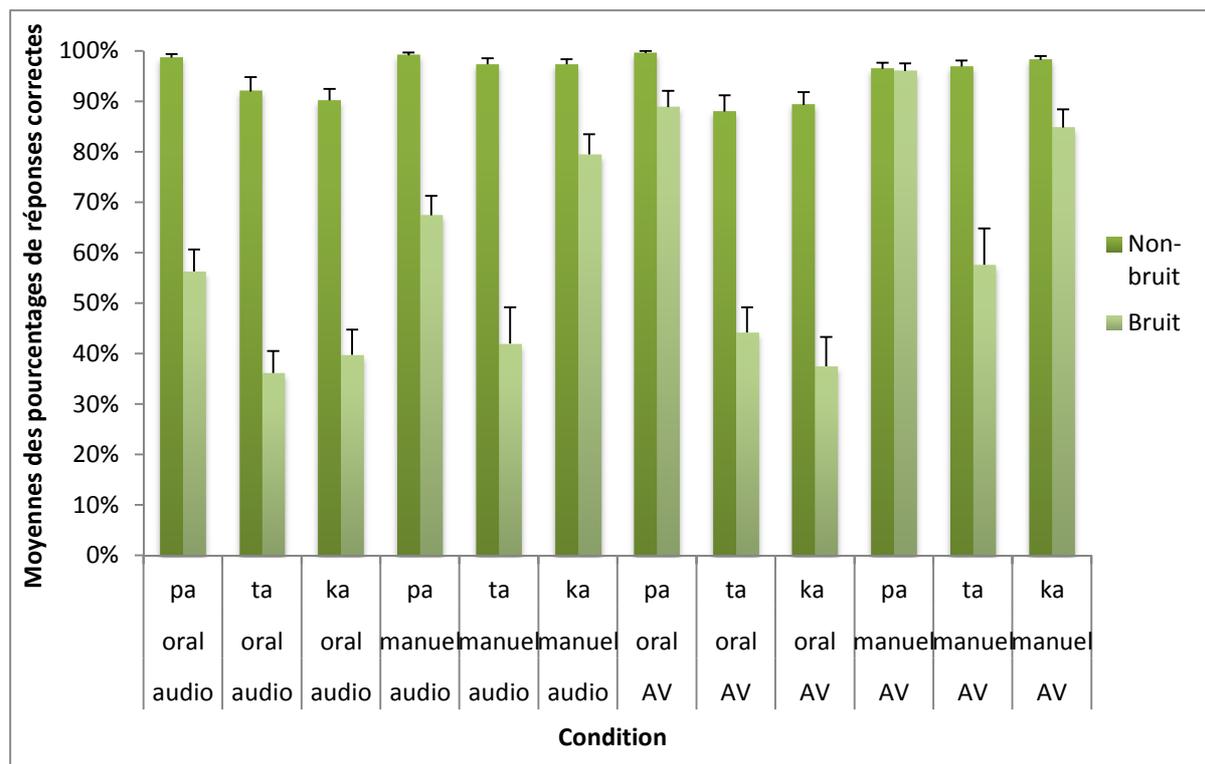


Figure 17 : Moyenne des pourcentages de réponses correctes en conditions bruitée et non-bruitée (les barres d’erreurs correspondent aux écart-types).

Concernant la tâche, une baisse importante des réponses correctes a été observée pour les réponses orales par rapport aux réponses manuelles (73% vs. 85%). Comme

l'interaction significative du groupe avec la tâche ( $F(1,27)=38,67$ ;  $p<0.001$ ) l'indique, cet effet est plus élevé pour les stimuli bruités (50% en réponse orale vs. 71% en réponse manuelle, différence significative) alors que la différence entre les réponses orales et manuelles pour les stimuli non-bruités (93% en réponse orale vs. 98% en réponse manuelle) n'est pas significative. Concernant la modalité, la modalité audiovisuelle apporte plus de réponses correctes que la modalité auditive (82% vs. 75%). Par contre, comme l'indique l'interaction significative 'groupe x modalité' ( $F(1,27)=72,36$ ;  $p<0.001$ ) la différence entre les deux modalités avec les stimuli non bruités est très faible et non significative (96% vs. 95%) alors qu'avec les stimuli bruités, la modalité audiovisuelle apporte plus de réponses correctes (68%) que la modalité auditive (53%, différence significative). Là encore, les résultats dépendent de la syllabe présentée. Si les 3 syllabes sont parfaitement identifiées en l'absence de bruit, quelle que soit la tâche et la modalité, en condition bruitée la syllabe « pa » apparaît la plus saillante à la fois auditivement et visuellement (voir Figure 17, Tableau 2 et Tableau 3).

Ainsi, si l'on obtient ici encore une confirmation d'un patron attendu (réponses plus précises en l'absence de bruit, et, dans le cas de stimuli bruités, en présence de la modalité visuelle), un résultat fort et inattendu doit être relevé : le fait que la réponse orale dégrade la précision de la réponse en cas de stimuli bruités. Une nouvelle fois, il n'apparaît pas d'interaction entre modalité et type de réponse.

#### 4. Discussion

Dans cette discussion, nous allons nous concentrer sur les effets associés aux deux composantes majeures de notre étude (le mode de réponse, c'est à dire les réponses orales vs. manuelles, et la modalité de présentation, c'est à dire auditive vs. audiovisuelle), et la façon dont ils impactent les réponses des participants lors des deux expériences de close-shadowing sans et avec bruit.

Globalement, comme nous l'avions supposé lors de nos hypothèses, les temps de réaction ont été plus longs et les pourcentages de réussite plus faibles pour les stimuli bruités par rapport aux stimuli non bruités (494 ms vs. 351 ms et 61% vs. 95%).

Concernant les effets de la réponse, conformément à notre première hypothèse, sans bruit, les temps de réaction ont été significativement plus rapides pour les réponses orales par rapport aux réponses manuelles (240 vs. 462 ms), avec une diminution non significative de la précision de réponse lors de la tâche orale (93% vs. 98%). Les temps de réaction lors des réponses orales sont cohérents avec ceux obtenus par Fowler et collègues en 2003 (248 ms) et par Porter et Castellanos en 1980 (223 ms) lors de leur tâche à choix multiples. Par contre, la précision lors des réponses orales est plus élevée dans notre étude par rapport à celle des études de Fowler et collègues (86%) et de Porter et Castellanos (77%). Cette différence peut être expliquée soit par la clarté de nos stimuli soit par le niveau sonore dans lequel la présentation a été faite (le shadowing de la voyelle initiale entraîne un son concurrent produit par le participant qui peut cacher dans une certaine mesure la perception de la consonne cible à identifier). L'interprétation de la rapidité de réponse lors des réponses orales proposée par Porter et

Castellanos (1980) puis par Fowler et collègues (2003) est faite en référence aux théories motrices de la perception de parole, dans laquelle l'input de parole serait associé à une représentation motrice (Liberman et Mattingly, 1985) ou directement perçu comme un geste orofacial (Fowler, 1986). Cela pourrait permettre au système orofacial de répondre de façon rapide, puisque le percept est déjà représenté dans un format moteur adéquat. Le système orofacial serait en conséquence plus rapide que le système manuel, celui-ci nécessitant un stade de transduction entre la décision et la réponse. Plus généralement, ces résultats apparaissent en ligne avec des effets de compatibilité stimulus-réponse qui suggèrent un codage commun de la perception et de l'action (pour une revue, voir Prinz, 1997, et Hommel et coll.. 2001).

Cependant les résultats observés lors des stimuli bruités apportent un nouvel éclairage à ce raisonnement. En effet, alors que les temps de réaction restent plus rapides lors des réponses orales (334 ms vs. 663 ms), la précision de réponse diminue drastiquement entre les réponses orales et manuelles (50% vs. 71%). Ces résultats nécessitent de modifier l'explication de Fowler et collègues et de Porter et Castellanos précédemment mentionnée. Nous allons proposer une tentative d'explication dans le cadre du modèle perceptivo-moteur introduit par Skipper et collègues en 2007.

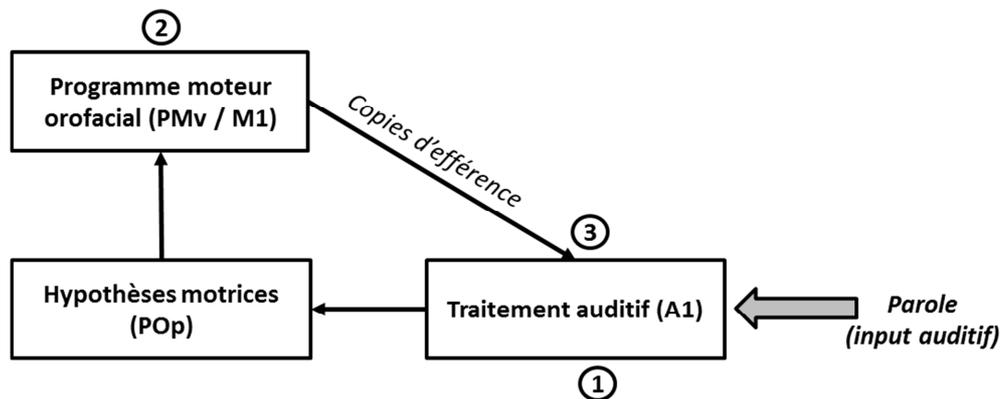


Figure 18 : Schéma représentant le modèle de Skipper et collègues (2007) pour la perception d'inputs de parole auditifs

Skipper et collègues (2007) proposent un modèle de perception de parole qu'ils réfèrent à l'approche dite « d'analyse par synthèse », présentée précédemment (Halle et Stevens, 1959; Stevens et Halle, 1967; voir une revue dans Bever et Poeppel, 2010). Ce modèle implique une boucle de traitement entre les aires auditives et motrices dans le cerveau humain (voir Figure 18). Après un premier stade de traitement auditif (dans le cortex auditif primaire, A1 et dans le cortex secondaire et les aires associées, stade 1 dans la figure), le cortex auditif générerait des hypothèses phonémiques associées à des buts articulatoires (dans la pars opercularis du gyrus frontal inférieur, POp). Ceci résulterait en une simulation des commandes motrices correspondant à ces prédictions initiales (dans le cortex prémoteur ventral, PMv, et dans le cortex moteur primaire, M1, stade 2 dans la figure), entraînant la production d'une copie d'efférence qui serait enfin

renvoyée dans le cortex auditif dans le but d'être comparée avec l'input auditif (stade 3 dans la figure).

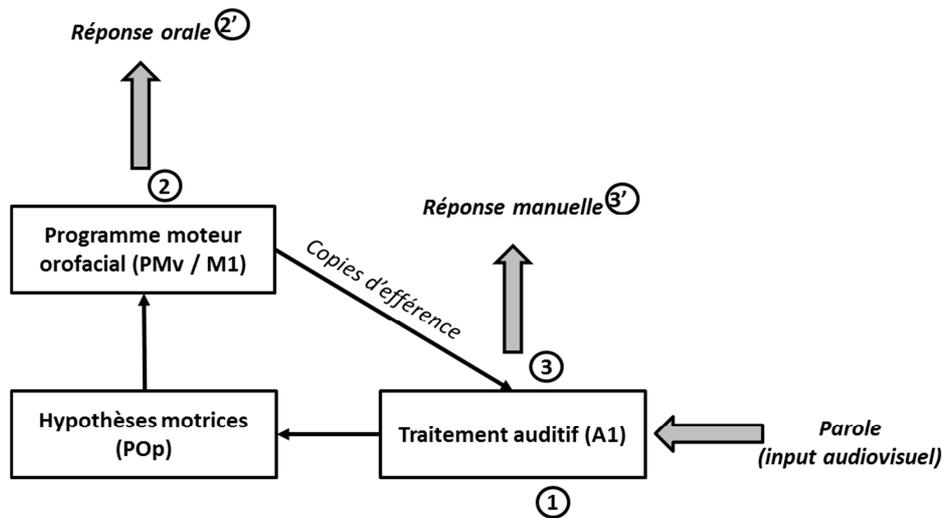


Figure 19 : Interprétation possible du mode de réponse à partir du modèle de Skipper et collègues (2007).

Ce modèle peut être utilisé comme une base pour l'interprétation de nos données (voir Figure 19). Dans ce but, nous supposons que les réponses orales et manuelles sont générées à deux stades différents dans la boucle perceptive. Les réponses orales seraient générées au stade 2, en lien avec les hypothèses de Porter et Castellanos (1980) et de Fowler et collègues (2003). Quand les informations provenant du cortex auditif sont transférées au POp et génèrent des commandes motrices dans le cortex moteur (système de commandes prédictives), le système orofacial, déjà pré-activé depuis le début de la tâche de close-shadowing pour permettre au participant de répondre le plus vite possible, produirait une réponse orale induite par ces commandes motrices (stade 2' de la figure). Ces réponses orales seraient alors plus rapides, mais également moins précises, puisque c'est seulement une première hypothèse (possiblement erronée) qui est susceptible d'être corrigée lors des stades suivants, ce qui est cohérent avec la proposition de Skipper et collègues. Au stade 2, en revanche, le système manuel ne recevrait pas de stimulation spécifique permettant de générer une réponse. Par contre, lors du prochain stade (stade 3), le transfert des commandes prédictives des informations articulatoires au cortex auditif, grâce aux copies d'efférence, permettrait par un mécanisme intégrateur de générer une réponse plus précise qui pourrait alors être transférée au système manuel pour la réponse (stade 3'). En conséquence, les temps de réaction pour les réponses manuelles seraient plus longs que pour les réponses orales, mais les réponses seraient plus précises, puisque contrairement au processus utilisé dans notre hypothèse pour les réponses orales, dans le mode de décision manuel, les prédictions seraient confirmées et ajustées dans le cortex auditif avant que la décision finale ne soit envoyée aux commandes motrices manuelles (c'est-à-dire avant l'appui de la touche adéquate).

Bien sûr, cette explication est probablement trop simple pour expliquer tous les aspects de nos données. L'augmentation des temps de réaction pour les stimuli bruités, un résultat classique dans toutes les expériences de catégorisation, nécessiterait d'introduire des dynamiques temporelles à différents stades de la boucle présentée dans la Figure 19. A noter que le fait que cette augmentation de temps de réponse est la même dans les tâches orales et manuelles (puisque'il n'y a aucune interaction entre la tâche et le groupe pour les temps de réaction) suggère que le ralentissement dû au bruit serait essentiellement généré aux stades 1 et 2 plutôt qu'au stade 3, même si des variantes peuvent certainement être suggérées.

L'aspect crucial de nos résultats est qu'une hypothèse de réponses motrices pures, typique des théories motrices, bien que compatible avec des temps de réaction plus rapides lors des réponses orales par rapport aux réponses manuelles, n'explique pas la diminution associée de la précision de réponse. Au contraire, cet effet de diminution de la performance associée à la génération rapide de réponses orales semble plutôt compatible avec des théories perceptivo-motrices de la perception de parole comme celle proposée par Skipper et collègues (2007) (voir également l'implémentation computationnelle d'une théorie perceptive-motrice dans Moulin-Frier et collègues, 2012).

Enfin, concernant les effets de la modalité, comme nous l'avons envisagé, les effets de la modalité sont uniquement présents lors des stimuli bruités. Dans la modalité auditive, les temps de réaction sont plus lents et la proportion de réponses correctes est plus faible que dans la modalité audiovisuelle. Ces résultats étendent au cas du *close-shadowing* le bénéfice de l'addition de la modalité visuelle à l'input auditif, ce qui est cohérent avec toutes les études précédentes depuis Sumbly et Pollack (1954) qui exposent un bénéfice audiovisuel lors de la reconnaissance de parole dans des conditions bruitées. Dans notre étude, l'avantage audiovisuel est présent uniquement pour la syllabe /apa/, un résultat classique en lien avec la saillance visuelle de la consonne bilabiale /p/, et le haut degré de confusion entre les mouvements visuels des consonnes /t/ et /k/, qui sont généralement considérées comme des visèmes. Ces effets de la modalité ne sont pas présents pour des stimuli non bruités, probablement parce que pour ces stimuli les temps de réaction en modalité auditive sont déjà très courts et les proportions de réponses correctes sont déjà trop élevées pour être améliorés par l'addition de la modalité visuelle (effet plancher).

Un point intéressant est qu'il n'y a pas d'interaction significative entre la modalité et la tâche, c'est-à-dire que la diminution des temps de réaction et l'amélioration des pourcentages de réponses correctes de la modalité audio vers la modalité audiovisuelle sont les mêmes pour les réponses orales et manuelles. Encore une fois, nous allons tenter d'interpréter cette absence d'interaction grâce au modèle proposé par Skipper et collègues (2007).

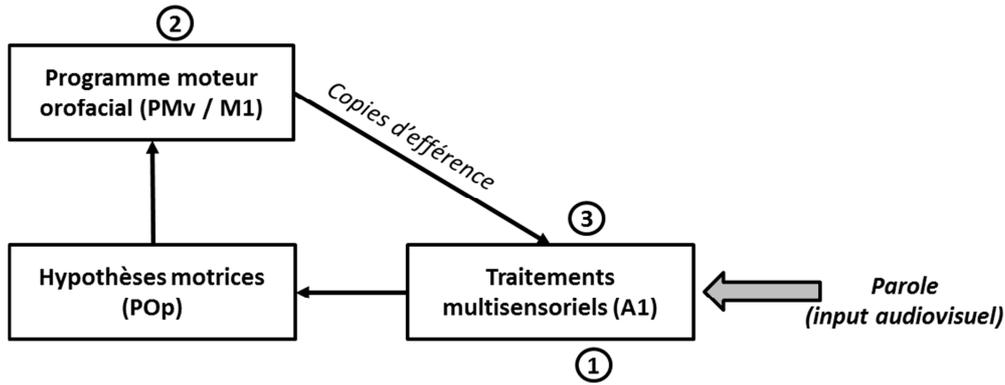


Figure 20 : Schéma représentant le modèle de Skipper et collègues (2007) pour la perception d'inputs de parole multisensoriels.

Dans leur modèle (voir Figure 20), Skipper et collègues proposent que les informations auditives et audiovisuelles, après un premier stade de traitement uni-sensoriel dans les aires auditives et visuelles, convergent vers l'aire multi-sensorielle STp dans le cortex temporal postéro-supérieur (stade 1 de la figure). Ainsi, dans le cas d'input multi-sensoriel, la première hypothèse serait multi-sensorielle plutôt qu'uniquement auditive. A partir de là, comme précédemment, une hypothèse phonémique associée aux buts articulatoires serait générée dans le POp, et produirait des commandes motrices dans PMv/M1 (stade 2) puis les copies d'efférence conduiraient à une prédiction auditive dans STp, alors comparée à l'input multisensoriel dans le stade final 3. Dans notre étude, les interactions audiovisuelles au stade 1 pourraient affiner le traitement sensoriel pour produire des hypothèses phonémiques plus rapides et plus précises au stade 2, qui est le stade où, selon notre interprétation, les réponses orales sont générées (stade 2'). Un gain semblable de rapidité et de précision dû à l'input visuel serait propagé vers le stade 3 où les réponses manuelles seraient générées (stade 3'). Ainsi, il n'y a pas de raison d'attendre des différences de gain visuel entre les réponses orales et manuelles, ce gain étant essentiellement déterminé précocement dans le stade 1 du modèle.

En résumé, les résultats de notre étude suggèrent que les réponses orales et manuelles sont générées à deux stades différents de la boucle perceptive. A la lumière d'une approche « d'analyse par synthèse », nous proposons que les réponses manuelles seraient disponibles seulement à la fin de la boucle perceptive, au stade d'intégration entre les hypothèses multi-sensorielles associées aux prédictions motrices, via des commandes motrices simulées, et la prise en compte fine de l'input multi-sensoriel. Au contraire, les réponses orales seraient disponibles lors d'un stade plus précoce où les commandes motrices sont générées, ce qui entrainerait des réponses plus rapides mais moins précises. L'input visuel améliorerait la rapidité et la précision pour des phonèmes suffisamment visibles (comme le /p/), dans le cas de conditions d'écoute difficiles (comme dans le bruit). Encore une fois, il est important de préciser que d'autres interprétations peuvent être proposées. Cependant, globalement, nous pensons que l'ensemble des résultats de cette étude semblent nécessiter une théorie perceptivo-motrice de la

perception de parole dans laquelle les voies auditives et visuelles sont intégrées, et où des représentations motrices internes sont générées et fournies au mécanisme de décision finale.

#### **4. Expérience d'idiosyncrasies**

##### **1. Principe et hypothèses**

L'expérience dite d'« idiosyncrasies » a été créée dans le but de tester d'une part les idiosyncrasies lors de la production et de la perception de voyelles, séparément, et d'autre part, les idiosyncrasies couplées entre la production et la perception de voyelles. Nous nous sommes basés sur les études réalisées par Ménard et coll. (2008) et Ménard et Schwartz (2014) qui ont mesuré les variations du premier formant F1 lors de tâches de production et de perception des voyelles orales du français. Rappelons que les auteurs ont observé une variabilité inter-sujets des valeurs de F1 aussi bien en production qu'en perception, et que pour un même participant ces valeurs étaient corrélées entre les deux tâches. Ménard et Schwartz (2014) proposent que cette corrélation puisse être interprétée comme une conséquence du couplage perceptivo-moteur qui se traduirait par un lien entre prototypes phonémiques perceptuels et moteurs dans le cerveau.

Nous avons choisi de reproduire cette expérience pour pouvoir l'utiliser chez des personnes sourdes porteuses d'un implant cochléaire. Cependant, nous avons dû raccourcir cette expérience pour pouvoir l'intégrer à notre protocole expérimental général et éviter que la durée totale de passation ne soit trop longue.

Notre expérience d'idiosyncrasies consistait donc en deux tâches de perception et de production de voyelles, l'ordre de ces deux tâches étant contrebalancé entre les participants. Contrairement à Ménard et Schwartz (2014), et pour la raison de durée de l'expérience globale mentionnée ci-dessus, nous nous sommes limités, pour la tâche de perception des voyelles, aux voyelles antérieures non-arrondies du français, autrement dit /i e ε a/.

Notre hypothèse essentielle ici était que les valeurs de F1 des voyelles perçues seraient corrélées entre les participants avec les valeurs des voyelles produites.

##### **2. Matériel et méthodes**

###### **1. Participants**

Nous avons enregistré un premier groupe de participants pour l'expérience d'idiosyncrasies, mais nous avons dû supprimer les enregistrements de sept participants à cause d'un problème de saturation de signal lors de l'enregistrement. Au total, 27 participants (20 femmes, 7 hommes) âgés de 19 à 41 ans (moyenne d'âge 25 ans) ont pu être testés. Tous les participants étaient normo-entendants, de langue maternelle française, sans déficits auditifs ni déficits de la vision et sans troubles du langage.

## 2. Stimuli

Pour la partie perceptive de l'expérience d'idiosyncrasie, nous souhaitons obtenir des stimuli répartis le plus adéquatement possible dans l'espace auditif. <sup>2</sup>Nous avons choisi de nous référer à un espace perceptif F1/F<sup>2</sup> plutôt qu'à un espace acoustique F1/F2/F3/F4. Pour ce faire, nous avons tout d'abord déterminé un espace maximum avec les valeurs de formants F1, F2, F3, F4 en Hz sur le modèle articulatoire VLAM (Variable Linear Articulatory Model). VLAM est un modèle du conduit vocal qui permet de passer de configurations articulatoires vers des représentations acoustiques. Il a été mis en œuvre à partir d'un modèle de conduit vocal adulte, créé par Maeda en 1979. VLAM utilise des configurations articulatoires contrôlées par sept paramètres du conduit vocal : protrusion et ouverture des lèvres, hauteur de la mandibule, pointe, dos et corps de la langue et hauteur du larynx. En se basant sur ces configurations, le modèle peut générer des voyelles susceptibles d'être produites par des locuteurs allant de la naissance à l'âge adulte. Ces voyelles sont caractérisées par leur spectre à partir duquel on peut estimer les formants, puis ce que l'on appelle le « second formant effectif » F<sup>2</sup>, qui représente une sorte de centre de gravité perceptif des formants F2, F3 et F4 et permet, en combinaison avec F1, une caractérisation efficace du timbre des spectres vocaliques (Carlson et coll., 1970 ; Bladon et Fant, 1978).

Pour chaque point de l'espace maximal généré par le modèle VLAM, c'est-à-dire l'ensemble des possibilités acoustiques associées aux configurations vocaliques dans l'espace des formants, nous avons calculé les valeurs de F1/F<sup>2</sup> en Barks, en utilisant, pour calculer les valeurs de formants en Barks, la formule de Schroeder et collègues (1979).

$$F(\text{Bark}) = 7 \operatorname{asinh}\left(\frac{F(\text{Hz})}{650}\right)$$

et pour déterminer la position du second formant effectif F<sup>2</sup> à partir des formants F2, F3 et F4, le modèle proposé par Mantakas et coll. (1986 ; voir aussi Schwartz et coll., 1997).

Par la suite, nous avons dessiné les bords d'un triangle vocalique hypothétique dans l'espace F1/F<sup>2</sup>, puis proposé une distribution régulière de 84 points dans ce triangle (voir Figure 21). Pour chacun des points du triangle, nous avons calculé les valeurs F2, F3 et F4 en Barks, compatibles avec chaque valeur de F<sup>2</sup>. Ces valeurs ont ensuite été retransformées en Hertz selon la formule de Schroeder inversée.

$$F(\text{Hz}) = 650 \times \sinh\left(\frac{F(\text{Barks})}{7}\right)$$

Enfin, nous avons calculé les bandes passantes des quatre formants par une formule proposée par Fant (voir formules ci-dessous), puis les stimuli ont été générés par synthèse.

---

<sup>2</sup> Les stimuli utilisés pour cette expérience d'idiosyncrasies ont été créés en collaboration avec Lucie Ménard, que nous souhaitons remercier pour son aide.

$$B1 = 15 \times \left(\frac{500}{F1}\right)^2 + 20 \times \sqrt{\frac{F1}{500}} + 5 \times \left(\frac{F1}{500}\right)^2$$

$$B2 = 22 + 16 \times \left(\frac{F1}{500}\right)^2 + \frac{12000}{F3 - F2}$$

$$B3 = 25 \times \left(\frac{F1}{500}\right)^2 + 4 \times \left(\frac{F2}{500}\right)^2 + 10 \times \frac{F3}{F4 - F3}$$

$$B4 = 150 \text{ Hz}$$

Enfin, pour limiter la durée de l'expérience, nous avons choisi de nous concentrer sur les voyelles antérieures seulement, et nous avons conservé 13 stimuli entre /i/ et /a/, ayant des valeurs de F1 s'échelonnant entre 2.2 et 7.0 Barks, par pas de 0.4 Barks (voir Figure 21, Tableau 4 Tableau 5).<sup>3</sup>

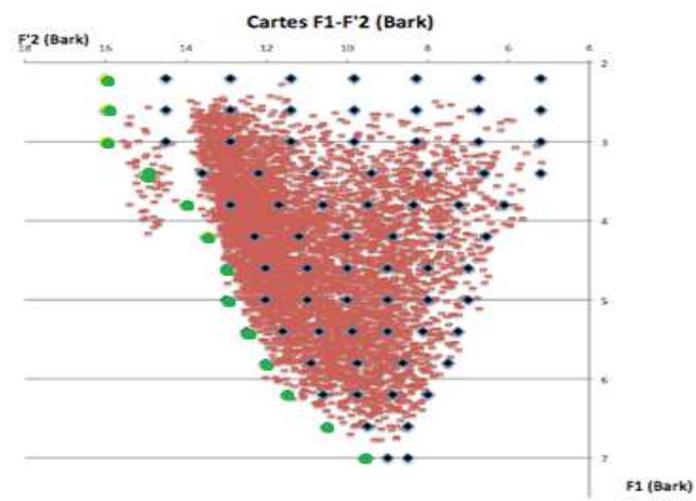


Figure 21: Représentation graphique sur l'espace F1/F2' des stimuli utilisés pour la tâche de perception de voyelles de l'expérience d'idiosyncrasies (les stimuli utilisés sont représentés par les points verts sur le graphique – ils sont numérotés dans les Tableaux 3 et 4 par leur numérotation d'origine parmi les 84 stimuli de départ).

<sup>3</sup> Les stimuli ont été sélectionnés en périphérie de l'espace perceptif et ce en raison des limites imposées aux paramètres articulatoires et géométriques lors de la génération de l'espace.

Tableau 4 : Valeurs des formants pour les stimuli utilisés dans la tâche de perception (en Barks et en Hertz)

Stimulus	F1 (Hz)	F2 (Hz)	F3 (Hz)	F4 (Hz)	F1 (Bk)	F2 (Bk)	F3 (Bk)	F4 (Bk)	F'2 (Bk)
<b>1</b>	208	2300	2980	3560	2,2	13,8	15,6	16,8	16,0
<b>9</b>	247	2300	2980	3560	2,6	13,8	15,6	16,8	16,0
<b>17</b>	287	2300	2980	3560	3	13,8	15,6	16,8	16,0
<b>25</b>	328	2610	3000	3560	3,4	14,7	15,6	16,8	15,0
<b>33</b>	370	2289	2500	3560	3,8	13,8	14,4	16,8	14,0
<b>41</b>	414	2047	2500	3560	4,2	13,1	14,4	16,8	13,5
<b>48</b>	459	1830	2500	3560	4,6	12,3	14,4	16,8	13,0
<b>55</b>	505	1830	2500	3560	5	12,3	14,4	16,8	13,0
<b>62</b>	553	1695	2600	3560	5,4	11,8	14,7	16,8	12,5
<b>69</b>	602	1660	2700	3560	5,8	11,7	14,9	16,8	12,0
<b>74</b>	654	1617	2750	3560	6,2	11,5	15,0	16,8	11,5
<b>79</b>	708	1384	2500	3560	6,6	10,5	14,4	16,8	10,5
<b>82</b>	764	1179	2500	3560	7	9,5	14,4	16,8	9,5

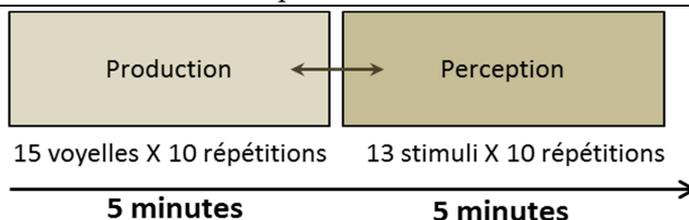
Tableau 5 : Valeurs des bandes passantes pour les stimuli utilisés dans la tâche de perception (en Hertz)

Stimulus	B1_Fant(Hz)	B2_Fant(Hz)	B3_Fant(Hz)	B4 (Hz)
<b>1</b>	100,7	42,4	140,3	150
<b>9</b>	76,7	43,6	142,1	150
<b>17</b>	62,3	44,9	144,3	150
<b>25</b>	53,2	59,7	173,3	150
<b>33</b>	47,3	87,7	121,1	150
<b>41</b>	43,5	59,5	107,8	150
<b>48</b>	41,2	53,4	98,2	150
<b>55</b>	39,9	56,2	102,6	150
<b>62</b>	39,4	54,8	103,6	150
<b>69</b>	39,5	56,8	111,8	150
<b>74</b>	40,2	60,0	118,6	150
<b>79</b>	41,3	64,8	104,3	150
<b>82</b>	42,8	68,4	104,2	150

### 3. Procédure expérimentale

Lors de la tâche de perception, les participants devaient catégoriser manuellement des stimuli acoustiques en appuyant sur une touche de clavier, chaque touche correspondant à une des quatre voyelles antérieures non-arrondies du français /i e ε a/. Chaque stimulus était présenté dix fois dans un ordre randomisé. Les participants avaient tout le temps pour répondre, seul l'appui d'une touche déclenchant le stimulus suivant. Lors de la tâche de production de voyelles, les participants devaient produire quinze répétitions des dix voyelles orales du français /i y u e ø o ɔ œ ε a/ selon une instruction présentée visuellement (consigne orthographique). Les voyelles cibles étaient présentées dans un ordre randomisé. La durée totale de l'expérience était d'environ 10 minutes, l'ordre de passation des tâches étant contrebalancé entre les participants (voir Figure 22).

Figure 22 : Procédure expérimentale de l'expérience "Idiosyncrasies"



#### 4. Analyses

##### 1. Tâche de production

Pour la tâche de production de voyelles, grâce au logiciel Praat, nous avons segmenté les productions des participants selon les voyelles produites, puis nous avons extrait les valeurs de F1 de chaque voyelle, mesurées au milieu de la voyelle. A partir de ces valeurs, nous avons calculé la moyenne des F1 pour chaque type de voyelles produites.

Enfin, nous avons calculé les valeurs normalisées des voyelles /e/ et /ɛ/ ( $Y_v =$  Valeur normalisée de la voyelle donnée v) par rapport aux voyelles /a/ et /i/, pour chaque participant, en reprenant fidèlement la méthodologie proposée par Ménard et coll. (2014). Pour ce faire, nous avons soustrait la valeur moyenne de F1 de /i/ ( $M_1$ ) à la valeur moyenne de F1 pour la voyelle cible (/e/ ou /ɛ/)(Moyenne<sub>v</sub>), le tout divisé par la valeur moyenne de F1 de /a/ ( $M_4$ ) moins la valeur moyenne de F1 de /i/ ( $M_1$ ) (toutes valeurs en Barks, calculées par la formule de Schroeder) :

$$Y_v = (Moyenne_v - M_1) / (M_4 - M_1)$$

##### 2. Tâche de perception

Pour la tâche de perception de voyelles, nous avons là encore repris la méthodologie proposée par Ménard et coll. (2014). Nous avons d'abord calculé la moyenne de F1 de chaque voyelle à percevoir, à partir des valeurs de F1 de chaque stimulus catégorisé comme la voyelle correspondante (respectivement /e/, /a/, /i/ ou /ɛ/). Ensuite, nous avons calculé les valeurs normalisées de /e/ et de /ɛ/ par rapport aux valeurs de /i/ et /a/ comme pour la tâche de production.

##### 3. Corrélation entre les tâches

Nous avons alors analysé la corrélation entre les deux tâches et testé la significativité du coefficient de corrélation r, en calculant

$$t = \frac{r \times \sqrt{n-2}}{\sqrt{1-r^2}}$$

puis en comparant cette valeur à la loi de Student avec (n-2) degrés de liberté.

#### 3. Résultats

##### 1. Tâche de production

Dans la tâche de production, les moyennes des valeurs de F1 pour /i/, /e/, /ɛ/ et /a/ en Barks ont été respectivement de 3.1 (avec une étendue de 2.6 à 3.6), 4.4 (avec une étendue de 3.4 à 4.5), 5.9 (avec une étendue de 4.4 à 7.1) et 7.3 (avec une étendue de 6.2

à 8.5). Les idiosyncrasies ont été faibles pour /e/ avec des valeurs de distance normalisée à partir de /i/ allant de 0.19 à 0.46 Barks mais avec un écart-type réduit à 0.07 (voir Figure 23).

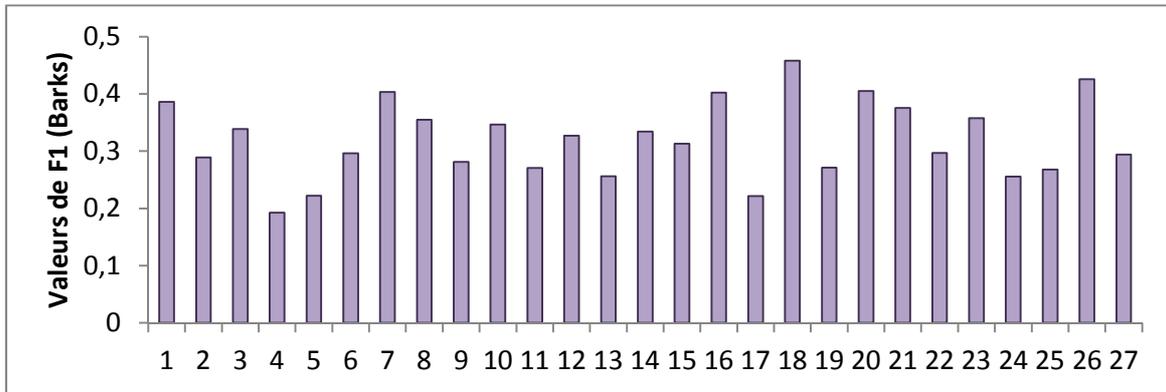


Figure 23 : Valeurs normalisées de F1 pour /e/ produites par chaque participant.

Par contre, les idiosyncrasies pour /ε/ ont été plus fortes, avec des valeurs de distance normalisées à partir de /i/ allant de 0.35 à 0.88 Barks, avec un écart-type de 0.15 (voir Figure 24).

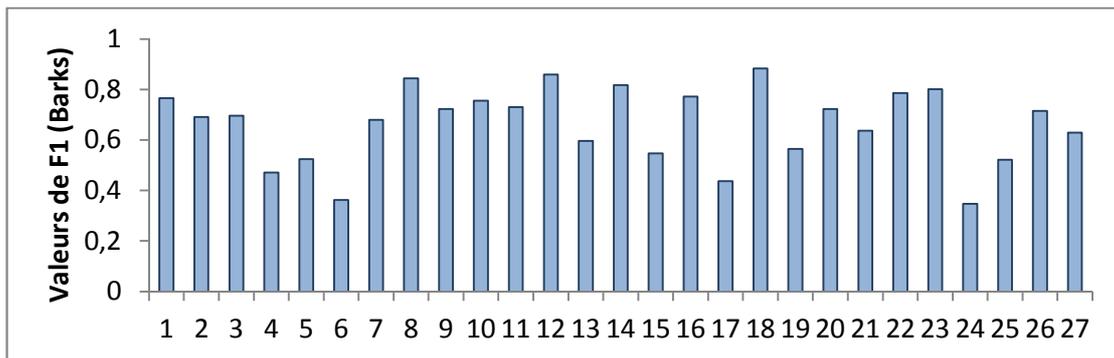


Figure 24 : Valeurs normalisées de F1 pour /ε/ produite par chaque participant.

## 2. Tâche de perception

Dans la tâche de perception, les moyennes des valeurs de F1 en Barks pour /i/, /e/, /ε/ et /a/ ont été respectivement de 2.8 (avec une étendue de 2.6 à 3.9), 4.2 (avec une étendue allant de 3.8 à 4.5), 5.5 (avec une étendue allant de 5.3 à 5.7) et 6.8 (avec une étendue allant de 6.6 à 7.0). La variabilité en perception a été extrêmement faible, ce qui explique qu'aucune idiosyncrasie n'a été trouvée entre les participants.

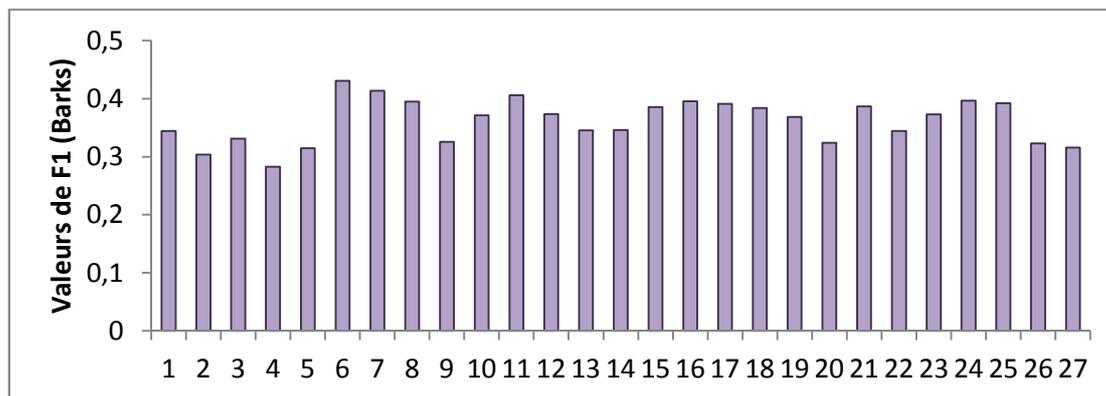


Figure 25 : Valeurs normalisées de F1 lors de la perception de la voyelle /e/, par participant.

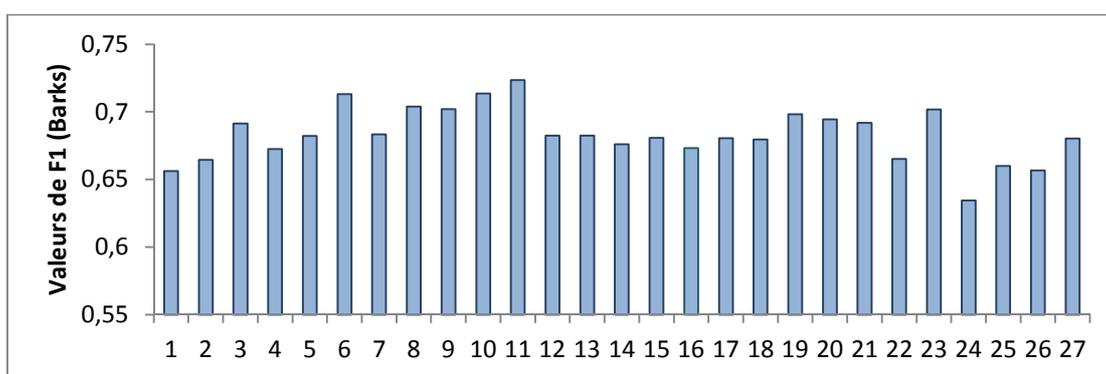


Figure 26 : Valeurs normalisées de F1 lors de la perception de la voyelle /e/, par participant.

Les valeurs de distance normalisée pour la voyelle /e/ à partir de /i/ se distribuent entre 0.28 à 0.43 Barks avec un écart-type de 0.04 (voir Figure 25). Pour la voyelle /e/ les valeurs de distance normalisées à partir de /i/ varient de 0.63 à 0.72 Barks, avec un écart-type très faible de 0.02 (voir Figure 26).

### 3. Corrélation perception/production

Etant donnée l'absence presque totale d'idiosyncrasies en perception, nous n'avons pas obtenu de corrélation individuelle entre production et perception, ni pour la voyelle /e/ ( $r=.2$ ,  $t=1.06$ ,  $p >.2$ , voir Figure 27) ni pour la voyelle /ɛ/ ( $r=.19$ ,  $t=1$ ,  $p>.3$ , voir Figure 28).

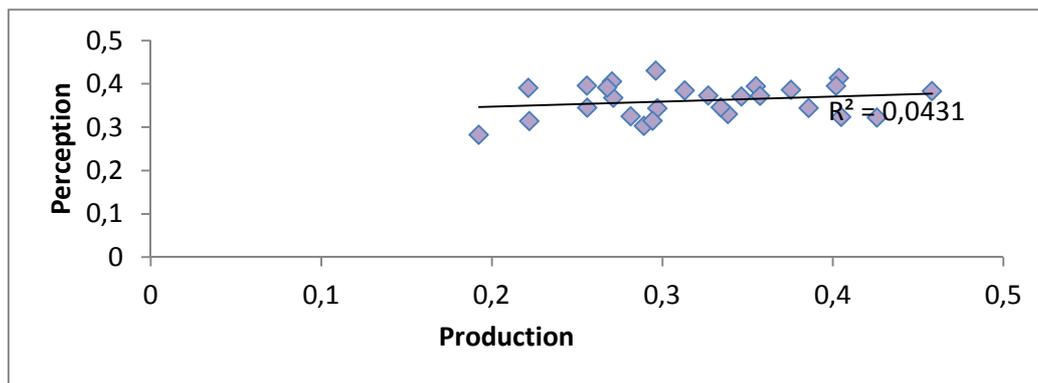


Figure 27 : Corrélation pour la voyelle /e/

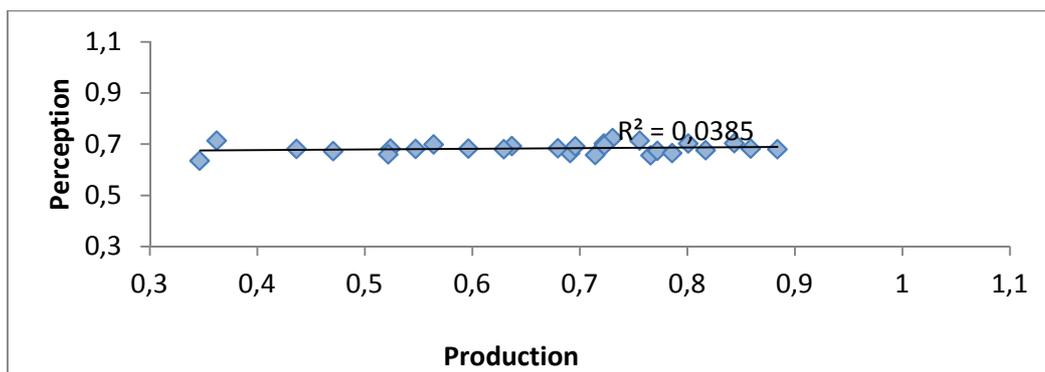


Figure 28 : Corrélation pour la voyelle /ε/

Par contre, à partir des valeurs moyennes des quatre voyelles, nous avons obtenu une corrélation très forte des valeurs acoustiques non normalisées entre les voyelles produites et perçues, avec une pente moyenne pour tous les participants de .93, allant de 0.7 à 1.3 selon les participants,  $r=.94$  (voir Figure 29)

Il n'est évidemment pas surprenant qu'il y ait une corrélation entre valeurs de F1 en perception et en production pour 4 voyelles différant en degré de hauteur. Le point qui nous semble remarquable ici est la régularité quasi parfaite de la distribution des voyelles, tant en perception qu'en production, et qui se traduit effectivement par une corrélation très proche de 1 pour tous les participants.

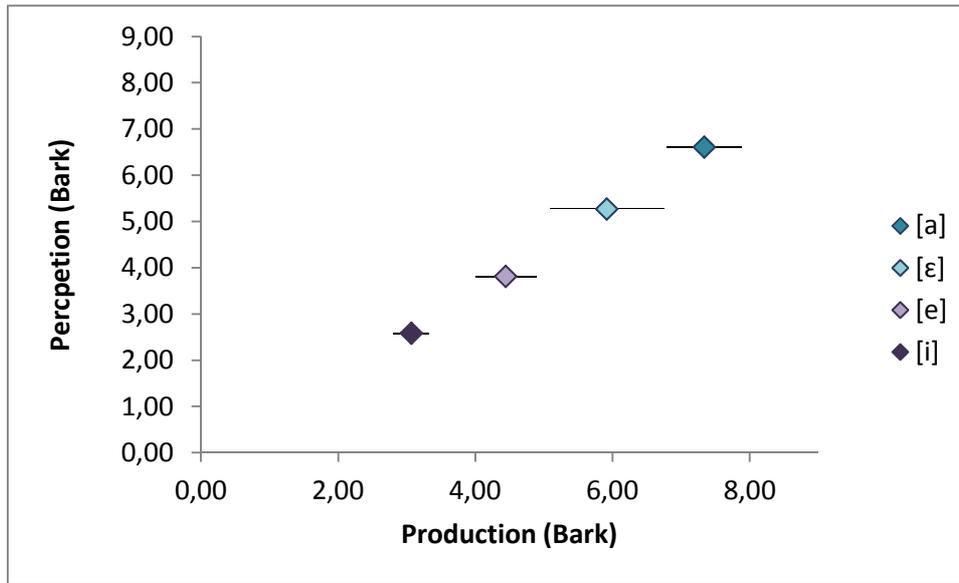


Figure 29 : Corrélation entre les valeurs produites et les valeurs perçues (les barres d'erreurs correspondent aux écart-types).

#### 4. Autres analyses

Parallèlement aux analyses principales, nous avons voulu tout d'abord vérifier si l'absence de corrélation entre les valeurs perçues et produites n'était pas influencée par l'ordre de passation des deux tâches. Nous avons donc séparé nos participants en deux groupes : « ordre production-perception » et « ordre perception-production », comprenant respectivement douze et quinze participants. Aucune corrélation significative n'a été révélée dans aucun des groupes, que ce soit pour la voyelle /e/ ou la voyelle /ε/ (voir Tableau 6 : Tests de significativité par voyelle selon l'ordre de passation des tâches).

ordre	voyelle	valeur t	significativité	r
production- perception	/e/	1.20	0.2	0.35
production- perception	/ε/	0.33	0.7	0.1
perception- production	/e/	0.73	0.3	0.2
perception- production	/ε/	0.94	0.4	0.2

Tableau 6 : Tests de significativité par voyelle selon l'ordre de passation des tâches.

Nous avons également voulu vérifier que l'absence de corrélation pour la voyelle /e/ uniquement n'était pas due aux participants ayant des écart-types trop élevés dans la tâche de production de voyelle. Pour cela, nous avons créé deux groupes : un groupe (nommé A) où nous avons écarté 10 participants ayant des écart-types supérieurs à 50 Hz pour /a/ et supérieurs à 40 Hz pour /e/ et /i/ (ce groupe comprenant 19 participants) et un groupe (nommé B) où nous avons écarté 14 participants ayant des écart-types supérieurs à 50 Hz pour /a/ et à 30 Hz pour /e/ et /i/ (ce groupe comprenant 15 participants). Dans les deux groupes, aucune corrélation ne s'est révélée significative (Groupe A :  $r=.33$ ,  $t= 1.47$ ,  $p>.1$  ; Groupe B :  $r=.45$ ,  $t = 1.83$ ,  $p>.05$ ). Nous n'avons pas proposé d'effectuer cette vérification pour la voyelle /ε/ en raison de la variabilité extrêmement réduite pour cette voyelle lors de la tâche de perception ; en effet, même en sélectionnant des participants moins variables lors de la tâche de production, aucune corrélation n'aurait pu être observée.

Cette absence de corrélation est évidemment fortement liée à la très grande stabilité des réponses perceptives des sujets. Nous avons donc également essayé de mieux caractériser cette stabilité surprenante. Une hypothèse possible était la tendance, classique dans les expériences de choix forcé à faible nombre de réponses possibles, à égaliser le nombre de réponses dans les 4 catégories. Pour évaluer cette hypothèse, nous avons donc estimé le nombre de stimuli par catégorie, ceci afin de vérifier si la stabilité des catégories perceptives n'était pas causée par une répartition égale des stimuli entre chaque catégorie. En moyenne, 15% des stimuli ont été catégorisés comme /a/, avec une variation entre 8 et 21% ; 33% des stimuli ont été catégorisés comme /e/ avec une variation entre 25 et 45% ; 23% des stimuli ont été catégorisés comme /ε/ avec une variation de 17 à 35% ; et 29% des stimuli ont été catégorisés comme /i/ avec une variation de 21 à 35%. (voir Figure 30). Ainsi que le montre la Figure 31, la proportion relative de stimuli catégorisés comme /e/ et /ε/ est très variable selon les participants. Il apparaît donc que le nombre de réponses varie fortement d'une catégorie à l'autre et d'un sujet à l'autre, ce qui montre clairement que la stabilité des moyennes perceptives n'est pas liée à une équi-répartition du nombre de réponses par catégorie.

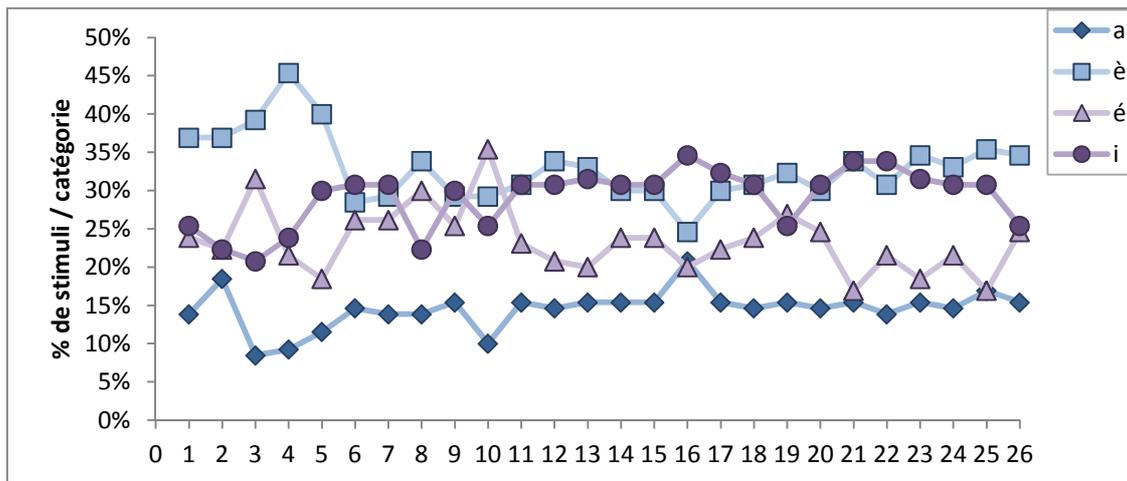


Figure 30 : Pourcentages de stimuli par catégorie, chez chaque participant normo-entendant junior.

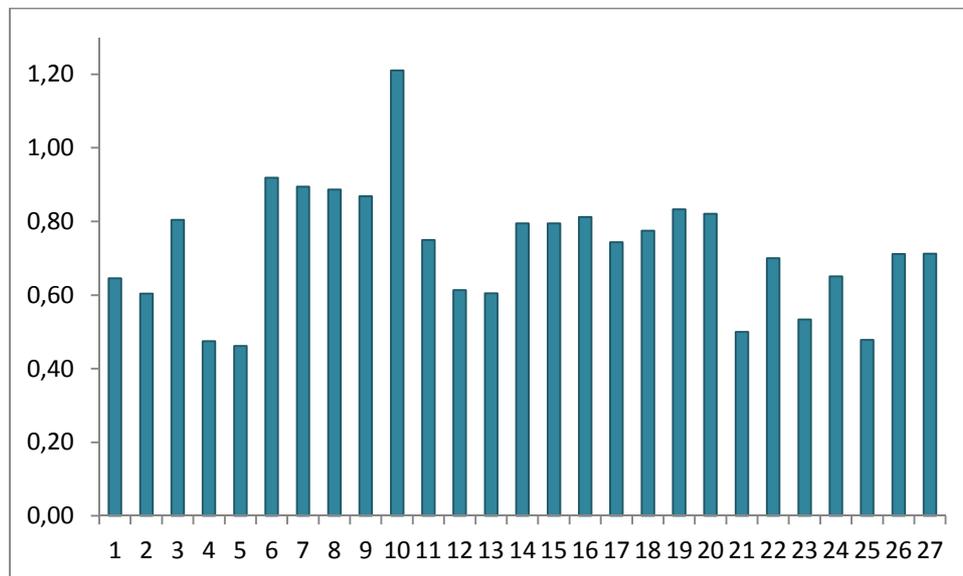


Figure 31 : Rapport entre les pourcentages relatifs de stimuli catégorisés comme les voyelles /e/ et /ɛ/, par participant normo-entendant junior.

#### 4. Discussion

En résumé, contrairement à l'étude de Ménard et Schwartz (2014), nous n'avons pas observé de corrélation individuelle entre la production et la perception des voyelles mi hautes et mi basses /e/ et /ɛ/.

Notre étude fait en réalité apparaître deux résultats majeurs. D'une part, notre expérience se caractérise par l'absence de variabilité inter-participants dans la tâche de perception. Notre hypothèse est que cette stabilité perceptive est due au nombre réduit de voyelles dans notre étude. En effet, Ménard et Schwartz (2014) ont d'une part utilisé un nombre bien plus important de stimuli (avec 114 stimuli présentés à chaque participant) et d'autre part, ils ont présenté à chaque groupe de participants des stimuli

synthétisés correspondant à des conduits vocaux de deux âges différents ce qui ajoute de la variabilité aux données et permet peut-être de mieux rendre compte des spécificités de chacun en perception. Dans notre expérience, le nombre réduit de stimuli et de réponses aurait ainsi pu permettre aux sujets de prendre globalement la mesure de leur espace perceptif, et de le partitionner de manière optimale en distances perceptives entre catégorie. Cette partition est bien une partition en distances et non en nombre de réponses comme nous venons de le voir.

Dans le domaine de la production, par contre, on observe, comme chez Ménard et coll. (2008), des idiosyncrasies chez les participants, avec une variation inter-individuelle des moyennes de F1. Mais le fait majeur ici est que nous avons mis en évidence que, au delà de la variabilité individuelle, il y a une structure remarquable des positions moyennes : là encore, elles se répartissent selon un principe de dispersion optimale, c'est-à-dire de distances homogènes entre les quatre voyelles. Bien évidemment, ceci résulte en une corrélation très forte entre les valeurs acoustiques moyennes produites et perçues des voyelles. Le jeu de distances égales entre voyelles consécutives, en perception et en production, fournit un écho frappant à la théorie de la dispersion maximale de Lindblom et collègues (1972, 1986), qui suppose que les phonèmes des langues sont organisés de telle manière à ce qu'ils soient suffisamment distinct pour être perçus. En un certain sens, on peut dire que les sujets de nos expériences sont des « phonéticiens optimaux », distribuant parfaitement leurs voyelles antérieures dans l'espace acoustique disponible – au sens des distances perceptives en Barks – et ce tant en perception – à l'échelle individuelle – qu'en production – à l'échelle du groupe.

Les interactions perceptivo-motrices dans cette étude peuvent se résumer ainsi : si lors de la tâche de perception, les locuteurs ne peuvent se démarquer de prototypes très ancrés à cause du protocole simplifié, lors de la tâche de perception, la production est libre, donc les locuteurs vont varier, et ce autour de gabarits communs. Ces gabarits dans leurs positions moyennes sont parfaitement identiques en perception et en production, et organisés selon un principe de distribution perceptive optimale.

## **5. Expérience de convergence phonétique et d'imitation volontaire**

### **1. Principe et hypothèses**

La dernière expérience de notre protocole expérimental consistait en une expérience de convergence phonétique et d'imitation volontaire de voyelles. Nous nous sommes appuyés sur les expériences de Sato et collaborateurs (Garnier et coll., 2013; Sato et coll., 2013) qui ont proposé à des participants de produire des voyelles à partir de cibles auditives de fréquences fondamentales variables, une première fois avec une simple consigne de production de voyelles (sans consigne de répétition ou d'imitation), et une seconde fois avec une consigne explicite d'imitation de la voix perçue. Lors de ces deux expériences, il est apparu que les participants avaient tendance à converger vers la cible (c'est-à-dire à ajuster la fréquence fondamentale de leurs productions à celles des cibles) dans les deux conditions, même lorsque la tâche n'impliquait pas de consigne

explicite d'imitation – quoique la tendance à la convergence ait été nettement plus forte dans la tâche d'imitation explicite.

Ces capacités de convergence et d'imitation mettent en jeu de facto des mécanismes de couplage perceptivo-moteur, et impliquent l'intervention de processus de plasticité perceptivo-motrice rapide dans les phénomènes de production de voyelle étudiés. Nous avons donc choisi d'inclure ce paradigme dans notre programme expérimental.

Notre expérience de convergence et imitation consistait, comme dans les expériences de Garnier et coll. (2013) et Sato et coll. (2013), en trois tâches de production de voyelles : une première tâche de production avec cible visuelle, suivie de deux tâches de production avec cible auditive, avec ou sans consigne d'imitation.

Nos hypothèses concernant cette troisième expérience étaient que d'une part les participants imiteraient clairement la cible présentée lors de la tâche d'imitation volontaire, et que, d'autre part, ils convergeraient – moins fortement – vers la cible lors de la tâche de production sans consigne imitative.

## **2. Matériel et méthodes**

### **1. Participants**

Seize participants (10 femmes et 6 hommes), âgés de 20 à 40 ans (moyenne d'âge : 29 ans) ont participé à l'expérience de convergence et imitation. Tous les participants étaient normo-entendants, de langue maternelle française, sans déficits auditifs ni déficits de la vision et sans troubles du langage.

### **2. Stimuli**

Pour créer les stimuli auditifs utilisés dans l'expérience de convergence et imitation, un homme et une femme, de langue maternelle française, ont été enregistrés dans une chambre sourde lors de la production des voyelles /e ø o/. Une production clairement articulée de chaque voyelle par participant a été sélectionnée et numérisée à 44.1 kHz. Par la suite, en utilisant le logiciel Praat, chaque voyelle a été coupée à partir du début jusqu'à la fin de la vocalisation. Le pitch de chaque voyelle a été modifié artificiellement par pas de 5 Hz. Cette manipulation du pitch a été réalisée en utilisant le module PSOLA intégré dans Praat, qui permet de modifier le pitch sans affecter les formants des voyelles. Finalement, pour le locuteur, un panel de 21 stimuli par type de voyelle a été créé avec une étendue de F0 de 70 à 180 Hz, et pour la locutrice, un panel de 41 stimuli par type de voyelle avec une étendue de F0 de 150 à 350 a été obtenu.

Ces jeux de stimuli ont permis de présenter à chaque participant un ensemble de neuf stimuli par voyelle (/e ø o/) ces stimuli ayant une fréquence fondamentale variant entre -20% et +20% autour de la valeur de F0 du participant (-20%, -15%, -10% -5%, 0, +5%, +10%, +15%, +20%). Les valeurs de F0 pour les participants hommes pouvaient donc s'étendre entre 90Hz (avec les neuf valeurs modifiées s'échelonnant entre 70 et 110 Hz) et 150 Hz (avec les neuf valeurs modifiées s'échelonnant entre 120 et 180 Hz), alors que pour les participantes femmes, ces valeurs pouvaient s'étendre entre 190 Hz et 290 Hz

(avec les neuf valeurs modifiées s'échelonnant entre 150 et 230 Hz pour une extrême et entre 230 et 350 Hz pour l'autre extrême) (voir Tableau 7).

homme/pas	-20%	-15%	-10%	-5%	0%	5%	10%	15%	20%
<b>90</b>	70	75	80	85	<b>90</b>	95	100	105	110
<b>95</b>	75	80	85	90	<b>95</b>	100	105	110	115
<b>100</b>	80	85	90	95	<b>100</b>	105	110	115	120
<b>105</b>	85	90	95	100	<b>105</b>	110	115	120	125
<b>110</b>	90	95	100	105	<b>110</b>	115	120	125	130
<b>115</b>	90	100	105	110	<b>115</b>	120	125	130	140
<b>120</b>	95	100	110	115	<b>120</b>	125	130	140	145
<b>125</b>	100	105	115	120	<b>125</b>	130	140	145	150
<b>130</b>	105	110	115	125	<b>130</b>	135	145	150	155
<b>135</b>	110	115	120	130	<b>135</b>	140	150	155	160
<b>140</b>	110	120	125	135	<b>140</b>	145	155	160	170
<b>145</b>	115	125	130	140	<b>145</b>	150	160	165	175
<b>150</b>	120	130	135	145	<b>150</b>	160	165	175	180

femme/pas	-20%	-15%	-10%	-5%	0%	5%	10%	15%	20%
<b>190</b>	150	160	170	180	<b>190</b>	200	210	220	230
<b>195</b>	155	165	175	185	<b>195</b>	205	215	225	235
<b>200</b>	160	170	180	190	<b>200</b>	210	220	230	240
<b>205</b>	165	175	185	195	<b>205</b>	215	225	235	245
<b>210</b>	170	180	190	200	<b>210</b>	220	230	240	250
<b>215</b>	170	185	195	205	<b>215</b>	225	235	245	260
<b>220</b>	175	185	200	210	<b>220</b>	230	240	255	265
<b>225</b>	180	190	205	215	<b>225</b>	235	250	260	270
<b>230</b>	185	195	205	220	<b>230</b>	240	255	265	275
<b>235</b>	190	200	210	225	<b>235</b>	245	260	270	280
<b>240</b>	190	205	215	230	<b>240</b>	250	265	275	290
<b>245</b>	195	210	220	235	<b>245</b>	255	270	280	295
<b>250</b>	200	215	225	240	<b>250</b>	265	275	290	300
<b>255</b>	205	215	230	240	<b>255</b>	270	280	295	305
<b>260</b>	210	220	235	245	<b>260</b>	275	285	300	310
<b>265</b>	210	225	240	250	<b>265</b>	280	290	305	320
<b>270</b>	215	230	245	255	<b>270</b>	285	295	310	325
<b>275</b>	220	235	250	260	<b>275</b>	290	305	315	330
<b>280</b>	225	240	250	265	<b>280</b>	295	310	320	335
<b>285</b>	230	240	255	270	<b>285</b>	300	315	330	340
<b>290</b>	230	245	260	275	<b>290</b>	305	320	335	350

Tableau 7 : Valeurs de F0 pour chacune des neuf voyelles présentées au participant, selon son sexe et sa moyenne de F0

### 3. Procédure expérimentale

Les trois tâches de production de voyelles ont été présentées dans un ordre fixe pour tous les participants. Tout d'abord, les participants devaient produire les voyelles /e/ /ø/ et /o/ selon une cible orthographique visuelle. Cette première consigne fournissait une condition contrôle de production « non influencée », et permettait à l'expérimentateur de mesurer la fréquence fondamentale F0 moyenne du participant.

Dans la tâche suivante, appelée tâche de convergence, on demandait aux participants de produire les trois voyelles en leur fournissant cette fois des cibles auditives. Cependant, aucune instruction d'imitation ou de répétition des cibles auditives n'était donnée aux participants.

Enfin, la troisième tâche, appelée tâche d'imitation, consistait également en une répétition de voyelles selon une cible auditive, mais cette fois, on demandait au participant d'imiter intentionnellement les caractéristiques de la voix du locuteur perçu.

Les cibles auditives présentées à chaque participant consistaient en 27 stimuli (trois voyelles x neuf fréquences fondamentales F0), sélectionnés à partir de la base de données de voyelles présentée dans le paragraphe 2 ; les neuf fréquences F0 quantifiées variant entre -20% et +20%, par pas de 5% autour de la moyenne F0 du participant mesurée lors de la première tâche de production avec cible visuelle.

Au total, l'expérience de convergence et imitation durait environ douze minutes, en comptant l'extraction de la moyenne de F0 du participant après la tâche de production avec cible visuelle (voir Figure 32).

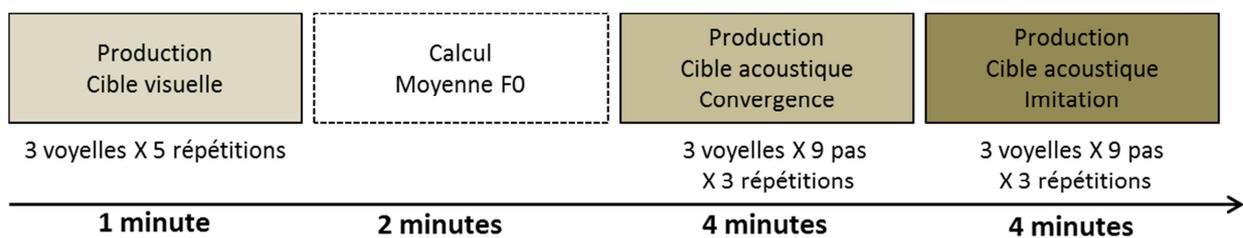


Figure 32 : Procédure expérimentale de l'expérience de Convergence et Imitation

#### 4. Analyses

Pour chaque tâche de production avec cible auditive, toutes les analyses acoustiques ont été réalisées avec le logiciel Praat. Une procédure semi-automatique a été utilisée pour segmenter les 81 voyelles produites par les participants. Pour chaque participant, la procédure impliquait une segmentation automatique de chaque voyelle, basée sur un algorithme de détection de l'intensité et de la durée. Si nécessaire, la segmentation a été corrigée manuellement, en se basant sur les informations de l'onde sonore et du spectrogramme, afin d'identifier correctement le début et la fin de chaque voyelle. Après une extraction individuelle des fichiers sons de chaque voyelle, les omissions et les hésitations ont été identifiées manuellement et retirées des analyses. Finalement, pour chaque voyelle, la valeur de F0 a été calculée à partir d'une période définie à plus ou moins 25 ms à partir du pic d'intensité maximum du fichier son.

Pour chaque participant, les valeurs de F0 de chaque voyelle produite dans les tâches de « Convergence » et « d'Imitation » ont été calculées, et exprimées par un pourcentage de changement par rapport à la moyenne de F0 du participant (calculée préalablement à partir de la tâche de production avec cible visuelle). A partir de ces données, un panel de 2x27 points de corrélation (3 voyelles X 9 valeurs X 2 tâches) a été calculé pour chaque participant, en reliant la variation de F0 produite par le participant à la variation de F0 imposée dans le stimulus. Nous avons calculé les coefficients de pente et les coefficients de corrélation pour chaque participant et nous avons réalisé une analyse de la variance (ANOVA) sur ces coefficients avec la tâche (convergence vs. imitation) comme variable intra-participant. Lorsque cela était possible, nous avons

réalisé des tests post-hoc avec correction de Bonferroni et fixé le seuil de significativité à  $p=.05$  ; également, nous avons vérifié la sphéricité des données grâce au test de Mauchly, et lorsque nécessaire, nous avons effectué des corrections de Greenhouse-Geiser (Greenhouse et Geiser, 1959).

### 3. Résultats

Comme attendu, des changements imitatifs ont été observés dans les deux tâches, même s'ils étaient plus importants dans la tâche d'imitation volontaire (voir Figure 33). Les coefficients de pente étaient significativement différents de zéro à la fois dans la tâche dite de « convergence » ( $t(14)=5.98$ ;  $p<0.001$ ) et dans la tâche d'imitation volontaire ( $t(14)=35.78$ ;  $p<0.001$ ) (voir Figure 34) ; et, les coefficients de pente étaient plus importants dans la tâche d'imitation volontaire (0.87) par rapport à la tâche dite de « convergence » (0.45) ( $t(14)=6.2$ ;  $p<0.001$ ) (voir Figure 35).

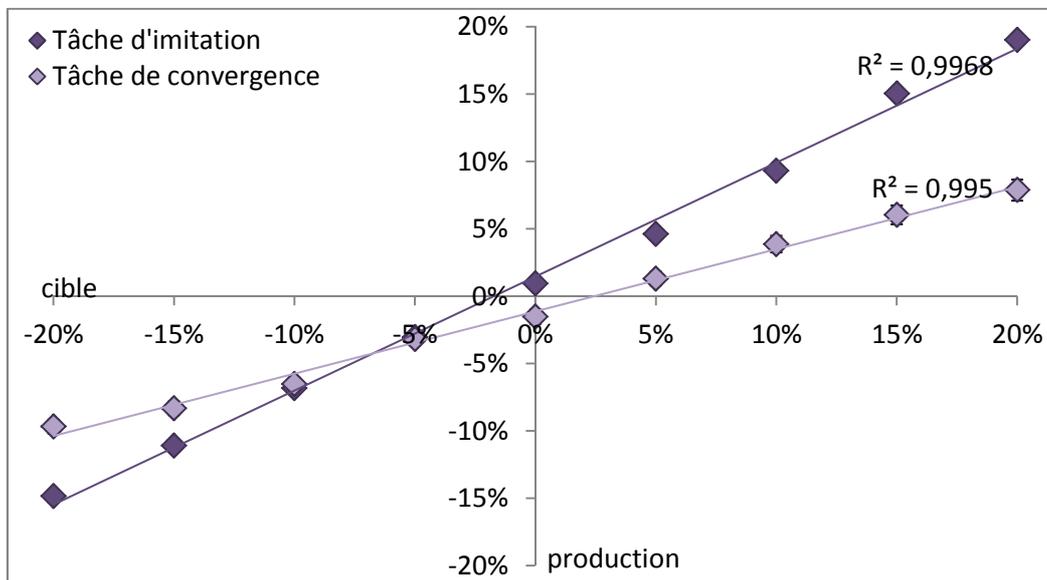


Figure 33 : Moyenne des pourcentages de changement lors des tâches de convergence et d'imitation (les barres d'erreurs correspondent aux erreurs standards).

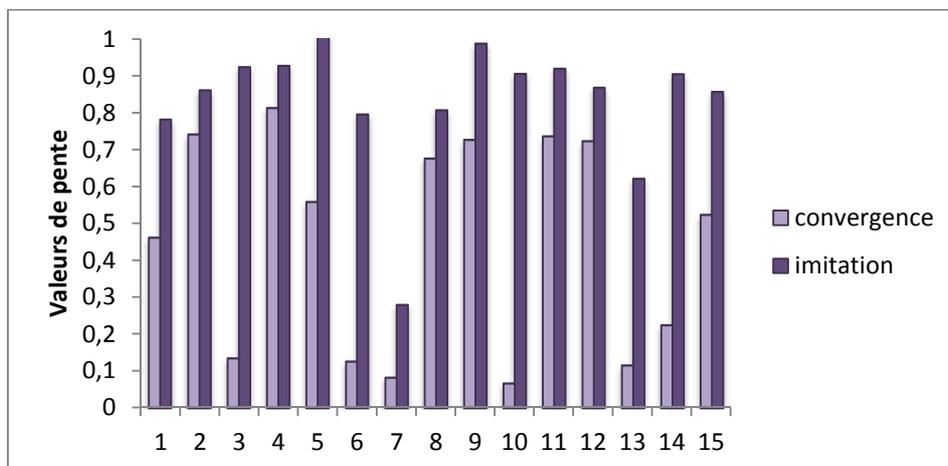


Figure 34 : Valeurs de pente par participant, dans les deux tâches.

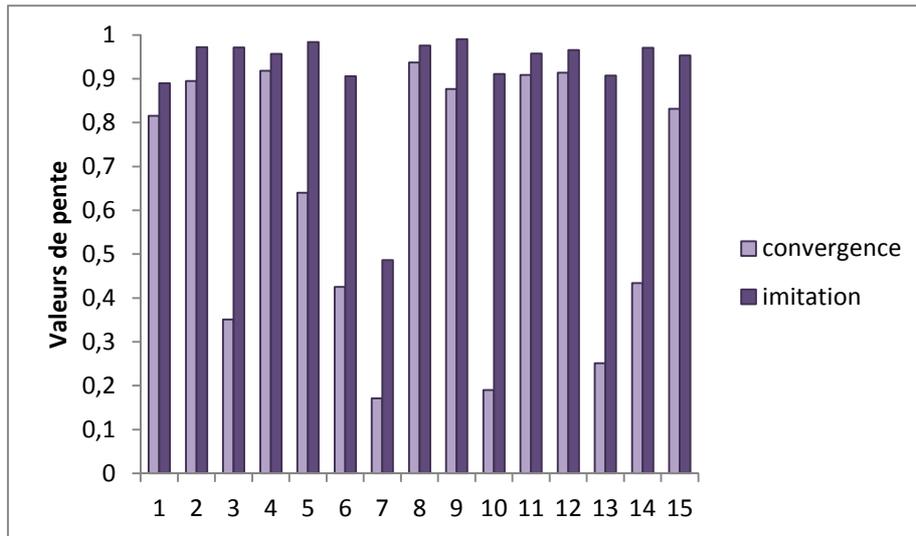


Figure 35 : Valeurs de coefficient de corrélation par participant, dans les deux tâches.

De même, les coefficients de corrélation étaient significativement différents de zéro à la fois dans la tâche de convergence ( $t(14)=8.3$ ;  $p<.001$ ) et dans celle d'imitation ( $t(14)=93.34$  ;  $p<.001$ ). Ces coefficients de corrélation étaient également plus élevés dans la tâche d'imitation ( $r=0.94$ ) par rapport à ceux de la tâche de convergence ( $r=0.64$ ) ( $t(14)=4.3$  ;  $p<.001$ ). Par contre, aucune corrélation individuelle n'a été observée entre la tâche de convergence et la tâche d'imitation ( $r^2=0.12$ ) (voir Figure 34, Figure 35).

Enfin, nous avons effectué une corrélation entre les valeurs de pente et les résultats obtenus lors du test de discrimination de fréquence JND réalisé en fin de passation, ceci afin de vérifier l'influence de la discrimination de fréquence sur les capacités d'imitation de fréquence fondamentale. Les résultats obtenus pour le test JND par les participants Juniors étaient en moyenne de 1,06 Hz (allant de 0,14 à 3,88 Hz). Aucune corrélation n'a été obtenue (voir Figure 37), ni par rapport aux valeurs de pente de la tâche de convergence ( $r^2 = 0,0042$ ), ni par rapport à celles de la tâche d'imitation ( $r^2 = 0,0014$ ).

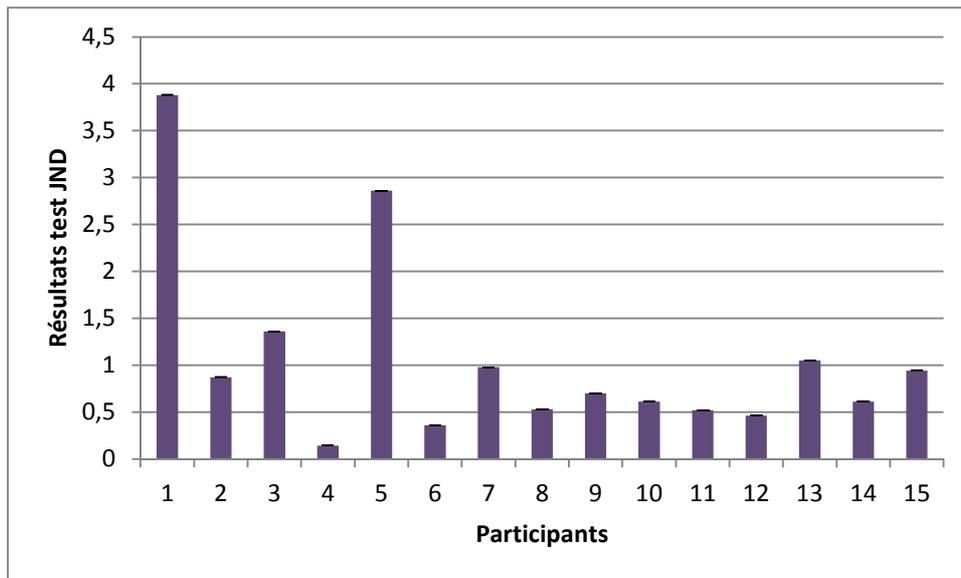


Figure 36 : Résultats obtenus lors du test JND par les participants Juniors

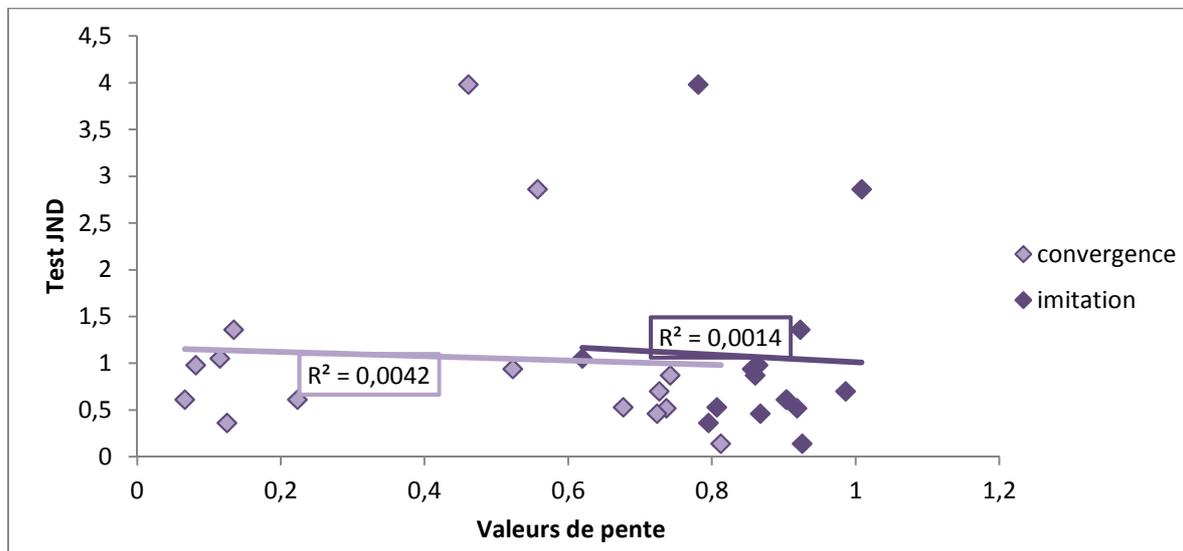


Figure 37 : Corrélation entre les résultats du test de discrimination de fréquence et les valeurs de pente des tâches d'imitation et de convergence, chez les participants normo-entendants juniors

#### 4. Discussion

Conformément à nos hypothèses, nous avons obtenu une réplication des résultats des études de Sato et collègues (Garnier et coll. 2013, Sato et coll. 2013). Lors de la tâche d'imitation, les participants ont clairement convergé vers la cible auditive, réalisant une imitation quasi parfaite de la F0 ; lors de la tâche de production sans consigne imitative, les participants ont également convergé vers la cible. Cependant, comme nous l'avions supposé, la convergence a été plus élevée dans la tâche d'imitation.

Par contre, également comme chez Garnier et coll. (2013) et Sato et coll. (2013), nous n'avons pas observé de corrélation individuelle entre le degré de convergence et celui d'imitation, c'est-à-dire que les participants les plus performants lors de la tâche

d'imitation n'ont pas été ceux qui ont le mieux convergé vers la cible lors de la tâche de production sans consigne d'imitation. Cette absence de corrélation peut être expliquée par le fait que les participants ont tous obtenu une convergence extrêmement forte lors de la tâche d'imitation, ce qui entraîne une quasi absence de variabilité des résultats (erreur standard de 0.02).

En résumé les participants ont donc été capables d'imiter, même non intentionnellement, une cible auditive, c'est-à-dire qu'ils ont ajusté la fréquence fondamentale de leur productions par rapport à celle de la cible. L'observation de l'imitation volontaire et involontaire de la fréquence fondamentale de voyelles isolées, en dehors d'une situation de communication interactive, nous permet de montrer une nouvelle fois que lors de la production de parole, le locuteur peut faire appel à ses capacités d'apprentissages perceptivo-motrices adjoignant à ses connaissances préalables des informations sur l'environnement de parole externe. A ce titre, cette expérience de convergence et d'imitation nous fournit bien un paradigme de mise en évidence de mécanismes d'interaction sensori-motrice lors de la production de parole.

## **6. Conclusion**

Dans ce chapitre, nous avons introduit le protocole expérimental que nous souhaitons utiliser pour tester le lien perceptivo-moteur lors de la perception et de la production de parole et présenté notre méthodologie ainsi que les résultats obtenus par notre population contrôle de jeunes adultes normo-entendants.

Ainsi, le premier paradigme comportemental proposé consiste en une expérience de close-shadowing, basée sur les études précédentes de Porter et coll. (1980) et Fowler et coll. (2003). Deux tâches de perception de parole, présentée auditivement ou audiovisuellement, sont proposées aux participants : une première tâche à réponse manuelle où le participant doit identifier le plus rapidement possible un stimulus de parole, en choisissant une touche de clavier selon le stimulus reconnu ; et une seconde tâche de répétition rapide des syllabes perçues. Lors de ces deux tâches la proportion de réponses correctes et la rapidité de réponse ont été mesurées. Deux groupes de jeunes adultes ont été enregistrés, un premier groupe répondant à des stimuli non-bruités, puis un second groupe répondant à des stimuli bruités, ceci afin de faire ressortir un avantage possible de la présentation audiovisuelle par rapport à la présentation auditive.

Dans le groupe répondant à des stimuli non-bruités, aucun avantage de la modalité audiovisuelle n'a été constaté, en raison d'un effet plafond sur la vitesse et la précision de réponse. En revanche, comme attendu, dans le groupe répondant aux stimuli bruités, les réponses aux stimuli audiovisuels ont été plus rapides et plus précises par rapport à la modalité auditive. De manière plus importante, dans les deux groupes, nous avons constaté des différences entre le mode de réponse : en effet, les réponses orales ont été plus rapides mais moins précises que les réponses manuelles. Nous avons proposé d'interpréter ces différences à la lumière de la boucle perceptivo-

motrice de la perception de parole introduite par Skipper et coll. (2007), en suggérant que les réponses orales seraient générées à un stade précoce de la boucle perceptive, entraînant des réponses rapides, mais moins précises que les réponses manuelles, qui, elles, seraient générées en fin de la boucle perceptive.

Le second paradigme expérimental proposé, basé sur les études précédentes de Ménard et coll. (2008 et 2014), vise à observer des idiosyncrasies au sein de deux tâches de production et de perception de voyelles, en espérant obtenir une corrélation à travers les participants entre la production et la perception des voyelles, comme cela avait été le cas chez Ménard et Schwartz (2014). Cependant, les résultats ont été inattendus : en effet, en raison probablement d'un nombre réduit de stimuli lors de notre tâche de perception, la variabilité inter participants a été extrêmement réduite et, en conséquence, aucune corrélation n'a été constatée entre la production et la perception des voyelles à travers les participants. Par contre, nous avons obtenu une corrélation globale très forte entre les valeurs acoustiques produites et perçues et avons proposé de l'interpréter dans le cadre de la théorie de la dispersion maximale de Lindblom et collègues (1972, 1986) : lors de la tâche de perception, en raison du nombre réduit de stimuli, les participants ont caractérisé les voyelles selon des prototypes communs équirépartis dans l'espace perceptif, et lors de la tâche de production, malgré la variabilité, les participants vont converger autour de gabarits communs, ces gabarits étant basés sur des moyennes similaires à celles de la tâche de perception, illustrant des liens étroits entre la perception et la production, comme suggéré par les théories perceptivo-motrices.

Enfin, le troisième paradigme, reproduction fidèle des précédentes études de Sato et collaborateurs (Garnier et coll., 2013, Sato et coll., 2013), vise à observer des phénomènes d'imitation volontaire et involontaire de cibles auditives. Pour cela, deux tâches de production de voyelles selon une cible auditive ont été proposées aux participants, avec ou sans consigne d'imitation. Comme lors des études précédentes, nous avons pu observer des capacités d'imitation consciente des participants lors de la tâche avec consigne d'imitation, mais surtout, les participants ont également convergé inconsciemment vers la cible auditive lors de la tâche sans consigne d'imitation. Ces capacités d'imitation conscientes et inconscientes nous offrent une illustration des relations sensori-motrices lors de la production de parole puisqu'elles démontrent l'adaptation des commandes motrices à l'environnement de parole extérieur, à partir de l'input sensoriel.

Pris ensembles, ces trois expériences suggèrent fortement un couplage perceptivo-moteur fonctionnel lors de la perception et de la production de parole et apportent donc de nouvelles évidences pour une nature sensori-motrice des représentations de parole.

En conclusion de cette partie, nous validons notre protocole expérimental, composé de ces trois paradigmes comportementaux, qui va être utilisé maintenant pour tester les relations perceptivo-motrices lors de la perception et de la production de parole chez une population de normo-entendants séniors ainsi que chez un groupe de personnes sourdes porteuses d'un implant cochléaire.





## Partie II : Relations perceptivo- motrices chez les personnes normo- entendantes Séniors

---



## 1. Introduction

Cette seconde partie consiste en l'évaluation des liens perceptivo-moteurs au sein d'une population normo-entendante âgée. L'objectif était, d'une part, de mesurer l'influence du déclin des fonctions cognitives et langagières chez les séniors, et d'autre part de constituer un second groupe contrôle plus âgé, dont les résultats pourront être comparés par la suite avec ceux d'une population sourde implantée, constituée majoritairement de personnes plus âgées que celles de notre groupe de participants normo-entendants juniors.

Nous allons tout d'abord proposer une revue de la littérature portant sur les capacités de production et de perception de parole ainsi que sur le lien perceptivo-moteur chez les séniors. Ensuite, nous présenterons les résultats obtenus par notre population normo-entendante sénior.

## 2. Organisation générale de l'expérience et hypothèses globales

### 1. L'enjeu des « séniors »

Le mot « sénior » provient du latin « senior » qui peut être traduit par « plus âgé ». Cependant, il s'agit d'un terme général, qui peut, selon le contexte, définir des personnes de plus de 35 ans, lors de contexte sportif, alors qu'utilisé dans le contexte médical, ce terme caractérisera les personnes âgées de plus de 70 ans. Dans cette thèse, nous utiliserons ce terme pour parler des personnes de plus de 60 ans.

Chaque année, la proportion de personnes séniors est plus élevée, grâce à l'augmentation de l'espérance de vie d'une part, mais aussi sous l'effet de l'avancée en âge des générations du baby-boom (1947-1973). En France, le nombre de personnes de plus de 60 ans a augmenté de 20,7% entre 2001 et 2011, et si cette tendance se poursuit, en 2060, une personne sur trois sera âgée de plus de 60 ans<sup>4</sup>. Ainsi, la recherche sur le vieillissement revêt une importance de plus en plus capitale au cours du temps, étant donnée la part de plus en plus élevée des personnes séniors dans la société. Plus spécifiquement, la recherche sur la communication orale est également nécessaire, puisqu'on estime que 12% des personnes séniors sont sujettes à des troubles de la communication, comme par exemple la perception de la parole dans le bruit qui s'avère difficile dans cette population.

Dans cette partie, nous proposons donc de tester les capacités sensori-motrices chez un groupe de personnes séniors normo-entendantes et ce grâce au protocole expérimental préalablement décrit, en comparant leurs données avec celles de notre population normo-entendante junior. Également, explorer l'évolution de la boucle perceptivo-motrice au cours de l'âge nous sera utile pour déterminer si les résultats obtenus pour notre groupe de personnes sourdes implantées sont influencés par l'âge de nos participants (la moyenne d'âge du groupe de participants sourds implantés étant de 59 ans).

---

<sup>4</sup> [http://www.insee.fr/fr/themes/document.asp?ref\\_id=T11F036](http://www.insee.fr/fr/themes/document.asp?ref_id=T11F036)

## 2. Perception de la parole chez les personnes normo-entendantes seniors

### 1. Perception auditive

Il est bien connu que les capacités auditives des adultes se dégradent avec l'âge. En effet, 20 à 40% des seniors sont sujets à des pertes auditives, cependant nous nous intéresserons ici aux capacités auditives et plus précisément aux capacités de perception de parole au cours du vieillissement « normal », c'est-à-dire lorsque les pertes auditives ne dépassent pas 20dB (voir Partie III paragraphe 2. 1. 1. pour une présentation rapide de l'audiométrie tonale) dans les fréquences inférieures à 6000 Hz (Fullgrabe, 2014).

Cependant, malgré des capacités auditives correctes permettant à tous les sons de la parole de pouvoir être perçus, les seniors semblent avoir des difficultés à comprendre certains sons de parole, particulièrement lors de conditions adverses, dans un environnement bruité par exemple.

Gelfand et collègues (1985) ont mesuré les capacités de reconnaissance de syllabes chez 62 participants normo-entendants âgés de 20 à 65 ans, en condition d'écoute la plus confortable (Most Comfortable Listening Level, MCL) et à plus et moins 8dB par rapport à ce niveau MCL. Ils ont constaté une baisse significative des performances de reconnaissance au cours de l'âge, et ce particulièrement lors de la condition où le niveau d'audibilité était le plus bas. Dans une seconde étude (Gelfand et collègues, 1986), les mêmes participants ont effectué des tests d'identification de syllabes en condition bruitée avec un rapport signal sur bruit de +5 et +10 dB. Les participants les plus âgés ont obtenu des scores significativement plus bas que ceux des participants les plus jeunes, par contre, le même patron de performance a été obtenu, c'est-à-dire que les confusions phonémiques restaient globalement similaires entre les participants jeunes et âgés.

Par la suite, Ohde et Abou-Khalil (2001) ont voulu déterminer le rôle des indices statiques (comme le burst de la consonne) et dynamiques (comme les transitions formantiques) dans la perception de consonnes plosives et de voyelles chez trois groupes de 8 adultes différant en âge (respectivement 20-26 ans, 52-59 ans et 70-76 ans). Les participants ont dû identifier des syllabes CV modifiées quant à leurs transitions formantiques, l'absence ou la présence du burst et la durée du voisement. Aucune différence n'a été observée entre les trois groupes concernant la perception des voyelles. Concernant les consonnes, les deux groupes les plus jeunes ont identifié correctement les consonnes même en l'absence d'un des indices (burst ou transitions formantiques). Par contre, le groupe de personnes seniors semble avoir besoin de l'intégration des indices statiques et dynamiques (c'est-à-dire des transitions formantiques et du burst) pour identifier le lieu d'articulation des consonnes. De plus, il a été observé que la durée du voisement a aidé les deux groupes les plus jeunes et non les seniors pour l'identification des consonnes quand le burst était absent. Les auteurs ont fait l'hypothèse que ces résultats pourraient être expliqués par une baisse de la vitesse de traitement de la parole au cours de l'âge, rendant moins efficace les mécanismes d'intégration d'indices multiples.

Plus récemment, Fullgrabe et collègues (2013) ont testé les performances d'intelligibilité de parole chez les seniors au cours du vieillissement normal. Pour cela, un groupe de 9 jeunes adultes (de 18 à 27 ans) et un groupe de 21 seniors (âgés de plus de 60 ans) ont réalisé des tests d'identification de séquences VCV différant entre elles par le type de consonne et ce, avec plusieurs niveaux de bruit ou en condition d'écoute non-bruitée. Ils ont également été soumis à des tests de reconnaissance de phrases soit dans le silence soit en présence de deux locuteurs parlant en même temps que le locuteur cible (locuteur prononçant le stimulus à identifier), ces locuteurs pouvant se trouver au même endroit que le locuteur-cible (condition « co-localisée »), ou dans des endroits différents (condition « séparée »). Que ce soit pour l'identification de syllabes ou de phrases en condition non bruitée, les deux groupes de participants ont atteint des scores plafond de près de 100%, même si une différence faible mais significative a été observée entre les deux groupes de participants concernant l'identification de syllabes (baisse de pourcentage d'identification correcte de 1,7% chez le groupe de seniors). Concernant l'identification de syllabes en condition bruitée, une baisse significative des performances de reconnaissance a été observée entre les deux groupes de participants, et ce à tous niveaux de bruit, comme cela avait été obtenu par Gelfand et collègues (1985 et 1986). De même concernant l'identification de phrases en condition bruitée, les participants seniors ont obtenu des scores de reconnaissance plus bas que les jeunes participants. Enfin, de manière intéressante, la différence entre les deux groupes s'est révélée plus importante lors de la condition « co-localisée » par rapport à la condition « séparée », ce qui montre que les seniors sont plus perturbés lorsque le son à percevoir provient de la même direction que les sons concurrents. En résumé, il est donc important de noter que même si les seniors ne présentent pas de problèmes auditifs majeurs, leurs capacités de perception de parole semblent dégradées, particulièrement en condition bruitée.

Enfin, quelques études ont été menées pour évaluer l'influence du déclin des fonctions cognitives sur la perception de parole. Tout d'abord, Cienkowski (2010) et Cienkowski et Vasil (2006) ont mené une expérimentation chez trois groupes de participants : un groupe de jeunes adultes sans perte auditive, et deux groupes de seniors avec et sans pertes auditives. Ils ont corrélé leurs résultats lors de tests de perception de parole dans le bruit avec trois tests permettant d'évaluer les fonctions cognitives (la rapidité de processus et la mémoire de travail). Les résultats obtenus ont montré une faible corrélation entre les performances de reconnaissance de parole dans le bruit et les tests cognitifs, suggérant une influence très modérée du déclin des fonctions cognitives sur la perception de parole. En revanche, dans une revue menée sur les contributions cognitives à l'audition chez les personnes âgées, Moore et Füllgrabe (2012) ont reporté une corrélation positive entre les performances d'identification de parole dans le bruit et les performances dans divers tests cognitifs, et plus particulièrement lors de tests d'intelligence fluide et de mémoire de travail verbale. En résumé, la perception de parole chez les seniors pourrait être influencée par le déclin des fonctions cognitives, même si pour le moment, cette corrélation n'a pas encore été complètement investiguée.

## 2. Perception audiovisuelle

Si les capacités de perception de parole chez les personnes âgées sont réduites, qu'en est-il des performances d'intégration audiovisuelle ? Autrement dit, la question est de savoir si une personne senior est autant capable d'associer la lecture labiale à un signal de parole auditif qu'un jeune adulte. Plusieurs études ont tenté de répondre à cette question.

Cienkowski et collègues (2002) ont évalué les capacités d'intégration audiovisuelle chez trois groupes de personnes : un groupe de 12 jeunes adultes (de 18 à 35 ans), un groupe de 12 seniors (de 65 à 74 ans) et un groupe de 12 jeunes adultes « contrôles » à qui un bruit de conversation a été ajouté lors de l'écoute des stimuli, pour tenter de les mettre à égalité avec les capacités purement auditives, et détériorées, des seniors. Les trois groupes ont dû identifier des syllabes CV présentées en condition audio seul, en condition visuel seul, et en condition audio-visuelle incongruente avec des stimuli de type McGurk (visuel /gi/ + audio /bi/ par exemple). Les participants seniors ont obtenu des scores de réponse « fusion » (correspondant pour l'exemple précédemment cité à la réponse /di/) similaires à ceux des jeunes adultes, ce qui sous-entend des capacités d'intégration audiovisuelle équivalentes chez les seniors. En revanche, en l'absence de fusion des informations auditive et visuelle, les seniors ont eu tendance à donner des réponses se basant sur les indices visuels des stimuli (correspondant pour l'exemple précédemment cité à la réponse /gi/) alors que les jeunes adultes se sont basés sur la partie auditive des stimuli (correspondant pour l'exemple précédemment cité à la réponse /bi/), ce qui pourrait s'expliquer par le besoin de sélectionner une réponse à partir de la modalité la moins ambiguë pour les participants.

Par la suite, Sommers et collègues (2005) ont réalisé une étude sur l'identification de parole en environnement bruité par un groupe de 44 seniors (moyenne d'âge de 70 ans) et un groupe de 38 jeunes adultes (moyenne d'âge de 20 ans). Les différentes tâches proposées consistaient en l'identification de syllabes VCV, de mots et de phrases en condition audio seul (A), visuel seul (V) et audiovisuel (AV), avec un ajout de bruit, celui-ci étant ajusté aux capacités auditives des participants afin de permettre un score d'environ 50% en condition audio seul pour chaque participant. En conditions visuel seul et audiovisuel, les seniors ont obtenu des scores plus bas que les jeunes adultes. Les capacités de « visual enhancement » (VE) ou avantage du visuel (calculé typiquement comme suit :  $VE = (\text{scoreAV} - \text{scoreA}) / (1 - \text{scoreA})$ , où les scores sont les proportions de réponses correctes calculées sur l'ensemble du corpus, entre 0 et 1) et d'« auditory enhancement » (AE) ou avantage de l'audio (calculé typiquement comme suit :  $AE = (\text{scoreAV} - \text{scoreV}) / (1 - \text{scoreV})$ ) ont également été mesurées. Aucune différence significative n'a été observée entre les deux groupes de participants, ce qui montre que les participants seniors n'ont pas de problème spécifique pour combiner les signaux auditifs et visuels pour identifier la parole en condition bruitée. Ainsi leur score inférieur en condition audiovisuelle serait le résultat d'un déficit de capacité de lecture labiale et non la conséquence d'une baisse des capacités d'intégration audiovisuelle.

Très récemment, Stevenson et collègues (2012) ont étudié comment l'âge affecte le bénéfice de la multi-sensorialité lors de la perception de parole dans le bruit, en fonction de deux niveaux de complexité : un niveau basique, c'est-à-dire la perception des phonèmes, ainsi qu'un niveau complexe, c'est-à-dire la perception de mots. Pour cela, 18 jeunes adultes de 19 à 38 ans, et 16 adultes plus âgés, de 45 à 68 ans, ont été testés. Les participants devaient écouter des mots et les retranscrire sur un clavier. Les stimuli à percevoir étaient présentés à différents niveaux de bruit, en condition visuel seul, audio seul et audiovisuel. Les réponses des participants ont été analysées en termes de précision au niveau de la globalité du mot (c'est-à-dire quand le mot était entièrement reconnu, avec tous les phonèmes corrects) et au niveau des phonèmes, et ce pour chaque condition et chaque niveau de complexité (mot et phonèmes). Le gain potentiel de la multi-sensorialité a été calculé par la formule :  $\text{gainAV} = \text{scoreA} + \text{scoreV} - (\text{scoreAxscoreV})$ , en considérant que gainAV « représente une hypothèse nulle caractérisant quelle serait la réponse en condition audiovisuelle si les informations auditives et visuelles étaient traitées indépendamment »<sup>5</sup>). Concernant la reconnaissance des mots, le groupe d'adultes les plus vieux ont obtenu un gain de la multi-sensorialité plus élevé que celui des jeunes adultes lorsque les stimuli étaient présentés à des niveaux élevés de bruit, alors qu'à des niveaux de bruits plus faibles, les adultes les plus âgés ont présenté un gain de la multi-sensorialité plus faible que les jeunes adultes. En revanche au niveau des phonèmes, les deux groupes ont montré un gain de la multi-sensorialité similaire. Ces résultats suggèrent donc à la fois un bénéfice de l'audiovisuel chez les séniors aux niveaux de traitement élevés (décodage lexical) et un impact accru de la modalité visuelle en cas de bruit élevé.

Enfin, Sekiyama et collègues (2014) ont également évalué l'influence de la vision sur la perception de parole dans le bruit chez les personnes âgées en comparaison avec de jeunes adultes. Deux expériences en condition audio seul, visuel seul et audiovisuel ont été conduites : une première expérience où les deux groupes de participants (jeunes adultes et séniors) ont dû écouter des stimuli bruités avec le même rapport signal-sur-bruit, et une seconde expérience, où le rapport signal-sur-bruit a été adapté aux participants, pour obtenir le même niveau de performances de reconnaissance de parole en condition auditive seule. Les résultats ont montré que, malgré des performances similaires en condition visuelle seule, donc en lecture labiale, l'influence de la vision s'est révélée plus importante chez les personnes séniors, et ce, même lors de la deuxième expérience, alors que les performances en condition auditive seule étaient similaires chez les deux groupes de participants (grâce à l'adaptation du rapport signal-sur-bruit).

En conclusion, à travers ces différentes études, il semble donc que dans tous les cas, les séniors tirent parti de l'addition de la modalité visuelle lors de la perception de signaux perturbés, au moins autant voire plus que les sujets moins âgés, même si l'intégration audiovisuelle pourrait être moins performante à des niveaux élevés de

---

<sup>5</sup>Traduction à partir de Stevenson et collègues (2014)

complexité, plus particulièrement lorsque la nécessité de s'aider des indices visuels est moins pertinente, par exemple en condition d'écoute peu perturbée.

### **3. Production de la parole chez les personnes normo-entendantes seniors**

#### **1. Contrôle sensori-moteur de la parole chez les seniors**

Les personnes seniors sont reconnues pour avoir une production de parole plus lente que les personnes jeunes. Ramig (1983) a mesuré la vitesse de lecture et la vitesse de parole de 48 hommes répartis en trois groupes d'âge (25 à 35 ans, 45-55 ans et 65-75 ans). Les résultats ont montré que les productions des participants les plus âgés ont été significativement plus lentes que celles des participants les plus jeunes. Cette baisse de la vitesse de parole peut être en partie liée aux capacités motrices des personnes âgées.

En effet, Routhier et coll. (2012) ont réalisé une étude portant sur la production de parole chez les personnes seniors. Ils ont enregistré un groupe de personnes juniors (entre 20 et 34 ans) et un groupe de personnes seniors (entre 61 et 74 ans) dans une tâche de répétition de syllabes simples (CV) ou complexes (CCV), présentées visuellement, lors de séquences simples (composées de trois syllabes identiques) et complexes (composées de trois syllabes différentes). Les pourcentages d'erreurs de production, la durée des séquences ainsi que les temps de réaction entre la présentation des syllabes à produire et leur production ont été mesurés. Aucune différence de pourcentages d'erreurs n'a été observée entre les deux groupes. En revanche, les temps de réaction ont été plus longs chez les personnes seniors par rapport aux personnes juniors, lors de la production de séquences complexes, ce qui montre que la planification de mouvements complexes s'avère plus difficile chez ces participants. Les durées de production ont été plus longues chez les personnes seniors, à la fois pour les séquences simples et les séquences complexes, avec une différence plus importante entre les deux groupes pour les séquences complexes. Routhier et collègues supposent que ces différences ne peuvent être interprétées seulement par un ralentissement sensorimoteur général, mais plutôt par des difficultés à planifier les mouvements complexes tels que la coarticulation par les personnes seniors.

Par la suite, Bilodeau-Mercure et collègues (2015) ont mesuré l'évolution de la capacité de production de mouvements de parole, de mouvements oro-faciaux ou de gestes manuels au cours du vieillissement. Soixante-seize participants âgés de 22 à 93 ans, répartis en quatre groupes d'âges, ont dû réaliser différentes tâches : production de syllabes CCVC (parole), production de quatre mouvements oro-faciaux sans rapport avec la parole, et production de mouvements manuels (une couleur était associée à un doigt et lorsque la couleur était présentée au participant, il devait appuyer sur une touche avec le doigt correspondant), chaque type de mouvement étant produit lors de séquences simples (trois mouvements identiques) ou complexes (trois mouvements différents). Les stimuli utilisés pour déterminer quel mouvement devait être effectué étaient présentés selon une consigne visuelle ou auditive. Les pourcentages d'erreurs ont été mesurés pour chaque type de mouvement, et pour les mouvements manuels, les temps de réponses ont également été mesurés. Aucune différence liée à l'âge n'a été observée en termes

d'erreurs pour les productions de gestes manuels. Par contre, les temps de réponse ont été affectés par l'âge avec des temps de réaction plus longs pour les trois groupes les plus âgés, particulièrement lors de la production de séquences complexes de gestes manuels. Concernant la production de mouvements oro-faciaux, les participants plus âgés ont fourni plus d'erreurs que les participants jeunes, particulièrement lors de la production de séquences complexes. De la même manière, lors de la production de parole, une diminution significative de la performance a été observée avec l'âge, ce déclin étant plus important pour les séquences complexes.

En résumé, les séniors semblent donc présenter des déficits lors de la production de mouvements oro-faciaux, qui pourraient impacter la production de parole.

## 2. Production des voyelles chez les séniors

Plusieurs études ont été menées portant sur des analyses acoustiques de la production des voyelles chez les personnes séniors. Les premières études menées dans les années 80 sur ce sujet ont montré une centralisation du triangle vocalique chez les séniors très âgés (Benjamin, 1982; Liss et coll., 1990; Rastatter et Jacques, 1990, cités dans Torre et Barlow, 2009), suggérant une baisse de la différenciation entre les voyelles. Cependant, les changements formantiques dus à l'âge pourraient dépendre du sexe du locuteur, puisque Rastatter et coll. (1997) ont obtenu des valeurs de formants différentes selon l'âge chez un groupe d'hommes mais pas chez un groupe de femmes.

Torre et Barlow (2009) ont par la suite mesuré les caractéristiques de la parole en fonction de l'âge et du sexe. Ils ont enregistré 27 participants jeunes (de 20 à 35 ans, 15 femmes et 12 hommes) et 59 participants plus âgés (de 60 à 89 ans, 32 femmes et 27 hommes), lors de la production de syllabes CVC, et ont mesuré les formants des voyelles produites. Concernant F1, les femmes âgées ont produit un F1 plus bas par rapport aux femmes jeunes pour toutes les voyelles, alors que, par rapport aux hommes jeunes, les hommes âgés ont produit des valeurs de F1 plus élevées pour les voyelles /i/ et /u/ et plus basses pour les voyelles /ɪ æ ʌ ε/. Concernant les valeurs de F2 et F3, les productions des participants étant très variables selon les voyelles, aucune significativité n'a été obtenue concernant l'âge ou le sexe. Enfin, une augmentation significative de l'écart-type a été observée avec l'âge que ce soit chez les hommes ou chez les femmes, montrant une grande variabilité de production inter-participant.

Plus récemment encore, Ferland-Longpré et collègues (2013) ont étudié la production de voyelles par deux groupes de participantes juniors (de 25 à 37 ans) et séniors (de 68 à 80 ans). Les participantes ont dû produire des syllabes CV contenant les voyelles /i e u a/ dans différents contextes consonantiques, les syllabes étant insérées dans une phrase porteuse. Les moyennes des valeurs de F1 et F2 ont été mesurées pour chaque voyelle afin de déterminer la taille de l'espace vocalique. Les participantes juniors ont produit des espaces vocaliques plus grands que les séniors, et les séniors ont présenté plus de variabilité de production que les juniors, confirmant que le contrôle de l'articulation est moins stable chez les séniors.

### 3. Contrôle de la fréquence fondamentale chez les séniors

Dans les années 80/90, plusieurs études ont été menées sur le contrôle de la fréquence fondamentale chez des personnes séniors. Deux études ont ainsi mesuré la moyenne de production de F0 chez des personnes âgées. Tout d'abord, Morgan et Rastatter, en 1986, ont mesuré la moyenne de F0 produite par trois groupes de participantes (âgées en moyenne de 21, 72 et 84 ans) lors d'une tâche de narration. Ils ont observé que les deux groupes de participantes les plus âgées ont produit des valeurs de F0 moyen plus basses que les participantes jeunes. Par la suite, Russell et collègues (1995), ont apporté une nouvelle illustration à cette baisse de F0, en comparant des enregistrements de voix de femmes datant de 1943 avec des enregistrements des mêmes femmes réalisés en 1993. Ils ont mesuré la fréquence fondamentale et ont constaté une baisse de la fréquence moyenne lors de l'enregistrement de 1993 par rapport à celui réalisé 50 ans auparavant.

Concernant la variabilité de production de fréquence fondamentale, Fisher et Linville (1985) ont mesuré l'écart-type de F0 lors de la production de la voyelle /æ/ par trois groupes de participantes (de 25 à 35 ans, de 45 à 55 ans et de 70 à 80 ans) et ont observé une baisse de la stabilité des valeurs de F0 intra-participante chez les groupes plus âgés, révélant un contrôle laryngal de la phonation détérioré avec l'âge. De même, Linville et Korabic (1987) ont enregistré 18 femmes séniors (de 69 à 90 ans) lors de la production tenue des voyelles /i u a/ avec pour consigne de produire les voyelles les plus stables possibles. L'écart-type de F0 et le jitter (petites et rapides variations autour de la valeur de référence au cours de la production) ont été mesurés et dans certains cas, la variabilité intra-participante a été très importante, suggérant, encore une fois, une difficulté de contrôle de la voix par les séniors.

Une étude plus récente de Lortie et collègues (2015) a obtenu des résultats en partie différents de ceux des études de Morgan et Rastatter (1986) et de Russell et collègues (1995). Lortie et collègues ont réalisé une étude portant sur la fréquence et l'amplitude de la voix en fonction de l'âge, en enregistrant 80 participants répartis en trois groupes (juniors, âge moyen, séniors) lors d'une tâche de narration. A partir des productions, les valeurs de F0 et d'amplitude ainsi que le jitter (c'est-à-dire la perturbation à court terme de la fréquence fondamentale du signal) et le shimmer (c'est-à-dire la perturbation à court terme de l'amplitude du signal) ont été mesurés. Contrairement aux résultats observés chez Morgan et Rastatter (1986) et Russell et collègues (1995), aucune différence n'a été observée entre les trois groupes concernant les valeurs de F0 ou d'amplitude. En revanche, comme dans les études de Fisher et Linville (1985) et de Linville et Korabic (1986), les participants séniors de Lortie et collègues ont obtenu des valeurs de jitter et de shimmer plus élevées par rapport aux deux autres groupes. Ceci confirme à nouveau la tendance de moindre stabilité du contrôle de la voix chez les séniors.

Prises ensembles la majorité des études menées sur la production de parole chez les personnes seniors font état d'une production parfois perturbée, mais surtout de plus grande instabilité et variabilité.

#### **4. Lien perceptivo-moteur chez les personnes normo-entendantes seniors**

Les études sur le lien perceptivo-moteur chez les seniors sont peu nombreuses. Tout d'abord, deux études portant sur l'influence du feedback auditif sur la production de parole peuvent être mentionnées. Liu et collègues (2010) ont proposé une première étude de production de parole avec un feedback auditif modifié. Ils ont comparé trois groupes de dix participants (un groupe d'enfants de 7 à 12 ans, un groupe de jeunes adultes de 19 à 21 ans, et un groupe de seniors de 60 à 73 ans), qui ont dû produire la voyelle /a/ pendant 5 secondes, au cours desquelles le feedback auditif était perturbé à cinq reprises pendant un très court laps de temps. Cette perturbation consistait à augmenter ou diminuer la fréquence fondamentale du feedback par rapport à celle de la production réelle, et à évaluer l'amplitude d'éventuelles corrections apportées par les locuteurs en réponse à cette perturbation (soit une diminution de F0 en réponse à une augmentation induite par la perturbation du feedback acoustique). L'influence du feedback a été mesurée en termes d'ampleur de changement de production, et de temps de latence entre le moment de la modification du feedback et la réponse dans la voix du participant. Les résultats ont montré que les participants seniors ont été plus influencés que les jeunes adultes et les enfants, puisque l'amplitude de leur réponse corrective était plus importante. En revanche les enfants ont présenté des temps de latence plus longs que ceux des deux autres groupes, alors qu'aucune différence n'a été observée entre les deux groupes d'adultes.

Par la suite, Liu et collègues (2011) ont étendu cette recherche, en comparant cette fois différents groupes d'adultes. Soixante participants ont participé à cette étude, répartis en cinq groupes de 12 participants, en fonction de leur âge (19-25 ans, 31-40 ans, 41-50 ans, 51-59 ans et 61-75 ans). Ils ont dû produire la voyelle /u/, leur feedback auditif étant modifié comme précédemment, par variation de la fréquence fondamentale, de manière positive ou en négative, à deux niveaux ( $\pm 50$  cents ou  $\pm 100$  cents ; 100 cents = 1 semi ton). Comme dans l'étude précédente, l'influence du feedback était mesurée via l'amplitude et la latence de la réponse. Aucune différence significative n'a été observée entre les différents groupes d'âge sur les temps de latence. En revanche, l'ampleur de la réponse variait avec l'âge, croissant avec l'âge jusqu'à atteindre un pallier maximum dans le groupe d'âge 51-59 ans puis baissant dans le groupe des 61-75 ans. Par ailleurs, le groupe d'âge de 51 à 59 ans semble être plus influencé par la direction du changement (positif ou négatif) et le niveau de perturbation.

En résumé, pour ces deux études, il semble que les seniors soient capables d'adapter leur production en fonction du feedback auditif perçu, ce qui implique que la boucle perceptivo-motrice reste active au cours du vieillissement.

Sur ce point, Tremblay et collègues (2013) ont mesuré les activations cérébrales lors de la perception et de la production de parole chez des personnes âgées, en

comparaison avec des jeunes adultes, en utilisant l'imagerie par résonance magnétique fonctionnelle IRMf). Six régions ont présenté un degré d'activation moindre chez les personnes âgées : le sillon temporal transverse (STT), situé dans le cortex auditif primaire/secondaire; une aire auditive associative, le planum temporale (PT) ; une aire d'intégration audiovisuelle et de traitement phonologique, le sillon temporal supérieur (STS), ainsi que trois régions classiquement associées aux mécanismes de production de la parole, le cortex prémoteur ventral (PMV), l'aire motrice supplémentaire (AMS) et le gyrus frontal supérieur (GFS). Plus particulièrement, la sous-activation chez les personnes âgées du PMV, impliqué dans la planification motrice et la mémoire de travail verbale, et de l'AMS, impliquée dans l'initiation et le contrôle du mouvement, est apparue principalement lors de la tâche de production. En revanche, le GSF, impliqué dans la planification et le contrôle des mouvements, a été suractivé lors de la tâche de perception et non lors de la tâche de production, ce qui pourrait suggérer qu'il apporterait des informations sensori-motrices pouvant être utilisées pour l'identification des sons de parole. Globalement, c'est donc à ce que les auteurs appellent un « vieillissement du système neural sensori-moteur de la parole » que l'on assiste, et qui apparaît compatible avec les lentes dégradations des performances en production et en perception que nous avons passées en revue précédemment.

## **5. Proposition et hypothèses expérimentales**

Face à ces études, il nous a semblé important d'explorer plus avant l'évolution des boucles perceptivo-motrices mises en œuvre lors de la perception et de la production de parole chez un groupe de personnes âgées. De plus, l'examen de ce groupe de participants apparaît crucial pour déterminer dans quelle mesure les résultats observés auprès du groupe de participants implantés sont fonction de leur âge ou de leur implantation.

Pour ce faire, le même protocole expérimental décrit lors de la première partie de cette thèse a été réalisé chez un groupe de participants seniors. Rappelons brièvement que ce protocole incluait une première expérience de close-shadowing, lors de laquelle on compare la rapidité et la précision de réponse à des stimuli de parole auditifs et audiovisuels selon une réponse orale ou manuelle, suivie d'une seconde expérience dite « d'idiosyncrasies » permettant d'étudier une possible organisation perceptivo-motrice des voyelles du Français lors de tâches de perception et de production de parole, enfin, une dernière expérience de convergence et d'imitation, où les capacités d'imitation automatique ou volontaire des participants sont évaluées. Les conditions expérimentales chez les seniors ont été en tous points similaires à celles des participants juniors (voir partie I paragraphe 2. 4.).

### **1. Participants**

Dix participants seniors (4 femmes et 6 hommes), âgés de 63 à 78 ans (moyenne d'âge : 69 ans) ont été enregistrés. Les participants étaient tous de langue maternelle française, sans trouble de la vision et sans trouble de langage. Préalablement aux expériences, comme pour les participants juniors, nous avons réalisé un audiogramme

afin de vérifier leurs capacités auditives. Ces audiogrammes ont révélé, dans les hautes fréquences (au-dessus de 3000 Hz), une perte moyenne de 20 dB (selon les participants : entre 10 et 35 dB de perte), phénomène classique lié à l'âge. Comme pour les participants juniors, le protocole expérimental a été validé par le CERNI (Comité d'éthique pour les recherches non interventionnelles) et tous les participants ont donné par écrit leur accord en signant, préalablement aux expériences, un formulaire de consentement.

## **2. Procédures expérimentales communes aux trois expériences**

Les participants seniors ont réalisé les trois expériences dans les mêmes conditions que les participants juniors, suivant la même procédure expérimentale et le même ordre de passation (voir partie I.2.4.).

### **3. Expérience de close-shadowing**

#### **1. Principes et hypothèses**

Comme pour les participants juniors, la première expérience réalisée chez les seniors était celle de close-shadowing. La procédure expérimentale proposée était identique à celle utilisée auprès des participants juniors lors de l'expérience incluant des stimuli non-bruités (voir partie I.2). Deux tâches de catégorisation des séquences /apa/ /ata/ et /aka/, présentées soit de manière auditive soit de manière audiovisuelle, étaient proposées selon un ordre contrebalancé entre les participants : une tâche de répétition rapide des syllabes cibles et une tâche de catégorisation manuelle où les participants devaient appuyer le plus vite possible sur la touche correspondant à la séquence perçue.

Rappelons que les résultats sur les participants normo-entendants lors des expériences de close-shadowing bruitées et non-bruitées ont montré un apport de la modalité visuelle uniquement dans le bruit, les réponses aux stimuli audiovisuels étant alors plus rapides que les réponses aux stimuli auditifs. Concernant la tâche, les réponses orales ont été plus rapides mais moins précises que les réponses manuelles, à la fois dans le bruit et sans bruit. Pour expliquer cette différence de temps de réaction et de précision de réponse entre les deux tâches, nous avons proposé d'interpréter nos résultats en relation avec les théories perceptivo-motrices de perception de la parole, notamment celle de Skipper et coll. (2007), et supposé que les réponses orales et manuelles seraient générées lors de deux stades différents d'une boucle sensori-motrice. Les réponses manuelles seraient générées au stade final de la boucle, lors de la comparaison de prédictions motrices avec les hypothèses sensorielles initiales fonction de l'input de parole. Les réponses orales, en revanche, seraient produites lors d'un stade précoce où les commandes motrices seraient générées avant que les hypothèses induites n'aient été comparées et vérifiées avec l'input de parole. En conséquence, les réponses orales seraient donc plus rapides, mais moins précises que les réponses manuelles.

En regard des études précédentes, montrant un déficit de la perception de parole chez les seniors malgré des capacités auditives correctes, nous nous attendions à ce qu'en général, notre groupe de participants seniors présente des résultats inférieurs à ceux des

participants juniors, que ce soit au niveau des temps de réaction ou des taux de précision de réponse. De plus, nous nous attendions à obtenir de meilleurs résultats pour les stimuli audiovisuels par rapport aux stimuli uniquement auditifs pour ce groupe. Enfin, plus généralement, notre principale interrogation concernait le maintien de la boucle perceptivo-motrice chez les personnes seniors, c'est-à-dire l'obtention de la différence entre les réponses orales et manuelles, avec des réponses plus rapides mais moins précises lors des réponses orales par rapport à la catégorisation manuelle.

## **2. Méthodologie**

### **1. Stimuli**

Les stimuli utilisés pour l'expérience de close-shadowing chez la population senior ont été ceux utilisés lors de l'expérience de close-shadowing sans bruit avec les normo-entendants (voir partie I.3.2.1.)

### **2. Procédure expérimentale**

Les participants seniors ont réalisé l'expérience dans les mêmes conditions et selon la même procédure expérimentale que les participants normo-entendants « juniors » (voir partie I.2.4. Population normo-entendante « junior » pour la procédure générale aux trois expériences, et partie I.3.2.1. pour la procédure spécifique à l'expérience de close-shadowing).

### **3. Analyses**

Les analyses réalisées sur les réponses de la population senior ont été identiques à celles décrites dans la partie I.3.2.1. Concernant les analyses statistiques, nous avons tout d'abord analysé les données des participants seniors séparément, en réalisant deux ANOVA à mesures répétées (une ANOVA pour les temps de réaction et une ANOVA pour les pourcentages de réponses correctes, avec les variables intra-participants modalité, tâche, syllabe). Nous avons ensuite comparé les données du groupe de seniors avec celles des deux groupes de juniors (expérience non-bruitée et bruitée), en réalisant deux ANOVA à mesures répétées (temps de réaction et pourcentages de réponses correctes) avec la variable groupe comme variable inter-sujet (seniors, juniors non-bruités, juniors bruités) et les variables tâche, modalité et syllabe comme variables intra-sujets. Du fait des comparaisons multiples lors de cette troisième série d'analyses, nous avons fixé le seuil de significativité à  $p = .017$  (sur ce point, il est de plus à noter que tous les effets et interactions observés avaient un seuil de significativité inférieur à  $p = .002$ ). Egalement, nous avons vérifié la sphéricité des données grâce au test de Mauchly, et lorsque nécessaire, nous avons effectué des corrections de Greenhouse-Geiser (Greenhouse et Geiser, 1959).

## **3. Résultats seniors seuls**

### **1. Temps de réaction**

Aucune différence significative n'a été observée entre les stimuli audiovisuels et les stimuli uniquement auditifs (414 ms vs. 422 ms ; voir Figure 38). Par contre, comme

attendu, les réponses orales ont été plus rapides que les réponses manuelles (268 ms vs. 568 ms) ( $F(1,9)=255,31$  ;  $p<.001$ ). Enfin, un effet de la syllabe a également été observé ( $F(1,9)=12,67$  ;  $p<.001$ ), avec la syllabe /pa/ plus rapidement reconnue que la syllabe /ka/, elle-même entraînant des temps de réaction plus rapides que la syllabe /ta/ (respectivement 374 ms, 429 ms et 451 ms). Aucune interaction n'a été observée entre les trois facteurs.

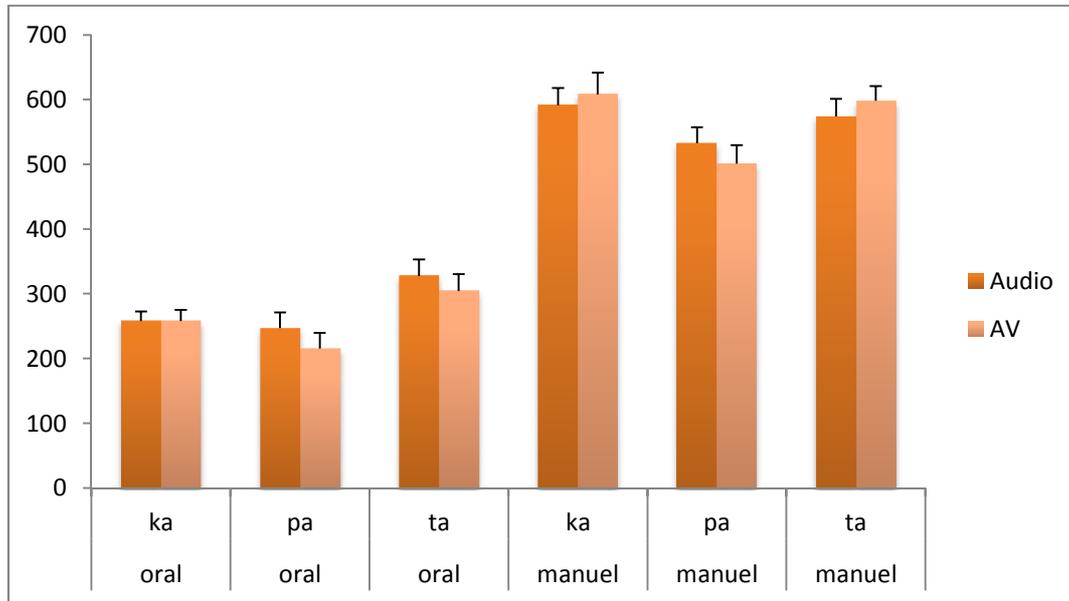


Figure 38 : Moyenne des temps de réaction chez les participants seniors (les barres d'erreurs correspondent aux écart-types).

## 2. Pourcentage de réponse correctes

Comme pour les temps de réaction, aucune différence significative n'a été observée entre les stimuli auditifs et audiovisuels (88% vs. 90% ; voir Figure 39). Par contre, l'effet de la réponse s'est révélé significatif, avec les réponses manuelles plus précises que les réponses orales (98% vs. 81%) ( $F(1,9)=25,5$  ;  $p<.001$ ). De même, l'effet de la syllabe a été significatif ( $f(2,18)=7,9$  ;  $p<.005$ ), avec de plus une interaction réponse x syllabe, où aucune différence n'a été observée entre les trois syllabes lors des réponses manuelles, alors que lors des réponses orales, les participants seniors ont fait moins d'erreurs pour la syllabe /pa/ (97%) par rapport aux deux autres syllabes, et pour la syllabe /ka/ (78%) par rapport à la syllabe /ta/ (67%) ( $F(2,18)=7,9$  ;  $p<.005$ ).

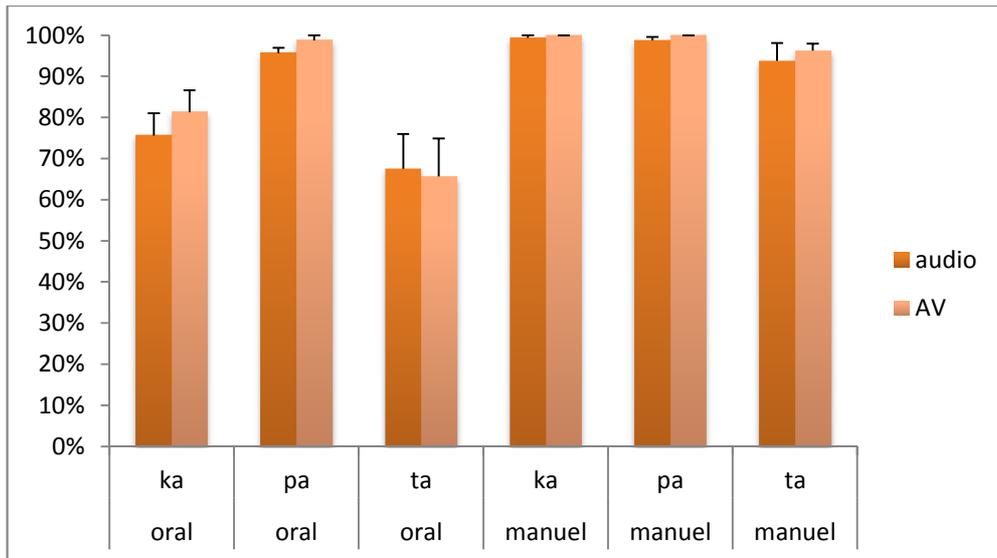


Figure 39 : Moyenne des pourcentages de réussite chez les participants séniors (les barres d'erreurs correspondent aux écart-types)

#### 4. Comparaison avec la population implantée et les populations normo-entendante « junior » dans le bruit et sans bruit

Nous avons dans un second temps comparé les résultats obtenus sur les sujets séniors avec ceux des sujets normo-entendants juniors dans le bruit et sans bruit.

##### 1. Temps de réaction

Comme attendu, l'effet principal du groupe est significatif ( $F(2,36)=9,77$  ;  $p<.001$ ) : aucune différence n'a été observée entre les participants juniors sans bruit et les séniors, ni entre les séniors et les participants juniors dans le bruit. En revanche, les participants juniors sans bruit ont obtenu des temps de réaction plus courts que ceux des participants juniors dans le bruit (voir Figure 40 ; Juniors sans bruit : 351 ms ; Seniors : 418 ms ; Juniors dans le bruit : 484 ms). Les effets principaux de la tâche ( $F(1,36)=266,26$  ;  $p<.001$ ) et de la modalité ( $F(1,36)=12,05$  ;  $p<.002$ ) ont aussi été significatifs. Pour la tâche, les réponses orales étaient plus rapides que les réponses manuelles (en moyenne 273 ms vs. 542 ms). Par rapport à la modalité, les temps de réaction étaient plus courts dans la modalité audiovisuelle par rapport à la modalité auditive (416 ms vs. 400 ms). Une interaction significative entre groupe et modalité ( $F(2,36)=11,62$  ;  $p<.001$ ) montre que l'effet bénéfique de la présentation audiovisuelle est présent chez les participants juniors en condition bruitée (461ms vs. 507ms) alors qu'aucune différence n'a été observée chez les participants séniors (414ms vs. 422ms) ni chez les participants normo-entendants juniors en condition non-bruitée (354ms vs. 349ms). Aucune autre interaction ne s'est révélée significative.

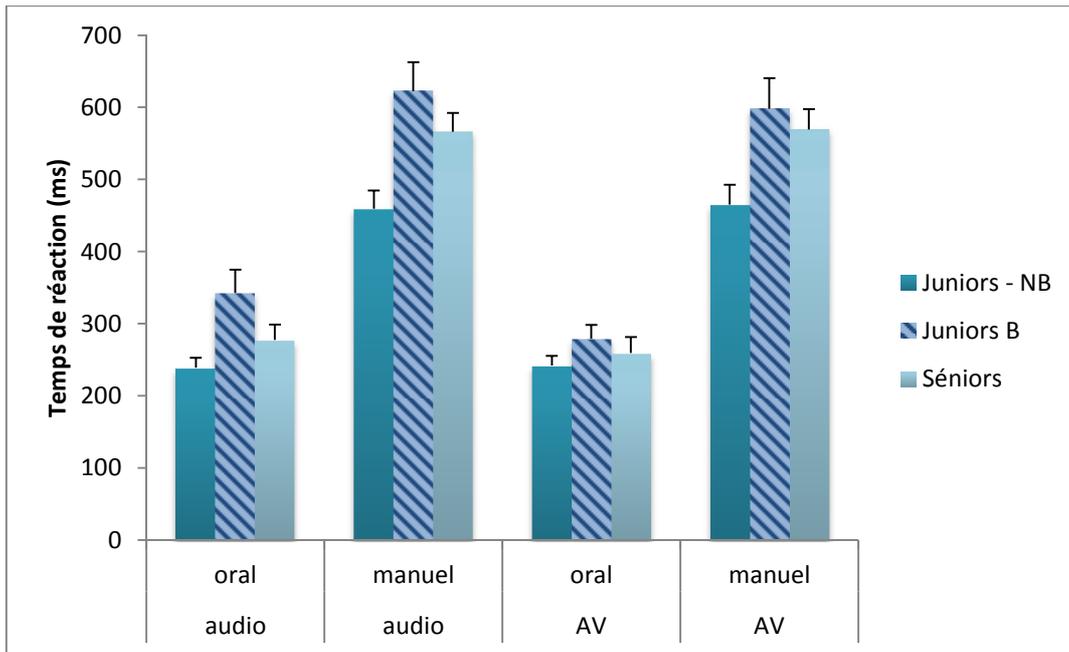


Figure 40: Moyennes des temps de réaction chez les participants juniors en condition bruitée (Juniors – B) et non-bruitée (Juniors - NB), ainsi que chez les participants séniors (Séniors) (en ms, les barres d'erreurs correspondent aux écart-types).

## 2. Pourcentages de réponses correctes

L'effet principal du groupe a été significatif ( $F(2,36)=126,24$  ;  $p<.001$ ) avec une proportion de réponses correctes plus élevée pour les participants juniors en condition non-bruitée (95%) ainsi que pour les participants normo-entendants séniors (90%) par rapport aux participants juniors en condition bruitée (61%). Aucune différence significative n'a été observée entre les participants juniors en condition non-bruitée et les participants séniors (voir Figure 41).

Egalement, l'effet principal de la tâche a été significatif ( $F(1,36)=95,96$  ;  $p<.001$ ) avec une baisse des réponses correctes observée pour les réponses orales par rapport aux réponses manuelles (75% vs. 88%). De même, l'effet principal de la modalité a été significatif ( $F(1,36)=39,84$  ;  $p<.001$ ) avec des réponses plus précises pour les stimuli audiovisuels par rapport aux stimuli uniquement auditifs (84% vs. 79%).

Une interaction 'groupe x tâche' a été obtenue ( $F(2,36)=13,19$  ;  $p<.001$ ) montrant que la différence de précision entre les réponses orales et les réponses manuelles n'est présente que chez les participants séniors (81% vs. 98%) et les participants juniors dans le bruit (50% vs. 71%). De plus, une interaction 'groupe x modalité' ( $F(2,36)=39,71$  ;  $p<.001$ ) montre que l'effet bénéfique de la modalité n'est présent que chez les participants normo-entendants dans le bruit (audiovisuel : 68% ; audio : 53%).

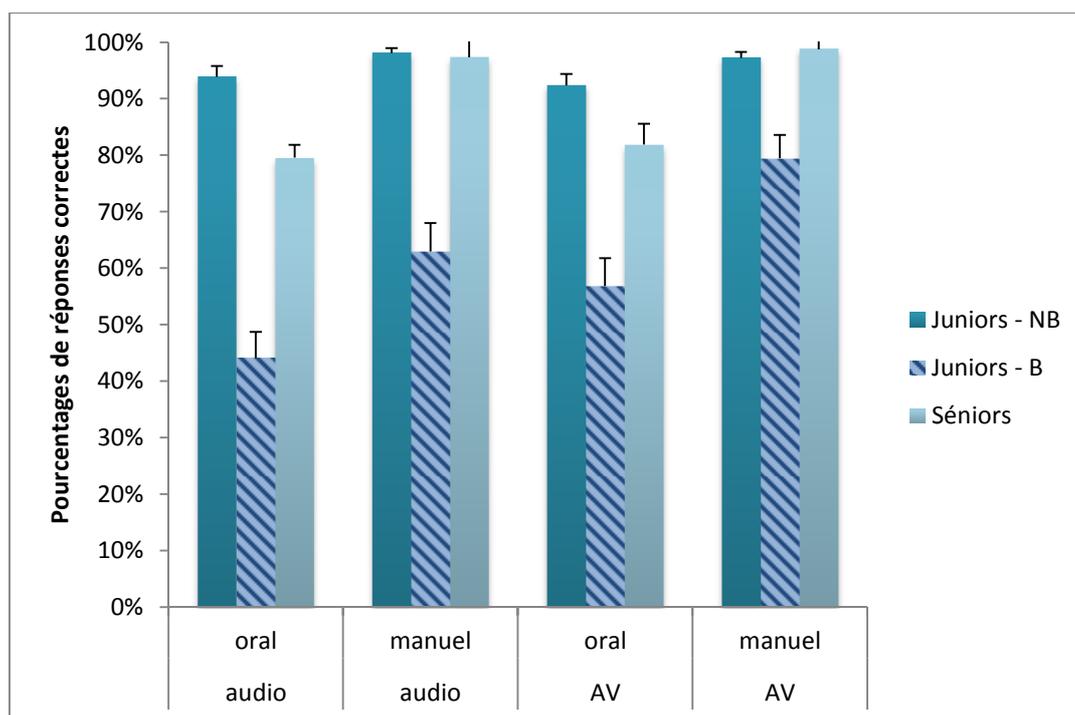


Figure 41 : Moyenne des pourcentages de réponses correctes les participants normo-entendants juniors en condition bruitée (Juniors – B) et non-bruitée (Juniors-NB) et les participants seniors (les barres d'erreurs correspondent aux écart-types).

Les résultats dépendent de la syllabe présentée. En effet une interaction 'groupe x modalité x syllabe' a été obtenue ( $F(4,72)=11,4$  ;  $p<.001$ ), montrant que pour les participants juniors en condition bruitée, l'effet de la modalité n'est présent que pour la syllabe /pa/ (audiovisuel : 92% ; audio : 61%) et la syllabe /ta/ (audiovisuel : 51% ; audio : 39%). Enfin, une interaction 'groupe x tâche x syllabe' ( $F(4,72)=9,65$  ;  $p<.001$ ) montre que chez les participants juniors en condition non-bruitée les syllabes sont toutes parfaitement reconnues ; chez les participants seniors, lors des réponses manuelles, aucune différence n'a été observée entre les trois syllabes alors que lors des réponses orales, la syllabe /pa/ (97%) a été mieux reconnue que les syllabes /ka/ (78%) et /ta/ (66%); et enfin chez les participants normo-entendants en condition bruitée, lors des réponses orales la syllabe /pa/ (72%) est mieux reconnue que les syllabes /ta/ (40%) et /ka/ (39%) tandis que lors des réponses manuelles, les syllabes /pa/ (82%) et /ka/ (82%) sont mieux reconnues que la syllabe /ta/ (50%) (voir Figure 17, Figure 15 et Figure 39).

## 5. Discussion

En résumé, les participants seniors présentent des performances légèrement dégradées par rapport à celles des participants juniors dans les mêmes conditions (c'est-à-dire sans bruit), soit un peu inférieures en termes de reconnaissance et sensiblement ralenties en termes de réponse – mais des performances qui restent nettement supérieures à celles des juniors en condition bruitée.

Il est cependant important de noter que les participants seniors ont présenté des patrons de réponses extrêmement similaires à ceux des participants juniors de

l'expérience non-bruitée. En effet, les seniors n'ont pas présenté de bénéfice de la modalité visuelle puisqu'aucune différence de temps de réaction ou de taux de précision n'a été observée entre les réponses aux stimuli auditifs et celles aux stimuli audiovisuels. Cette absence de bénéfice de la modalité visuelle peut être mise en lien avec les résultats obtenus lors des études précédentes portant sur la perception audiovisuelle chez les seniors. En effet, la majorité des études confirment les capacités d'intégration audiovisuelle chez les seniors (Cienkowski et collègues 2002, Sommers et collègues 2005), particulièrement dans les niveaux basiques de perception de parole (Stevenson et collègues 2014), même si les seniors semblent s'aider de la modalité visuelle principalement lors d'une écoute très perturbée (Stevenson et collègues 2014), ce qui pourrait expliquer l'absence de bénéfice de l'addition de la modalité visuelle chez nos participants seniors.

Enfin, de manière intéressante, comme chez les participants juniors en condition bruitée, nous avons retrouvé une baisse importante de rapidité de réponse allée à plus de précision lors des réponses manuelles par rapport aux réponses orales chez nos participants seniors. Cette différence semble donc indiquer que chez les participants seniors, la boucle perceptivo-motrice est parfaitement effective, comme cela a été suggéré par Liu et collègues (2010, 2011).

#### **4. Expérience d'idiosyncrasies**

##### **1. Principes et hypothèses**

La seconde expérience réalisée chez les participants seniors consistait en l'expérience dite d'idiosyncrasies, visant à corréliser l'organisation perceptivo-motrice de voyelles du Français perçues et produites. Comme chez les participants juniors, l'expérience consistait en une tâche de catégorisation de stimuli auditifs entre les voyelles /a ε e i/ et une tâche de production des voyelles orales du français.

Rappelons que chez les participants normo-entendants juniors, contrairement à nos hypothèses de départ, nous n'avons pas observé de corrélation individuelle entre la production et la perception des voyelles /e/ et /ε/, ce en raison d'une quasi absence de variation perceptive dans cette tâche. Nous avons interprété ce fait en lien avec le nombre réduit de voyelles utilisées lors de la tâche de perception, qui pourrait avoir conduit les participants à catégoriser ces voyelles tous de la même manière, selon une répartition « optimale » dans l'espace perceptif F1 en Barkes. En revanche, malgré l'absence d'idiosyncrasies en perception, nous avons observé des idiosyncrasies en production, avec des variations autour de positions moyennes respectant également un principe d'optimalité perceptive entre les quatre voyelles à produire. La double répartition optimale en perception comme en production a comme conséquence une corrélation très forte entre les valeurs acoustiques produites et perçues des voyelles. En premier lieu, nous avons proposé d'expliquer cette équidistance à la lumière de la théorie de la dispersion maximale de Lindblom et collègues (Liljencrants et Lindblom, 1972 ; Lindblom, 1986), qui propose une organisation optimale des phonèmes des langues du

monde afin que ceux-ci soient suffisamment distincts pour être différenciés. Subsidiairement, nous avons proposé que l'équidistance entre les quatre voyelles en perception et en production puisse être un argument en faveur de cibles de gabarit perceptivo-moteur, puisque les participants, lors de la tâche de production, vont varier autour de gabarits communs, ces gabarits correspondant à ceux représentés lors de la tâche de perception.

Pour les participants séniors, nous nous attendions à observer une plus grande variabilité de production par rapport aux participants juniors, comme cela avait été obtenu lors d'études précédentes (Torre et Barrow, 2009). Egaleme nt, en raison des possibles réductions de capacités de perception (Ohde et Abou-Khalil, 2001 ; Gelfand et collègues, 1985) nous supposons que les participants séniors seraient peut-être plus variables dans la catégorisation des stimuli. Enfin, plus généralement, nous nous attendions à retrouver la corrélation globale entre les valeurs acoustiques moyennes produites et perçues des voyelles.

## **2. Méthodologie**

### **1. Stimuli**

Les stimuli utilisés pour l'expérience d'idiosyncrasies chez la population sénior ont été les mêmes que ceux utilisés lors de l'expérience chez la population normo-entendante junior (voir partie I.4.2.2.)

### **2. Procédure expérimentale**

Les participants séniors ont réalisé l'expérience dans les mêmes conditions et selon les mêmes conditions expérimentales que les participants normo-entendants juniors (voir partie I.2. Population normo-entendante « junior » pour la procédure générale aux trois expériences, et partie I.4.2.3. pour la procédure spécifique à l'expérience d'idiosyncrasies).

### **3. Analyses**

Les analyses réalisées sur les réponses de la population séniors ont été identiques à celles décrites dans la partie I.4.2.4.

## **3. Résultats population séniors**

### **1. Tâche de production**

Lors de la tâche de production, les moyennes des valeurs de F1 pour /i/, /e/, /ɛ/ et /a/ en Barks ont été respectivement de 2,91 (avec une étendue de 2,4 à 3,65), 3,81 (avec une étendue de 3,2 à 4,94), 5,17 (avec une étendue de 4,42 à 5,98) et de 7,58 (avec une étendue de 5,96 à 8,88).

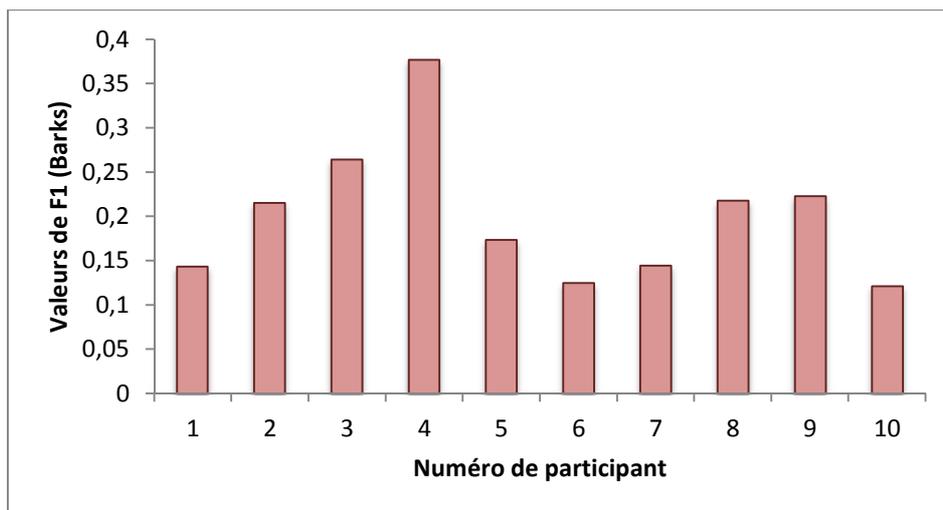


Figure 42 : Valeurs normalisées de F1 (en Barks) pour /e/ produites par chaque participant sénior.

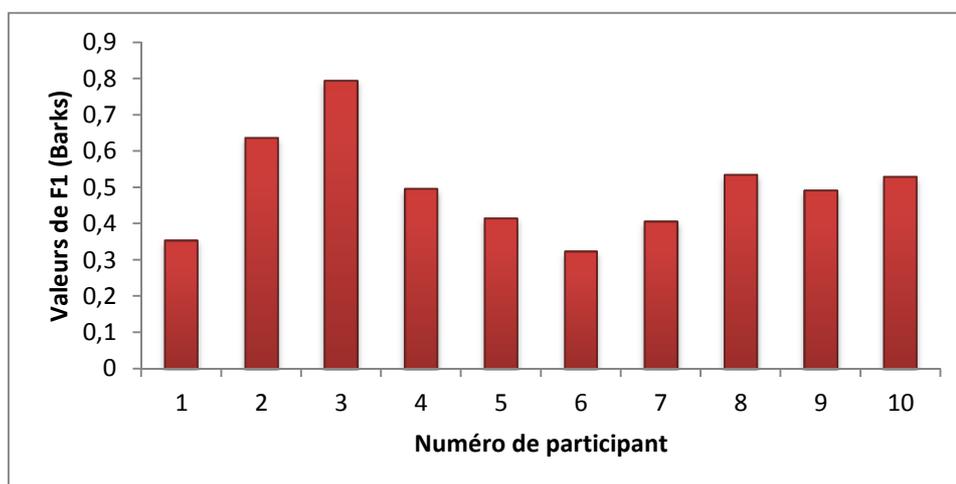


Figure 43 : Valeurs normalisées de F1 (Barks) pour /ε/ produites par chaque participant sénior.

Les idiosyncrasies ont été faibles pour /e/ avec des valeurs de distance normalisée à partir de /i/ allant de 0,12 à 0,38 Barks mais avec un écart-type réduit à 0,07 (voir Figure 42). Par contre, les idiosyncrasies pour /ε/ ont été plus élevées, avec des valeurs de distance normalisées à partir de /i/ allant de 0,32 à 0,79 Barks, avec un écart-type de 0,14 (voir Figure 43).

## 2. Tâche de perception

Lors de la tâche de perception, les moyennes des valeurs de F1 pour /i/, /e/, /ε/ et /a/ ont été respectivement de 2,72 (avec une étendue de 2,59 à 2,8), 4,06 (avec une étendue allant de 3,67 à 5,06), 5,31 (avec une étendue allant de 4,94 à 5,59) et 6,79 (avec une étendue allant de 6,71 à 6,87).

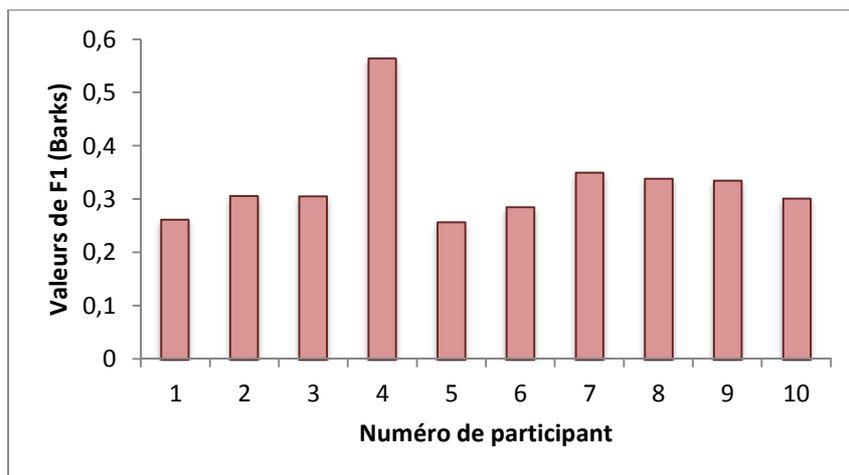


Figure 44 : Valeurs normalisées de F1 lors de la perception de la voyelle /e/, pour chaque participant séniors.

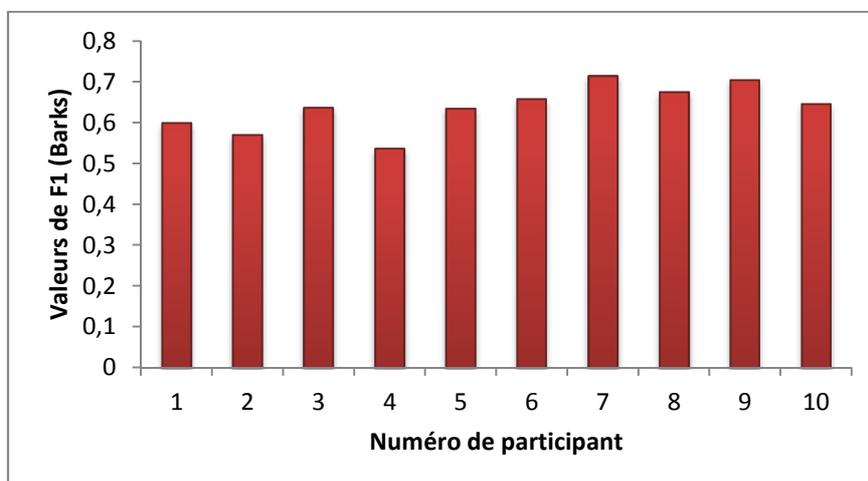


Figure 45 : Valeurs normalisées de F1 lors de la perception de la voyelle /ε/, pour chaque participant séniors.

La variabilité étant faible en perception, de très faibles idiosyncrasies ont été observées entre les participants. Les valeurs de distance normalisée pour la voyelle /e/ à partir de /i/ étaient de 0,25 à 0,56 Barks avec un écart-type de 0,08 (voir Figure 44). Pour la voyelle /ε/ les valeurs de distance normalisées à partir de /i/ étaient de 0,54 à 0,71 Barks, avec un écart-type faible de 0,05 (voir Figure 45). En étudiant séparément chaque participant, on constate qu'un des dix participants (senior4) obtient des valeurs normalisées quasiment identiques pour la voyelle /e/ (0,56) et la voyelle /ε/ (0,54). La distance entre ces deux voyelles a été calculée pour les deux tâches chez ce participant, ce qui nous a permis de démontrer que ce participant ne différencie pas les voyelles /e/ et /ε/, ni en production (distance /e/-/ε/ égale à 0,12 Barks) ni en perception (distance /e/-/ε/ égale à 0,03 Barks), nous avons donc choisi de l'exclure des analyses suivantes.

### 3. Corrélation perception/production

Etant donnée l'absence de différenciation entre les voyelles /e/ et /ɛ/ du participant « senior 4 », nous avons choisi de l'exclure de ces analyses. Considérant l'absence presque totale d'idiosyncrasies en perception chez les neuf autres participants, nous n'avons pas obtenu de corrélation entre production et perception, ni pour la voyelle /e/ ( $r=.36$ ,  $t=0,94$ ,  $p > .05$ , voir Figure 46) ni pour la voyelle /ɛ/ ( $r=.24$ ,  $t=0,7$ ,  $p > .05$ , voir Figure 47).

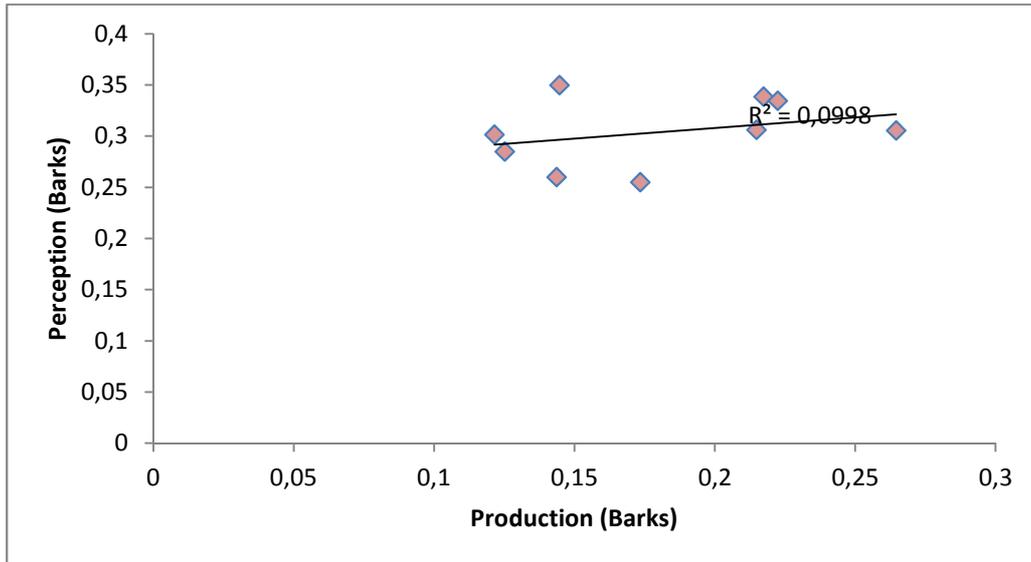


Figure 46 : Corrélation perception-production pour la voyelle /e/ chez les participants seniors.

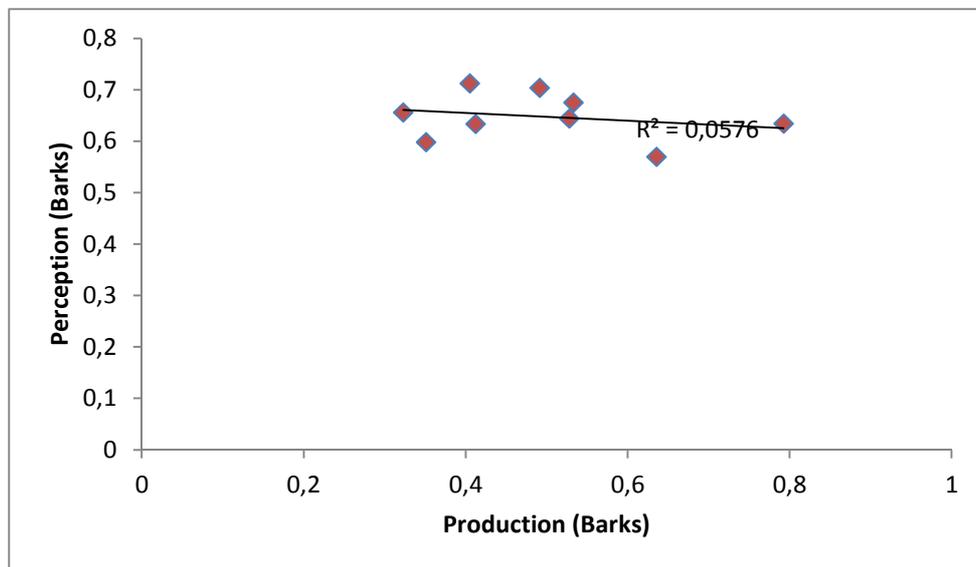


Figure 47 : Corrélation perception-production pour la voyelle /ɛ/ chez les participants seniors.

Par contre, à partir des valeurs moyennes des quatre voyelles, nous avons obtenu une corrélation forte des valeurs acoustiques entre les voyelles produites et perçues, avec une pente moyenne pour tous les participants de 0,87, allant de 0,63 à 1,17,  $r=.9$  (voir

Figure 48). Néanmoins, l'alignement et équidistance quasi parfaits entre valeurs moyennes, que nous avons obtenus pour la population d'adultes juniors, sont moins clairs ici, essentiellement à cause de valeurs en production pour /e/ et /ɛ/ décalées vers les valeurs basses : nous allons y revenir dans la section suivante.

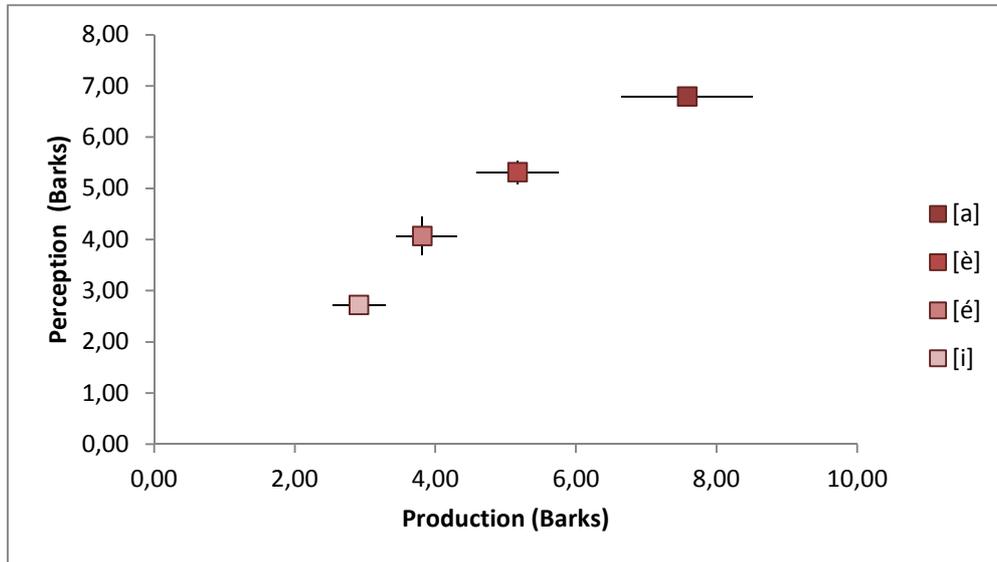


Figure 48 : Corrélation entre les valeurs acoustiques produites et perçues par les participants seniors (les barres d'erreurs correspondent aux écart-types).

Comme chez les participants normo-entendants juniors (voir Partie 1. 4.3 4., nous avons finalement mesuré le nombre de stimuli par catégorie. En moyenne, 15% des stimuli ont été catégorisés comme /a/, avec une variation entre 12 et 17% ; 38% des stimuli ont été catégorisés comme /è/ avec une variation entre 26 et 51% ; 19% des stimuli ont été catégorisés comme /é/ avec une variation de 8 à 33% ; et 28% des stimuli ont été catégorisés comme /i/ avec une variation de 22 à 31% (voir Figure 49). Ainsi comme illustré par la Figure 50, la proportion relative de stimuli catégorisés comme /è/ et /é/ est très variable selon les participants, comme chez les participants normo-entendants juniors, ce qui ne peut expliquer la stabilité des valeurs des catégories perceptives.

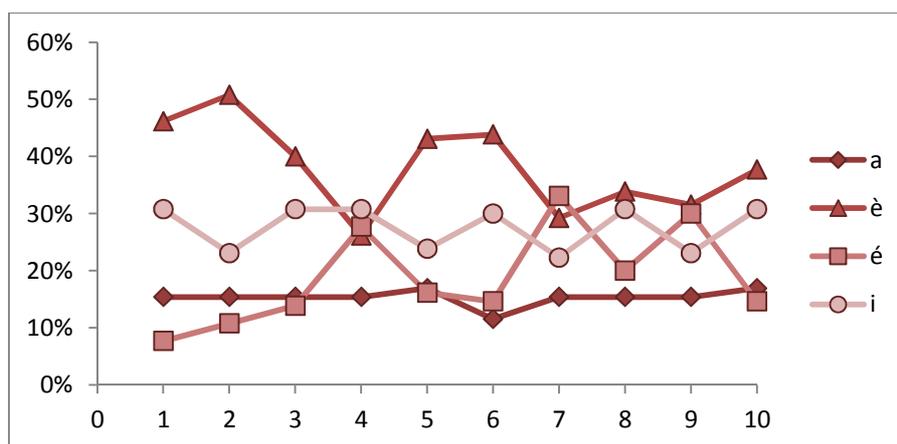


Figure 49 : Pourcentages de stimuli par catégorie, chez chaque participant normo-entendant séniors

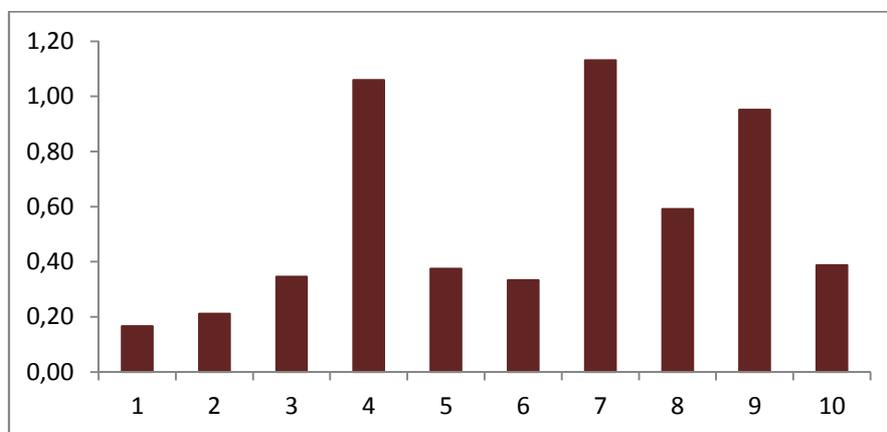


Figure 50 : Rapport entre les pourcentages relatifs de stimuli catégorisés comme les voyelles /e/ et /ɛ/ par participant normo-entendant séniors.

#### 4. Comparaison entre la population séniors et la population junior

Les valeurs comparées des données de perception et de production pour les séniors et les juniors sont présentées dans le Tableau 8

Tableau 8 : Valeurs moyennes de F1 pour chaque tâche et chaque groupe (en Barks).

		Juniors	Séniors
Production	/a/	7,32	7,58
	/ɛ/	5,95	5,17
	/e/	4,43	3,81
	/i/	3,05	2,91
	/e/ normalisé	.32	.20
	/ɛ/ normalisé	.67	.50
Perception	/a/	6,80	6,79
	/ɛ/	5,52	5,31
	/e/	4,21	4,06
	/i/	2,76	2,72
	/e/ normalisé	.36	.33
	/ɛ/ normalisé	.68	.64

Afin de comparer la population seniors avec la population junior, nous avons réalisé une ANOVA à mesure répétée avec la variable groupe (Seniors, Juniors) comme variable inter-participant, et les variables tâche (Perception vs. Production) et voyelles (avec les valeurs des moyennes de F1 perçues et produites pour les voyelles /i e ε a/, ainsi que les valeurs normalisées calculées pour les voyelles /e/ et /ε/), comme variables intra-participants. Lorsque cela a été possible, nous avons réalisé des tests post-hoc de Bonferroni. Du fait des comparaisons multiples lors de cette seconde série d'analyses, nous avons fixé le seuil de significativité à  $p = .025$ . Egalement, nous avons vérifié la sphéricité des données grâce au test de Mauchly, et lorsque nécessaire, nous avons effectué des corrections de Greenhouse-Geiser (Greenhouse et Geiser, 1959).

L'effet du groupe s'est révélé significatif ( $F(1,34)=7,93$  ;  $p<.01$ ), avec des valeurs moyennes de F1 (perçues et produites, toutes voyelles confondues) plus basses pour les participants seniors par rapport aux participants juniors (seniors : 3,33 Barks ; juniors : 3,5 Barks). L'effet de la tâche est également significatif ( $F(1,34)=6,26$  ;  $p<.02$ ) avec des valeurs moyennes de F1 produites plus élevées que les valeurs de F1 perçues (3,55 vs. 3,37). Enfin, l'effet principal de la voyelle est significatif ( $F(5,170)=4491,66$  ;  $p<.001$ ), toutes les moyennes obtenues pour les voyelles étant significativement différentes entre elles (en Barks, /i/=2,88 ; /e/=4,21 ; /ε/=5,60 ; /a/=7,09 ; valeurs normalisées moyennes : /e/= .32 et /ε/= .65). Une interaction groupe x voyelle ( $F(5,170)=7,92$  ;  $p<.001$ ) révèle qu'en moyenne les participants seniors ont obtenu des valeurs de F1 plus basses pour les voyelles /e/ et /ε/. Egalement, une interaction entre les trois facteurs ('groupe x tâche x voyelles') a été observée ( $F(5,170)=4,38$  ;  $p<.001$ ), révélant que les différences principales apparaissent dans la tâche de production, avec, chez les participants seniors, des voyelles /e/ et /ε/ produites avec des valeurs de F1 nettement plus basses que pour les participants juniors. En perception au contraire, les différences entre les deux groupes sont extrêmement faibles. Aucune autre interaction ne s'est révélée significative.

Nous avons également étudié la variabilité des participants lors des deux tâches. Pour ce faire, nous avons réalisé deux ANOVAs : une première ANOVA avec les écart-types de chaque participant lors de la production des quatre voyelles, puis une seconde ANOVA avec les écart-types mesurés lors de la tâche de perception. Plus précisément, pour chaque participant et pour chaque catégorie vocalique dans l'expérience de perception, nous avons calculé l'écart type de la distribution des réponses perceptives. Des écarts types très élevés sont le signe de distributions très plates, c'est-à-dire de capacité de discrimination très faible entre les catégories.

Aucune différence n'apparaît entre les deux groupes de participants en regard de la variabilité de production des voyelles ( $F(1,34)=0,24$  ;  $p=.62$ ). En revanche, nous avons obtenu un effet de la voyelle ( $F(3,102)=11,65$  ;  $p<.001$ ), où les écart-types mesurés pour les voyelles /a/ et /ε/ ont été significativement plus élevés par rapport aux voyelles /e/ et /i/ (Valeurs en Barks : /a/ : .42 ; /ε/ : .40 ; /e/ : .23 ; /i/ : .23).

Egalement, concernant la variabilité lors de la catégorisation des stimuli, aucune différence n'a été observée entre les deux groupes ( $F(1,34)=3,08$  ;  $p=.09$ ). Par contre, l'effet de la voyelle s'est révélé significatif ( $F(3,102)=82,67$  ;  $p<.001$ ). Enfin, de manière plus intéressante, une interaction groupe x voyelles ( $F(3,102)=3,9$  ;  $p<.02$ ) a été observée, montrant que les participants seniors ont présenté moins de variabilité pour la catégorisation perceptive de /e/ que les participants juniors, alors qu'aucune différence n'a été observée pour les voyelles /a/, /i/ et /ɛ/ (Valeurs en Barks : Juniors : /a/=.21 /ɛ/=.60 /e/=.46 /i/=.44 ; Seniors : /a/=.22 /ɛ/=.65 /e/=.35 /i/=.40).

Enfin, afin de comparer la corrélation des valeurs acoustiques entre les voyelles produites et perçues, nous avons réalisé une ANOVA sur les valeurs de pente entre les deux groupes de participants. Aucune différence significative n'a été observée entre les participants seniors et les participants implantés ( $F(1,34)=1,29$  ;  $p=.26$ ) (seniors :  $r=.93$  ; Juniors :  $r=.94$ ).

## 5. Discussion

En résumé, lors de la tâche de production, une légère baisse du F1 a été observée chez les seniors pour les voyelles /e/ et /ɛ/. En revanche, contrairement aux résultats des études précédentes menées sur la production de voyelles (Torre et Barlow, 2009, Fernand-Longpré et collègues, 2013), aucune différence de variabilité de production n'a été observée entre les deux groupes de participants. Dans la tâche de perception, les participants seniors ont présenté moins de variabilité de catégorisation intra-participant que les participants juniors pour la voyelle /e/.

En dehors de ces légères différences entre les deux groupes, les participants seniors ont présenté des patrons de résultats similaires à ceux des participants juniors. En effet, lors de la tâche de production, de la variabilité inter-participant a été obtenue pour chaque voyelle, entraînant des idiosyncrasies pour les valeurs normalisées des voyelles /e/ et /ɛ/. En revanche, comme chez les participants juniors, les idiosyncrasies ont été très faibles en perception chez les seniors, à cause de la quasi-absence de variabilité inter-participants lors de cette tâche. En conséquence, aucune corrélation entre les deux tâches n'a été observée ni pour la voyelle /e/ ni pour la voyelle /ɛ/, comme chez les participants juniors.

Si l'on retrouve la quasi-optimalité des distributions des valeurs moyennes en perception chez les seniors, avec des valeurs normalisées pour /e/ et /ɛ/ proches de 1/3 (0.33) et 2/3 (0.67) comme chez les juniors, les choses sont un peu différentes en production. Les positions de /i/ et /a/ sont très proches pour les deux populations, ce qui est peu surprenant car ces valeurs sont des valeurs extrêmes, déterminées essentiellement par la taille du conduit vocal qui ne varie pas avec l'âge. Les positions de /e/ et /ɛ/ au contraire sont nettement abaissées chez les seniors, et du coup la distribution des positions moyennes en production s'écarte de l'optimalité, avec des valeurs normalisées à 0.20 et 0.50. Globalement ces résultats sont compatibles avec ceux de

Torre et Barlow (2009), qui n'obtiennent pas de modification significative de F1 avec l'âge pour /i/, mais des diminutions significatives pour les autres voyelles – non inclus /a/, ne faisant pas partie du système vocalique de l'anglais américain.

Enfin, de manière générale, la corrélation des valeurs acoustiques globales produites et perçues a été retrouvée chez les participants seniors, montrant que les participants seniors conservent des gabarits perceptivo-moteurs communs, autour desquels ils varient lors de la tâche de production – même si les corrélations sont moins fortes, étant données les modifications de valeurs moyennes.

## **5. Expérience de convergence et imitation**

### **1. Principes et hypothèses**

Enfin, la dernière expérience réalisée chez les participants seniors était l'expérience de convergence et d'imitation. Comme pour les participants juniors, l'expérience consistait en trois tâches de production des voyelles /e ø o/ : une première tâche de production selon cible orthographique visuelle, permettant de mesurer la moyenne de F0 du participant, suivie de deux tâches de production selon une cible auditive variant autour de la fréquence fondamentale moyenne du participant, soit sans qu'aucune consigne d'imitation de la voix perçue (tâche de convergence) ne soit donnée au participant, soit avec pour consigne d'imiter volontairement la voix perçue (tâche d'imitation).

Pour rappel, chez les participants juniors, comme lors des études précédentes (Garnier et collègues 2014 ; Sato et collègues 2014) nous avons observé une imitation quasi-parfaite de F0 lors de la tâche d'imitation, et lors de la tâche de convergence, les participants ont convergé de manière automatique vers la cible présentée, cette convergence se révélant moins élevée par rapport à celle de la tâche d'imitation. Les capacités d'imitation volontaire et involontaire d'une cible auditive montrent que les participants, lors de la production de parole, vont adapter leurs commandes motrices à leur environnement, à partir de l'input sensoriel qu'ils reçoivent, ce qui met en jeu des interactions sensori-motrices.

Chez les participants seniors, nous nous attendions à observer une plus grande variabilité de production. Considérant, les capacités d'imitation et de convergence vers la cible auditive, les résultats déjà mentionnés de Liu et al. (2010) montrant une plus grande correction de l'altération du feedback pouvait conduire à prédire des capacités de convergence maintenues voire renforcées – mais on pouvait aussi interpréter ces résultats comme indiquant un ancrage renforcé des participants seniors sur leurs valeurs cibles (sur-apprises au cours de la vie) ce qui pouvait aussi suggérer au contraire un affaiblissement des capacités de convergence et imitation chez ces participants.

## **2. Méthodologie**

### **1. Stimuli**

Les stimuli utilisés pour l'expérience de convergence et imitation chez la population senior ont été les mêmes que ceux utilisés lors de l'expérience chez la population « junior » (voir partie I.5.2.2.)

### **2. Procédure expérimentale**

Les participants seniors ont réalisé l'expérience dans les mêmes conditions et selon les mêmes procédures expérimentales que les participants normo-entendants « juniors » (voir partie I.2. Population normo-entendante « junior » pour la procédure générale aux trois expériences, et partie I.5.2.3. pour la procédure spécifique à l'expérience de convergence et imitation)

### **3. Analyses**

Les analyses réalisées sur les réponses de la population senior ont été identiques à celles décrites dans la partie I.5.2.4.

### **3. Résultats population seniors**

Des changements imitatifs ont été observés dans les deux tâches, même s'ils étaient plus importants dans la tâche d'imitation volontaire. Les coefficients de pente étaient significativement différents de zéro à la fois dans la tâche dite de « convergence » ( $t(9)=3,52$ ;  $p<01$ ) et dans la tâche d'imitation volontaire ( $t(9)=8,5$ ;  $p<0.001$ ). Egalement, les coefficients de pente étaient plus importants dans la tâche d'imitation volontaire (0,75) par rapport à la tâche dite de « convergence » (0,33) ( $t(9)=4,97$ ;  $p<.001$ ). De même, les coefficients de corrélation étaient significativement différents de zéro à la fois dans la tâche de convergence ( $t(9)=5,96$ ;  $p<.001$ ) et dans celle d'imitation ( $t(9)=11,78$  ;  $p<.001$ ). Ces coefficients de corrélation étaient également plus élevés dans la tâche d'imitation ( $r=0,84$ ) par rapport à ceux de la tâche de convergence ( $r=0,54$ ) ( $t(9)=4,12$  ;  $p<.001$ ). La corrélation entre la tâche de convergence et la tâche d'imitation apparaît élevée mais n'est en réalité pas significative eu égard au faible nombre de sujets. ( $r=.59$   $F(1,8)=4,31$  ;  $p=.07$ ) (Voir Figure 51, Figure 52, Figure 53).

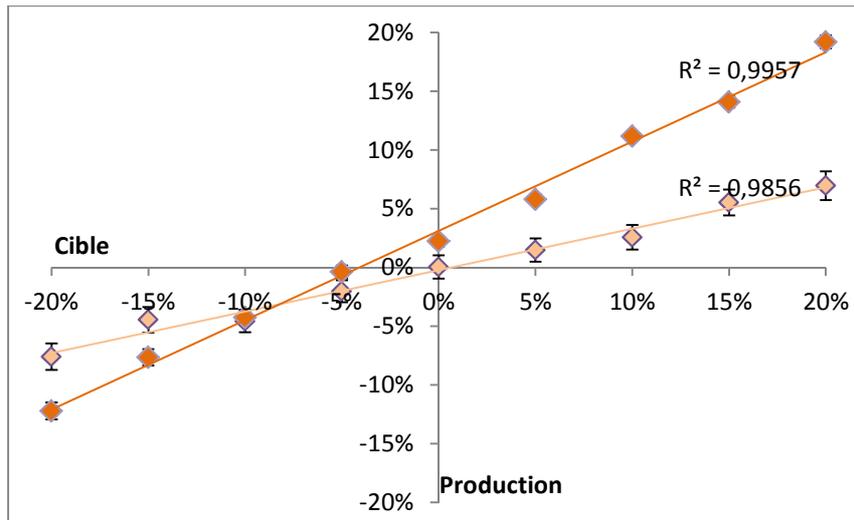


Figure 51 : Moyenne de pourcentage de changement lors des tâches de convergence et d'imitation chez les participants seniors (les barres d'erreurs correspondent aux écart-types).

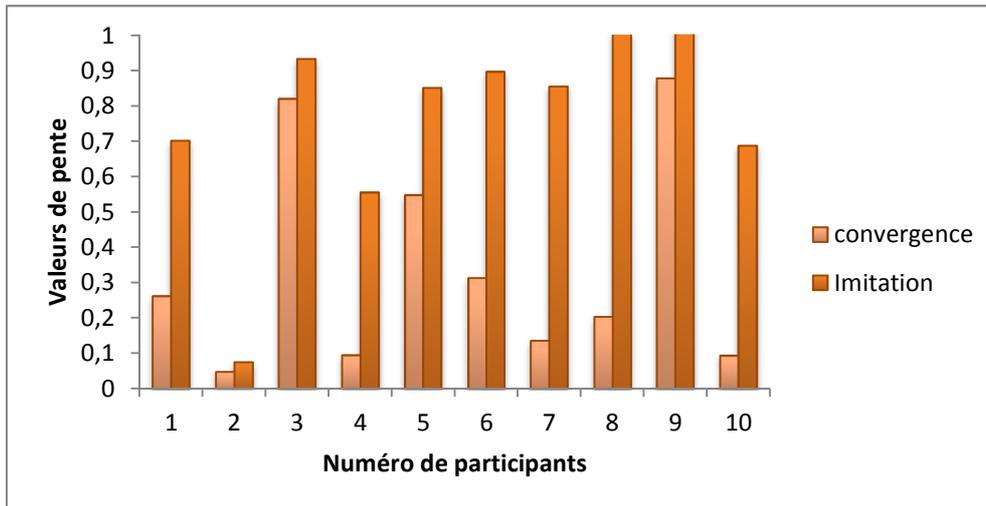


Figure 52 : Valeurs de pente par participant senior, dans les deux tâches.

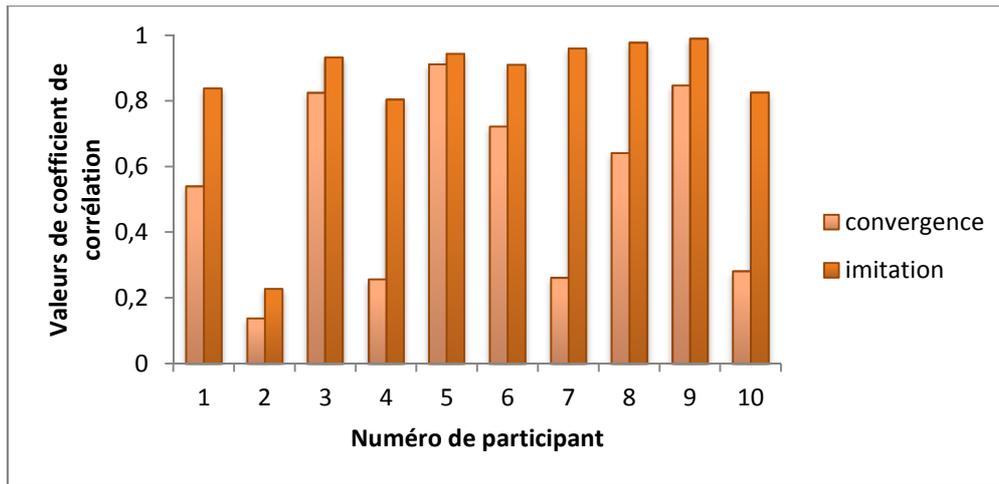


Figure 53 : Valeur de coefficient de corrélation par participant sénior, dans les deux tâches.

Enfin comme chez les participants normo-entendants juniors, nous avons corrélié les valeurs de pente et les résultats obtenus lors du test de discrimination de fréquence JND réalisé, afin de vérifier l'influence de la discrimination de fréquence sur les capacités d'imitation de fréquence fondamentale. Les résultats obtenus pour le test JND par les participants seniors étaient en moyenne de 0,94 Hz (allant de 0,43 à 1,8 Hz). Aucune corrélation n'a été obtenue, ni par rapport aux valeurs de pente de la tâche de convergence ( $r^2 = 0,03$ ), ni par rapport à celles de la tâche d'imitation ( $r^2 = 0,009$ ) (voir Figure 55).

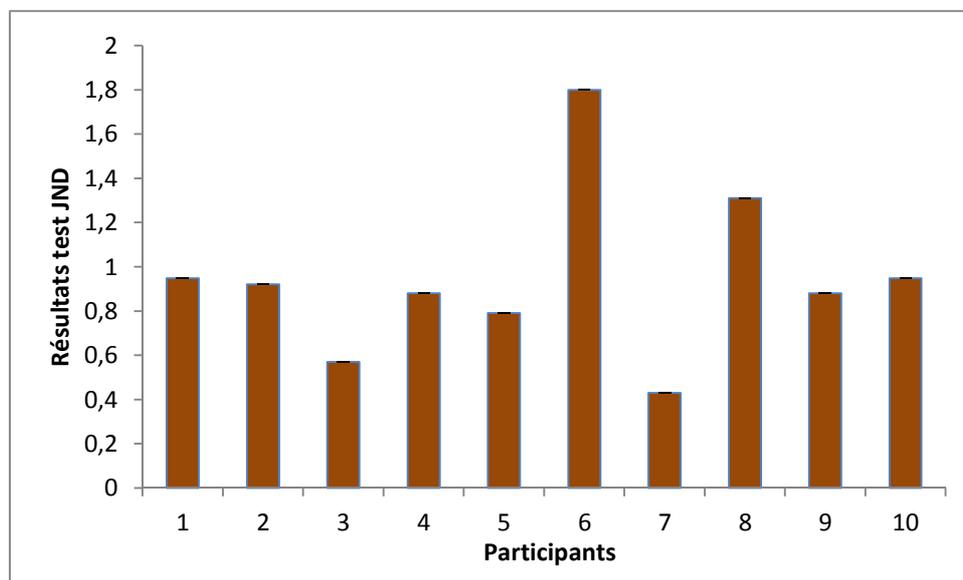


Figure 54 : Résultats obtenus lors du test JND par les participants seniors

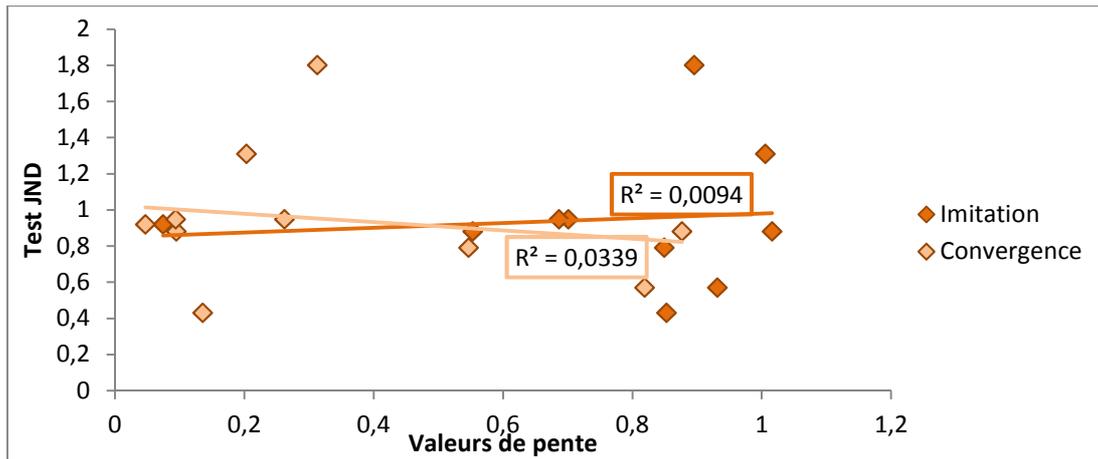


Figure 55 : Corrélacion entre les résultats du test de discrimination de fréquence et les valeurs de pente lors des tâches de convergence et d'imitation, chez les participants normo-entendants séniors

#### 4. Comparaison avec la population junior

Nous présentons, sur la Figure 56, les valeurs de pentes comparées dans les deux tâches et pour les deux groupes (juniors et séniors). Afin de comparer les résultats de la population séniors avec ceux de la population juniors, nous avons réalisé une ANOVA, avec la variable groupe (Juniors vs. Séniors) comme facteur inter-participants et la variable tâche (convergence vs. imitation) comme facteur intra-participants. Du fait des comparaisons multiples lors de cette seconde série d'analyses, nous avons fixé le seuil de significativité à  $p = .025$ . Egalement, nous avons vérifié la sphéricité des données grâce au test de Mauchly, et lorsque nécessaire, nous avons effectué des corrections de Greenhouse-Geiser (Greenhouse et Geiser, 1959).

Nous avons obtenu un effet de la tâche ( $F(1,23)=58,47$  ;  $p<.001$ ) où lors de la tâche de convergence (.40), les coefficients de pentes ont été significativement plus élevés par rapport à la tâche d'imitation (.82). En revanche, l'effet du groupe n'a pas été significatif ( $F(1,23)=1,65$  ;  $p=.21$ ) et aucune interaction significative n'a été observée entre les variables groupe et tâche ( $F(1,23)=0,00$  ;  $p=.97$ ). On observe globalement des patrons proches, avec des pentes plus fortes en imitation qu'en convergence. Il faut noter que les participants séniors ont une erreur standard plus élevée que les participants juniors lors de la tâche de convergence (séniors : 0,1 ; Juniors : 0,07) ainsi que lors de la tâche d'imitation (Séniors : 0,09 ; Juniors : 0,02).

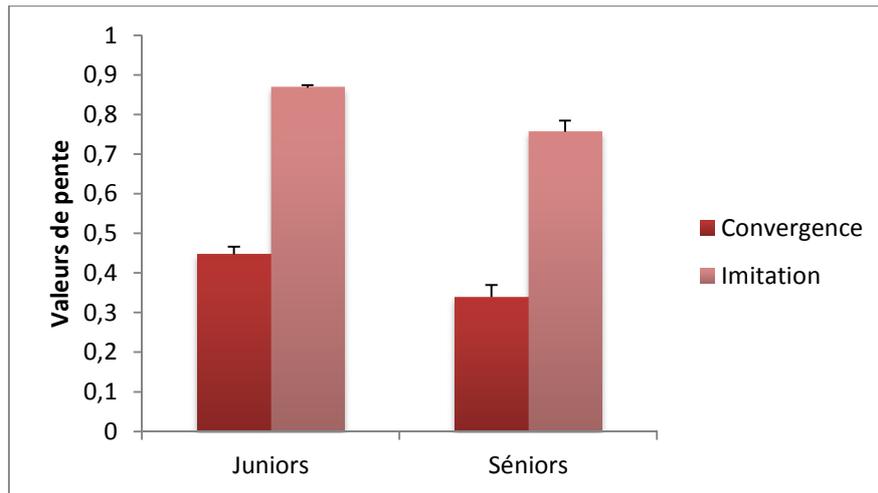


Figure 56 : Coefficients de pente lors des tâches de convergence et d'imitation chez les participants juniors et les participants seniors (les barres d'erreurs correspondent aux écart-types).

## 5. Discussion

Les participants seniors ont présenté des résultats similaires à ceux des participants juniors. Les participants seniors ont donc été capables d'imiter la cible auditive, lors de la tâche d'imitation volontaire ; de même, ils ont convergé vers la cible de manière plus ou moins inconsciente lors de la tâche de convergence. Egalement, comme chez les participants juniors, aucune corrélation n'a été obtenue entre les deux tâches, c'est-à-dire que les participants ayant le mieux imité la cible lors de la tâche d'imitation n'ont pas forcément obtenu les meilleurs taux de convergence. En revanche, la variabilité inter-participant s'est révélée plus élevée chez les participants seniors lors des deux tâches. Ces résultats sont en adéquation avec les résultats d'études précédentes menées sur la production de la fréquence fondamentale chez les seniors, qui montrent que, bien que leur voix soit moins stable que celle des jeunes adultes, les seniors sont capables de moduler leur fréquence fondamentale (Fisher et Linville, 1985 ; Linville et Korabic, 1996 ; Lortie et collègues, 2015). En résumé, les capacités d'imitation consciente et inconsciente des seniors nous permettent de démontrer l'influence de l'environnement externe de parole sur leur production, ce qui illustre donc le maintien de l'intégration sensori-motrice lors de la production de parole chez ce groupe de personnes seniors.

## 6. Conclusion

Dans ce chapitre, nous avons proposé de tester les interactions sensori-motrices au sein d'un groupe de participants normo-entendants seniors et nous avons comparé les résultats obtenus par cette population avec ceux de notre groupe de participants normo-entendants juniors.

Le premier paradigme expérimental proposé était celui de close-shadowing, composé d'une tâche de catégorisation manuelle et d'une tâche de répétition rapide de stimuli de parole auditifs ou audiovisuels, où nous avons mesuré la rapidité de réponse et la proportion de réponses correctes. Nos hypothèses premières étaient que les

participants séniors présenteraient des réponses dégradées par rapport à celles des participants juniors, soit au niveau de la précision de réponse soit au niveau de la rapidité de réponse. Nous supposons également que ces participants tireraient bénéfice de la présentation audiovisuelle par rapport à la présentation auditive. Par contre, notre interrogation principale concernait le maintien ou non des liens perceptivo-moteurs, illustrés, dans ce paradigme, par des réponses plus rapides mais moins précises lors de la tâche de répétition rapide. Conformément à nos hypothèses, nous avons effectivement obtenu des réponses dégradées par rapport aux participants juniors, au niveau de la précision de réponse, d'une part, mais plus particulièrement au niveau de la rapidité de réponse. Par contre, comme chez les participants juniors répondant aux stimuli non-bruités, aucun avantage de la modalité audiovisuelle n'a été observé. De manière plus importante, les participants séniors ont présenté des réponses orales plus rapides mais moins précises que les réponses manuelles, suggérant un lien perceptivo-moteur toujours actif lors de la perception de parole chez ces participants.

Le second paradigme visait à observer une corrélation entre la production et la perception de voyelles. Chez notre population sénior, nous supposons observer, d'une part, une variabilité de production et de perception plus importante par rapport aux jeunes adultes, et d'autre part nous nous attendions à retrouver, comme chez les participants juniors, une corrélation globale entre les valeurs acoustiques produites et perçues. Contrairement à notre première hypothèse, nous n'avons pas observé de différence de variabilité de production entre les séniors et les juniors, et lors de la tâche de perception, nous avons au contraire constaté moins de variabilité intra-participant. Par contre, comme chez les participants juniors, nous avons pu observer des idiosyncrasies uniquement lors de la tâche de production chez la population séniors. En raison de cette absence de variabilité inter-participants, aucune corrélation individuelle n'a été obtenue entre la production et la perception de voyelles. En revanche, encore une fois comme chez la population junior, nous avons pu observer une corrélation globale entre les valeurs acoustiques produites et perçues, illustrant une nouvelle fois des gabarits perceptivo-moteurs et donc un lien sensori-moteur toujours actif chez ces participants.

Enfin, le troisième paradigme consistait en des tâches d'imitation consciente et inconsciente de cibles auditives. Nos hypothèses concernant les personnes séniors prédisaient une plus grande variabilité de production, ainsi que des capacités d'imitation réduites chez ces personnes. Comme attendu, la variabilité de production a été plus importante chez les participants séniors par rapport aux participants juniors. En revanche, aucune différence n'a été observée entre les deux groupes de participants, ni lors de l'imitation consciente, ni lors de l'imitation inconsciente. Les participants séniors ont donc été capables d'imiter volontairement et involontairement une cible auditive, montrant que leur production peut être influencée par un input de parole externe, tout aussi bien que les participants juniors. Ainsi, en regard des résultats obtenus lors de nos trois paradigmes par notre population normo-entendante âgée, nous pouvons supposer

que chez ces personnes, les liens sensori-moteurs semblent être toujours actifs et efficaces.



# Partie III : Relations perceptivo- motrices chez les patients sourds post-linguaux porteurs d'un implant cochléaire

---



## 1. Introduction

Notre troisième et dernière partie visait à déterminer l'impact d'une déprivation sensorielle et d'un réapprentissage sensorimoteur chez des patients sourds post-linguaux puis implantés. L'objectif était de déterminer comment une personne, devenue sourde tardivement puis recevant de nouvelles stimulations sensorielles via un implant cochléaire, va organiser les liens entre ses représentations sensorielles anciennes et nouvelles et son système moteur.

Nous proposons tout d'abord d'introduire la question de la surdité et de l'implant cochléaire, avant d'effectuer une revue de la littérature portant sur la perception et la production de parole chez les patients sourds implantés. Enfin, nous présenterons les résultats de notre population implantée, en les comparant avec ceux des populations de jeunes adultes et de séniors normo-entendants.

## 2. Introduction : surdité et implant cochléaire

### 1. Surdité

#### 1. Les différents types de surdité dans le monde et en France

Entre 10 et 20% de la population mondiale est atteinte de déficience auditive, autrement dit de surdité. La population âgée est majoritairement touchée par la surdité, avec un tiers des personnes de plus de 65 ans souffrant de perte d'audition incapacitante, c'est-à-dire une perte d'audition supérieure à 40 dB dans la meilleure oreille pour une personne adulte<sup>6</sup>.

La déficience auditive est classifiée selon trois paramètres. Le premier paramètre est le moment d'apparition de la surdité : on va parler de surdité congénitale lorsque la personne atteinte est née sourde, de surdité pré-linguale lorsque la personne est devenue sourde entre sa naissance et l'acquisition de la parole (typiquement après trois ans) et de surdité post-linguale lorsque la personne est devenue sourde après l'acquisition critique de la parole. Le second paramètre permettant de classifier les types de déficience auditive est la région du système auditif touchée et entraînant la surdité. Les personnes sont atteintes de surdité de transmission lorsque l'oreille externe ou moyenne est touchée (tympan, conduit auditif externe, osselets), de surdité de perception lorsque la cochlée ou le nerf auditif sont altérés, ou de surdité centrale lorsque ce sont les centres nerveux auditifs (cortex temporal ou voies sous-corticales) qui sont déficients. Enfin, on parle de surdité mixte lorsque la personne sourde est atteinte à la fois de surdité de transmission et de perception. La surdité est également classifiée selon que son évolution a été lente, entraînant des pertes auditives de plus en plus élevées dans le temps (le terme de surdité « progressive » étant alors utilisé) ou que la surdité est survenue brutalement, sans évolution dans le temps (surdité « brusque »).

---

<sup>6</sup> Statistique de l'Organisation Mondiale de la Santé : <http://www.who.int/mediacentre/factsheets/fs300/fr/>

On distingue quatre niveaux de surdité en fonction des degrés de perte d'audition« en décibels par rapport à l'oreille normale (dB H.L.) en référence aux normes « ISO »<sup>7</sup>. Pour évaluer le degré de surdité, on fait la moyenne des pertes d'audition de la meilleure oreille pour les fréquences de 500, 1000 et 2000 Hz. On considère que l'audition reste normale entre 0 et 20 dB de perte. Entre 21 et 40 dB de perte, il s'agit d'une surdité légère, où la parole est perçue correctement à voix normale, mais où elle est difficilement perçue à voix basse ou lointaine. A partir de 30 dB de perte, la personne peut recevoir un appareillage classique. On considère qu'une personne est atteinte de surdité moyenne lorsqu'elle présente une perte de 41 à 70 dB. La personne ne perçoit la parole que lorsqu'elle est forte, et la lecture labiale devient nécessaire. On parle de surdité sévère entre 71 et 90 dB de perte. La parole est perçue à voix très forte près de l'oreille et les bruits forts restent perceptibles. Entre 91 et 120 dB de perte, la surdité est qualifiée de surdité profonde. Plus aucune perception de la parole n'est possible, et seuls les bruits de l'environnement très puissants sont perceptibles. Enfin, à partir de 120 dB de perte, on parle de surdité totale, où plus aucune perception de sons n'est possible.

Les causes de la surdité, autrement dit l'étiologie, sont très diverses. Nous nous intéresserons ici à l'étiologie des surdités pré- ou post-linguales et non congénitales, celles-ci étant généralement génétiquement causées. La surdité de transmission peut être causée par des otites à répétition, des tumeurs ou des malformations congénitales ou par des causes « mécaniques » c'est-à-dire suite à l'introduction d'un corps étranger dans le conduit auditif, suite à un traumatisme crânien ou encore suite à des différences de pression (par exemple lors de voyages en avion ou de plongée sous-marine). Concernant la surdité de perception, la cause principale est le vieillissement, produisant la presbycusie, qui caractérise une détérioration lente et progressive de l'audition liée au vieillissement naturel de l'organe de Corti. La presbycusie atteint une grande majorité des personnes âgées, et débute en général entre 45 et 55 ans. La seconde cause de surdité de perception est le traumatisme, que ce soit un traumatisme dû à une exposition prolongée ou non à un très haut niveau sonore, ou un traumatisme crânien. La surdité mixte, quant à elle, peut être avoir des causes médicamenteuses ou génétiques, mais elle peut également être due à des maladies, des traumatismes crâniens ou des malformations de l'oreille interne<sup>8</sup>. Enfin, la surdité centrale peut être due à plusieurs causes : une atteinte neurologique causée par des tumeurs, céphalites, méningites ou traumatismes crâniens, un retard de neuromaturation du système nerveux auditif central, une atteinte développementale ou encore une atteinte due à une privation sensorielle (privation causée par des otites répétées par exemple)<sup>9</sup>.

## **2. Les différents modes de communication des personnes atteintes de surdité**

Deux moyens de communication sont disponibles pour les personnes sourdes : la langue des signes et l'oralisation.

---

<sup>7</sup> Selon BIAP (Bureau International d'Audiophonologie) : <http://www.biap.org/>

<sup>8</sup><http://www.hear-it.org/mixed-hearing-loss>

<sup>9</sup><http://lemedecinquebec.org/Media/70556/043-048Blais1001.pdf>

Les langues signées sont des langues gestuelles/visuelles, elles possèdent leur grammaire et phonologie propre, et, linguistiquement, elles peuvent être décrites similairement aux langues orales, en termes de pragmatique, morphologie, syntaxe, phonologie ou encore sémantique. Un signe peut être défini comme une combinaison de plusieurs sources différentes d'informations, articulées autour du corps du signeur. Phonologiquement, un signe va être décrit par trois paramètres : la configuration manuelle, le mouvement et la localisation, chaque paramètre pouvant être décomposé - par exemple, pour la configuration manuelle, par rapport aux différentes parties de la main impliquées dans la production du signe.

Historiquement, les langues des signes se sont développées et formalisées au gré des rencontres entre Sourds (ici, nous utiliserons le terme Sourd, avec un S majuscule, en référence au sens sociologique du terme, qui désigne les individus s'identifiant et participant à la culture, la société et la langue de la communauté Sourde<sup>10</sup>), chaque pays possédant sa propre langue des signes, comme la LSF (Langue des Signes Française), la BSL (British Sign Language) ou encore la LSFb (Langue des Signes de Belgique Francophone, à distinguer de la VGT, la langue des signes flamande). La Langue des Signes Internationale (ISL) est, en revanche, une langue artificielle/construite (au même titre que l'Espéranto pour les langues orales), créée à partir de composants des différentes langues signées, essentiellement européennes, et elle est principalement utilisée lors de congrès internationaux réunissant des Sourds ou des professionnels en contact avec les Sourds signeurs ou encore lors de congrès de recherche axés sur les langues signées.

En France, la Langue des Signes Française a été reconnue comme une « langue à part entière » seulement depuis 2005, dans la loi 2005-102 qui modifie le code de l'éducation<sup>11</sup>. D'autre part, cette même loi impose la liberté de choix d'une communication bilingue (LSF et français oral) ou unilingue (uniquement en français oral) lors de l'éducation et du parcours scolaire des jeunes sourds<sup>12</sup>. Actuellement, la LSF est utilisée par plusieurs centaines de milliers de Sourds, majoritairement sourds pré-linguaux, et de plus en plus d'interprètes sont formés pour accompagner les Sourds signeurs dans leur scolarité ou dans leurs démarches administratives.

Le second moyen de communication pour les personnes sourdes est constitué des méthodes oralistes, qui visent à faire rentrer la personne sourde dans la communication orale, par le biais de la lecture labiale et/ou de la Langue Parlée Complétée.

---

<sup>10</sup> Mottez, B. (2006) *Les Sourds existent-ils ?*, éd. L'Harmattan.

<sup>11</sup> Article L312-9-1 du Code de l'éducation : « La langue des signes française est reconnue comme une langue à part entière. »

<sup>12</sup> Article L. 312-9-1 du Code de l'éducation : « Tout élève concerné doit pouvoir recevoir un enseignement de la langue des signes française. Le Conseil supérieur de l'éducation veille à favoriser son enseignement. Il est tenu régulièrement informé des conditions de son évaluation. Elle peut être choisie comme épreuve optionnelle aux examens et concours, y compris ceux de la formation professionnelle. Sa diffusion dans l'administration est facilitée. »

La lecture labiale (référée par le terme de « lipreading » en anglais), aussi appelée lecture oro-faciale (ou « speech reading »), consiste à s'appuyer sur les indices visuels de la partie basse du visage (lèvres, langue, muscles oro-faciaux, mâchoires) de son interlocuteur pour décoder le message. La lecture labiale est presque systématiquement utilisée par les personnes sourdes pour comprendre leurs interlocuteurs entendants ; la lecture labiale va s'acquérir souvent progressivement dans le cas des surdités progressives ou tardives, puisque la personne malentendante va s'aider de la lecture labiale pour compenser la perte de son audition. Par contre, dans le cas de surdité brutale, chez les enfants ou encore pour renforcer la compréhension, la lecture labiale est enseignée par les orthophonistes.

Cependant, la lecture labiale ne permet pas un décodage parfait des messages, en raison des cas d'ambiguïté visuelle entre deux ou plusieurs phonèmes (appelés « visèmes »), qui ont une image labiale/faciale similaire, comme les visèmes /p b m/ ou /t d n/ par exemple. De plus, certains phonèmes restent peu saillants visuellement comme par exemple /k g/, articulés à l'intérieur du tractus vocal. Pour ces deux raisons, le contexte sémantique, syntaxique et l'environnement phonologique sont indispensables pour tenter de compléter les informations ambiguës, et l'apprentissage orthophonique va souvent se révéler nécessaire pour entraîner la personne sourde au décodage labial.

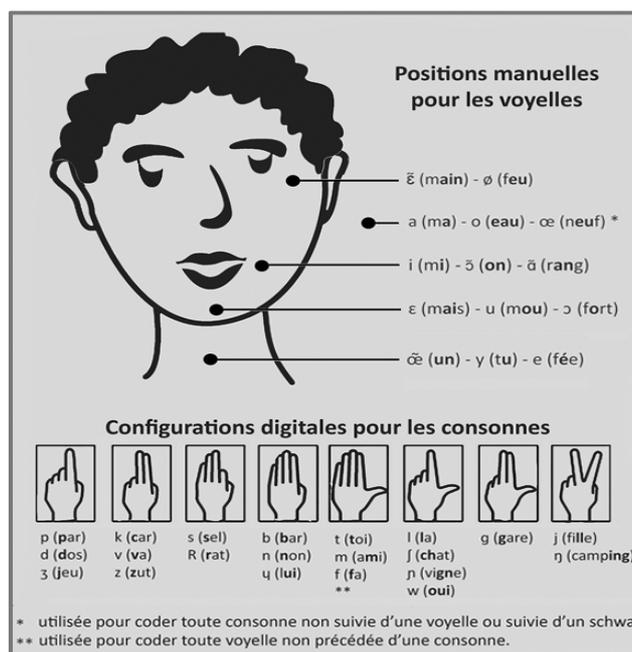


Figure 57 : Clés manuelles de la LPC pour le Français (Bayard, 2014).

La Langue Parlée Complétée (LPC ou cued-speech), développée par Cornett (Cornett, 1967) est un code manuel artificiel qui complète la lecture labiale, et permet de pallier le manque d'information provenant de la lecture labiale seule. Chaque syllabe est codée par la position de la main autour du visage pour représenter les voyelles et par la configuration des doigts pour représenter les consonnes (voir Figure 57). Une même position vs. configuration regroupe plusieurs voyelles vs. consonnes qui ont la caractéristique d'être contrastées aux lèvres. Ainsi avec ce système la vue conjointe des

lèvres et de la main permet de décoder le phonème produit. Ce système particulièrement efficace permet la perception complète de la parole et, pour les enfants sourds l'utilisant depuis le plus jeune âge, il assure des représentations complètes du système phonologique et l'acquisition du langage avec des compétences en lecture et écriture comparables à celles des enfants entendants (Leybaert et coll., 2000).

## 2. Prothèses auditives et implant cochléaire

### 1. Prothèses auditives

Différents types de prothèses peuvent être utilisées par les personnes sourdes. Les prothèses ossiculaires permettent de remplacer les osselets lorsque ceux-ci sont déficients tandis que les prothèses auditives externes (dites « aides auditives classiques ») permettent de renforcer l'audition en amplifiant les parties du son correspondant aux fréquences mal perçues.

Il existe également trois types d'implant qui stimulent de manière mécanique ou électrique les organes déficients : l'implant d'oreille moyenne qui va stimuler la chaîne ossiculaire, les implants osseux qui stimulent l'oreille interne via des vibrations mécaniques des os de la voûte crânienne, et enfin les implants cochléaires, qui stimulent électriquement les fibres nerveuses auditives de la cochlée.

### 2. Principe de l'Implant cochléaire

L'implant cochléaire est une technologie permettant une restitution partielle de l'audition pour les personnes atteintes de surdité sévère à profonde, grâce à une stimulation électrique des terminaisons nerveuses situées dans la cochlée. L'implant cochléaire est indiqué dans le cas de surdité cochléaire, c'est-à-dire lorsque l'organe de Corti est déficient et rend l'utilisation d'une prothèse auditive classique inutile pour la restitution de l'audition<sup>13</sup>.

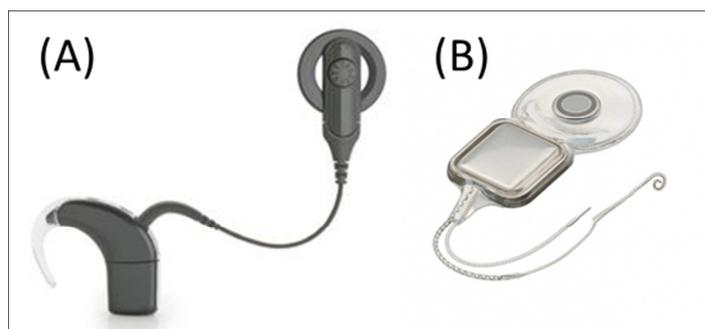


Figure 58 : Partie externe (A) et interne (B) d'un implant cochléaire de marque Cochlear®<sup>14</sup>.

L'implant cochléaire est composé d'une partie interne et d'une partie externe (voir Figure 58). La partie externe comprend en général deux composants principaux : un contour d'oreille et une antenne émettrice. Le contour d'oreille contient une batterie

<sup>13</sup>Actuellement, quatre fabricants d'implants cochléaires sont répertoriés dans le monde : *MXM* en France, *Advanced Bionics Corporation* en Amérique du Nord, *Cochlear Corporation* en Australie et *MedEl Corporation* en Autriche.

<sup>14</sup>[http:// www.cochlear.com](http://www.cochlear.com)

permettant de faire fonctionner l'implant, un micro et un processeur. Le processeur contient un ou des logiciels permettant au porteur de l'implant de s'adapter à la situation ; les logiciels adaptent l'apport de l'implant, en fonction du bruit environnant et de la situation de communication. L'antenne émettrice, présente sur chaque modèle d'implant, est placée au-dessus de l'oreille au moyen d'un aimant et transmet les signaux électriques envoyés par le processeur à la partie interne. La partie interne (Figure 58) comporte deux éléments principaux : le récepteur situé au-dessus du pavillon de l'oreille, et le tube d'électrodes, qui est inséré dans la cochlée. Le récepteur est une petite capsule électronique, d'une épaisseur de 4 à 8 mm, constituée de céramique, de titane et d'un aimant permettant de connecter le récepteur à l'antenne. Avec les modèles d'implants les plus récents, les patients implantés ont la possibilité de bénéficier des techniques d'imagerie par résonance magnétique (IRM anatomique et fonctionnelle), l'aimant étant adapté et compatible avec cette technologie. Le tube peut contenir jusqu'à 24 électrodes, qui stimulent les cellules du nerf auditif, en compensant la déficience des cellules ciliées de la cochlée. Selon le type de fréquences touchées par la surdité et le degré d'ossification de la cochlée, le porte-électrodes peut être plus ou moins long, et peut être inséré à différents endroits de la cochlée.

Concernant le fonctionnement général de l'implant cochléaire (voir Figure 59), l'information auditive reçue par le micro est convertie en signal électrique et envoyée au processeur (1). Le processeur traduit le signal électrique en une série d'impulsions électriques, alors transmises au récepteur interne (3), via l'antenne émettrice (2) pour parvenir enfin aux électrodes insérées dans la cochlée (4). Celles-ci stimulent les fibres nerveuses de la cochlée (les neurones primaires) qui envoient des impulsions électriques au cerveau (5) qui les interprétera en tant que son.

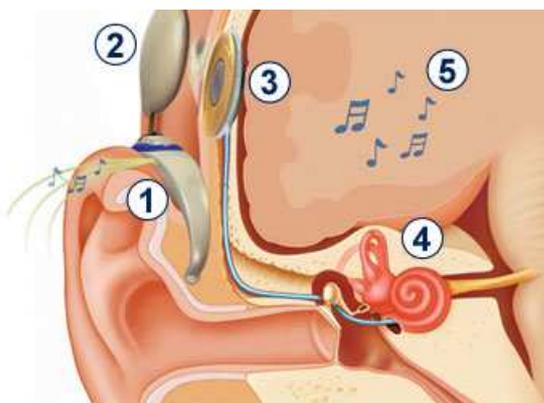


Figure 59 : Fonctionnement d'un implant cochléaire<sup>15</sup>.

<sup>15</sup><http://tpe-ondes-sonores-lasource.e-monsite.com/medias/images/ci-how-it-works.jpg>

### **3. Production et perception de la parole chez les patients porteurs d'un implant cochléaire**

#### **1. Perception**

##### **1. Facteurs affectant les compétences de perception de parole**

De nombreuses études ont été menées pour déterminer les facteurs influençant les capacités de perception de parole chez les patients porteurs d'un implant cochléaire. Ces études montrent avant tout la considérable variabilité des situations et des performances, selon les différentes étiologies ou encore à l'intérieur d'une même catégorie étiologique.

Un premier point, majeur, est l'efficacité de l'implantation et la dynamique de récupération perceptive. Ainsi, Hamzavi et collègues (2003) ont suivi longitudinalement 66 patients sourds post-linguaux pendant six ans, en leur faisant réaliser des tests de reconnaissance de mots, de monosyllabes et de phrases. Ils ont pu observer que les performances des patients s'amélioreraient considérablement au cours du temps, particulièrement dans les douze premiers mois suivant l'implantation.

Faisant suite à une première étude de cohorte menée en 1996, Blamey et collègues (2012), lors d'une étude de très large envergure dans divers centres de suivi de patients implantés cochléaires autour du monde, ont récolté les résultats de tests de reconnaissance de parole (phonèmes, mots monosyllabiques et dissyllabiques, phrases) de 2251 patients sourds post-linguaux âgés de plus de 15 ans, implantés adultes après 2003. Ils ont pu ainsi déterminer un classement des cinq facteurs influençant le plus les résultats des patients : l'expérience de l'implant, l'âge de début de surdité, l'âge d'implantation, la durée de surdité et enfin l'étiologie de la surdité. Cependant, la variabilité résiduelle non expliquée par ces 5 facteurs est considérable : elle recouvre 90% de la variation dans cette étude de grande ampleur.

Ainsi, même s'il semble clair que la durée d'utilisation de l'implant a son importance dans les performances des patients, il est important de rappeler que la variabilité inter-individuelle a une importance encore plus élevée.

##### **2. Perception des paramètres supra-segmentaux**

L'estimation de la fréquence fondamentale ou de l'intonation est difficile pour une personne implantée cochléaire, et ce pour plusieurs raisons : tout d'abord, l'implant cochléaire est composé de 12 à 22 électrodes, et dans les basses fréquences, une même électrode peut capturer plusieurs fréquences, ce qui impacte leur discrimination par le patient ; de plus, en raison de la proximité des électrodes, lorsqu'une électrode est activée, il peut arriver qu'elle stimule des fibres nerveuses ne correspondant pas à la fréquence censée être transmise, ce qui entraîne une restitution très imparfaite de la tonotopie.

Différentes études ont été menées pour évaluer la capacité des patients implantés à estimer l'intonation ou la fréquence fondamentale de leur interlocuteur. Parmi celles-ci, Chatterjee et Peng (2009) ont mesuré les capacités de discrimination d'intonation par

dix participants implantés, huit d'entre eux étant sourds post-linguaux, en comparaison avec quatre entendants participant au test avec des simulations d'implant cochléaire obtenues par un vocodeur. Préalablement au test de discrimination, les participants ont passé un test de mesures psychophysiques visant à estimer les capacités de discrimination de fréquence des participants via les indices temporels. Les stimuli utilisés pour le test de discrimination d'intonation ont été créés synthétiquement en modifiant un stimulus initial composé du mot « popcorn » produit par une femme. Les valeurs initiales de F0 de chaque stimulus ont été modifiées pour atteindre une valeur basse (120 Hz) ou haute (200 Hz), de façon à associer à chaque stimulus une intonation descendante, plate ou montante. Les participants devant déterminer si les stimuli étaient des questions ou des déclarations. Les résultats ont montré que les participants implantés ont obtenu des performances très variables dans les deux tests. Néanmoins, deux points importants sont à noter. En premier lieu, les performances de reconnaissance d'intonation étaient positivement corrélées avec les performances aux tests de mesures psychophysiques. Ensuite, les performances des participants implantés au test de reconnaissance d'intonation étaient meilleures lorsque la valeur initiale de F0 était basse plutôt qu'élevée, et ce patron de résultats était similaire à celui des participants entendants répondant aux stimuli spectralement réduits par le vocodeur.

Cullington et Zeng (2011) ont réalisé par la suite une étude portant sur la discrimination de musique, de prosodie et d'identification de locuteur par des participants sourds implantés soit bilatéralement (référés par le terme d'utilisateurs bilatéraux), soit unilatéralement avec une prothèse auditive classique dans l'oreille non-implantée (référés par le terme d'utilisateurs bimodaux). Leur hypothèse était que les utilisateurs bimodaux obtiendraient de meilleures performances que les utilisateurs bilatéraux lorsque les tests réalisés nécessitent une bonne reconnaissance du pitch. Les participants de cette étude ont dû répondre à plusieurs tests : reconnaissance dans le bruit de phrases produites par des locuteurs homme, femme ou enfant, test de musique permettant d'évaluer la perception du pitch, du rythme et du timing ainsi que la mémoire musicale, test de reconnaissance de prosodie affective et de sarcasme, et enfin, test de reconnaissance de locuteurs. Encore une fois, une grande variabilité a été observée dans les résultats, et ce malgré une pré-sélection des participants via des tests de reconnaissance de parole claire. Cependant, dans tous les tests, les participants implantés bilatéraux et bimodaux ont obtenu des scores similaires, même si leurs performances étaient plus faibles que celle des normes observées chez des participants normo-entendants, ces résultats permettant donc de réfuter l'hypothèse du bénéfice de l'audition résiduelle via les prothèses auditives classiques.

Enfin, Marx et collègues (2014) ont étudié les capacités de perception de la prosodie par les personnes implantées, avec ou sans restes auditifs. Pour cela, ils ont réalisé un test de reconnaissance d'intonation (question ou déclaration) avec ou sans modulation de la F0, et un test de discrimination de fréquence (nommé F0DL, où le participant doit déterminer, entre trois sons purs, lequel est le plus haut). Quinze participants normo-entendants, dix participants porteurs d'un implant cochléaire, sans

restes auditifs (groupe implanté) et 11 participants implantés ayant une audition résiduelle et portant une prothèse auditive dans l'oreille non-implantée (groupe combiné) ont pris part aux tests. Les participants du groupe combiné ont présenté des résultats systématiquement supérieurs à ceux du groupe implanté et inférieurs à ceux du groupe normo-entendant que ce soit lors de la tâche de reconnaissance d'intonation ou lors de la tâche de discrimination de fréquences. Lors de la tâche de reconnaissance d'intonation, dans la condition où la F0 était stable, les performances du groupe combiné et du groupe normo-entendants ont chuté pour atteindre celles du groupe implanté. Enfin, pour le groupe combiné, une corrélation positive en ce qui concerne la durée d'expérience de l'implant a été obtenue. En conclusion, cette étude permet de montrer l'importance de l'aide auditive résiduelle chez les patients implantés en ce qui concerne la perception de F0 et d'intonation.

En résumé ces trois études suggèrent que les personnes sourdes post-linguales implantées semblent être capable d'extraire des informations sur la fréquence fondamentale, bien que de manière dégradée, et ce malgré que les implants ne transmettent qu'imparfaitement ces informations au patient, aussi bien en ce qui concerne les limites du codage temporel de la fréquence – avec des stratégies de codage diverses et souvent peu fidèles dans le codage temporel fin – qu'en ce qui concerne les limites du codage tonotopique, puisqu'à une même électrode peut être associé un spectre de fréquences assez large, rendant difficile l'analyse spectrale fine et, partant, l'estimation fine de F0.

### **3. Perception des consonnes et des voyelles**

Nous avons vu précédemment qu'il semble y avoir plusieurs facteurs influençant les capacités de perception de parole chez les patients implantés cochléaires. Nous proposons maintenant d'observer plus précisément les performances de perception des patients implantés selon les phonèmes.

Valimaa (2002) a testé les performances de discrimination de phonème chez 20 patients sourds post-linguaux, entre 4 jours et 24 mois post-implantation. Les patients devaient identifier des voyelles et des consonnes au sein de syllabes CV (C : consonne, V : voyelle), CVV et VC, ces syllabes étant insérées dans une phrase porteuse. Les résultats ont montré que les patients identifiaient mieux les voyelles que les consonnes, les capacités de discrimination s'améliorant au cours du temps. Plus précisément, concernant les voyelles, à un mois post-implantation, elles ont été identifiées à 50%, à 6 mois post-implantation, à 68%, et à partir de 18 mois post-implantation, le taux de reconnaissance a atteint 80%. Concernant les consonnes, le taux d'identification était de 40% à un mois post-implantation, 58% à 6 mois, et 71% à deux ans post-implantation. Il a été également observé que les consonnes labiales étaient plus difficilement identifiées par rapport aux consonnes dentales ou alvéolaires (53% vs. 80%, à deux ans post-implantation).

Ces résultats sont cohérents avec ceux des études réalisées précédemment. En effet, il est montré en général que les voyelles sont mieux décodées (scores moyens entre

60 et 80%) que les consonnes (scores entre 50 et 75%) (Kompis et collègues, 1999 ; Pelizzone et collègues, 1999 ; Skinner et collègues, 1999 ; VanVierigen et Wouters, 1999 ; Kiefer et collègues, 2001 ; pour une revue, voir Valimaa, 2001).

Cependant, même si les voyelles semblent donc être plus facilement reconnaissables par les patients sourds implantés, les performances dépendent du contexte dans lequel la voyelle est produite. Hanna (2011) a étudié l'influence des transitions formantiques lors de la discrimination de voyelles chez 8 patients implantés sourds post-linguaux et 12 participants normo-entendants. Les stimuli utilisés étaient des syllabes /dVd/, avec trois conditions : une première condition « naturelle », une seconde condition où la voyelle à identifier avait des formants stables, et une dernière condition où la voyelle comprenait des transitions formantiques synthétisées. Les résultats ont montré que les participants implantés obtenaient des taux d'identification plus faibles que les normo-entendants, cette différence étant plus importante lorsque les voyelles avaient des formants stables, ce qui suggère que les personnes sourdes porteuses d'un implant cochléaire semblent avoir particulièrement besoin des transitions formantiques et donc d'un contexte syllabique pour identifier correctement les voyelles.

#### **4. Perception audiovisuelle**

Nous avons précédemment montré le bénéfice de l'addition de la modalité visuelle à un stimulus de parole auditif chez les personnes normo-entendantes. L'avantage de l'audiovisuel a aussi été exploré au sein des populations sourdes : tout d'abord chez les enfants sourds prélinguaux, où plusieurs études ont montré une amélioration des performances en lecture labiale et en perception audiovisuelle après trois ans post-implantation (Geers et Brenner, 1994 ; Tyler et collègues, 1997 ; Bergeson et collègues, 2005). Également, différentes études sur ce sujet ont été également menées chez les sujets sourds post-linguaux implantés.

Tout d'abord, Goh et collègues (2001) ont étudié les capacités d'intégration audiovisuelle chez un patient sourd post-lingual implanté, en comparaison avec des participants contrôles normo-entendants. Comme attendu, par comparaison avec le sujet implanté, les participants normo-entendants ont obtenu de meilleurs résultats en condition audio seule. Par contre, les résultats de celui-ci ont été meilleurs que ceux des entendants en condition visuelle seule, ainsi qu'en condition audiovisuelle, ce qui suggère une capacité de décodage visuel importante chez ce patient.

Par la suite, Kirk et collègues (2002) puis Kaiser et collègues (2003) ont testé l'interaction audiovisuelle chez 41 participants sourds post-linguaux implantés, en comparaison à 21 participants normo-entendants. Les participants ont dû identifier des mots monosyllabiques en conditions audio seule, visuelle seule et audiovisuelle, les stimuli étant organisés en plusieurs listes : soit des listes produites par un seul locuteur, soit des listes produites par plusieurs locuteurs. Dans les deux groupes, les stimuli audiovisuels ont été mieux reconnus que les stimuli audio, eux-mêmes plus précisément identifiés que les stimuli visuels. Comme chez Goh et collègues (2001) les participants normo-entendants ont obtenu des scores plus élevés que les participants implantés en

condition audio seule, tandis que les participants implantés ont mieux reconnu les stimuli présentés en condition visuelle. Également, dans les deux groupes, une corrélation entre les réponses aux stimuli audio seuls et les réponses aux stimuli audiovisuels a été observée, suggérant que les participants ayant les meilleurs taux de reconnaissance de stimuli audio ont également les meilleurs résultats en condition audiovisuelle. Enfin, de manière intéressante, chez les participants implantés, le bénéfice de l'entrée visuelle était plus important dans les listes uni-locuteur que dans les listes multi-locuteurs, ce qui n'était pas le cas pour les participants normo-entendants. Ce résultat suggère que les participants implantés ont de meilleures capacités d'extraction des informations visuelles inhérentes au locuteur pour des stimuli audiovisuels. En résumé, cette étude démontre donc une grande utilisation des indices visuels chez les personnes sourdes porteuses d'un implant cochléaire, et ce similairement aux personnes normo-entendantes.

Par la suite, Rouger et collègues (2007) ont étudié les capacités d'intégration audiovisuelle chez 97 patients sourds post-linguaux en comparaison avec un groupe de participants normo-entendants. Les performances de reconnaissance de mots disyllabiques en condition auditive, visuelle et audiovisuelle ont été évaluées dans une étude longitudinale de 8 ans, les patients étant testés pré-implantation et douze fois post-implantation (de T0, c'est-à-dire juste après l'activation de l'implant à 8 ans post-implantation). Comme attendu, les résultats ont montré de meilleures performances de lecture labiale pour les patients pré-implantation par rapport aux participants normo-entendants. En revanche, de façon surprenante, ces bonnes performances de lecture labiale n'ont pas été influencées par la récupération des capacités auditives : en effet, alors que les patients ont obtenu de hauts taux de reconnaissance de mot en condition auditive à partir de deux mois post-implantation (plus de 80% de reconnaissance à 4 mois post-implantation), leurs performances de reconnaissance en condition visuelle restent inchangées. Rouger et collègues ont également comparé les performances de reconnaissance audiovisuelle des patients à T0 avec celles de participants normo-entendants en condition masquée avec du bruit ou avec un vocodeur. Les résultats ont montré de meilleures performances pour les patients implantés, et il est proposé que les patients implantés seraient de meilleurs intégrateurs que les normo-entendants, c'est-à-dire qu'ils associent plus facilement les informations visuelles (lecture labiale) avec les informations auditives perturbées que leur envoie l'implant.

Cependant, il est intéressant de noter que les capacités d'interaction audiovisuelles ne semblent pas dépendre de la vitesse de perte d'audition chez les personnes sourdes implantées, c'est-à-dire selon si les personnes sont devenues sourdes soudainement ou si leur perte d'audition était progressive. En effet, Bergeson et collègues (2003) ont testé 16 participants devenus sourds soudainement puis implantés et 32 participants implantés atteints de surdité progressive, lors de tests de reconnaissance de phrases en condition audio seul, visuel seul, et audiovisuelle. Aucune différence significative n'a été observée entre les deux groupes dans les conditions audio seul et audiovisuelle. Ces résultats suggèrent donc que les capacités d'interaction

audiovisuelle ne dépendent pas de la vitesse de perte d'audition. Par contre, les participants atteints de surdité progressive ont obtenu de meilleurs résultats que les participants devenus sourds soudainement lors de la condition visuelle seule.

## **2. Production**

### **1. Contrôle de la fréquence fondamentale dans la production de l'intonation et le chant**

Très peu d'études ont été menées sur la production de chant ou sur le contrôle du pitch chez les patients sourds porteurs d'un implant cochléaire.

Kishon-Rabin et collègues (1999), dans leur étude portant sur la production de parole par des sourds post-linguaux implantés, analysent la production de F0. Ils ont enregistré cinq participants implantés âgés de 35 à 61 ans, lors de tâches de production de mots monosyllabiques ou de phrases, pré-implantation puis à différents moments post-implantation. Les propriétés segmentales (VOT et distribution spectrale pour les consonnes et F1 et F2 pour les voyelles) et supra-segmentales, comme la fréquence fondamentale et la durée de mots et des phrases ont été mesurées. La fréquence fondamentale moyenne des participants lors de la production de mots et de phrases a été comparée avec la moyenne produite par des participants entendants. Les résultats obtenus ont montré que les participants implantés présentent des F0 moyens plus élevés que la moyenne des participants entendants, comme cela avait déjà été obtenu lors de différentes études (Leder et collègues, 1987 ; Leder et Spitzer, 1993). Par contre, Kishon-Rabin et collègues montrent également que le F0 moyen des participants implantés diminue significativement post-implantation, puisque deux participants reviennent dans la norme après deux ans post-implantation. Cette diminution de la fréquence fondamentale post-implantation a également été démontrée lors d'autres études, comme chez Ball et Faulkner (1989), Lane et collègues (1995) ou encore Langereis et collègues (1998), qui ont observé une diminution significative du F0 après 12 mois post-implantation, chez un tiers de leur groupe de participants sourds post-linguaux. Kishon-Rabin et collègues supposent que cette amélioration de la F0 serait une des conséquences des changements de niveau de pression sonore (« SPL : Sound Pressure Level »), comme l'avaient proposé Perkell et collègues (1992), ces suppositions ne pouvant être confirmées en raison de l'absence de mesure d'intensité chez Kishon-Rabin et collègues.

Concernant la production de chant, les études existantes ont été menées chez des enfants sourds pré-linguaux implantés et non chez des adultes sourds post-linguaux. Xu et collègues (2010) ont enregistré sept enfants sourds pré-linguaux implantés lors de la production d'une chanson familière, et ont comparé des mesures de F0 pour chaque note avec celles produites par des enfants entendants. Leurs résultats ont montré que, malgré une grande variabilité entre les enfants, les enfants implantés ont présenté des performances de chant plus faibles que les enfants entendants lors de toutes les évaluations basées sur le pitch. Par la suite, Mao et collègues (2003) ont également mesuré des propriétés acoustiques lors de la production de chant chez des enfants sourds pré-linguaux porteurs d'implants cochléaires ou de prothèses auditives. Ils ont obtenu

des résultats similaires à ceux de l'étude de Xu et collègues, à savoir que les enfants implantés et ceux porteurs de prothèse auditive ont obtenu des résultats inférieurs à ceux d'enfants entendants. En revanche, chez les deux groupes d'enfants sourds, les performances étaient positivement corrélées avec la durée d'utilisation de leur aide auditive.

En résumé, si la production de fréquence fondamentale chez les patients sourds semble s'améliorer post-implantation, il est important de rappeler que les études menées sur ce sujet sont assez anciennes, particulièrement chez les adultes (plus de 15 ans), et ne peuvent donc prendre en compte les améliorations technologiques de l'implant cochléaire. De plus, ces études comportent de très petites cohortes de patients (cinq adultes chez Kishon-Rabin et collègues, 1999, et sept enfants chez Xu et collègues, 2010), et la variabilité inter-participant reste assez élevée. Il semblerait donc nécessaire d'évaluer sur une plus grande cohorte les capacités de production de la F0, afin de confirmer les résultats des études précédentes.

## 2. Production des voyelles

Plusieurs études ont été menées sur la production de parole chez les patients sourds post-linguaux porteurs d'un implant cochléaire.

Concernant la production de voyelles, Langereis et collègues (1997) ont enregistré 20 participants sourds post-linguaux pré-implantation puis à trois mois et douze mois post-implantation, avec implant soit fonctionnel soit éteint. La tâche consistait à produire des voyelles du néerlandais en contexte /hVt/, productions à partir desquelles les valeurs de F1 et de F2 ont été mesurées. Une amélioration générale des productions a été observée à 3 et 12 mois post-implantation, se traduisant par une meilleure séparation des productions dans l'espace formantique. Par contre, lorsque l'implant était éteint, les productions se sont trouvées dégradées. Enfin, chez certains individus, il a été observé une dégradation des voyelles à trois mois post-implantation, avant une nette amélioration à 12 mois post-implantation. Par rapport aux scores observés lors de la phase de pré-implantation, cette dégradation à trois mois post-implantation est supposée être due au nouveau feedback auditif apporté par l'implant cochléaire auquel les participants doivent s'habituer.

Par la suite, Lane et collègues (2001) ont enregistré 7 patients sourds post-linguaux pré-implantation et à 0, 1, 3, 6 et 12 mois post-implantation, leurs productions étant comparées avec celle de participants normo-entendants. Les participants ont dû produire les 8 voyelles de l'anglais américain en contexte /dVt/ et /bVt/ dans une phrase porteuse. La moyenne de l'espace vocalique (AVS : Average Mean Space) a été mesurée sur les dimensions F1/F2, et les mesures ont été réalisées à deux points des voyelles pour vérifier l'impact de la coarticulation. Il n'a été observé pratiquement aucun changement post-implantation, ni sur la moyenne de l'espace vocalique ni sur la coarticulation. Par contre, les participants implantés ont produit des voyelles plus longues que les participants normo-entendants, même si la durée des voyelles a par la suite été réduite à

6 mois et 12 mois post-implantation. Selon les auteurs, les participants implantés pourraient ainsi ralentir leur diction dans l'optique de la rendre plus intelligible.

Lane et collègues (2005) ont évalué l'évolution de contrastes vocaliques en perception et en production, chez huit participants sourds post-linguaux puis implantés, lors d'une tâche de production et de perception de paires de voyelles proches dans l'espace vocalique. Ils ont observé une amélioration de la production post-implantation, que ce soit en production ou en perception ; également une évolution conjointe de la perception et de la production a été constatée, puisque les participants ayant produit des paires de voyelles plus contrastées post-implantation par rapport à pré-implantation sont également ceux qui ont eu le plus d'évolution lors des tâches de perception post-implantation. Cette évolution conjointe est interprétée par le fait que l'amélioration du feedback auditif entraîne une production plus précise pour permettre aux locuteurs de se rendre le plus intelligible possible.

Plus récemment, en 2007, Ménard et collègues ont analysé les productions de voyelles de 13 participants sourds post-linguaux pré-implantation et à 1 mois et 12 mois post-implantation, en comparaison avec 10 participants normo-entendants. Les voyelles ont été produites en contexte /pVp/. Pour mesurer l'impact du feedback auditif sur la production de parole, des tests de perception de voyelles ont été également réalisés.

Les résultats ont montré une augmentation des contrastes entre voyelles dans l'espace vocalique à un an post-implantation par rapport aux productions réalisées pré-implantation et à 1 mois post-implantation. Les voyelles produites ont été plus longues chez les participants implantés en général, même si les valeurs de durée à un an post-implantation se rapprochent de celles produites par les participants normo-entendants. Concernant la perception, les résultats des participants implantés se sont améliorés significativement au cours du temps post-implantation, même s'ils restent inférieurs à ceux des participants entendants, comme nous l'avons déjà montré dans la partie 3. En revanche, aucune interaction entre les valeurs mesurées en production de parole et les résultats des tests de perception de parole n'a été observée, contrairement aux résultats obtenus par Vick et collègues (2001), bien que les données montrent également, comme dans les études précédentes, que la production est plus intelligible avec implant « on » que « off ».

### **3. Lien perceptivo-moteur**

Extrêmement peu d'études portent sur le lien perceptivo-moteur chez les participants implantés. Rouger et collègues (2012) ont réalisé une étude en Tomographie par Emission de Positrons (PET) sur la perception de parole visuelle par des participants sourds implantés. Ils ont enregistré dix participants sourds post-linguaux et ont comparé leurs résultats avec ceux de six participants normo-entendants. Les participants ont dû réaliser des tâches de reconnaissance de mots en condition visuelle uniquement, et ont été enregistrés une première fois le plus tôt possible après implantation (T0) puis une seconde fois lorsque le participant obtenait un taux de reconnaissance de mots supérieur à 60% lors de contrôles orthophoniques (T1). Les résultats comportementaux ont montré

que les participants implantés obtenaient des taux de reconnaissance de mots visuels supérieurs à ceux des participants entendants, montrant que pendant leur surdité, ces patients ont développé des capacités supra-normales de lecture labiale. Concernant les résultats en neuro-imagerie, dans les trois groupes (patients à T0, patients à T1, normo-entendants), les aires visuelles ont été activées pendant la lecture labiale, aucune différence n'ayant été observée entre les trois groupes. Chez les participants implantés, les aires auditives ont été également activées, contrairement aux participants normo-entendants, avec une plus forte activation de la partie postérieure du sillon temporal supérieur droit (STS, c'est-à-dire le siège des interactions cross-modales) à T0. Enfin, les participants implantés à T0 présentent une hypo-activation de l'aire de Broca alors qu'aucune différence significative n'a été observée entre les participants à T1 et les participants normo-entendants. De plus, une corrélation a été observée entre l'activité de l'aire de Broca et du STS, montrant qu'une plus haute activation de l'aire de Broca est associée à une plus basse activation du STS. Ces résultats suggèrent donc, chez les patients implantés après un temps d'adaptation à l'implant, une régression des activations cross-modales anormales du STS au profit d'une réactivation progressive des aires frontales normalement impliquées dans la production de la parole, ce qui d'une part, suggère des stratégies compensatoires mises en place par les patients implantés, et d'autre part apporte un argument en faveur d'une réactivation progressive du lien audio-visuo-moteur.

Vick et collègues (2002) ont réalisé une étude sur la compensation articulatoire selon la présence ou non de retour auditif et de bloqueurs de mâchoires (bite-blocks). Ils ont enregistré 8 participants sourds post-linguaux pré-implantation, à un mois post-implantation et à un an post-implantation, en comparaison avec 10 participants normo-entendants, durant la production de syllabes /hVd/ selon quatre conditions : avec ou sans feedback auditif associé à la présence ou l'absence de bite-block (lors de la session pré-implantation, seules les conditions avec ou sans bite-block ont été testées). Dans les conditions avec bite-block, les participants (implantés et entendants, lors des deux conditions avec ou sans feedback) ont réduit leur espace vocalique. De plus, lors des sessions avec feedback auditifs à un an post-implantation, les participants implantés ont augmenté leur espace vocalique et produit des voyelles moins dispersées par rapport aux sessions précédentes. Ces résultats permettent de démontrer l'importance du retour auditif pour les patients implantés, puisqu'il leur permet de corriger, lorsqu'elle est perturbée, leur trajectoire articulatoire en fonction du retour auditif qu'ils reçoivent, le tout afin d'adapter leur production pour la rendre plus intelligible.

Par la suite, Turgeon et collègues (2015) ont testé l'effet de la surdité sur la production de parole au travers d'une étude de perturbation motrice de la parole. Ils ont enregistré un groupe de participants normo-entendants et deux groupes de patients implantés soit sourds post-linguaux ou sourds pré-linguaux, lors de la production de la voyelle /u/ selon plusieurs conditions : pré-perturbation, perturbation (insertion d'un tube dans la bouche) et post-perturbation, à chaque fois avec ou sans retour auditif (implant ON ou OFF chez les sujets implantés, bruit blanc chez les normo-entendants).

A partir des productions, des mesures acoustiques (F0, F1, F2), articulatoires et l'intelligibilité de production ont été mesurées. Ces mesures leur ont permis de montrer l'influence du feedback auditif chez les sourds pré-linguaux uniquement (se traduisant par leur incapacité, lorsque l'implant est éteint, à ramener F1 à une valeur basse, indispensable à la réalisation du /u/, malgré la perturbation qui fait augmenter les valeurs des formants F1 et F2). L'interprétation des auteurs est que ces participants, n'ayant jamais eu d'expérience d'audition naturelle, ont une représentation acoustico-articulatoire incomplète des voyelles. En effet, puisque leur surdité a eu lieu avant l'acquisition de parole et donc avant la construction des commandes feed-forward (décrites par Perkell et collègues, 2012), le contrôle du F1 chez ces participants repose essentiellement sur le retour auditif qu'ils reçoivent, contrairement aux participants post-linguaux, qui, eux, ont pu construire et stocker leurs commandes feed-forward lors de l'acquisition de la parole. Les données articulatoires confirment la différence entre sourds pré-linguaux et post-linguaux, puisque les participants pré-linguaux ont produit les voyelles avec une antériorité de la langue en condition perturbée, alors que les participants sourds post-linguaux ou entendants ont pu au contraire maintenir une stabilité de la langue voire un mouvement postérieur, dû à la compensation articulatoire. Cette observation est reliée par les auteurs à des mesures articulatoires effectuées chez des enfants de 4 ans, qui présentent la même antériorité de langue lorsque leur production était perturbée : il est proposé que le développement de stratégies compensatoires constitue une étape clef dans le développement de l'enfant, étape qui n'aurait pas été atteinte chez les sourds pré-linguaux. Enfin, globalement, les trois groupes ont présenté des after-effects : F0 et F2 ont atteint des valeurs plus hautes en condition perturbée et en condition post-perturbation par rapport à la condition pré-perturbation ; également, les scores d'intelligibilité ont été plus faibles en condition perturbation et post-perturbation, et ce chez les trois groupes de participants. Ces after-effects peuvent révéler une perturbation des relations acoustico-articulatoires (ou perceptivo-motrices) lors de la condition perturbation qui se conserverait post-perturbation. En résumé, cette étude suggère, d'une part que les sourds pré- et post-linguaux implantés ont conservé ou récupéré une certaine forme de relations perceptivo-motrices, et d'autre part, que les sourds post-linguaux, contrairement aux pré-linguaux, ont pu bâtir des représentations acoustico-articulatoires lors de l'acquisition de la parole, qu'ils ont ensuite conservées pendant leur surdité.

Pour conclure, au travers de ces trois études, il semblerait que les personnes sourdes post-linguales puis implantées ont mis en place un lien effectif entre les nouvelles représentations sensorielles fournies par l'implant et les représentations motrices nécessaires au contrôle de la parole, même si la qualité précise de ce lien n'est pas clairement établie, et donc qu'une exploration plus poussée des interactions perceptivo-motrices reste à réaliser.

#### **4. Rôle des interactions perceptivo-motrices et leur réorganisation**

##### **1. Proposition et hypothèses expérimentales**

En France, plus de 1000 personnes sourdes reçoivent un implant cochléaire chaque année. En raison du vieillissement de la population, de plus en plus de personnes deviennent sourdes, ce qui impliquera, à terme, de plus en plus de personnes implantées. L'implant cochléaire représente donc un enjeu majeur de recherche, afin de mieux comprendre son fonctionnement, et l'adaptation du cerveau humain aux nouvelles représentations auditives reçues via l'implant cochléaire.

Un des buts de cette thèse était de tester le lien perceptivo-moteur chez des patients sourds post-linguaux porteurs d'un implant cochléaire.

Chez les normo-entendants, lors de la production de parole, le locuteur va élaborer des buts auditifs et moteurs à atteindre lors de la production du message, et lorsqu'il parle, il va comparer le feedback sensoriel qu'il reçoit à ses buts initiaux de manière à mieux contrôler sa production. Inversement, lors de la perception de parole, il va comparer le message qu'il perçoit à ses représentations motrices internes, stockées dans le cerveau depuis l'acquisition de la parole (voir Figure 60-haut).

Notre interrogation principale était de savoir comment une personne devenue sourde tardivement, puis implantée, qui reçoit donc de nouvelles représentations auditives via son implant, va faire le lien entre ses anciennes et ses nouvelles représentations auditives et son système moteur. (voir Figure 60-bas). Notre hypothèse centrale est que le lien perceptivo-moteur joue un rôle essentiel tant pour la production que pour la perception de parole chez un sujet normo-entendant, que ce lien a été mis en place chez le sourd post-lingual lors de son apprentissage de la parole avant surdité, puis qu'il s'est trouvé confronté successivement à la déprivation auditive associée au handicap sensoriel, puis à l'apparition d'un nouveau type d'afférences via l'implantation. La question expérimentale au centre de cette thèse, et particulièrement de ce chapitre, est de savoir comment se sont réorganisés les liens entre représentations motrices et sensorielles (pré- et post-implantation), autrement dit, comment une personne, devenue sourde puis implantée, va établir un lien entre ses anciennes représentations auditives, ses nouvelles représentations auditives apportées par l'implant cochléaire, et son système moteur. Egaleme nt, à minima, nous souhaitons déterminer s'il existe un lien audio-articulatoire fonctionnel et efficace chez le sourd implanté.

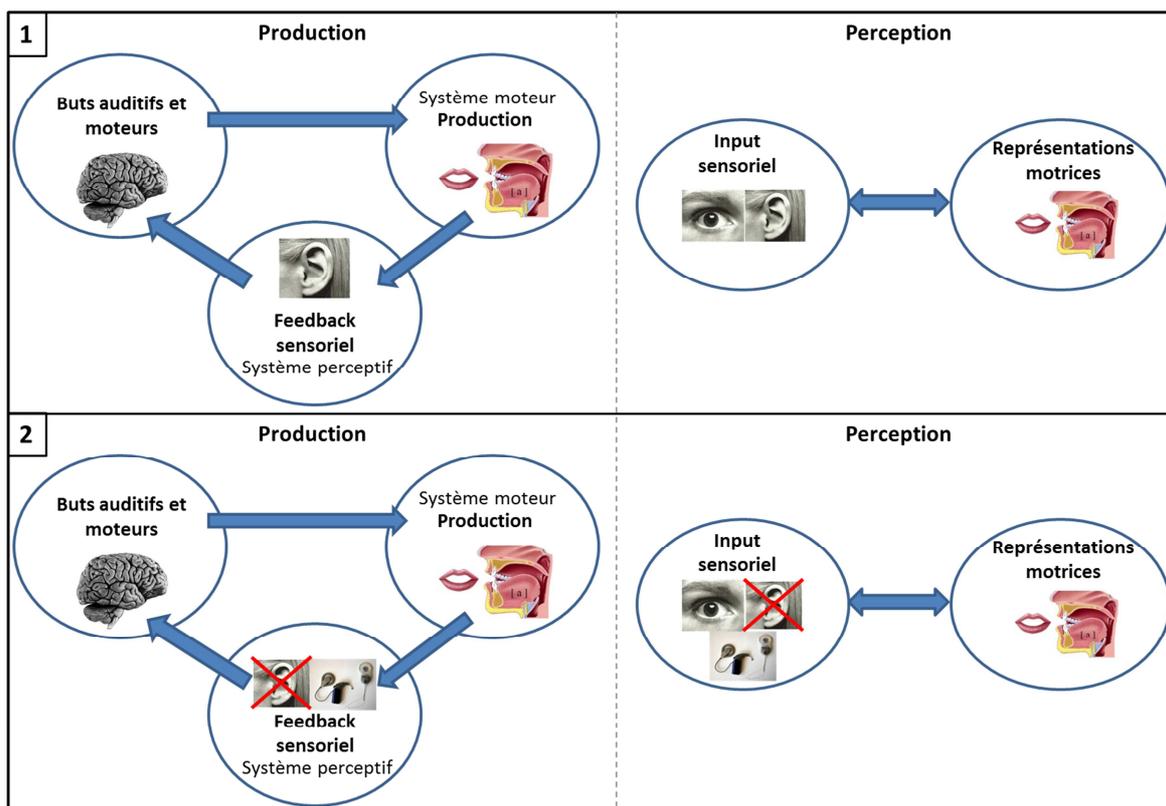


Figure 60 : Schéma des interactions perceptivo-motrices chez les normo-entendants (1) et schéma hypothétique des interactions perceptivo-motrices chez les patients sourds post-linguaux porteurs d'un implant cochléaire (2)

Afin de tester les interactions sensori-motrices de ces patients, nous avons choisi d'utiliser les trois paradigmes expérimentaux décrits dans les parties précédentes : l'expérience de close-shadowing, qui vise à comparer la rapidité et la précision de réponse à des stimuli de parole auditifs et audiovisuels entre une tâche de catégorisation manuelle et une tâche de répétition rapide, l'expérience dite « d'idiosyncrasies », qui a pour objectif de déterminer une possible organisation perceptivo-motrice des voyelles du Français perçues et produites et, enfin, l'expérience de convergence et d'imitation, qui implique des tâches d'imitation automatique et volontaire de la fréquence fondamentale lors de la production de voyelles.

Globalement, du fait du déficit auditif des patients implantés, nous nous attendions à obtenir des performances générales dégradées lors des trois expériences mais un apport accru de la modalité visuelle lors de la tâche de close-shadowing, en raison de l'expérience en lecture labiale des participants durant leur surdité. Mais aussi, et plus centralement, nous cherchions à déterminer grâce à ces expériences si les interactions sensori-motrices pouvaient se rétablir en perception et en production de parole chez les patients implantés, et ce malgré leur déficit auditif.

## 2. Participants

Dix sujets sourds post-linguaux porteurs d'un implant cochléaire ont participé à nos expériences. Les participants (7 hommes et 3 femmes), âgés de 27 à 76 ans (moyenne

d'âge : 58,9 ans), étaient tous de langue maternelle française, sans trouble de la vision non corrigé et sans troubles du langage (pour l'étiologie de chaque participant, voir Tableau 9).

	Sexe	Age (an)	Age début surdité	Date d'implantation	Prothèse contrôle	Durée de surdité	Expérience implant
CI1	H	65	7 ans	Févr-14	non	58	1 mois
CI2	H	56	progressive	Févr-14	oui	35	3 mois
CI3	F	66	32 ans	Oct-05	oui	25	9 ans
CI4	H	60	59 ans	Févr-13	non	1 mois	1 an 4 mois
CI5	F	43	20 ans	Avr-14	oui	13	2 mois
CI7	H	27	25 ans	Déc-12	implants bilatéraux	2 mois	2 ans 6 mois
CI8	F	67	65 ans	Avr-14	oui	2 ans	7 mois
CI9	H	72	40 ans	Juil-14	oui	30 ans	5 mois
CI10	H	76	48 ans	Janv-13	non	27 ans	3 ans 4 mois
CI11	H	57	48 ans	Juil-14	oui	8 ans	10 mois

Tableau 9 : Etiologie des participants implantés (le terme "prothèse contrôle" réfère à une prothèse auditive classique, portée dans l'oreille non-implantée). L'expérience de l'implant correspond à la durée depuis l'activation de l'implant, alors que la durée de surdité correspond à la période entre le début de la surdité et l'activation de l'implant, la passation des différents participants ayant été réalisée sur plus d'un an.)

Comme pour les participants juniors et séniors, le protocole expérimental a été validé par le CERNI (Comité d'éthique pour les recherches non interventionnelles) et tous les participants ont donné par écrit leur accord en signant, préalablement aux expériences, un formulaire de consentement. Les patients ont été recrutés au sein du service ORL du C.H.U. de La Tronche, grâce à la collaboration des orthophonistes du centre et du Pr. Schmerber.

### 3. Procédures expérimentales communes aux trois expériences

Les participants implantés ont réalisé les trois expériences dans les mêmes conditions que les participants juniors et séniors, suivant la même procédure expérimentale et le même ordre de passation.

#### 3. Expérience de close-shadowing

##### 1. Principes et hypothèses

La première expérience réalisée chez nos participants implantés était donc l'expérience de close-shadowing. La procédure expérimentale proposée aux participants implantés était la même que celle utilisée chez les participants normo-entendants lors des expériences non-bruitées. En effet, les participants implantés devaient réaliser deux tâches de catégorisation des séquences /apa/, /ata/ et /aka/: une tâche en répétition rapide des séquences cibles et une tâche de catégorisation manuelle, où les participants devaient appuyer sur une touche correspondant à la séquence perçue le plus vite possible.

Rappelons que les résultats sur les participants normo-entendants juniors lors des expériences de close-shadowing bruitées et non-bruitées ont montré un apport de la

modalité visuelle dans le bruit, puisque les réponses aux stimuli audiovisuels ont été plus rapides que les réponses aux stimuli auditifs. Concernant la tâche, les réponses orales ont été plus rapides mais moins précises que les réponses manuelles, à la fois dans le bruit et sans bruit. Les participants normo-entendants seniors ont présenté des performances légèrement dégradées par rapport à celles des participants juniors dans les mêmes conditions (c'est-à-dire sans bruit), soit un peu inférieures en termes de reconnaissance et sensiblement ralenties en termes de réponse – mais des performances qui restent nettement supérieures à celles des juniors en condition bruitée. Notamment, il est à noter l'absence de bénéfice de la modalité visuelle pour les deux groupes dans la condition non-bruitée.

Nous avons proposé une explication de ces résultats à la lumière des théories perceptivo-motrices, en supposant que les deux modes de réponses seraient générés à deux étapes d'une boucle sensori-motrice : alors que les réponses manuelles seraient générées au stade final, lorsque l'input de parole a été comparé aux représentations motrices internes, les réponses orales, en revanche seraient formulées à un stade plus précoce, lorsque les commandes motrices sont générées. Ainsi les réponses orales seraient plus rapides que les réponses manuelles, mais surtout moins précises, puisque qu'il s'agirait seulement d'une première hypothèse, susceptible d'être corrigée lors du stade final de comparaison avec l'input sensoriel.

Nous nous attendions à obtenir, chez les participants implantés et du fait du déficit auditif, des temps de réaction plus longs et des pourcentages d'erreurs plus importants par rapport aux participants normo-entendants en ce qui concerne la modalité auditive seule. A la lumière des études précédentes montrant des capacités d'intégration audiovisuelle élevées chez les patients implantés (Bergeson et coll., 2003 ; Goh et coll., 2001, Kirk et coll., 2002 ; Kaiser et coll. 2003 ; Rouger 2007), nous nous attendions également à observer un apport important de la modalité visuelle chez ces sujets, entraînant des temps de réaction plus rapides et moins d'erreurs pour les stimuli audiovisuels. Concernant les interactions perceptivo-motrices, nous supposons que si le lien perceptivo-moteur des participants implantés n'était pas complètement rétabli, ceci serait de nature à, peut-être, modifier d'une certaine manière le patron de relations entre réponses orales et manuelles, que ce soit en termes de temps de réaction ou de pourcentages de réussite.

## **2. Méthodologie**

### **1. Stimuli**

Les stimuli utilisés pour l'expérience de close-shadowing chez la population implantée ont été ceux utilisés lors de l'expérience de close-shadowing sans bruit avec les normo-entendants juniors et seniors (voir partie I.3.2.1.)

### **2. Procédure expérimentale**

Les participants implantés ont réalisé l'expérience dans les mêmes conditions et selon la même procédure expérimentales que les participants normo-

entendants juniors et séniors (voir la partie I.2.4. pour la procédure générale aux trois expériences, et partie I.3.2.1. pour la procédure spécifique à l'expérience de close-shadowing.)

### 3. Analyses

Les analyses réalisées sur les réponses de la population implantée ont été identiques à celles décrites dans la partie I.3.2.1.

#### 3. Résultats implantés seuls

##### 1. Temps de réaction

Comme attendu, les stimuli audiovisuels ont été plus rapidement reconnus par rapport aux stimuli uniquement auditifs (626 ms vs. 669 ms) ( $F(1,9)=7,10$  ;  $p=.02$ ). De manière importante, comme pour les sujets normo-entendants, les réponses orales ont été nettement plus rapides que les réponses manuelles (469 ms vs. 826 ms) ( $F(1,9)=36,27$  ;  $p<.001$ ). L'effet de la syllabe a également été significatif, avec les syllabes /pa/ et /ka/ plus rapidement reconnues par rapport à la syllabe /ta/ (respectivement 586 ms, 621 ms et 736 ms) ( $F(2,18)=17,39$  ;  $p<.001$ ). Une interaction significative entre la modalité et la syllabe ( $F(2,18)=6,05$  ;  $p=.009$ ) montre que l'effet bénéfique de la présentation audiovisuelle est présent uniquement pour la syllabe /ta/ (audio : 782 ms ; audiovisuel : 690 ms). Aucune autre interaction significative n'a été observée (Voir Figure 61, Tableau 10 et Tableau 11).

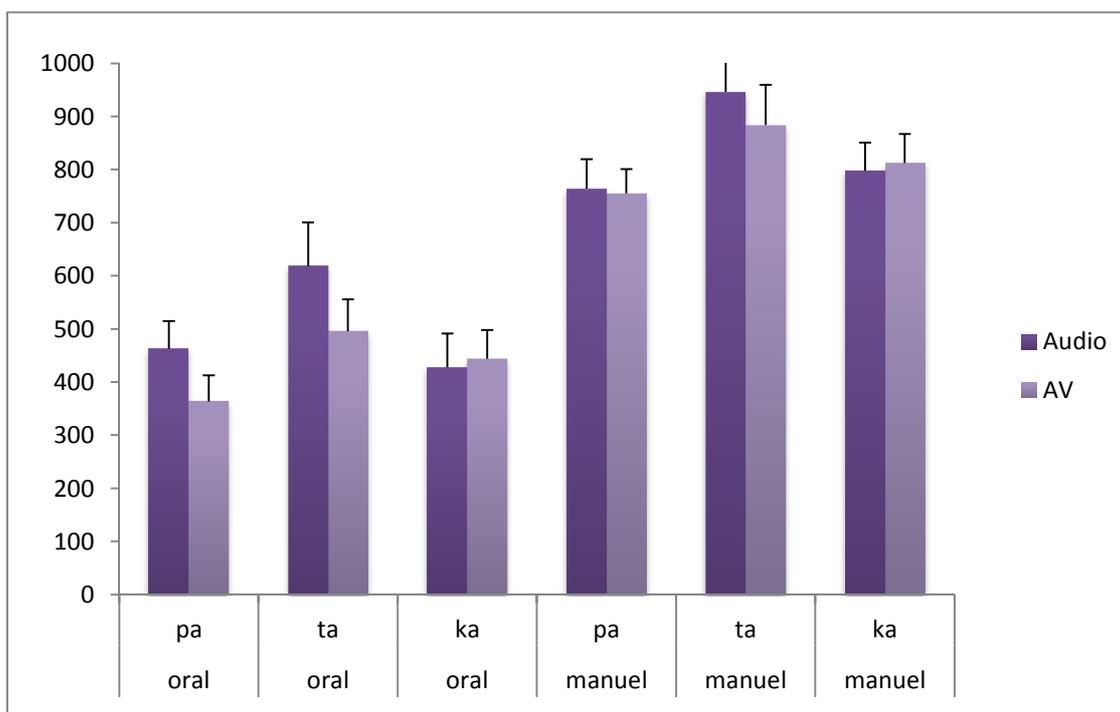


Figure 61 : Moyenne des temps de réaction pour les participants implantés (en ms, les barres d'erreurs correspondent aux écart-types).

	Temps de réaction	Proportion de réponses correctes
Modalité	p<.05	n.s.
Réponse	p<.001	n.s.
Syllabe	p<.001	p<.001
Modalité x Réponse	n.s.	n.s.
Modalité x Syllabe	p<.005	n.s.
Réponse x Syllabe	n.s.	n.s.
Modalité x Réponse x Syllabe	n.s.	n.s.

Tableau 10 : Effets significatifs et interactions pour les temps de réaction et les pourcentages de réponses correctes chez les participants implantés.

Modalité	Audio	Audio	Audio	Audio	Audio	Audio	AV	AV	AV	AV	AV	AV
Mode	oral	oral	oral	manuel	manuel	manuel	oral	oral	oral	manuel	manuel	manuel
Syllabe	pa	ta	ka	pa	ta	ka	pa	ta	ka	pa	ta	ka
RTs												
Implantés	462	618	428	764	946	798	364	496	444	755	884	812
% Implantés	73%	38%	79%	82%	45%	98%	96%	41%	78%	94%	51%	96%

Tableau 11 : Moyenne des temps de réaction (RTs, en ms) et des pourcentages de réponses correctes chez les participants implantés, par condition.

## 2. Pourcentages de réponse correctes

Contrairement à nos hypothèses, aucune différence significative n'a été observée entre les stimuli audio et audiovisuels (69% vs. 76%). De même, aucune différence significative n'a été observée entre les réponses orales et les réponses manuelles (67% vs. 78%). Par contre, les participants implantés ont fait moins d'erreurs pour les syllabes /pa/ (86%) et /ka/ (88%) par rapport à la syllabe /ta/ (44%) ( $F(2,18)=14,45$  ;  $p<.001$ ). Aucune interaction significative n'a été observée (voir Figure 62).

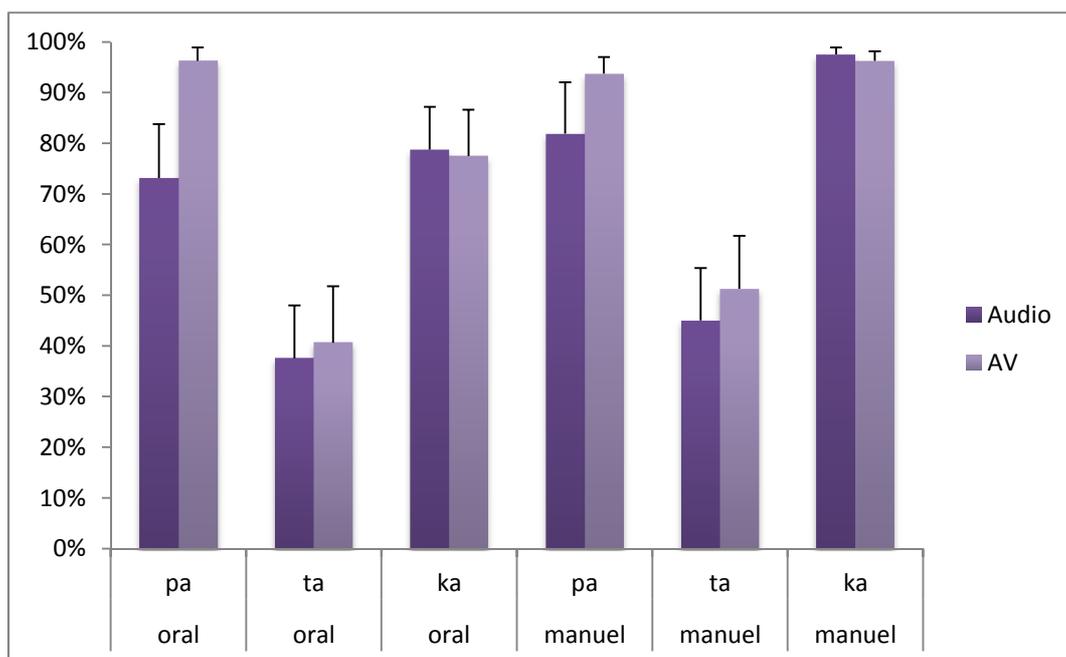


Figure 62 : Moyenne des pourcentages de réponses correctes pour les participants implantés (les barres d'erreurs correspondent aux écart-types).

### 3. Corrélation des résultats avec l'étiologie

En plus des analyses précédemment décrites, nous avons effectué des analyses de corrélation à partir des facteurs étiologiques afin de vérifier leur influence sur les résultats obtenus. Ainsi nous avons comparé les pourcentages des réponses correctes et les temps de réaction des tâches à réponses orales et manuelles avec l'âge des participants, la durée de surdité, l'âge d'implantation, l'âge de début de surdité et la durée d'expérience de l'implant. En raison des comparaisons multiples, une correction de Bonferroni a été utilisée, abaissant le seuil de significativité à  $p = .005$ . Aucune des corrélations ne s'est révélée significative, étant donné le petit nombre de participants, et la variabilité des résultats obtenus.

### 4. Comparaison avec les populations normo-entendantes

Nous avons dans un second temps comparé les résultats obtenus sur les sujets implantés avec ceux des sujets normaux-entendants juniors et séniors dans deux ANOVA à mesures répétées, en distinguant pour les sujets normo-entendants juniors la condition sans bruit (comparant donc les sujets sur une même tâche) et la condition bruitée (en testant d'éventuelles ressemblances entre sujets implantés testés sans bruit et sujets normo-entendants testés dans le bruit). Nous avons donc effectué deux ANOVA à mesures répétées sur ces données (une ANOVA pour les temps de réaction et la deuxième pour les pourcentages de réponses correctes), avec la variable groupe (participants implantés, participants normo-entendants juniors en condition non-bruitée, participants normo-entendants juniors en condition bruitée et participants normo-entendants séniors) comme variable inter-participant, et les variables tâche, modalité et syllabe comme variable intra-participants. Nous avons également réalisé des tests post-hoc avec correction de Bonferroni. Egalement, nous avons vérifié la sphéricité des

données grâce au test de Mauchly, et lorsque nécessaire, nous avons effectué des corrections de Greenhouse-Geiser (Greenhouse et Geiser, 1959). Enfin, du fait des comparaisons multiples lors de cette quatrième série d'analyses, nous avons fixé le seuil de significativité à  $p = .0125$  (sur ce point, il est de plus à noter que tous les effets et interactions observés avaient un seuil de significativité inférieur à  $p = .005$ ).

### 1. Temps de réaction

Comme attendu, l'effet principal du groupe est significatif ( $F(3,45)=24,43$  ;  $p<.001$ ), avec des temps de réaction pour les participants implantés plus longs que ceux des trois autres groupes. En revanche, aucune différence significative n'a été observée entre les participants séniors et les participants juniors bruités et non bruités, alors que les participants juniors non-bruités ont obtenu des temps de réaction significativement plus courts que ceux des participants juniors en condition bruitée (Implantés : 648 ms ; Juniors bruités : 484 ms ; Séniors : 418 ms ; Juniors non-bruités : 351 ms) (voir Figure 63 et Tableau 12).

Les effets principaux de la tâche ( $F(1,45)=263,95$  ;  $p<.001$ ) et de la modalité ( $F(1,45)=21,09$  ;  $p<.001$ ) sont aussi significatifs. Pour la tâche, les réponses orales étaient plus rapides que les réponses manuelles (313 ms vs. 600 ms). Par rapport à la modalité, les temps de réaction étaient plus courts dans la modalité audiovisuelle par rapport à la modalité auditive (446 ms vs. 467 ms).

Une interaction significative entre groupe et modalité ( $F(3,45)=6,93$  ;  $p<.001$ ) montre que l'effet bénéfique de la présentation audiovisuelle est présent chez les participants implantés (626 ms vs. 669 ms) et chez les participants normo-entendants juniors en condition bruitée (461 ms vs. 507 ms) alors qu'aucune différence n'a été observée chez les participants séniors (414 ms vs. 422 ms) ni chez les participants normo-entendants juniors en condition non-bruitée (354 ms vs. 349 ms). Ces effets semblent être dépendants des syllabes perçues. En effet, une interaction 'groupe x syllabes' montre que chez les participants normo-entendants juniors sans bruit, aucune différence n'a été observée entre les trois syllabes, alors que dans le bruit, la syllabe /pa/ a été plus rapidement reconnue que les syllabes /ta/ et /ka/ ; chez les normo-entendants séniors, la syllabe /pa/ a été plus rapidement reconnue que la syllabe /ta/, alors qu'aucune différence n'a été observée entre les syllabes /pa/ et /ka/, ni entre les syllabes /ka/ et /ta/ ; enfin, chez les participants implantés, les syllabes /pa/ et /ka/ ont été plus rapidement reconnues que la syllabe /ta/. Egalement, une interaction 'modalité x syllabes' a été constatée, montrant que l'effet de la modalité a été présent uniquement pour la syllabe /pa/. Enfin, une interaction 'groupe x modalité x syllabes' a été observée ( $F(6,90)=4,04$  ;  $p<.005$ ), montrant que l'effet de la modalité chez les participants implantés est présent uniquement pour la syllabe /ta/ (audio : 782 ms. ; audiovisuel : 690 ms) alors que chez les participants normo-entendants dans le bruit, l'effet de la modalité est présent uniquement pour la syllabe /pa/ (audio : 442 ms ; audiovisuel : 354 ms). Aucune autre interaction ne s'est révélée significative.

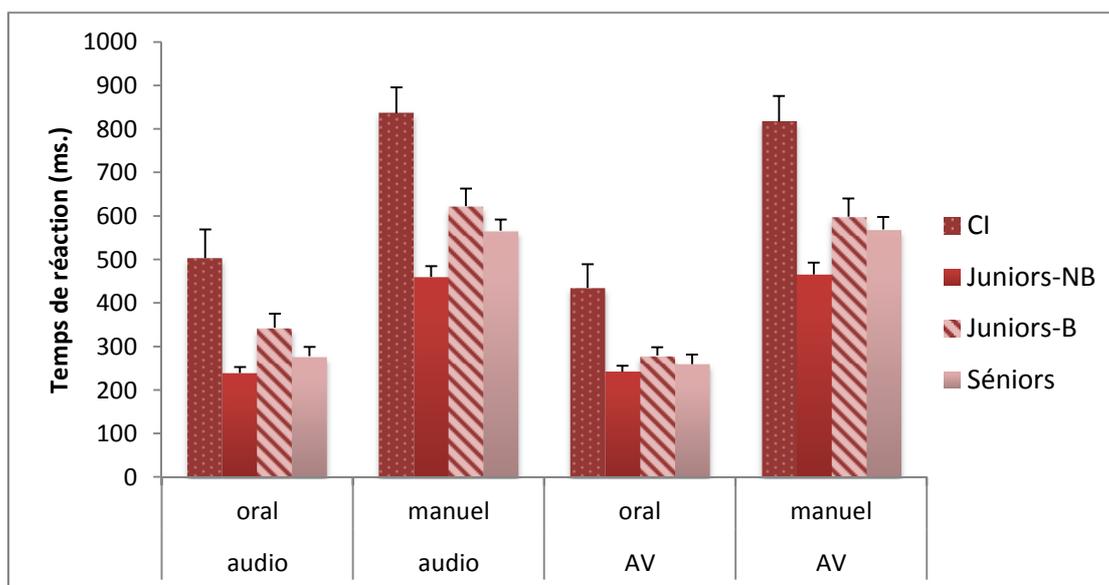


Figure 63 : Moyenne des temps de réaction chez les participants implantés (CI), chez les participants normo-entendants juniors en condition bruitée (Juniors – bruit) et non-bruitée (Juniors – non bruit), ainsi que chez les participants normo-entendants séniors (Séniors) (en ms, les barres d'erreurs correspondent aux écart-types).

	Temps de réaction	Proportion de réponses correctes
Groupe	p<.001	p<.001
Modalité	p<.001	p<.001
Réponse	p<.001	p<.001
Syllabe	p<.001	p<.001
Groupe x Modalité	P<.001	p<.001
Groupe x Réponse	n.s.	p<.005
Groupe x Syllabe	p<.005	p<.001
Modalité x Syllabe	p<.001	p<.001
Réponse x Syllabe	n.s.	p<.001
Groupe x Réponse x Syllabe	n.s.	p<.001
Groupe x Modalité x Syllabe	p<.005	p<.001
Groupe x Modalité x Réponse x Syllabe	n.s.	n.s.

Tableau 12 : Tableaux des effets significatifs et des interactions sur les temps de réaction et les pourcentages de réponses correctes lors de l'ANOVA comparant le groupe de participants implantés et les trois groupes de participants normo-entendants (Juniors bruités et non-bruités, Séniors)

### 3. Pourcentages de réponses correctes

L'effet principal du groupe a été significatif ( $F(3,45)=57,66$  ;  $p<.001$ ) avec une proportion de réponses correctes plus élevée pour les participants entendants juniors en condition non-bruitée (95%) ainsi que pour les participants normo-entendants séniors (89%) par rapport aux participants implantés (72%), et une proportion de réponses correctes plus élevées entre les participants implantés et les participants normo-entendants en condition bruitée (60%). Aucune différence significative n'a été observée entre les participants juniors en condition non-bruitée et les participants séniors (voir Figure 64).

Egalement, l'effet principal de la tâche a été significatif ( $F(1,45)=71,36$  ;  $p<.001$ ) avec une baisse des réponses correctes observée pour les réponses orales par rapport aux réponses manuelles (73% vs. 86%). De même, l'effet principal de la modalité a été significatif ( $F(1,45)=28,76$  ;  $p<.001$ ) avec des réponses plus précises pour les stimuli audiovisuels par rapport aux stimuli uniquement auditifs (82% vs. 77%).

Une interaction 'groupe x tâche' a été obtenue ( $F(3,45)=6,26$  ;  $p<.005$ ) montrant que la différence de précision entre les réponses orales et les réponses manuelles n'est présente que chez les participants séniors (81% vs. 98%) et les participants juniors dans le bruit (50% vs. 71%). De plus, une interaction 'groupe x modalité' ( $F(3,45)=12,97$  ;  $p<.001$ ) montre que l'effet bénéfique de la modalité n'est présent que chez les participants normo-entendants dans le bruit (audiovisuel : 68% ; audio : 53%).

Comme pour les temps de réaction, les résultats dépendent de la syllabe présentée. En effet, une interaction 'groupe x syllabe' a été observée, montrant que les participants séniors et juniors sans bruit n'ont présenté aucune différence entre les trois syllabes, alors que pour les participants juniors dans le bruit la syllabe /pa/ a été mieux reconnue que la syllabe /ta/, et aucune différence n'a été constatée entre les syllabes /pa/ et /ka/, ni entre les syllabes /ka/ et /ta/ ; chez les participants implantés, les syllabes /pa/ et /ka/ ont été mieux reconnues que la syllabe /ta/. Egalement, une interaction 'modalité x syllabes' a été constatée montrant que l'effet de la modalité audiovisuelle n'a été présent que pour la syllabe /pa/. De même, une interaction 'réponse x syllabes' a été observé, montrant que lors des réponses orales, la syllabe /pa/ a été mieux reconnue que la syllabe /ka/, elle-même mieux reconnue que la syllabe /ta/, alors que lors des réponses manuelles, aucune différence n'a été observée entre les syllabes /pa/ et /ka/ mais elles ont été mieux reconnues que la syllabe /ta/. Une interaction 'groupe x modalité x syllabe' a été obtenue ( $F(6,90)=4,72$  ;  $p<.001$ ), montrant que pour les participants juniors en condition bruitée, l'effet de la modalité n'est présent que pour la syllabe /pa/ (audiovisuel : 92% ; audio : 61%).

Enfin, une interaction 'groupe x tâche x syllabe' ( $F(6,90)=5,07$  ;  $p<.001$ ) montre que chez les participants juniors en condition non-bruitée les syllabes sont toutes parfaitement reconnues ; chez les participants séniors, lors des réponses manuelles, aucune différence n'a été observée entre les trois syllabes alors que lors des réponses orales, la syllabe /pa/ a été mieux reconnue que les syllabes /ka/ et /ta/, chez les

participants implantés, lors des réponses orales et manuelle, les syllabes /pa/ et /ka/ ont été mieux reconnues que la syllabe /ta/ ; et enfin chez les participants normo-entendants en condition bruitée, lors des réponses orales la syllabe /pa/ est mieux reconnue que les syllabes /ta/ et /ka/ tandis que lors des réponses manuelles, les syllabes /pa/ et /ka/ sont mieux reconnues que la syllabe /ta/.

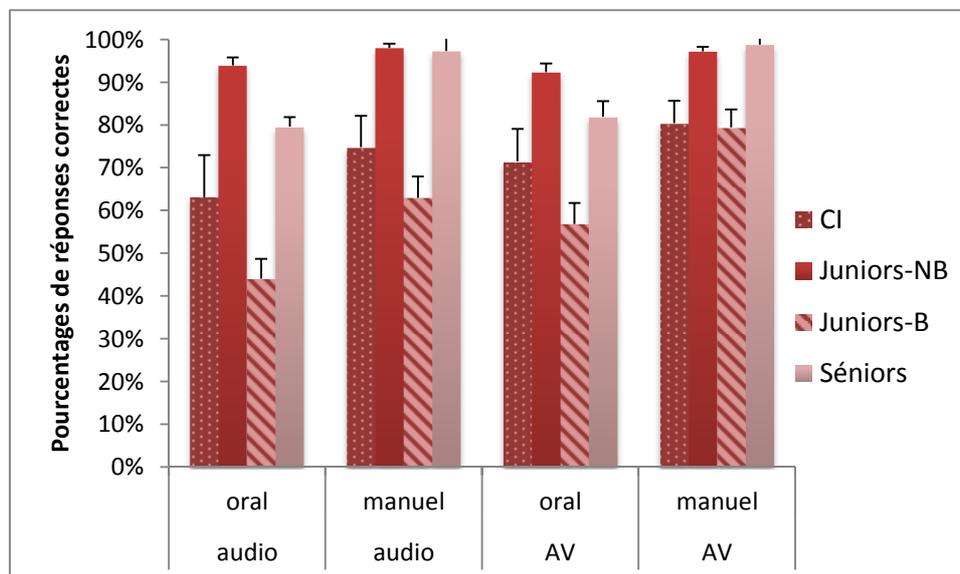


Figure 64 : Moyenne des pourcentages de réponses correctes chez les participants implantés (CI) et les participants normo-entendants juniors en condition bruitée (Juniors – B) et non-bruitée (Juniors-NB) et les participants séniors (les barres d'erreurs correspondent aux écart-types)

## 5. Discussion

En résumé, conformément à nos hypothèses, les réponses pour les participants implantés ont été moins rapides que celles des trois groupes de participants entendants (Juniors en condition bruitée et non-bruitée, Séniors). Leurs réponses se sont également révélées moins précises en comparaison de celles des normo-entendants en condition non-bruitée et des séniors. Par contre, les participants implantés ont produit moins d'erreurs que les participants juniors en condition bruitée.

Concernant la modalité de présentation des stimuli, nous supposons que les participants implantés bénéficieraient de l'aide de l'addition du visuel tant en termes de rapidité et qu'en précision de réponse. Cela a été partiellement le cas, puisque les participants implantés, comme les participants normo-entendants en condition bruitée, ont obtenu des réponses plus rapides pour les stimuli audiovisuels. Par contre, alors que les participants normo-entendants en condition bruitée ont présenté des réponses plus précises pour les stimuli audiovisuels, aucune différence entre les deux modalités de présentation n'a été observée sur le plan de la précision de réponse chez les participants implantés.

Ces résultats apparaissent en lien avec les études précédemment menées sur les capacités de lecture labiale des patients implantés (Goh et collègues, 2001 ; Kirk et

collègues, 2002 ; Bergeson et collègues, 2003 ; Kaiser et collègues, 2003 ; Rouger et collègues, 2012), qui montrent un rôle majeur de la vision chez des sujets implantés cochléaires et également la possibilité d'une meilleure capacité à intégrer les informations visuelles et auditives chez ces participants.

Enfin, les réponses orales ont été plus rapides que les réponses manuelles, et ce chez les quatre groupes de participants, entendants et implantés. La différence entre les deux modes de réponses en terme de précision ne s'est pas avérée significative chez les participants implantés, contrairement aux participants normo-entendants juniors dans le bruit, dont nous avons interprété dans le premier chapitre les réponses orales moins précises à la lumière de la boucle perceptivo-motrice de Skipper et collègues (2007). L'absence de différence significative pourrait provenir à la fois d'un relatif effet plafond (avec de meilleures performances globales des implantés par rapport aux juniors dans le bruit) et de la variabilité des performances des implantés, renforcée par le petit nombre de sujets dans le groupe étudié.

Reste que la différence de temps de réponse entre les réponses orales et manuelles suggère, chez les patients implantés, qu'un certain lien perceptivo-moteur semble avoir été restauré et/ou maintenu, même s'il est, peut-être, moins efficace que chez les personnes normo-entendantes.

#### **4. Expérience d'idiosyncrasies**

##### **1. Principes et hypothèses**

La seconde expérience réalisée chez les participants implantés était l'expérience dite d'idiosyncrasies. Comme les participants normo-entendants juniors et séniors, l'expérience comportait une tâche de production des voyelles orales du français, et une tâche de perception de stimuli auditifs, à catégoriser entre les voyelles /i e ε a/, l'ordre de ces tâches étant contrebalancé entre les participants.

Chez les participants normo-entendants juniors et séniors, nous n'avons pas observé de corrélation individuelle entre la production et la perception des voyelles /e/ et /ε/, en raison du manque de variabilité observée lors de la tâche de perception. Par contre, lors de la tâche de production, nous avons observé des idiosyncrasies entre les participants. Une quasi équité-répartition des voyelles sur l'axe F1 a été observée tant en production qu'en perception, conduisant à une corrélation très forte entre les valeurs acoustiques produites et perçues des voyelles. A noter enfin que ces corrélations étaient cependant moins fortes chez les normo-entendants séniors, étant données les modifications de valeurs moyennes et les augmentations de variance en production.

Deux éléments d'interprétation ont été proposés pour expliquer ces phénomènes. D'une part, l'équidistance entre les voyelles illustre le principe de dispersion maximale de Lindblom et collègues (1986), qui suppose qu'une organisation optimale des phonèmes a lieu dans les langues afin de fournir un contraste maximal entre les différents phonèmes. Ensuite, l'équidistance entre les voyelles attestée en production et en

perception suggère que la variation observée lors de la tâche de production serait organisée autour de gabarits communs, correspondant à ceux de la tâche de perception.

Chez les participants implantés, nous supposons observer une plus grande variabilité entre les participants lors des deux tâches indépendamment. Par contre, notre principale interrogation concernait la présence ou non d'une corrélation globale entre les valeurs acoustiques perçues et produites : nous nous sommes demandés si les participants implantés, malgré une plus grande variabilité, allaient pouvoir varier autour de gabarits communs, tant en perception qu'en production.

## **2. Méthodologie**

### **1. Stimuli**

Les stimuli utilisés pour l'expérience d'idiosyncrasies chez la population implantée ont été les mêmes que ceux utilisés lors de la même expérience chez la population normo-entendante « junior » (voir partie I.4.2.2.)

### **2. Procédure expérimentale**

Les participants implantés ont réalisé l'expérience dans les mêmes conditions et selon les mêmes procédures expérimentales que les participants normo-entendants juniors et séniors (voir partie I.2.4. pour la procédure générale aux trois expériences, et partie I.4.2.2. pour la procédure spécifique à l'expérience d'idiosyncrasies)

### **3. Analyses**

Les analyses réalisées sur les réponses de la population implantée ont été identiques à celles décrites dans la partie 4.2.4.

## **3. Résultats population sourde implantée**

### **1. Production**

Dans la tâche de production, les moyennes des valeurs de F1 pour /i/, /e/, /ɛ/ et /a/ en Barks ont été respectivement de 2,95 (avec une étendue de 2,09 à 3,53), 4,15 (avec une étendue de 3,3 à 5,13), 5,58 (avec une étendue de 3,32 à 6,54) et de 6,79 (avec une étendue de 6,01 à 7,5).

Les idiosyncrasies ont été faibles pour /e/ avec des valeurs de distance normalisée à partir de /i/ allant de 0,14 à 0,43 Barks mais avec un écart-type réduit à 0,09 (voir Figure 65). Par contre, les idiosyncrasies pour /ɛ/ ont été plus fortes, avec des valeurs de distance normalisées à partir de /i/ allant de 0,28 à 0,96 Barks, avec un écart-type de 0,23 (voir Figure 66).

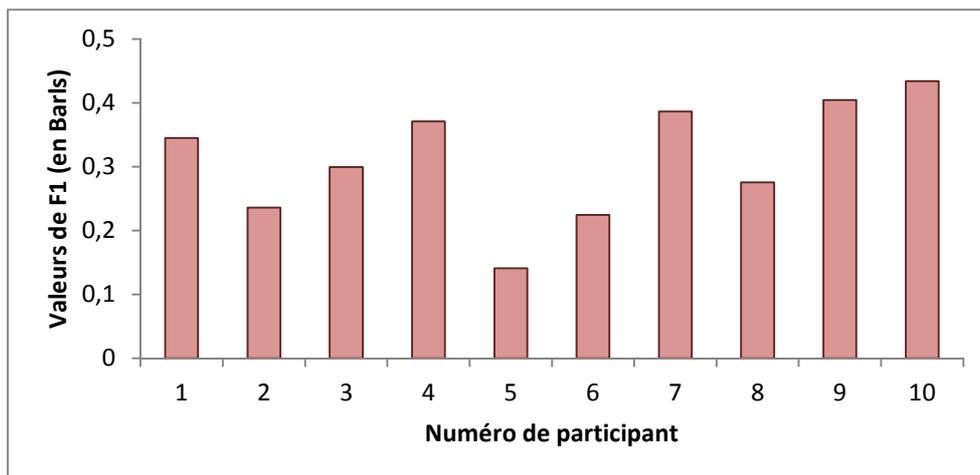


Figure 65 : Valeurs normalisées de F1 (en Barks) pour /e/ produite par chaque participant implanté

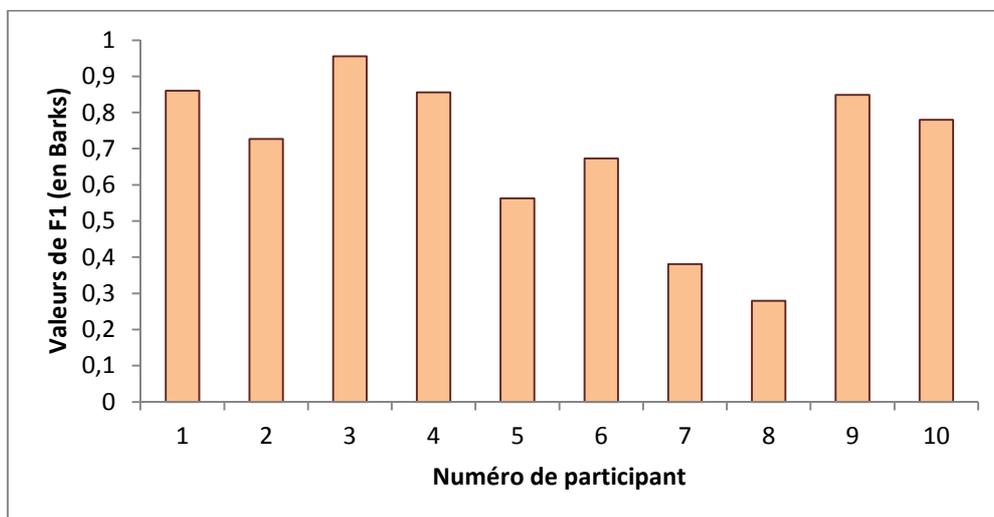


Figure 66 : Valeurs normalisées de F1 (en Barks) pour /ε/ produite par chaque participant implanté

## 2. Perception

Dans la tâche de perception, les moyennes des valeurs de F1 en Barks pour /i/, /e/, /ε/ et /a/ ont été respectivement de 2,81 (avec une étendue de 2,2 à 3,14), 4,03 (avec une étendue allant de 3,19 à 4,71), 5,25 (avec une étendue allant de 4,76 à 5,56) et 6,56 (avec une étendue allant de 6,1 à 6,8).

La variabilité s'est confirmée faible en perception quoique plus élevée chez les implantés que chez les normo-entendants. Les valeurs de distance normalisée pour la voyelle /e/ à partir de /i/ étaient de 0,11 à 0,45 Barks avec un écart-type de 0,11 (Voir Figure 67). Pour la voyelle /ε/ les valeurs de distance normalisées à partir de /i/ étaient de 0,54 à 0,7 Barks, avec un écart-type faible de 0,04 (voir Figure 68).

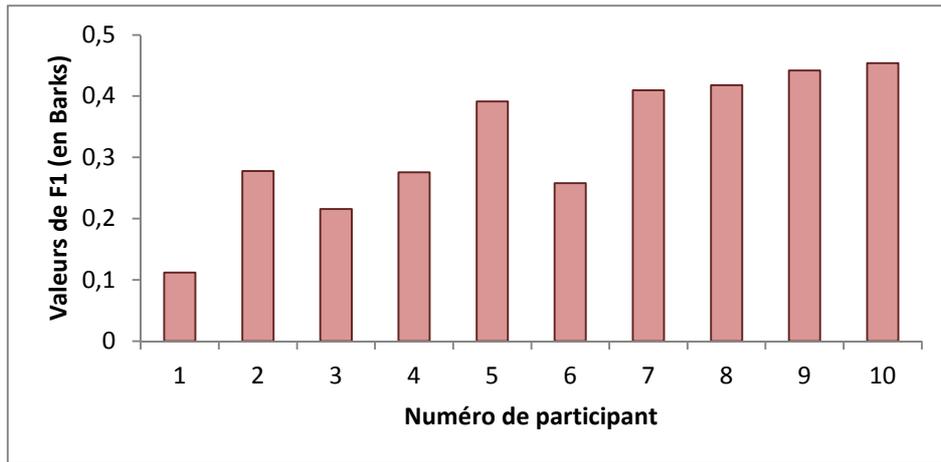


Figure 67 : Valeurs normalisées de F1 lors de la perception de la voyelle /e/, pour chaque participant implanté

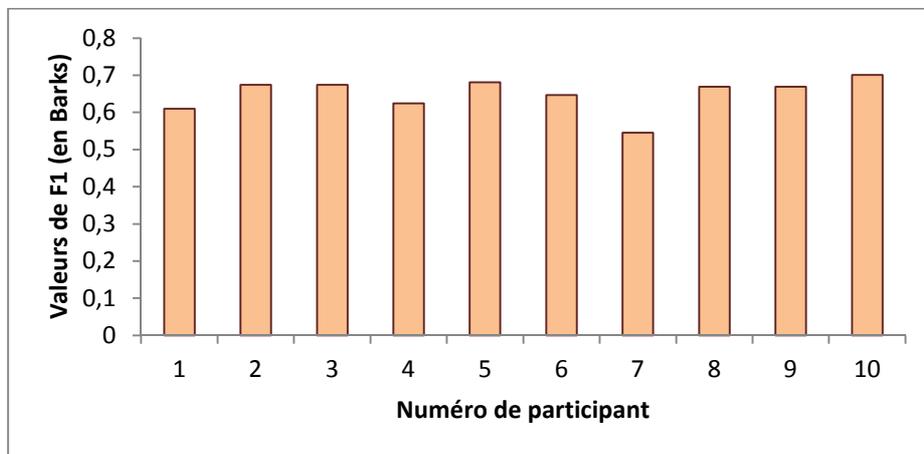


Figure 68 : Valeurs normalisées de F1 lors de la perception de la voyelle /ε/, pour chaque participant implanté

### 3. Corrélation perception/production

Comme pour les populations normo-entendantes, nous n'avons pas obtenu de corrélation entre production et perception, c'est-à-dire que les valeurs moyennes de F1 produites n'ont pas été corrélées avec les valeurs moyennes perçues, ni pour la voyelle /e/ ( $t=0,57$ ,  $p > .5$ ,  $r=.2$  voir Figure 69) ni pour la voyelle /ε/ ( $t=0,57$ ,  $p > .4$ ,  $r=.27$ , voir Figure 70).

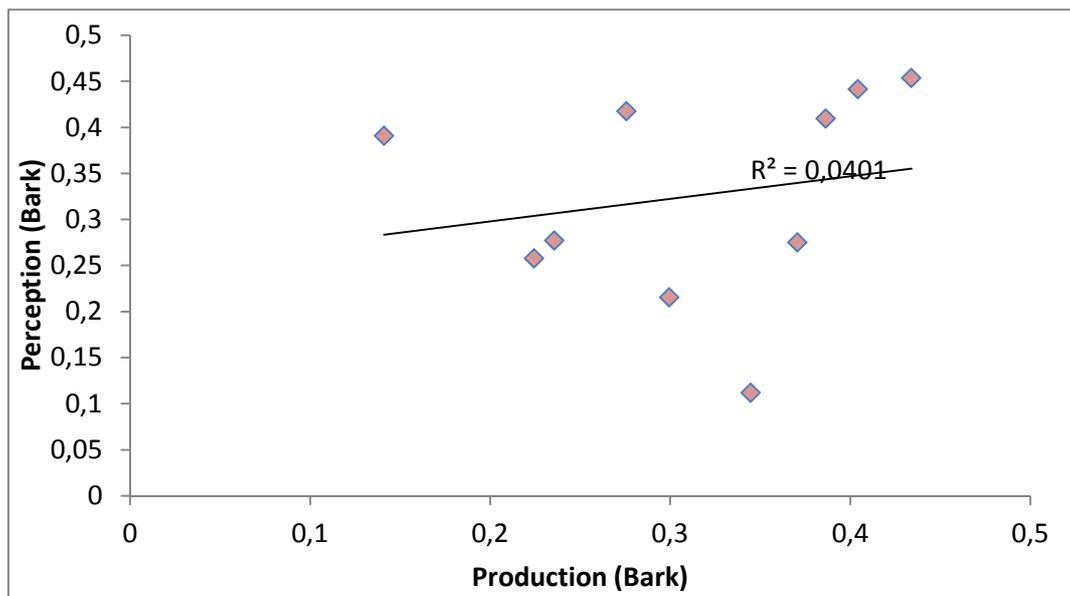


Figure 69 : Corrélacion perception-production pour la voyelle /e/ chez les participants implantés.

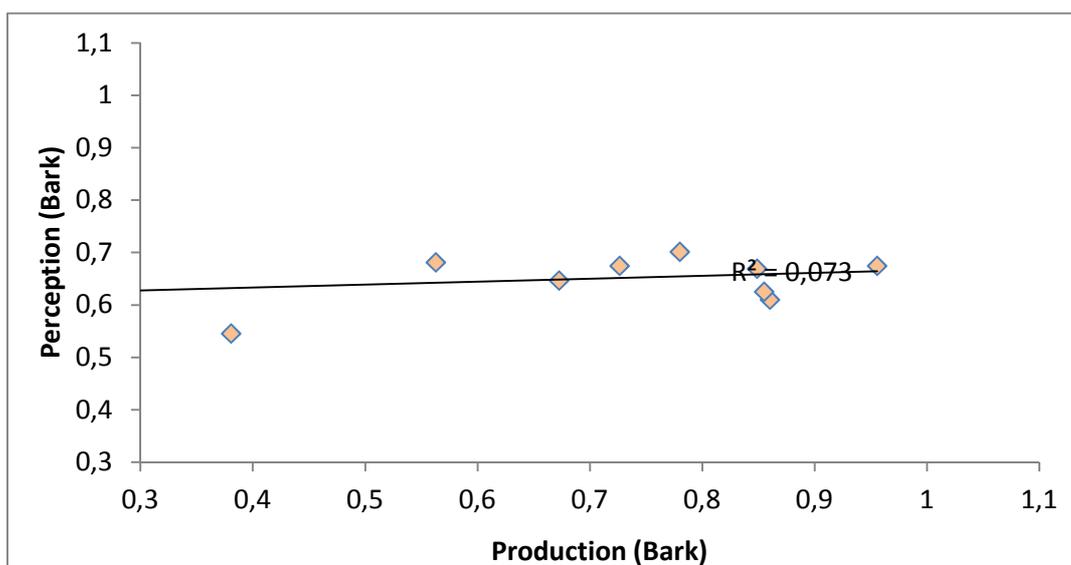


Figure 70 : Corrélacion perception-production pour la voyelle /ε/ chez les participants implantés.

Néanmoins, les variations des valeurs perceptives apparaissant pour les implantés cochléaires plus grandes que pour les normo-entendants, nous nous sommes demandés comment évoluait la distance entre les deux voyelles cibles, voyelles /e/ et /ε/, et il apparaît une corrélation entre cette distance en production et en perception ( $t(8)=3,46$  ;  $p<.01$ ) (voir Figure 71). Cette corrélation positive apparaît même si l'on écarte les deux sujets ne réalisant pas du tout de contraste /e/ et /ε/ en production (Figure 72). Elle semble indiquer que la réduction du contraste en production influe sur la manière dont les sujets positionnent leurs catégories en perception, en les conduisant à rapprocher aussi la voyelle mi-haute et la voyelle mi-basse en production. Ceci fournit un

élément de corrélation perceptivo-motrice inattendu mais très intéressant, sur lequel nous reviendrons dans notre discussion générale.

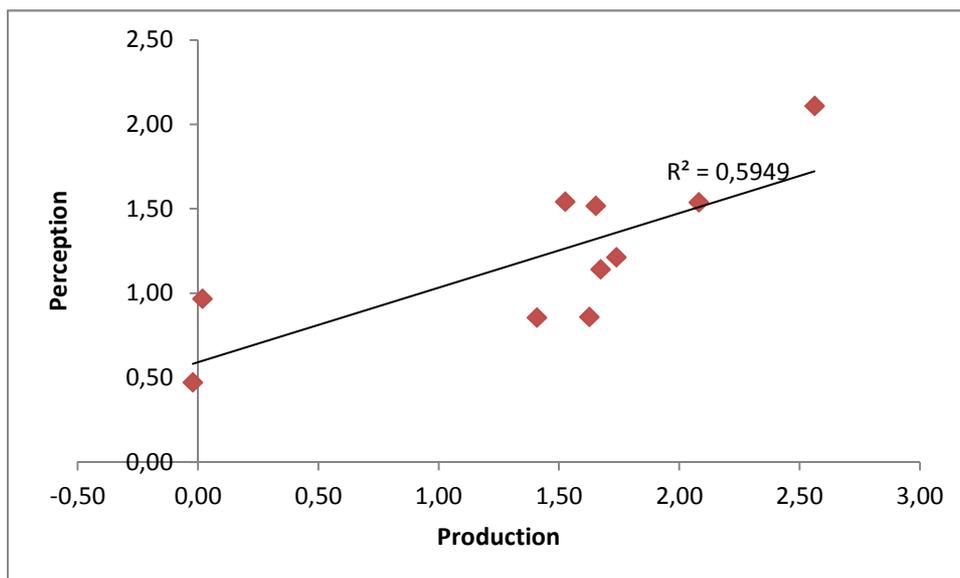


Figure 71 : Corrélation entre les distances /ε/ - /e/ lors des tâches de perception et de production.

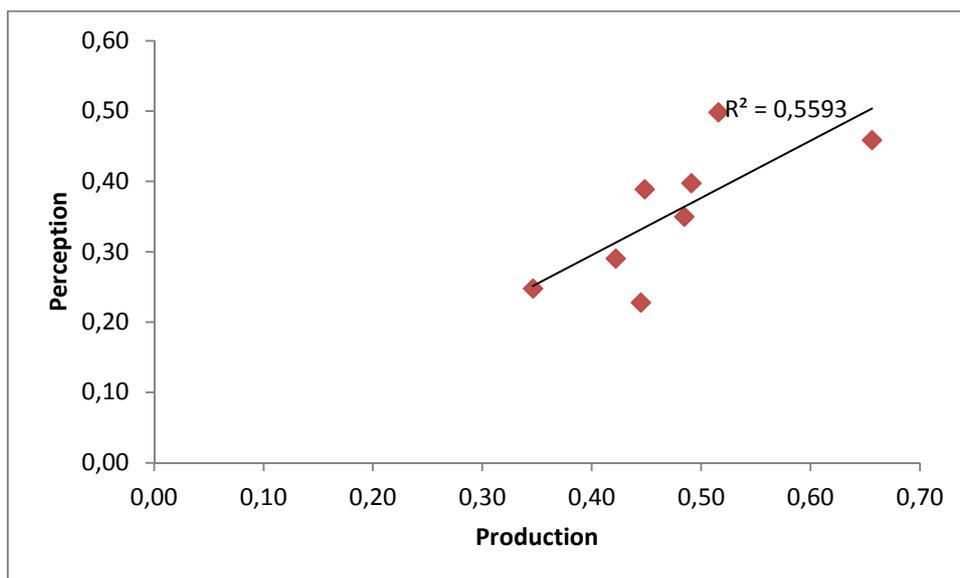


Figure 72 : Corrélation entre les distances /ε/ - /e/ lors des tâches de perception et de production sans les deux participants ne présentant pas de contrastes entre les deux voyelles lors de la tâche de production.

Enfin, à partir des valeurs moyennes des quatre voyelles, nous avons obtenu une corrélation forte des valeurs acoustiques entre les voyelles produites et perçues, avec une pente moyenne pour tous les participants de 0,93, allant de 0,73 à 1,2,  $r=0,9$  (voir Figure 73).

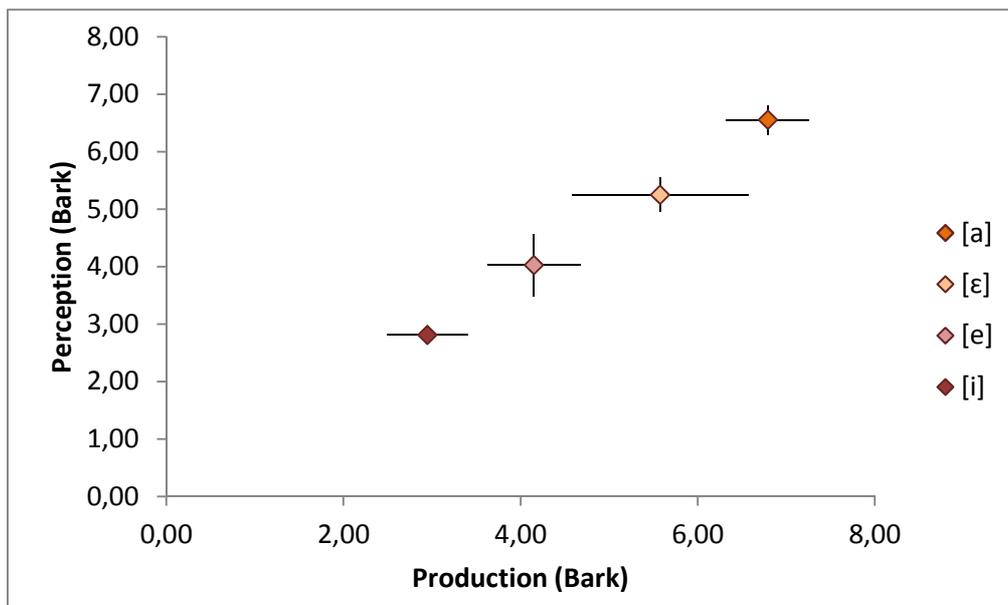


Figure 73 : Corrélation entre les valeurs acoustiques produites et perçues par les participants implantés (les barres d'erreurs correspondent aux écart-types)

Comme chez les participants normo-entendants, nous avons finalement mesuré le nombre de stimuli par catégorie. En moyenne, 23% des stimuli ont été catégorisés comme /a/, mais avec une variation entre 13 et 35% ; 26% des stimuli ont été catégorisés comme /ε/ mais avec une variation entre 13 et 41% ; 25% des stimuli ont été catégorisés comme /e/ mais avec une variation de 10 à 43% ; et 26% des stimuli ont été catégorisés comme /i/ mais avec une variation de 1 à 40%. (voir Figure 74). Ainsi comme illustré par la Figure 75, la proportion relative de stimuli catégorisés comme /e/ et /ε/ est très variable selon les participants, comme chez les participants normo-entendants juniors et séniors, ce qui ne peut expliquer la stabilité des valeurs des catégories perceptives.

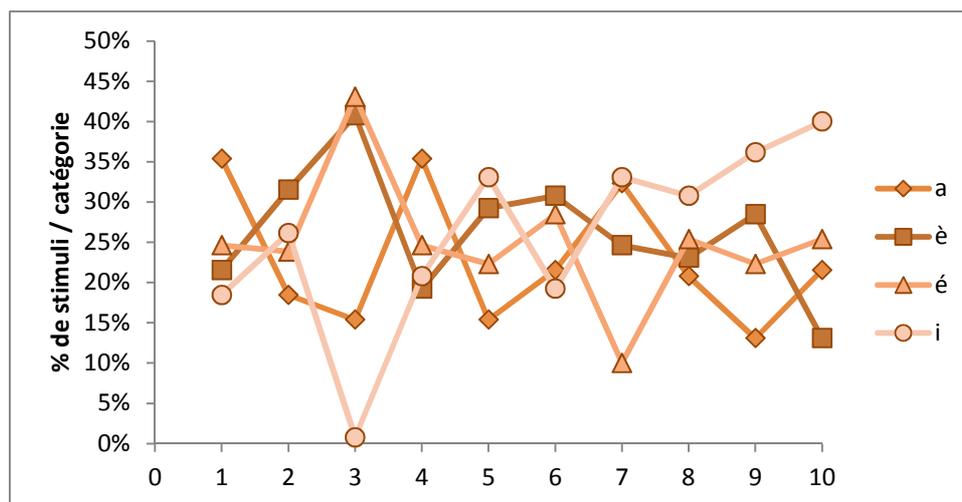


Figure 74 : Pourcentages de stimuli par catégorie, chez chaque participant implanté.

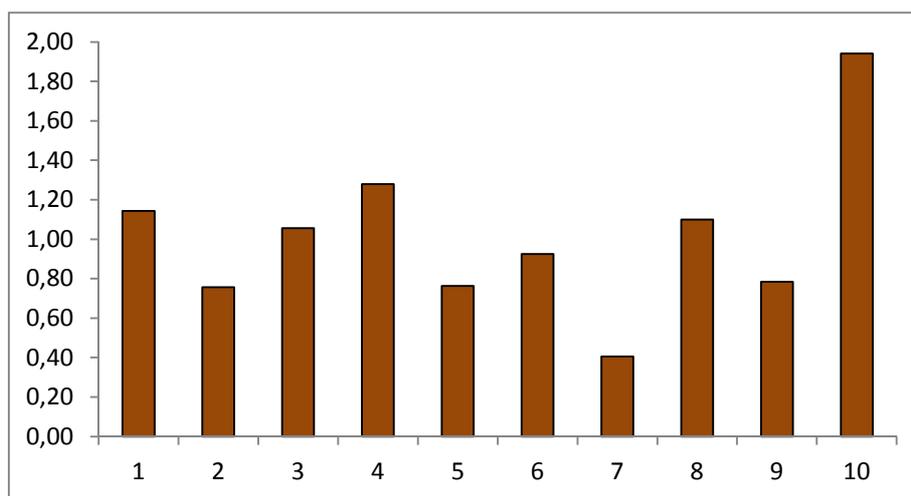


Figure 75 : Rapport entre les pourcentages relatifs de stimuli catégorisés comme les voyelles /e/ et /ɛ/, par participant implanté.

#### 4. Corrélation facteurs étiologiques

En plus des analyses précédemment décrites, nous avons effectué des analyses de corrélation à partir des facteurs étiologiques afin de vérifier leur influence sur les résultats obtenus. Ainsi nous avons comparé les valeurs de pentes des valeurs acoustiques produites et perçues et les valeurs de distance [e / ɛ] en production et en perception avec l'âge des participants, la durée de surdité, l'âge d'implantation, l'âge de début de surdité et la durée d'expérience de l'implant. En raison des comparaisons multiples, une correction de Bonferroni a été utilisée, abaissant le seuil de significativité à 0,005.

Aucune corrélation significative n'a été observée entre l'étiologie et les valeurs de pente des valeurs acoustiques produites et perçues par les participants implantés, les distances lors de la tâche de production ou les distances lors de la tâche de perception.

#### 4. Comparaison avec les populations normo-entendante « junior » et « sénior »

##### 1. Tâches de production et de perception

Afin de comparer la population implantée avec les populations normo-entendantes, nous avons réalisé une ANOVA à mesure répétée avec la variable groupe (Implantés vs. Juniors vs. Séniors) comme variable inter-participant, et les variables tâche (Perception vs. Production) et voyelles (avec les valeurs des moyennes des F1 perçus et produits pour les voyelles /i e ɛ a/, ainsi que les valeurs normalisées calculées pour les voyelles /e/ et /ɛ/), comme variable intra-participants. Lorsque nécessaire, nous avons réalisé des tests post-hoc avec correction de type Bonferroni. Egalement, nous avons vérifié la sphéricité des données grâce au test de Mauchly, et lorsque nécessaire, nous avons effectué des corrections de Greenhouse-Geiser (Greenhouse et Geiser, 1959). Du fait des comparaisons multiples lors de cette troisième série d'analyses, nous avons fixé le seuil de significativité à  $p = .017$ .

L'effet du groupe s'est révélé significatif ( $F(2,43)=5,64$  ;  $p<.01$ ), avec des valeurs moyennes de F1 (perçus et produits, toutes voyelles confondues) plus basses pour les participants implantés et séniors par rapport aux participants juniors (Implantés : 3,34 Barks ; séniors : 3,33 Barks ; juniors : 3,5 Barks). L'effet de la tâche est également significatif ( $F(1,43)=6,69$  ;  $p<.013$ ) avec des valeurs moyennes de F1 produites plus élevées que les valeurs de F1 perçues (3,52 Barks vs. 3,34 Barks). Enfin, l'effet principal de la voyelle est significatif ( $F(5,215)=5027,76$  ;  $p<.001$ ), toutes les moyennes obtenues pour les voyelles étant significativement différentes entre elles (en Barks, /i/=2,88 ; /e/=4,19 ; /ε/=5,55 ; /a/=7 ; valeurs normalisées moyennes : /e/=.32, /ε/=.65). Une interaction groupe x voyelle ( $F(10,215)=5,7$  ;  $p<.001$ ) révèle qu'en moyenne, les valeurs de F1 de la voyelle /a/ sont significativement plus basses chez les participants implantés par rapport aux participants séniors, alors que les valeurs de F1 pour les voyelles /i e ε/ ainsi que les valeurs normalisées des voyelles /e/ et /ε/ sont similaires entre les deux groupes. Par rapport aux participants juniors, les participants implantés ont obtenu des valeurs de F1 plus basses pour les voyelles /a/ et /ε/. Egalement, une interaction entre les trois facteurs ('groupe x tâche x voyelles') a été observée ( $F(10,215)=2,7$  ;  $p<.001$ ), révélant que lors de la tâche de production les participants séniors ont produit, pour la voyelle /a/ des valeurs de F1 plus hautes que les participants implantés (aucune différence pour les trois autres voyelles), alors qu'en perception, aucune différence n'a été observée entre les participants implantés et séniors. Par rapport aux participants juniors, aucune différence n'a été observée ni en perception ni en production.

Il apparaît donc (voir Tableau 13 et Figure 76) que, si les sujets implantés respectent la même double optimalité en perception et en production, ceci s'accompagne en production d'une tendance à la centralisation des valeurs produites, ce qui est conforme à la revue de littérature présentée précédemment. Aucune autre interaction n'a été observée ('groupe x tâche' :  $F(2,43)=1,12$  ;  $p=.33$ ).

		Implantés	Juniors	Séniors
Production	/a/	6,79	7,32	7,58
	/ɛ/	5,58	5,95	5,17
	/e/	4,15	4,43	3,81
	/i/	2,95	3,05	2,91
	/e/ normalisé	.31	.32	.20
	/ɛ/ normalisé	.69	.67	.50
Perception	/a/	6,56	6,80	6,79
	/ɛ/	5,25	5,52	5,31
	/e/	4,03	4,21	4,06
	/i/	2,81	2,76	2,72
	/e/ normalisé	.33	.36	.33
	/ɛ/ normalisé	.65	.68	.64

Tableau 13 : Valeurs moyennes de F1 pour chaque tâche et chaque groupe (en Barks).

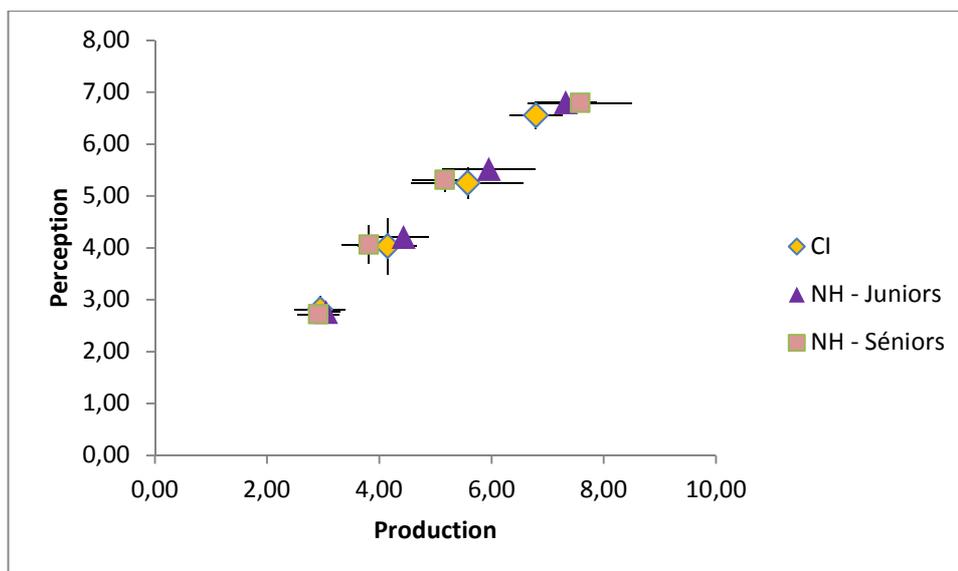


Figure 76 : Corrélation entre les valeurs acoustiques globales produites et perçues chez les trois groupes de participants (les barres d'erreurs correspondent aux écart-types).

Nous avons également voulu vérifier la variabilité des participants lors des deux tâches. Pour ce faire, nous avons réalisé deux ANOVAs : une première ANOVA avec les écart-types de chaque participant lors de la production des quatre voyelles, puis une seconde ANOVA avec les écart-types mesurés lors de la tâche de perception – voir Partie II 4.4. pour la définition de ces écarts-types.

Aucune différence n'a été observée entre les trois groupes de participants en regard de la variabilité de production des voyelles ( $F(2,43)=1,21$  ;  $p=.3$ ). En revanche, nous avons obtenu un effet de la voyelle ( $F(3,129)=15,69$  ;  $p<.001$ ), où les écart-types mesurés pour les voyelles /a/ et /ɛ/ ont été significativement plus élevés par rapport aux voyelles /e/ et /i/ (en Barks : /a/ : .41 ; /ɛ/ : .38 ; /e/ : .22 ; /i/ : .21).

Concernant la variabilité lors de la catégorisation des stimuli, l'effet du groupe est significatif ( $F(2,43)=14,89$  ;  $p<.001$ ), les participants implantés ayant présenté plus de variabilité que les participants entendants juniors et séniors (Implantés : .54 ; Juniors : .43 ; Séniors : .40). L'effet de la voyelle est aussi significatif ( $F(3,129)=60,47$  ;  $p<.001$ ). Enfin, de manière plus intéressante, une interaction groupe x voyelles ( $F(6,129)=4,32$  ;  $p<.001$ ) a été observée, montrant que les participants implantés ont présenté plus de variabilité pour les voyelles /a/ et /ɛ/ par rapport aux participants entendants juniors et séniors ; en revanche aucune différence avec les deux groupes d'entendants n'a été observée pour les voyelles /e/ et /i/. Aucune différence n'a été observée entre les groupes de participants juniors et séniors (Valeurs en Barks : Implantés : /a/=.39 /ɛ/=.60 /e/=.64 /i/=.53 ; Juniors : /a/=.22 /ɛ/=.60 /e/=.47 /i/=.39 ; Séniors : /a/=.22 /ɛ/=.65 /e/=.35 /i/=.40).

## 2. Corrélation production-perception

Afin de comparer la corrélation des valeurs acoustiques entre les voyelles produites et perçues, nous avons réalisé une ANOVA sur les valeurs de pente entre les trois groupes de participants. Concernant la corrélation des valeurs acoustiques entre les voyelles produites et perçues, aucune différence significative n'a été observée entre les participants séniors et les participants implantés ( $F(2,43)=0,73$  ;  $p=.5$ ) (juniors :  $r=.93$  ; implantés :  $r=.89$  ; séniors :  $r=.94$ ).

## 5. Discussion

En résumé, nous observons dans la tâche de production, chez les implantés, comme chez les entendants, des idiosyncrasies pour les valeurs normalisées des voyelles /e/ et /ɛ/. Ces idiosyncrasies respectent le même principe d'optimalité que pour les normo-entendants juniors ou seniors, c'est-à-dire avec une équi-répartition des positions moyennes des 4 voyelles sur l'axe F1. Par contre, cette similarité de distribution s'accompagne d'une centralisation des valeurs pour les sujets implantés, avec notamment des valeurs pour /a/ significativement réduites.

Lors de la tâche de perception, les participants implantés ont présenté plus de variabilité intra-participant par rapport aux participants normo-entendants. En conséquence, nous avons pu observer des idiosyncrasies en perception, ce qui n'avait pas

été possible chez les normo-entendants. Cependant, la variabilité en perception n'était pas assez forte pour pouvoir observer une corrélation entre la production et la perception pour aucune des deux voyelles testées, chez les participants implantés, tout comme cela avait été le cas chez les participants normo-entendants.

Par contre, de manière intéressante et inattendue, une corrélation sur la distance /e/-/ɛ/ entre production et perception a été constatée. Enfin, comme chez les participants normo-entendants juniors, nous avons obtenu une corrélation très forte entre les valeurs acoustiques globales produites et perçues, avec des voyelles réparties régulièrement sur l'axe F1, tant en production qu'en perception. Nous reviendrons sur ces différents points lors de la discussion générale.

## **5. Expérience de convergence et imitation**

### **1. Principes et hypothèses**

Comme chez les participants normo-entendants, la troisième expérience utilisée chez les participants implantés était donc l'expérience de convergence et d'imitation. Les participants ont dû réaliser une première tâche de production des voyelles /e o/ selon une consigne orthographique. Ensuite, les participants devaient produire les mêmes voyelles, cette fois selon une consigne auditive, une première fois sans qu'aucune instruction d'imiter la voix perçue ne leur soit donnée (tâche dite de convergence), puis une seconde fois en imitant volontairement la voix perçue (tâche d'imitation). Les stimuli utilisés lors de ces deux tâches étaient modifiés en fonction de la valeur moyenne de F0 du participant, calculée lors de la première tâche.

Pour rappel, chez les participants entendants juniors, comme chez les participants séniors, nous avons observé une imitation quasi-parfaite de la fréquence fondamentale lors de la tâche d'imitation, et lors de la tâche de convergence, les participants ont convergé involontairement vers la cible présentée, cette convergence se révélant moins élevée par rapport à celle de la tâche d'imitation. L'unique différence constatée entre les participants entendants juniors et séniors consistait dans la variabilité inter-participants, la convergence et l'imitation étant plus variables chez les séniors. Ces résultats ont permis de montrer l'adaptation des commandes motrices à l'environnement, c'est-à-dire en fonction de l'input sensoriel reçu par les participants, adaptation nécessitant des interactions sensori-motrices lors de la production de parole.

Par rapport aux participants implantés, la question était de savoir s'ils étaient capables d'estimer la fréquence fondamentale d'un stimulus, et d'adapter leur production à cette estimation. Le paradigme utilisé permettait d'évaluer la capacité des participants implantés à produire et à percevoir la fréquence fondamentale, mais surtout leur capacité à converger vers la valeur de F0 de la cible auditive, lorsqu'on leur demande d'imiter consciemment la voix perçue d'une part, et d'autre part, lorsqu'aucune instruction d'imitation ne leur est donnée, l'interrogation sous-jacente était de vérifier, encore une fois, si les participants implantés étaient capables d'associer des paramètres auditifs et moteurs.

## **2. Méthodologie**

### **1. Stimuli**

Les stimuli utilisés pour l'expérience de convergence et imitation chez la population implantée ont été les mêmes que ceux utilisés lors de la même expérience chez la population normo-entendante « junior » (voir partie I.5.2.2.).

### **2. Procédure expérimentale**

Les participants implantés ont réalisé l'expérience dans les mêmes conditions et selon les mêmes procédures expérimentales que les participants normo-entendants « juniors » (voir première partie. Population normo-entendante « junior » pour la procédure générale aux trois expériences, et partie I.5.2.3. pour la procédure spécifique à l'expérience de convergence et imitation).

### **3. Analyses**

Les analyses réalisées sur les réponses de la population implantée ont été identiques à celles décrites dans la partie I.5.2.4.

## **3. Résultats population sourde implantée**

### **1. Résultats généraux**

Des changements imitatifs ont été observés dans les deux tâches, même s'ils étaient plus importants dans la tâche d'imitation volontaire. Les coefficients de pente étaient significativement différents de zéro à la fois dans la tâche dite de « convergence » ( $t(9)=3,24$ ;  $p<0,02$ ) et dans la tâche d'imitation volontaire ( $t(9)=4,84$ ;  $p<0,001$ ).

Par contre, les coefficients de pente étaient plus importants dans la tâche d'imitation volontaire (0,38) par rapport à la tâche dite de « convergence » (0,14) ( $t(9)=3,53$ ;  $p<0,001$ ) (voir Figure 78, Figure 79 et Figure 80). De même, les coefficients de corrélation étaient significativement différents de zéro à la fois dans la tâche de convergence ( $t(9)=4,08$ ;  $p<0,01$ ) et dans celle d'imitation ( $t(9)=5,53$ ;  $p<0,001$ ). Ces coefficients de corrélation étaient également plus élevés dans la tâche d'imitation ( $r=0,49$ ) par rapport à ceux de la tâche de convergence ( $r=0,24$ ) ( $t(14)=2,95$ ;  $p<0,02$ ) (voir Figure 81).

Concernant le test JND, les participants implantés ont obtenu une moyenne de 14,06 Hz (les résultats allant de 1,71 à 45,35 Hz). (voir Figure 77)

Enfin, aucune corrélation n'a été observée entre la tâche de convergence et la tâche d'imitation ( $r^2=.28$ ). De même, aucune corrélation n'a été constatée entre les valeurs de pente et les résultats obtenus lors du test de discrimination de fréquence ni pour la tâche de convergence ( $r^2=.02$ ) ni pour la tâche d'imitation ( $r^2=.11$ ).

Les pourcentages d'erreurs de production par rapport à la voyelle cible ont été de 15% lors de la tâche de convergence et de 13% lors de la tâche d'imitation.

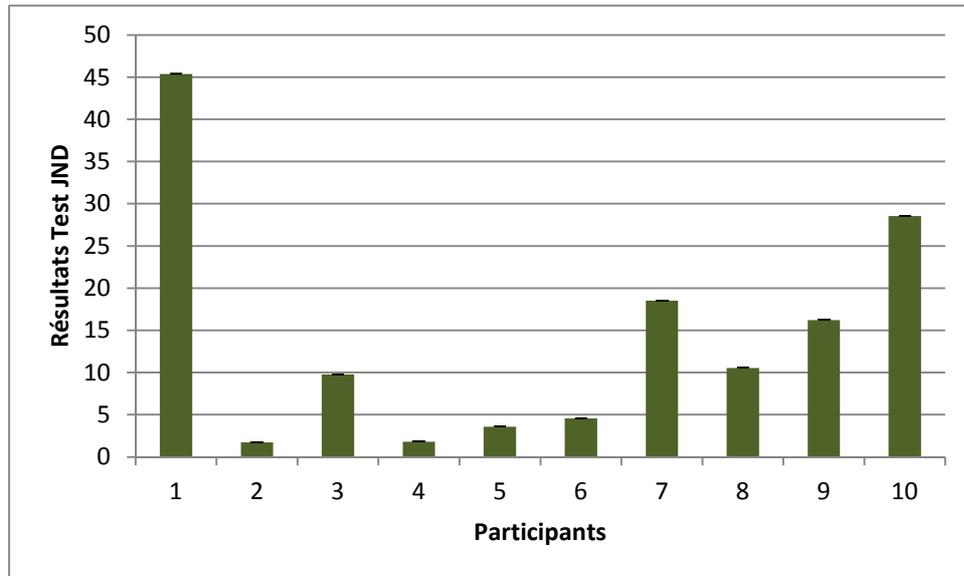


Figure 77 : Résultats obtenus lors du test JND chez les participants implantés

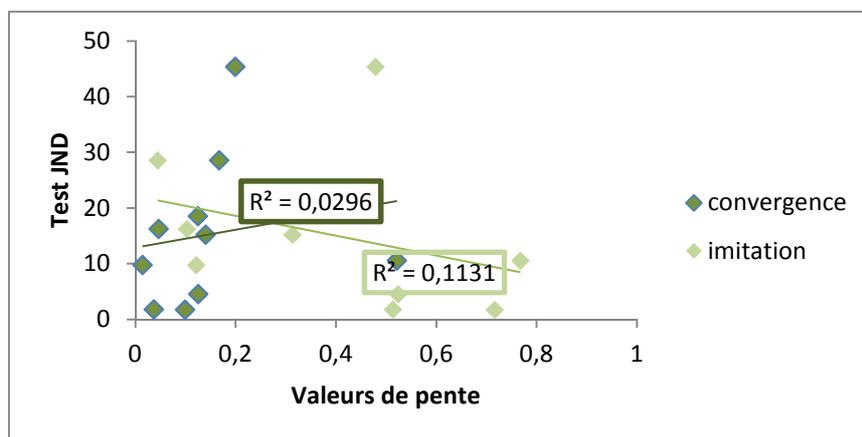


Figure 78 : Corrélation entre les valeurs de pente des deux tâches et les résultats du test de JND.

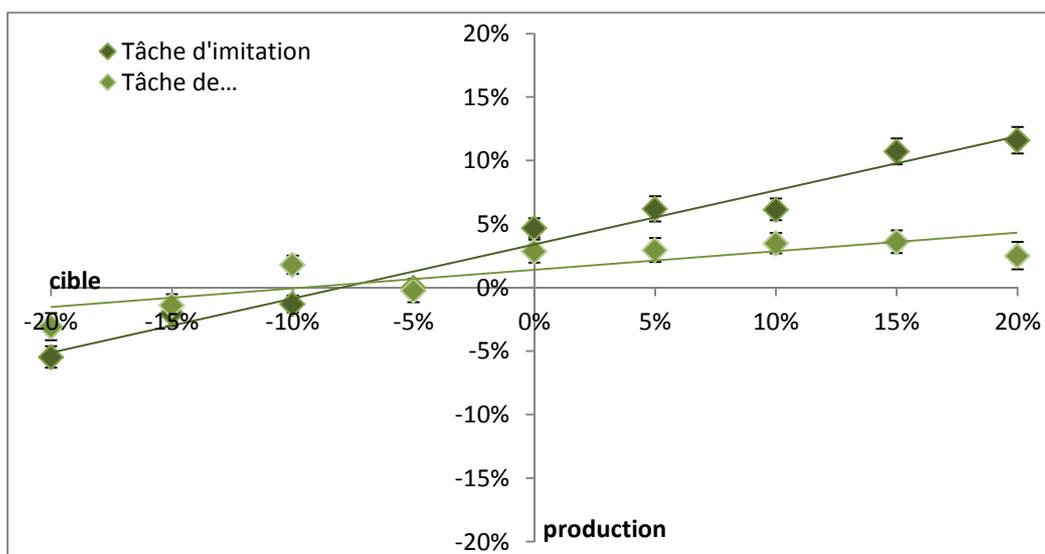


Figure 79 : Moyenne de pourcentage de changement lors des tâches de convergence et d'imitation chez les participants implantés (les barres d'erreurs correspondent aux écarts-types).

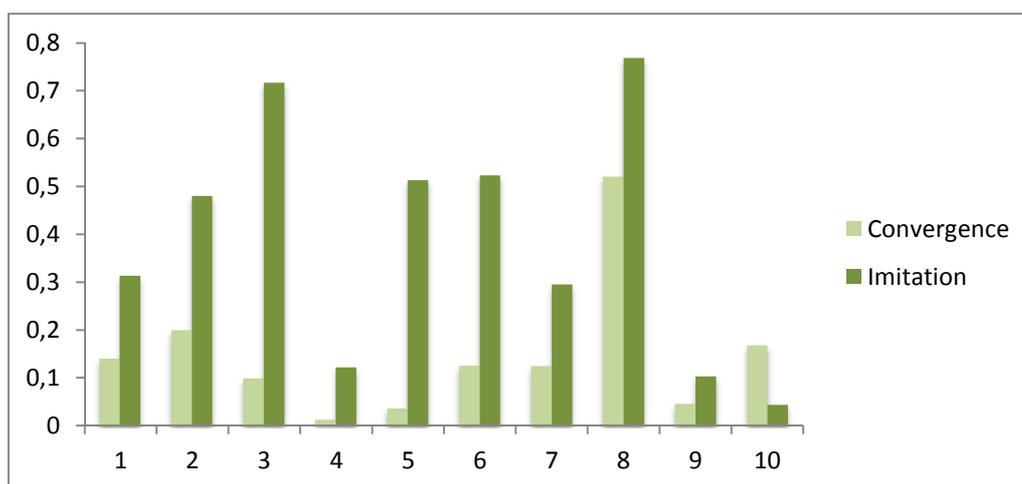


Figure 80 : Valeurs de pente par participant implanté, dans les deux tâches.

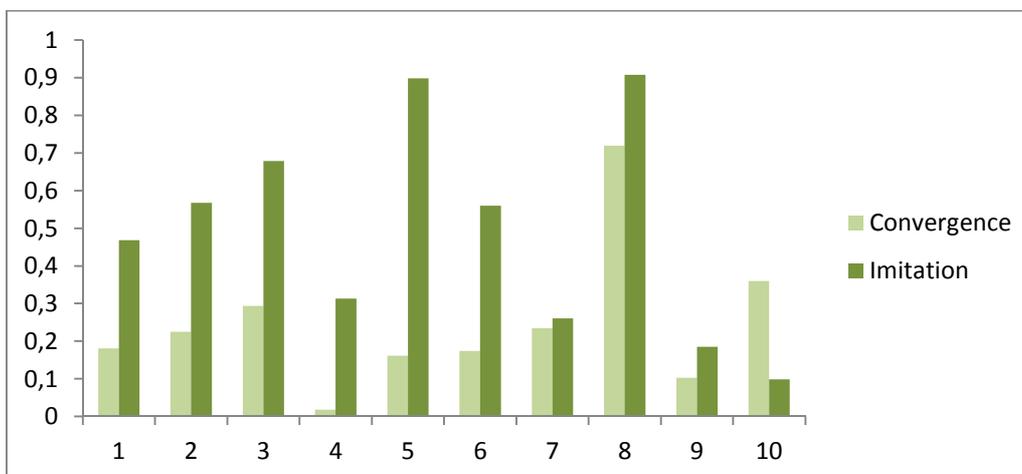


Figure 81 : Valeur de coefficient de corrélation par participant implanté, dans les deux tâches

## 2. Corrélation avec les facteurs étiologiques

En plus des analyses précédemment décrites, nous avons effectué des analyses de corrélation à partir des facteurs étiologiques afin de vérifier leur influence sur les résultats obtenus. Ainsi nous avons comparé les valeurs de pentes et les pourcentages d'erreurs lors des tâches de convergence et d'imitation avec l'âge des participants, la durée de surdité, l'âge d'implantation, l'âge de début de surdité et la durée d'expérience de l'implant. En raison des comparaisons multiples, une correction de Bonferroni a été utilisée, abaissant le seuil de significativité à 0,005.

Aucune corrélation significative n'a été observée entre l'étiologie et les valeurs de pente, ni lors de la tâche d'imitation ni lors de la tâche de convergence.

## 4. Comparaison avec les populations normo-entendantes

Afin de comparer les résultats de la population séniors ceux des populations normo-entendante « juniors » et implantée, nous avons réalisé une ANOVA, avec la variable groupe (implantés vs. Juniors vs. Séniors) comme facteur inter-participants et la variable tâche (Convergence vs. Imitation) comme facteur intra-participants. Du fait des comparaisons multiples lors de cette troisième série d'analyses, nous avons fixé le seuil de significativité à  $p = .017$ . Egalement, nous avons vérifié la sphéricité des données grâce au test de Mauchly, et lorsque nécessaire, nous avons effectué des corrections de Greenhouse-Geiser (Greenhouse et Geiser, 1959).

Nous avons obtenu un effet du groupe ( $F(2,32)=31,3$  ;  $p<.001$ ) où les participants implantés ont obtenu des valeurs de pentes moyennes plus basses que les participants entendants juniors et séniors (implantés : .26 ; juniors : .65 ; séniors : .55), mais aucune différence entre les participants juniors et séniors n'a été observée. Nous avons également obtenu un effet de la tâche ( $F(1,32)=67,29$  ;  $p<.001$ ) où lors de la tâche de convergence, les coefficients de pentes ont été significativement plus faibles par rapport à la tâche d'imitation (.33 vs. .69). En revanche, aucune interaction significative n'a été observée entre les variables groupe et tâche ( $F(2,32)=1,16$  ;  $p=.19$ ) (voir Figure 82).

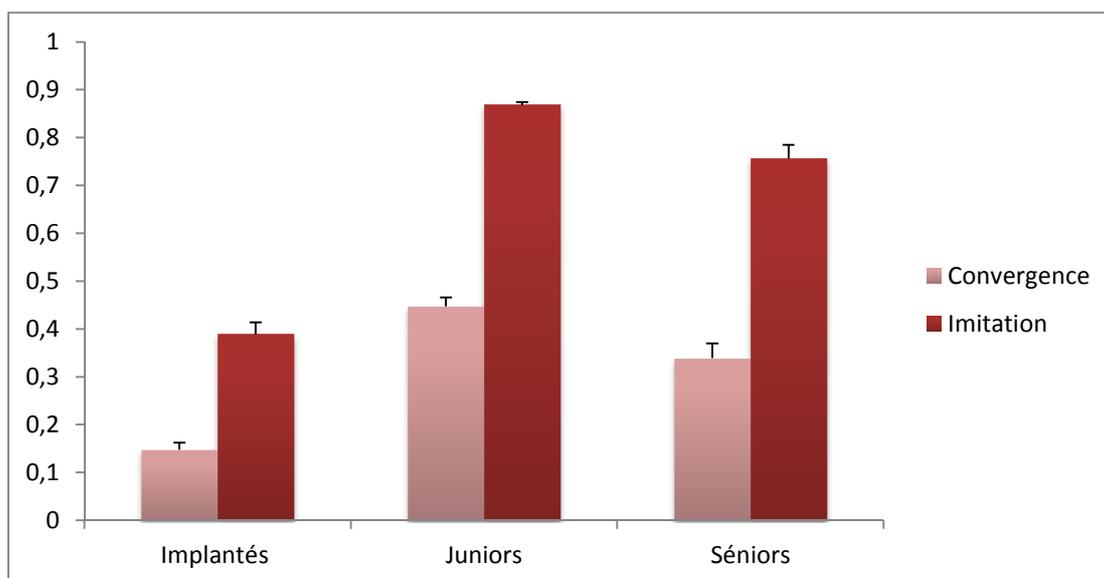


Figure 82 : Coefficients de pente lors des tâches de convergence et d'imitation chez les participants implantés, les participants normo-entendants juniors, et les participants séniors (les barres d'erreurs correspondent aux écart-types).

Une autre différence importante entre les participants implantés et les participants normo-entendants était la présence de pourcentages d'erreurs de production : alors que les participants normo-entendants n'en ont réalisé aucune, les participants implantés ont présenté 15 % d'erreurs lors de la tâche de convergence et 13 % lors de la tâche d'imitation.

Enfin, les participants implantés ont une erreur standard plus faible que les participants séniors, mais égale à celle des participants juniors lors de la tâche de convergence (Implantés : 0,06 ; séniors : 0,1 ; Juniors : 0,07) alors que lors de la tâche d'imitation les participants implantés ont eu autant de variabilité que les séniors, et les participants juniors ont eu une erreur type plus basse que celles des deux autres groupes (Implantés : 0,08 ; Séniors : 0,9 ; Juniors : 0,02).

## 5. Discussion

En résumé, comme les normo-entendants, les participants implantés ont convergé vers la cible auditive à atteindre lors des deux tâches. En revanche plusieurs différences séparent les groupes. Tout d'abord contrairement aux participants normo-entendants, les réponses des participants implantés n'ont pas toujours été correctes, ce qui signifie que la perception des voyelles isolée est difficile pour eux. En effet, il a été montré que les auditeurs implantés ont besoin des transitions formantiques pour catégoriser précisément une voyelle (Hana, 2001).

Ensuite, les participants implantés ont moins convergé vers les cibles par rapport aux deux groupes de participants entendants, et ce lors des deux tâches. Enfin, il est important de noter que la variabilité inter-participants de la population implantée a été importante dans les deux tâches, comme chez la population séniors, alors que chez les participants juniors, la variabilité a été faible lors de la tâche d'imitation, et plus élevée

lors de la tâche de convergence. Cette variabilité importante, associée aux taux de convergence plus faibles, suggère que les participants implantés ne maîtrisent pas complètement leur voix et donc leur production.

Cependant, le résultat majeur de cette expérience est donc que les participants implantés peuvent imiter et converger vers une cible acoustique, ce qui montre qu'ils sont donc capables d'estimer la fréquence fondamentale d'une voix cible et de gérer leur propre voix afin d'adapter leur production pour être le plus près possible de cette cible. Mieux encore, ces participants le font lors d'un paradigme de convergence, alors qu'aucune instruction d'imitation de la cible ne leur a été donnée. Ces capacités de convergence vers une cible auditive, conscientes et inconscientes, suggèrent, une nouvelle fois, que les participants implantés ont maintenu ou retrouvé une certaine sorte de lien perceptivo-moteur, que nous interpréterons dans la discussion générale.

## 6. Conclusion

Dans ce chapitre, nous avons proposé d'évaluer les liens perceptivo-moteurs au sein d'une population de patients sourds post-linguaux porteurs d'un implant cochléaire au travers du protocole expérimental testé auparavant chez deux groupes de jeunes adultes et de séniors normo-entendants.

Lors du premier protocole expérimental de close-shadowing, visant à explorer le format partiellement moteur de stimuli auditifs et audiovisuels de parole, nos hypothèses initiales concernant la population implantée étaient d'abord que les performances seraient dégradées, tant au niveau de la rapidité que de la précision de réponse. D'autre part, nous pensions observer un bénéfice de la modalité audiovisuelle par rapport à la modalité auditive, au travers de réponses soit plus précises soit plus rapides. Enfin, notre questionnement principal était de savoir si nous allions retrouver des différences entre les réponses orales et manuelles comme chez les participants normo-entendants, traduisant une activation du lien perceptivo-moteur. Conformément à nos hypothèses, nous avons obtenu des réponses moins précises et des temps de réaction plus lents chez la population implantée par rapport aux populations normo-entendantes. Également, nous avons obtenu des réponses plus rapides pour les réponses aux stimuli audiovisuels par rapport aux stimuli uniquement auditifs, démontrant un rôle important de la lecture labiale chez les participants implantés. Enfin, de manière frappante, les participants implantés ont présenté des réponses orales plus rapides que les réponses manuelles, suggérant une activation du lien sensori-moteur lors de la perception de parole chez ces participants.

Concernant le second paradigme expérimental, où nous proposons d'étudier des corrélations entre la production et la perception de voyelles, nous nous attendions à observer une plus grande variabilité chez les participants implantés, et nous souhaitions déterminer si cette population présenterait une corrélation entre les valeurs acoustiques produites et perçues. Contrairement à nos hypothèses, par rapport à la variabilité de production des voyelles, aucune différence n'a été observée entre les participants

implantés et les participants normo-entendants. Par contre, les participants implantés ont catégorisé les voyelles de manière plus variable que les participants normo-entendants. Ensuite, nous avons également obtenu un résultat inattendu au travers d'une corrélation entre les distances /e/-/ɛ/ entre perception et production. Enfin, malgré une variabilité plus importante lors de la perception des voyelles, nous avons constaté une corrélation très forte entre les valeurs acoustiques produites et perçues chez les participants implantés, comme chez les participants normo-entendants, permettant encore une fois de montrer des relations sensori-motrices actives chez ces participants.

Lors du troisième paradigme, composé de tâches d'imitation consciente et inconsciente de la fréquence fondamentale, nos prédictions consistaient d'une part en une plus grande variabilité de production chez la population implantée, et d'autre part en des capacités d'imitation plus faibles chez ces participants. Concernant les résultats, nous avons constaté des erreurs de production chez ce groupe de participants, alors qu'aucune erreur n'avait été observée chez les participants normo-entendants, suggérant une difficulté de perception des voyelles isolées au sein de la population implantée ; également, les participants implantés ont présenté plus de variabilité lors des deux tâches et de plus, la convergence vers la cible a été plus faible chez ces participants, traduisant une maîtrise incomplète de la voix. Cependant, même si les participants implantés semblent moins performants que les participants normo-entendants, ils ont présenté des capacités de convergence vers la cible auditive, que ce soit lors de l'imitation consciente ou inconsciente, ce qui suppose une adaptation de leurs commandes motrices à l'input auditif, et donc des liens sensori-moteurs actifs lors de la production de parole.





# Discussion générale

---



Au terme de ce travail de thèse, il nous semble important de rappeler le cadre général dans lequel il s'inscrit, ainsi que les objectifs principaux de la thèse. Cette thèse se situe dans le domaine des théories perceptivo-motrices de la communication parlée, supposant une collaboration étroite entre systèmes moteurs et auditifs aussi bien lors de la perception que de la production de parole. L'objectif général de ce travail était dans un premier temps d'évaluer ces possibles liens perceptivo-moteurs auprès d'une population de jeunes adultes normo-entendants, puis en fonction d'une déprivation sensorielle et d'un réapprentissage sensorimoteur chez des patients sourds post-linguaux implantés mais aussi en regard d'un déclin supposé des fonctions cognitives et langagières chez une population normo-entendante âgée. Pour évaluer les liens sensorimoteurs, un protocole expérimental spécifique a été élaboré, composé de trois paradigmes comportementaux portant sur la production et la perception de parole.

Lors de cette discussion, nous allons tout d'abord présenter une synthèse des principaux résultats, avant de proposer des éléments de perspectives de recherche. Nous distinguerons dans une première partie les résultats obtenus au sein de chacun des trois paradigmes expérimentaux utilisés lors de ce travail de thèse, et ce pour les trois populations. Ensuite, dans une seconde et une troisième partie, nous détaillerons plus particulièrement l'état des relations sensori-motrices lors de la perception de parole chez les seniors, d'une part, et chez les patients sourds post-linguaux porteurs d'un implant cochléaire, d'autre part.

## **1. Synthèse des principaux résultats**

### **1. Toutes populations confondues**

#### **1. Premier paradigme expérimental : close-shadowing**

Le premier paradigme proposé, basé principalement sur les études de Porter et Castellanos (1980) et de Fowler (2003) consistait en une tâche de close-shadowing, ou répétition rapide de syllabes, présentées de façon auditive ou audiovisuelle, où les temps de réaction et la proportion de réponses correctes étaient comparés avec ceux d'une tâche de catégorisation manuelle des mêmes syllabes. L'objectif de ce paradigme était de questionner le format partiellement moteur des représentations et processus perceptifs mis en œuvre lors du décodage du signal de parole, par la mise en évidence de temps de réaction plus courts lors des réponses orales par rapport aux réponses manuelles. Ce possible avantage perceptif était interprété par le fait que, lors de la perception de parole, les gestes articulatoires serviraient de guide et donc accélèreraient les temps de réponses orales. D'autre part, nous souhaitons, grâce à ce paradigme, pouvoir démontrer le bénéfice d'une présentation audiovisuelle par rapport à une présentation purement auditive des stimuli de parole, notamment chez les patients sourds implantés. Sur ce point, il est à noter qu'en ce qui concerne les participants normo-entendants juniors, nous avons enregistré deux groupes distincts, un premier groupe répondant à des stimuli non-bruités et un second groupe répondant à des stimuli bruités, ceci afin de tester plus

avant un possible avantage d'une présentation audiovisuelle par rapport à une présentation auditive des stimuli.

Chez les participants juniors répondant aux stimuli non-bruités, aucun avantage de la modalité audiovisuelle n'a été observé en raison d'un effet plafond, résultant en des réponses extrêmement rapides et précises lors des deux modalités. En revanche, lors de l'expérience bruitée, les participants ont produit des réponses plus rapides et plus précises pour les stimuli audiovisuels, un résultat en parfait accord avec la littérature portant sur la perception multisensorielle de la parole. Mais crucialement, nous avons découvert que, particulièrement dans la condition bruitée, l'augmentation de la rapidité de la prise de décision pour les réponses orales par rapport aux réponses manuelles allait de pair avec une diminution de la précision, et ce dans les deux modalités.

Chez les participants normo-entendants seniors, bien que globalement les réponses aient été dégradées par rapport à celles des participants juniors, les réponses orales ont également été plus rapides mais moins précises que les réponses manuelles, et aucune différence entre modalités n'a été observée (dans le cadre, nous le rappelons, de stimuli non-bruités).

Nous avons proposé d'interpréter ces résultats comme une nouvelle illustration des liens perceptivo-moteurs dans le cadre des théories de la perception de la parole, en proposant que les réponses orales seraient générées à un stade précoce d'une boucle sensori-motrice, alors que les réponses manuelles seraient formulées en toute fin de boucle entraînant des réponses plus lentes mais aussi plus précises.

Chez les participants implantés, conformément à nos hypothèses, nous avons mis en évidence un avantage perceptif de la modalité audiovisuelle par rapport à la modalité auditive, caractérisé par des réponses plus rapides. Nous pouvons noter que les participants implantés ont obtenu un patron de réponse rappelant celui des participants normo-entendants dans le bruit, comme illustré dans la Figure 83. Nous avons également observé des réponses orales bien plus rapides que les réponses manuelles chez les patients, suggérant un lien perceptivo-moteur actif chez ces participants.

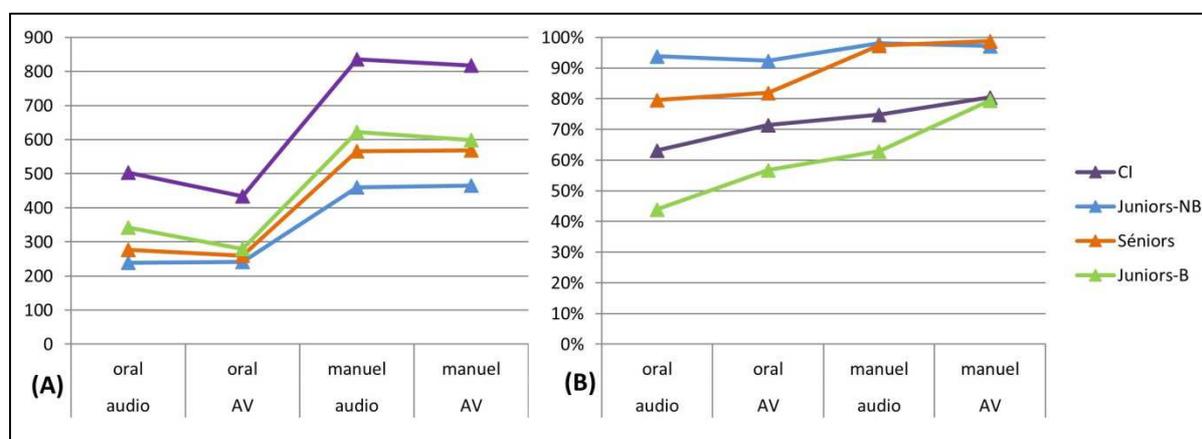


Figure 83 : Moyennes des temps de réaction (A) et des pourcentages de réponses correctes (B) par groupe et par condition.

## 2. Second paradigme expérimental : idiosyncrasies

L'objectif de notre second paradigme expérimental consistait à observer des idiosyncrasies lors de deux tâches de perception et de production de voyelles, et à partir de ces idiosyncrasies, à établir une corrélation entre les caractéristiques formantiques de ces voyelles produites et perçues. Nous avons proposé ce paradigme en nous basant sur une étude précédemment menée par Ménard, Schwartz et collègues (2008, 2014), qui ont observé, d'une part une variabilité inter-participants importante lors de tâches de perception et de production des voyelles orales du Français, mais également pour chaque participant, une corrélation entre valeurs acoustiques liées à la hauteur des voyelles produites et perçues, cette corrélation étant interprétée comme illustration d'un couplage perceptivo-moteur.

Nous avons donc proposé une version réduite de l'expérience de Ménard et Schwartz (2014) consistant en une tâche de production et une tâche de perception des voyelles antérieures du Français. Chez nos deux groupes de participants normo-entendants, juniors et séniors, le même patron de résultat a été observé. Comme lors de l'étude précédente de Ménard et Schwartz (2014), des idiosyncrasies sont apparues lors de la tâche de production ; en revanche, nous n'avons pas obtenu de variabilité inter-participant lors de la tâche de perception, un résultat interprété en raison du nombre réduit de voyelles à catégoriser. En raison de cette absence de variabilité, aucune corrélation individuelle n'a été observée entre les deux tâches. Cependant, nous avons constaté une corrélation forte entre les valeurs acoustiques globales produites et perçues. Nous avons proposé d'interpréter cette corrélation à la lumière de la théorie de la dispersion maximale de Lindblom et collègues (1972, 1986) : lors de la tâche de perception, les locuteurs caractérisent les voyelles au travers de prototypes similaires selon une organisation perceptive de dispersion optimale ; lors de la tâche de production, bien qu'une plus grande variabilité des comportements individuels soit observée, les réponses restent associées à des gabarits communs correspondant aux prototypes observés lors de la tâche de perception.

Chez notre groupe de patients implantés, en revanche, nous avons observé plus de variations intra- et inter-participants, résultant en des idiosyncrasies lors des deux tâches, contrairement aux participants normo-entendants. De manière intéressante, nous avons obtenu en regard de ces idiosyncrasies en perception et production, une corrélation entre les distances des voyelles / $\epsilon$ / et / $e$ /. La variabilité observée illustre selon nous une plus grande difficulté de régularité de production et de catégorisation chez ces participants. Ceux-ci ne se retrouvent donc pas dans la même situation que les normo-entendants et s'appuient possiblement sur les distances / $\epsilon$ -/ $e$ / comme moyen de différenciation des voyelles pour eux, lors de la perception, et pour leur interlocuteur, lors de la production. Enfin, il est important de noter également que malgré cette variabilité plus importante, comme chez les normo-entendants, nous avons observé une corrélation forte entre les valeurs acoustiques moyennes produites et perçues suggérant une nouvelle fois des gabarits perceptivo-moteurs communs (Figure 84).

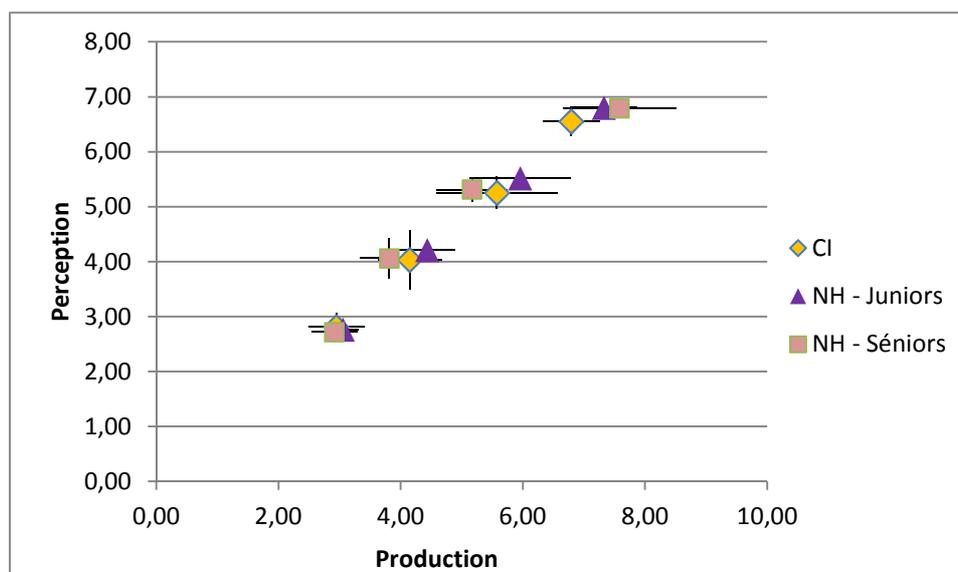


Figure 84 : Corrélations globales entre les valeurs produites et perçues chez les trois populations.

### 3. Troisième paradigme expérimental : convergence et imitation

L'objectif du troisième paradigme était de démontrer les interactions sensori-motrices lors de la production de parole, au travers de tâches d'imitation consciente (nommée tâche d'imitation) et involontaire (nommée tâche de convergence) de voyelles présentées auditivement, en reproduisant fidèlement le protocole expérimental proposé initialement par Sato et collaborateurs (2013, 2014).

Les participants normo-entendants, comme attendu, ont présenté des capacités d'imitation volontaires et involontaires, puisqu'ils ont convergé vers la cible auditive que ce soit avec ou sans consigne préalable d'imitation. Aucune différence concernant les capacités de convergence n'a été observée chez les participants normo-entendants juniors et séniors, ni lors de la tâche d'imitation, ni lors de la tâche de convergence ; néanmoins une plus grande variabilité entre participants a été observée chez les séniors, illustrant possiblement un contrôle vocal moins stable par rapport aux jeunes adultes. Enfin, étonnamment, les patients implantés ont clairement convergé vers la cible auditive lors des deux tâches, même si le degré de convergence était moins important que ceux des participants normo-entendants. Également, leur production s'est révélée plus variable mais aussi moins précise, puisque contrairement aux autres groupes nous avons observé des erreurs de répétition des voyelles perçues (voir Figure 85).

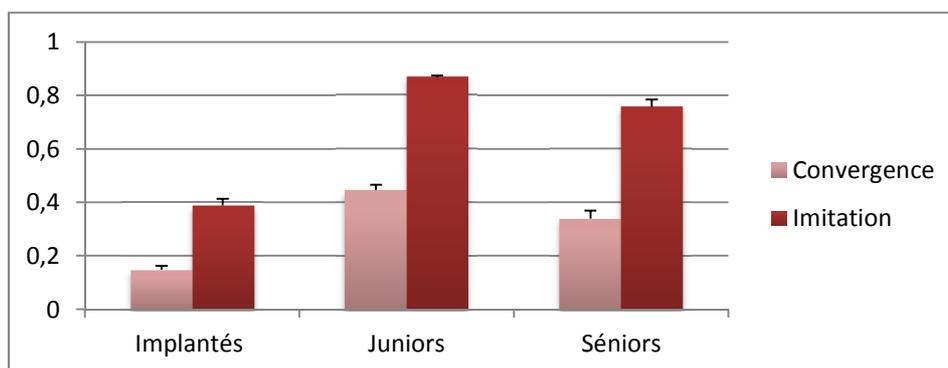


Figure 85 : Coefficients de pente des tâches de convergence et d'imitation chez les trois groupes de participants.

Ainsi ces résultats démontrent des capacités d'estimation de la fréquence fondamentale chez les patients implantés, alliées à des capacités d'adaptation de leurs propres productions en regard d'une cible auditive. De plus, les capacités d'imitation consciente et inconsciente, présentées par les trois groupes de participants, démontrent une adaptation des commandes motrices à l'environnement sensoriel, ce qui nécessiterait, comme proposé plus haut, des interactions sensori-motrices lors de la production de parole.

#### 4. Corrélations entre les paradigmes

Nous avons voulu vérifier si des corrélations pouvaient être établies entre les paradigmes de close-shadowing et de convergence et d'imitation, et ce pour les trois populations. Autrement dit, nous nous sommes demandé si nous pouvions attendre des différences au niveau du traitement auditif pur et également au niveau du lien sensori-moteur. Pour cela, nous avons supposé que la différence de temps de réaction entre les réponses orales et manuelles tout comme les valeurs de pente lors de la tâche de convergence pourraient être associées au lien sensori-moteur. Ainsi, nous avons souhaité corréler ces paramètres des deux paradigmes afin de déterminer si les participants ayant obtenu les meilleures performances lors du premier paradigme étaient également les meilleurs lors du troisième paradigme, supposant qu'une telle corrélation suggérerait des relations perceptivo-motrices plus actives. A noter que, du fait de leur nature, les résultats de l'étude d'idiosyncrasies perceptives et motrices n'ont pu être incluses dans ces analyses de corrélation. De plus, nous avons délibérément décidé d'écarter les pourcentages de réponses correctes de la tâche manuelle à cause de l'effet plafond qui induit une très faible variation de ces données ; pour la même raison, le degré de convergence lors de la tâche d'imitation volontaire a été écarté.

En raison des comparaisons multiples, une correction de Bonferroni a été utilisée, abaissant le seuil de significativité à 0,01 et aucune corrélation ne s'étant révélée significative (voir Tableau 14), plusieurs explications peuvent être proposées quant à cette absence de résultat. Tout d'abord, rappelons que, pour les populations sourde implantée et sénior, la taille réduite de notre corpus et les variations interindividuelles importantes ont pu induire des difficultés à révéler de possibles corrélations. Ensuite,

plus généralement, nos deux paradigmes impliquent des mécanismes multiples, tels que le mécanisme de perception des segments pour l'expérience de close-shadowing et la perception et le contrôle de la fréquence fondamentale pour le paradigme de convergence et d'imitation. Plus généralement, cette absence de corrélation souligne le caractère multiple des relations perceptivo-motrices et de leur mise en œuvre en regard du contexte, ici expérimental, et des stratégies individuelles.

Tableau 14 : Corrélation entre les paradigmes de close-shadowing et de convergence et d'imitation, chez les trois populations.

Corrélation	Degré convergence/RTs oral	Degré convergence/RTs oral – manuel	Degré de convergence / RTs manuel
NH – Juniors	$r^2 = 0,01$	$r^2 = 0,03$	$r^2 = 0,01$
NH – Séniors	$r^2 = 0,16$	$r^2 = 0,07$	$r^2 = 0,01$
CI	$r^2 = 0,05$	$r^2 = 0,25$	$r^2 = 0,1$

## 2. Relations sensori-motrices chez les séniors

Nos hypothèses premières supposaient des performances inférieures pour les personnes séniors par rapport aux personnes juniors, et ce pour les trois paradigmes : pour le paradigme de close-shadowing, nous supposions observer des temps de réaction plus lents et des pourcentages de réponses correctes plus faibles que ceux du groupe junior ; pour le paradigme d'idiosyncrasies, nous pensions observer plus de variabilité de production et de catégorisation chez les séniors ; enfin, pour le paradigme de convergence et imitation, nous nous attendions à obtenir, encore une fois, plus de variabilité de production, d'une part, et d'autre part, moins de convergence vers la cible lors des deux tâches par rapport aux participants juniors. Par contre, globalement, conformément à des études précédentes (Liu et coll., 2010, 2011), nous pensions que les résultats suggéreraient des relations perceptivo-motrices toujours actives chez les séniors.

Ces hypothèses ont été validées avec pour les trois paradigmes, un patron similaire des résultats des adultes séniors à celui du groupe junior mais des réponses partiellement dégradées. Plus précisément, lors du paradigme de close-shadowing, les réponses orales ont été généralement plus rapides et plus précises que les réponses manuelles et aucun avantage de la modalité audiovisuelle n'a été constaté. Par contre, les performances ont été légèrement dégradées par rapport aux participants juniors (pour des stimuli non-bruités), avec des temps de réaction plus lents et une proportion de réponses correctes légèrement inférieures chez les adultes séniors. Lors du paradigme d'idiosyncrasies, comme pour les participants juniors et contrairement à nos hypothèses, nous n'avons pas observé de différence de variabilité de production et de perception mais en revanche une corrélation entre les valeurs acoustiques globales produites et perçues. Enfin, lors du paradigme de convergence et imitation, aucune différence n'a été observée

entre les groupes juniors et séniors concernant le degré d'imitation ou de convergence vers la cible. Cependant, les participants séniors ont présenté plus de variabilité de production lors des deux tâches, suggérant un contrôle moins important de la production de la fréquence fondamentale.

Plus généralement, ces résultats suggèrent donc des relations sensori-motrices lors de la perception et de la production de parole toujours actives chez les séniors. Les études menées sur les relations perceptivo-motrices chez les séniors étant peu nombreuses et contradictoires, ce résultat nous semble important. Enfin, une question importante et à laquelle nos résultats ne permettent malheureusement pas de répondre est toutefois de savoir si une détérioration des capacités motrices et perceptives associée à l'âge peut être associée non pas à une simple stabilité des relations perceptives mais bien à un renforcement de ces relations.

### **3. Relations sensori-motrices chez les patients sourds implantés**

En ce qui concerne l'étude des liens perceptivo-moteurs chez des patients sourds post-linguaux implantés, l'interrogation majeure était de déterminer si une personne, ayant été privée de l'audition, puis ayant récupéré une certaine sorte d'input sensoriel au travers d'une implantation cochléaire, conserve ou rétablit des relations sensori-motrices actives entre ses représentations motrices et ses nouvelles représentations sensorielles fournies par l'implant, lors de la perception ou de la production de parole.

Nos hypothèses premières supposaient, du fait du déficit auditif évident chez ces personnes, des performances inférieures à celles de nos deux groupes de participants entendants, alliées éventuellement à une mise en évidence du bénéfice de la présentation audiovisuelle.

Conformément à ces hypothèses, lors du paradigme de close-shadowing, les participants implants ont fourni des réponses plus lentes et moins précises par rapport aux participants normo-entendants. Également, un avantage de l'audiovisuel par rapport à l'audio a été démontré, puisque les participants implantés ont obtenu des réponses plus rapides lors des stimuli audiovisuels. Enfin, de manière importante, nous avons retrouvé des temps de réaction sensiblement plus rapides lors des réponses orales, cette différence entre les réponses orales et manuelles pouvant suggérer une certaine activation du lien perceptivo-moteur lors de la perception de parole chez ces patients.

Lors du paradigme d'idiosyncrasies, si plus de variabilité a été observée chez les participants implantés par rapport aux participants entendants, la corrélation acoustique globale entre les voyelles produites et perçues a été retrouvée, dénotant une capacité des participants implantés à varier autour de gabarits perceptivo-moteurs communs, tant en production qu'en perception.

Enfin, lors du dernier paradigme, outre une plus grande variabilité des productions observée chez les participants implantés lors des deux tâches d'imitation, illustrant une plus grande difficulté à maîtriser leur voix par rapport aux participants normo-entendants juniors, une convergence moins importante vers la cible auditive a

également été observée. Néanmoins, et ce résultat nous semble revêtir une certaine importance, nous avons pu observer des aptitudes d'imitation consciente et inconsciente chez notre population implantés, traduisant une habilité à adapter leur production par rapport à l'environnement, grâce à des mécanismes sensori-moteurs.

Globalement, au travers de nos trois paradigmes expérimentaux, il semblerait que le lien perceptivo-moteur chez ces patients serait donc fonctionnel malgré des performances perceptives dégradées. Cette conclusion est à mettre en lien avec les études précédentes, en comportemental, comme celles de Vick et collègues (2002) et de Turgeon et collègues (2015) portant sur les mécanismes de compensation articulaire et de perturbation motrice, ainsi qu'en imagerie cérébrale, chez Rouger et collègues (2012), dont les résultats ont suggéré un lien sensori-moteur actif mais perturbé chez des patients sourds post-linguaux porteurs d'un implant cochléaire, nos données permettant d'apporter une nouvelle illustration de l'activation de la relation sensori-motrice chez ces patients.

## **2. Conclusion générale et perspectives**

Cette recherche avait pour objectif global d'observer dans quelles mesures les interactions entre systèmes sensoriels et moteur s'établissent chez des patients sourds post-linguaux après implantation cochléaire et, notamment, de déterminer la possible implication du système moteur lors de tâches de perception et de production de la parole chez ces patients.

Au travers de ce travail de thèse, rappelons que nos résultats s'inscrivent dans la littérature très importante portant sur les liens perceptivo-moteurs et soulignent la nature sensori-motrice de la parole. Nous avons donc créé et validé un protocole expérimental composé de trois paradigmes comportementaux, permettant de tester les interactions sensori-motrices, et pouvant être utilisé chez différentes populations, que ce soit des populations saines ou des populations sujettes à des troubles de la parole ou de l'audition.

D'autre part, nous avons pu mettre en évidence des relations sensori-motrices toujours actives chez les séniors, cette observation permettant de clarifier les études précédentes et d'offrir des pistes de perspectives de recherche importantes.

Enfin, notre recherche a permis de démontrer qu'un lien perceptivo-moteur semble actif ou réactivé chez les patients sourds post-linguaux implantés, bien qu'il soit peut-être dégradé. Nous pouvons rappeler qu'il s'agit d'un résultat nouveau et important pour permettre de comprendre les mécanismes de perception et de production de parole chez ces patients.

Cependant, plusieurs pistes d'amélioration et d'extrapolation de notre recherche peuvent être proposées. Tout d'abord, il serait intéressant de confirmer les résultats obtenus chez les séniors au travers d'une plus large population, et éventuellement avec une étendue d'âge plus importante.

D'autre part, nous proposons d'enregistrer plus de participants sourds post-linguaux implantés pour éventuellement créer plusieurs groupes de patients en fonction de leur âge post-implantation, ceci afin de vérifier s'il pourrait y avoir une évolution des résultats, et par là même du lien sensori-moteur lors de la production et la perception de parole.

Egalement, l'évaluation des relations sensori-motrices chez une population sourde pré-linguale implantée nous semble être une perspective intéressante. En effet, chez ces personnes n'ayant jamais eu l'expérience d'un input de parole naturel, et donc d'un lien audio-articulatoire fonctionnel, la question serait de déterminer si des relations perceptivo-motrices se mettent en place après l'implantation, et ce grâce au nouvel input sensoriel apporté par l'implant cochléaire.

Enfin, au-delà de l'augmentation des populations étudiées, il nous semblerait utile de faire le lien avec la rééducation de parole. En effet, si, contrairement à nos hypothèses initiales, le lien sensori-moteur semble être très actif, nous pouvons rappeler que nos résultats lors du paradigme de convergence et d'imitation suggèrent une certaine dégradation du contrôle vocal par les patients sourds implantés, il pourrait donc être important de prendre en compte la maîtrise de la voix lors de la rééducation.

Egalement, différentes pistes de travail de recherche restent à explorer. Concernant la population sénior, il a été précédemment suggéré le développement d'un surinvestissement sensori-moteur au cours du vieillissement, qui pourrait être exploré lors d'études futures, afin de déterminer si une rééducation du lien perceptivo-moteur peut être envisageable pour permettre une meilleure communication.

Concernant la population implantée, puisqu'a priori, les relations perceptivo-motrices semblent être fonctionnelles lors de nos trois paradigmes, des études pourraient être menées étendant nos recherches sur le lien sensori-moteur et la perception ou la production de la parole, vers d'autres composantes de la communication humaine associées par exemple à l'expressivité, la reconnaissance des émotions dans la voix, ou encore sur la perception de la musique.

A travers ces diverses perspectives, c'est selon nous dans la double dimension du lien perception-action et de la mise en relation de paradigmes expérimentaux variés qu'il conviendrait d'étudier à l'avenir les compétences perceptives, motrices, langagières et communicationnelles de populations diverses comme celles que nous avons étudiées dans cette thèse. Nous espérons que les résultats de cette thèse, qui nous semblent marqués d'une forte cohérence, pourront susciter de tels développements.



# Bibliographie

---



- Ball, V., Faulkner, A., & Fourcin, A. (1990). The effects of two different speech-coding strategies on voice fundamental frequency control in deafened adults. *British Journal of Audiology*, 24(6), 393-409.
- Bayard, C. (2014). *Perception de la Langue Française Parlée Complétée : intégration du trio lèvres-main-son* (Thèse de doctorat). Université Libre de Bruxelles.
- Beautemps, D., Cathiard, M.-A., & Le Borgne, Y. (2003). Benefit of audiovisual presentation in close shadowing task. In *Proceedings of the 15th International Congress on Phonetic Sciences* (p. 841-844).
- Bell-Berti, F., Raphael, L. J., Pisoni, D. B., & Sawusch, J. R. (1979). Some Relationships between Speech Production and Perception. *Phonetica*, 36(6), 373-383.
- Benjamin, B. J. (1982). Phonological performance in gerontological speech. *Journal of Psycholinguistic Research*, 11(2), 159-167.
- Bergeson, T. R., Pisoni, D. B., & Davis, R. A. O. (2003). A Longitudinal Study of Audiovisual Speech Perception by Children with Hearing Loss Who have Cochlear Implants. *The Volta review*, 103(4), 347-370.
- Bergeson, T. R., Pisoni, D. B., & Davis, R. A. O. (2005). Development of audiovisual comprehension skills in prelingually deaf children with cochlear implants. *Ear and Hearing*, 26(2), 149-164.
- Bever, T. G., & Poeppel, D. (2010). Analysis by Synthesis: A (Re-)Emerging Program of Research for Language and Vision. *BIOLINGUISTICS*, 4(2-3), 174-200.
- Bilodeau-Mercure, M., Ouellet, C., & Tremblay, P. (2015). Vieillesse oro-faciale et production de la parole. *Affiche présentée au Congrès québécois de recherche en adaptation-réadaptation. Boucherville, Qc, Canada.*
- Binder, J. R., Liebenthal, E., Possing, E. T., Medler, D. A., & Ward, B. D. (2004). Neural correlates of sensory and decision processes in auditory object identification. *Nature Neuroscience*, 7(3), 295-301.
- Bladon, R. A. W., & Fant, G. (1978). A two formant model and the cardinal vowels. *Speech Transmission Laboratory Quarterly Progress and Status Report*, 19(1), 1-8.
- Blamey, P., Arndt, P., Bergeron, F., Bredberg, G., Brimacombe, J., Facer, G., ... Whitford, L. (1996). Factors affecting auditory performance of postlinguistically deaf adults using cochlear implants. *Audiology & Neuro-Otology*, 1(5), 293-306.
- Blamey, P., Artieres, F., Başkent, D., Bergeron, F., Beynon, A., Burke, E., ... Lazard, D. S. (2013). Factors affecting auditory performance of postlinguistically deaf adults using cochlear implants: an update with 2251 patients. *Audiology & Neuro-Otology*, 18(1), 36-47.
- Blumstein, S. E., & Stevens, K. N. (1980). Perceptual invariance and onset spectra for stop consonants in different vowel environments. *The Journal of the Acoustical Society of America*, 67(2), 648-662.
- Boersma, P. (2001). Praat, a system for doing phonetics by computer, *Glott International*, 5(9/10), 341-345.
- Browman, C. P., & Goldstein, L. M. (1986). Towards an articulatory phonology. *Phonology*, 3(01), 219-252.

- Browman, C. P., & Goldstein, L. M. (1989). Articulatory gestures as phonological units. *Phonology*, 6(2), 201-251.
- Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: an overview. *Phonetica*, 49(3-4), 155-180.
- Callan, D. E., Jones, J. A., Callan, A. M., & Akahane-Yamada, R. (2004). Phonetic perceptual identification by native- and second-language speakers differentially activates brain regions involved with acoustic phonetic processing and those involved with articulatory-auditory/orosensory internal models. *NeuroImage*, 22(3), 1182-1194.
- Carlson, R., Granström, B., & Fant, G. (1970). Some studies concerning perception of isolated vowels. *Speech Transmission Laboratory Quarterly Progress and Status Report*, 2(2-3), 19-35.
- Chatterjee, M., & Peng, S.-C. (2008). Processing F0 with cochlear implants: Modulation frequency discrimination and speech intonation recognition. *Hearing Research*, 235(1-2), 143-156.
- Chon, H., Kraft, S. J., Zhang, J., Loucks, T., & Ambrose, N. G. (2013). Individual variability in delayed auditory feedback effects on speech fluency and rate in normally fluent adults. *Journal of Speech, Language, and Hearing Research: JSLHR*, 56(2), 489-504.
- Cienkowski, K. M., & Carney, A. E. (2002). Auditory-visual speech perception and aging. *Ear and Hearing*, 23(5), 439-449.
- Cienkowski, K. M., & Vasil-Dilaj, K. (2010). The Role of Central Auditory and Cognitive Processing Ability in Speech Perception for Elderly Listeners. *Perspectives on Hearing and Hearing Disorders Research and Diagnostics*, 14(2), 30.
- Cornett, R.O. (1967). Cued-speech. *American Annals of the Deaf*, 112, 3-13.
- Cotton, J. C. (1935). Normal « visual hearing ». *Science*, 82, 592-593.
- Cowie, R., & Douglas-Cowie, E. (1983). *Postlingually Acquired Deafness: Speech Deterioration and the Wider Consequences*. Ed. De Gruyter.
- Cullington, H. E., & Zeng, F.-G. (2011). Comparison of bimodal and bilateral cochlear implant users on speech recognition with competing talker, music perception, affective prosody discrimination, and talker identification. *Ear and Hearing*, 32(1), 16-30.
- D'Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., & Fadiga, L. (2009). The Motor Somatotopy of Speech Perception. *Current Biology*, 19(5), 381-385.
- Davis, C., & Kim, J. (2001). Repeating and Remembering Foreign Language Words: Implications for Language Teaching Systems. *Artificial Intelligence Review*, 16(1), 37-47.
- Delvaux, V., & Soquet, A. (2007). The influence of ambient speech on adult speech productions through unintentional imitation. *Phonetica*, 64(2-3), 145-173.
- Diehl, R. L., Lotto, A. J., & Holt, L. L. (2004). Speech perception. *Annual Review of Psychology*, 55, 149-179.
- Dole, M. (2012). *Perception de la parole dans le bruit et Dyslexie: Approches comportementale, neuroanatomique et fonctionnelle*. (Thèse de doctorat). Université Lumière Lyon II.
- Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *The European Journal of Neuroscience*, 15(2), 399-402.

- Feng, Y., Gracco, V. L., & Max, L. (2011). Integration of auditory and somatosensory error signals in the neural control of speech movements. *Journal of neurophysiology*, *106*(2), 667-679.
- Ferland-Longpré, V., Martel, V., Gadbois, M., Bilodeau-Mercure, M., Ouellet, C., & Tremblay, P. (2013). Vieillesse normale et production des voyelles.
- Fowler, C. A. (1986). An event approach to the study of speech perception for a direct-realist perspective. *Journal of Phonetics*, *14*.
- Fowler, C. A., & Smith, M. R. (1986). Speech perception as vector analysis: An approach to the problems of segmentation and invariance. *Invariance and Variability in Speech Processes*.
- Fowler, C. A., & Dekle, D. J. (1991). Listening with eye and hand: cross-modal contributions to speech perception. *Journal of Experimental Psychology. Human Perception and Performance*, *17*(3), 816-828.
- Fowler, C. A. (1996). Listeners do hear sounds, not tongues. *The Journal of the Acoustical Society of America*,
- Fowler, C. A., Brown, J. M., Sabadini, L., & Weihing, J. (2003). Rapid access to speech gestures in perception: Evidence from choice and simple response time tasks. *Journal of memory and language*, *49*(3), 396-413.
- Fox, R. A. (1982). Individual variation in the perception of vowels: implications for a perception-production link. *Phonetica*, *39*(1), 1-22.
- Füllgrabe, C. (2013). Age-Dependent Changes in Temporal-Fine-Structure Processing in the Absence of Peripheral Hearing Loss. *American Journal of Audiology*, *22*(2), 313.
- Fullgräbe, C. (2014). Why speech perception declines across the adult lifespan: Effects of age on audition and cognition. *Acoustics Bulletin*, (39), 38-40.
- Gallantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic bulletin & review*, *13*(3), 361-377.
- Garnier, M., Lamalle, L., & Sato, M. (2013). Neural correlates of phonetic convergence and speech imitation. *Cognitive Science*, *4*, 600.
- Geers, A., & Brenner, C. (1994). Speech Perception Results: Audition and Lipreading Enhancement. *Volta Review*, *96*(5), 97-108.
- Gelfand, S. A., Piper, N., & Silman, S. (1985). Consonant recognition in quiet as a function of aging among normal hearing subjects. *The Journal of the Acoustical Society of America*, *78*(4), 1198-1206.
- Gelfand, S. A., Piper, N., & Silman, S. (1986). Consonant recognition in quiet and in noise with aging among normal hearing listeners. *The Journal of the Acoustical Society of America*, *80*(6), 1589-1598.
- Guenther, F. H., Hampson, M., & Johnson, D. (1998). A theoretical investigation of reference frames for the planning of speech movements. *Psychological review*, *105*(4), 611.
- Guenther, F. H., & Vladusich, T. (2012). A neural theory of speech acquisition and production. *Journal of neurolinguistics*, *25*(5), 408-422.
- Goh, W. D., Pisoni, D. B., Kirk, K. I., & Remez, R. E. (2001). Audio-Visual Perception of Sinewave Speech in an Adult Cochlear Implant User: A Case Study. *Ear and hearing*, *22*(5), 412-419.

- Goldstein, L. M., & Fowler, C. A. (2003). Articulatory phonology: A phonology for public language use. In *Phonetics and phonology in language comprehension and production: Differences and similarities* (p. 159-207).
- Gonseth, C. (2013). *Multimodalité de la communication langagière humaine : interaction geste/parole et encodage de distance dans le pointage* (Thèse de doctorat). Université de Grenoble.
- Gould, J., Lane, H., Vick, J., Perkell, J. S., Matthies, M. L., & Zandipour, M. (2001). Changes in speech intelligibility of postlingually deaf adults after cochlear implantation. *Ear and Hearing, 22*(6), 453-460.
- Grabski, K. (2012). *Les cartes sensorimotrices de la parole : Corrélats neurocognitifs et couplage fonctionnel des systèmes de perception et de production des voyelles du Français*. (Thèse de doctorat). Université de Grenoble.
- Grabski, K., Schwartz, J.-L., Lamalle, L., Vilain, C., Vallée, N., Baciou, M., LeBas, J.F. & Sato, M. (2013a). Shared and distinct neural correlates of vowel perception and production. *Journal of Neurolinguistics, 26*(3), 384-408.
- Grabski, K., Tremblay, P., Gracco, V. L., Girin, L., & Sato, M. (2013b). A mediating role of the auditory dorsal pathway in selective adaptation to speech: a state-dependent transcranial magnetic stimulation study. *Brain Research, 1515*, 55-65.
- Gracco, V. L., & Abbs, J. H. (1985). Dynamic control of the perioral system during speech: kinematic analyses of autogenic and nonautogenic sensorimotor processes. *Journal of Neurophysiology, 54*(2), 418-432.
- Guenther, F. H., Hampson, M., & Johnson, D. (1998). A theoretical investigation of reference frames for the planning of speech movements. *Psychological Review, 105*(4), 611-633.
- Guenther, F. H. (2006). Cortical interactions underlying the production of speech sounds. *Journal of Communication Disorders, 39*(5), 350-365.
- Halle, M., & Stevens, K. (1959). Analysis by synthesis. In *Proceeding of the Seminar on Speech Compression and Processing* (Vol. Vol. II, p. Paper D7).
- Hamzavi, J., Baumgartner, W.-D., Pok, S. M., Franz, P., & Gstoettner, W. (2003). Variables affecting speech perception in postlingually deaf adults following cochlear implantation. *Acta Oto-Laryngologica, 123*(4), 493-498.
- Hanna, N. (2011). *Vowel Identification by Cochlear Implant Users: Comparison of Vowel Edges and Vowel Centers*. Honors College Thesis, University of South Florida.
- Hickok, G. & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences, 4*(4), 131-138.
- Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition, 92*(1-2), 67-99.
- Holt, L. L., & Lotto, A. J. (2010). Speech perception as categorization. *Attention, perception & psychophysics, 72*(5), 1218-1227.
- Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The Theory of Event Coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences, 24*(05), 849-878.

- Houde, J. F., & Jordan, M. I. (2002). Sensorimotor adaptation in speech production. *Science*, 279(5354), 1213-1216.
- Houde, J. F., Nagarajan, S. S., Sekihara, K., & Merzenich, M. M. (2002). Modulation of the auditory cortex during speech: an MEG study. *Journal of Cognitive Neuroscience*, 14(8), 1125-1138.
- Jones, J. A., & Striemer, D. (2007). Speech disruption during delayed auditory feedback with simultaneous visual feedback. *The Journal of the Acoustical Society of America*, 122(4), EL135-141.
- Kaiser, A. R., Kirk, K. I., Lachs, L., & Pisoni, D. B. (2003). Talker and Lexical Effects on Audiovisual Word Recognition by Adults With Cochlear Implants. *Journal of speech, language, and hearing research : JSLHR*, 46(2), 390-404.
- Kiefer, J., Hohl, S., Stürzebecher, E., Pfennigdorff, T., & Gstöettner, W. (2001). Comparison of speech recognition with different speech coding strategies (SPEAK, CIS, and ACE) and their relationship to telemetric measures of compound action potentials in the nucleus CI 24M cochlear implant system. *Audiology: Official Organ of the International Society of Audiology*, 40(1), 32-42.
- Kirk, K. I., Pisoni, D. B., & Lachs, L. (2002). Audiovisual integration of speech by children and adults with cochlear implants. *Proceedings : ICSLP. International Conference on Spoken Language Processing, 2002*, 1689-1692.
- Kishon-Rabin, L., Taitelbaum, R., Tobin, Y., & Hildesheimer, M. (1999). The effect of partially restored hearing on speech production of postlingually deafened adults with multichannel cochlear implants. *The Journal of the Acoustical Society of America*, 106(5), 2843-2857.
- Kompis, M., Vischer, M. W., & Häusler, R. (1999). Performance of compressed analogue (CA) and continuous interleaved sampling (CIS) coding strategies for cochlear implants in quiet and noise. *Acta Oto-Laryngologica*, 119(6), 659-664.
- Lametti, D. R., Nasir, S. M., & Ostry, D. J. (2012). Sensory preference in speech production revealed by simultaneous alteration of auditory and somatosensory feedback. *The Journal of Neuroscience*, 32(27), 9351-9358.
- Lane, H., Wozniak, J., Matthies, M., Svirsky, M., Perkell, J., O'Connell, M., & Manzella, J. (1997). Changes in sound pressure and fundamental frequency contours following changes in hearing status. *The Journal of the Acoustical Society of America*, 101(4), 2244-2252.
- Lane, H., Matthies, M., Perkell, J., Vick, J., & Zandipour, M. (2001). The effects of changes in hearing status in cochlear implant users on the acoustic vowel space and CV coarticulation. *Journal of Speech, Language, and Hearing Research: JSLHR*, 44(3), 552-563.
- Lane, H., Denny, M., Guenther, F. H., Matthies, M. L., Menard, L., Perkell, J. S., Stockmann, E., Tiede, M., Vic, J. & Zandipour, M. (2005). Effects of bite blocks and hearing status on vowel production. *The Journal of the Acoustical Society of America*, 118(3 ; 1), 1636-1646.
- Langereis, M. C., Bosman, A. J., van Olphen, A. F., & Smoorenburg, G. F. (1998). Effect of cochlear implantation on voice fundamental frequency in post-lingually deafened adults. *Audiology: Official Organ of the International Society of Audiology*, 37(4), 219-230.
- Laurent, R. (2014). *COSMO : un modèle bayésien des interactions sensori-motrices dans la perception de la parole* (Thèse de doctorat). Université de Grenoble.

- Lee, B., S. (1990). Effects of delayed speech feedback. *The Journal of the Acoustical Society of America*, 22(824).
- Lelong, A. (2012). *Convergence phonétique en interaction Phonetic convergence in interaction*. (Thèse de Doctorat) Université de Grenoble.
- Leybaert, J. (2000). Phonology acquired through the eyes and spelling in deaf children. *Journal of Experimental Child Psychology*, 75(4), 291-318.
- Lieberman, A. M., Delattre, P., & Cooper, F. S. (1952). The role of selected stimulus-variables in the perception of the unvoiced stop consonants. *The American Journal of Psychology*, 65(4), 497-516.
- Lieberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, 74(6), 431-461.
- Lieberman, A. M., & Mattingly, I. G. (1985). The Motor Theory of Speech Perception Revised. *Cognition*, 21(1), 1-36.
- Liljencrants, J., & Lindblom, B. (1972). Numerical Simulation of Vowel Quality Systems: The Role of Perceptual Contrast. *Language*, 48(4), 839-862.
- Lindblom, B. (1986). Phonetic universals in vowel systems. *Experimental phonology*, 13-44.
- Lindblom, B., & Maddieson, I. (1988). Phonetic Universals in Consonant Systems. In *Language, Speech and Mind* (p. 62-78). London: Routledge.
- Lindblom, B. (1990). Explaining Phonetic Variation: A Sketch of the H&H Theory. In W. J. Hardcastle & A. Marchal (éd.), *Speech Production and Speech Modelling* (p. 403-439). Springer Netherlands.
- Linville, S. E., & Fisher, H. B. (1985). Acoustic characteristics of women's voices with advancing age. *Journal of Gerontology*, 40(3), 324-330.
- Linville, S. E., & Korabic, E. W. (1987). Fundamental frequency stability characteristics of elderly women's voices. *The Journal of the Acoustical Society of America*, 81(4), 1196-1199.
- Liss, J. M., Weismer, G., & Rosenbek, J. C. (1990). Selected acoustic characteristics of speech production in very old males. *Journal of Gerontology*, 45(2), P35-45.
- Liu, H., Wang, E. Q., Chen, Z., Liu, P., Larson, C. R., & Huang, D. (2010). Effect of tonal native language on voice fundamental frequency responses to pitch feedback perturbations during sustained vocalizations. *The Journal of the Acoustical Society of America*, 128(6), 3739-3746.
- Liu, H., Meshman, M., Behroozmand, R., & Larson, C. R. (2011). Differential effects of perturbation direction and magnitude on the neural processing of voice pitch feedback. *Clinical neurophysiology: official journal of the International Federation of Clinical Neurophysiology*, 122(5), 951-957.
- Liu, P., Chen, Z., Jones, J. A., Huang, D., & Liu, H. (2011). Auditory Feedback Control of Vocal Pitch during Sustained Vocalization: A Cross-Sectional Study of Adult Aging. *PLoS ONE*, 6(7), e22791.
- Lombard, E. (1911). Le signe de l'élévation de la voix. *Annales des maladies de l'oreille et du larynx*, 37, 101-119.
- Lortie, C. L., Thibeault, M., Guitton, M. J., & Tremblay, P. (2015). Effects of age on the amplitude, frequency and perceived quality of voice. *Age (Dordrecht, Netherlands)*, 37(6), 117.

- Luce, D. R. (1986). *Response Times: Their Role in Inferring Elementary Mental Organization* (1 edition). New York : Oxford: Oxford University Press.
- Maeda, S. (1979). An articulatory model of the tongue based on a statistical analysis. *The Journal of the Acoustical Society of America*, 65(S1), S22-S22.
- Mantakas, M., Schwartz, J.-L., & Escudier, P. (1986). Modèle de prédiction du 'deuxième formant effectif F<sup>2</sup> - application à l'étude de la labialité des voyelles avant du français. In *Proceedings of 15th Journées d'étude sur la parole* (p. 157-161).
- Mao Yitao, & Xu Li. (2013). Music and Cochlear Implants. *Journal of Otology*, 45(1).
- Max, L., Daniels, J., Curet, K., & Cronin, K. (2008) Modulation of auditory and somatosensory processing during the planning of speech movements Proceedings of the 8th International Seminar on Speech Production, 41-44, Strasbourg, France.
- Marx, M., Gutierrez, D., Lepage, B., Khoury, E., Laborde, M.-L., Deguine, O., & Fraysse, B. (2011). Cochlear implants in elderly: performance outcomes in the long term. *Revue De Laryngologie - Otologie - Rhinologie*, 132(4-5), 187-191.
- Marx, M., James, C., Foxton, J., Capber, A., Fraysse, B., Barone, P., & Deguine, O. (2015). Speech prosody perception in cochlear implant users with and without residual hearing. *Ear and Hearing*, 36(2), 239-248.
- Massida, Z., Marx, M., Belin, P., James, C., Fraysse, B., Barone, P., & Deguine, O. (2013). Gender categorization in cochlear implant users. *Journal of Speech, Language, and Hearing Research: JSLHR*, 56(5), 1389-1401.
- Matthies, M. L., Guenther, F. H., Denny, M., Perkell, J. S., Burton, E., Vick, J., Lane, H., Tiede, M. & Zandipour, M. (2008). Perception and production of /r/ allophones improve with hearing from a cochlear implant. *The Journal of the Acoustical Society of America*, 124(5), 3191-3202.
- Mcgurk, H., & Macdonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264(5588), 746-748.
- Meister, I. G., Wilson, S. M., Deblieck, C., Wu, A. D., & Iacoboni, M. (2007). The essential role of premotor cortex in speech perception. *Current Biology: CB*, 17(19), 1692-1696.
- Ménard, L., Polak, M., Denny, M., Burton, E., Lane, H., Matthies, M. L., ... Vick, J. (2007). Interactions of speaking condition and auditory feedback on vowel production in postlingually deaf adults with cochlear implants. *The Journal of the Acoustical Society of America*, 121(6), 3790-3801.
- Ménard, L., Schwartz, J.-L., & Aubin, J. (2008). Invariance and variability in the production of the height feature in French vowels. *Speech Communication*, 50(1), 14-28.
- Ménard, L., & Schwartz, J.-L. (2014). Perceptuo-motor biases in the perceptual organization of the height feature in French vowels. *Acta Acustica united with Acustica*, 100(4), 676-689.
- Moore, D. R., & Fullgräbe, C. (2013). Cognitive contributions to hearing in older people. *Journal of Hearing Science*, 2(4), 58-60.
- Morgan, E. E., & Rastatter, M. (1986). Variability of voice fundamental frequency in elderly female speakers. *Perceptual and Motor Skills*, 63(1), 215-218.
- Möttönen, R., Järveläinen, J., Sams, M., & Hari, R. (2005). Viewing speech modulates activity in the left SI mouth cortex. *NeuroImage*, 24(3), 731-737.

- Möttönen, R., & Watkins, K. E. (2012). Using TMS to study the role of the articulatory motor system in speech perception. *Aphasiology*, *26*(9), 1103-1118.
- Möttönen, R., Dutton, R., & Watkins, K. E. (2013). Auditory-Motor Processing of Speech Sounds. *Cerebral Cortex (New York, NY)*, *23*(5), 1190-1197.
- Moulin-Frier, C. (2011). *Rôle des relations perception-action dans la communication parlée et l'émergence des systèmes phonologiques : étude, modélisation computationnelle et simulations* (Thèse de doctorat). Université de Grenoble.
- Moulin-Frier, C., Laurent, R., Bessière, P., Schwartz, J.-L., & Diard, J. (2012). Adverse conditions improve distinguishability of auditory, motor and perceptuo-motor theories of speech perception: an exploratory Bayesian modeling study. *Language and Cognitive Processes*, *27*(7-8), 1240-1263.
- Nasir, S. M., & Ostry, D. J. (2006). Somatosensory precision in speech production. *Current Biology*, *16*(19), 1918-1923.
- Nahorna, O. (2013). *Analyse de scènes de parole multisensorielle : Mise en évidence et caractérisation d'un processus de liage audiovisuel préalable à la fusion* (Thèse de doctorat). Université de Grenoble.
- Nguyen, N., Sato, M., Postma-Nilsenova, M., Pardo, J., & Babel, M. (2013). *Speech imitation: the cognitive underpinnings of adaptive vocal behaviour*. *Frontiers in Psychology*.
- Ohde, R. N., & Abou-Khalil, R. (2001). Age differences for stop-consonant and vowel perception in adults. *The Journal of the Acoustical Society of America*, *110*(4), 2156-2166.
- Öhman, S. E. (1967). Numerical model of coarticulation. *The Journal of the Acoustical Society of America*, *41*(2), 310-320.
- Pelizzone, M., Cosendai, G., & Tinembart, J. (1999). Within-patient longitudinal speech reception measures with continuous interleaved sampling processors for ineraid implanted subjects. *Ear and Hearing*, *20*(3), 228-237.
- Peng, S.-C., Lu, N., & Chatterjee, M. (2009). Effects of cooperating and conflicting cues on speech intonation recognition by cochlear implant users and normal hearing listeners. *Audiology & Neuro-Otology*, *14*(5), 327-337.
- Peng, S.-C., Chatterjee, M., & Lu, N. (2012). Acoustic cue integration in speech intonation recognition with cochlear implants. *Trends in Amplification*, *16*(2), 67-82.
- Perkell, J., Lane, H., Svirsky, M., & Webster, J. (1992). Speech of cochlear implant patients: A longitudinal study of vowel production. *The Journal of the Acoustical Society of America*, *91*(5), 2961-2978.
- Perkell, J. S., Matthies, M. L., Svirsky, M. A., & Jordan, M. I. (1995). Goal-based speech motor control: A theoretical framework and some preliminary data. *Journal of Phonetics*, *23*(1), 23-35.
- Perkell, J., Matthies, M., Lane, H., Guenther, F., Wilhelms-Tricarico, R., Wozniak, J., & Guiod, P. (1997). Speech motor control: Acoustic goals, saturation effects, auditory feedback and internal models. *Speech Communication*, *22*(2-3), 227-250.
- Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Perrier, P., Vick, J., Wilhelms-Tricarico, R. & Zandipour, M. (2000). A theory of speech motor control and supporting data from speakers with normal hearing and with profound hearing loss. *Journal of Phonetics*, *28*(3), 233-272.

- Perkell, J., Numa, W., Vick, J., Lane, H., Balkany, T., & Gould, J. (2001). Language-specific, hearing-related changes in vowel spaces: a preliminary study of English- and Spanish-speaking cochlear implant users. *Ear and Hearing, 22*(6), 461-470.
- Perkell, J. S. (2012). Movement goals and feedback and feedforward control mechanisms in speech production. *Journal of neurolinguistics, 25*(5), 382-407.
- Perrier, P., Payan, Y., Zandipour, M., & Perkell, J. (2003). Influences of tongue biomechanics on speech movements during the production of velar stop consonants: a modeling study. *The Journal of the Acoustical Society of America, 114*(3), 1582-1599.
- Perrier, P. (2005a). Control and representations in speech production. *ZAS Papers in Linguistics, 40*, 109-132.
- Perrier, P., Ma, L. & Payan, Y. (2005b) Modeling the production of VCV sequences via the inversion of a biomechanical model of the tongue. Proceedings of Interspeech 2005, Lisbon, Portugal, pp. 1041-1044.
- Perrier, P. (2012). Gesture planning integrating knowledge of the motor plant's dynamics: A literature review from motor control and speech motor control. In D. P. & P. P. Melanie Weirich Susanne Fuchs (éd.), *Speech Planning and Dynamics* (p. 191-238). Peter Lang Publishers.
- Porter, R. J., & Castellanos, F. X. (1980). Speech-production measures of speech perception: Rapid shadowing of VCV syllables. *The Journal of the Acoustical Society of America, 67*(4), 1349-56.
- Prinz, W. (1997). Perception and Action Planning. *European Journal of Cognitive Psychology, 9*(2), 129-154.
- Ramig, L. A., & Ringel, R. L. (1983). Effects of physiological aging on selected acoustic characteristics of voice. *Journal of Speech and Hearing Research, 26*(1), 22-30.
- Rastatter, M. P., & Jacques, R. D. (1990). Formant frequency structure of the aging male and female vocal tract. *Folia Phoniatica, 42*(6), 312-319.
- Rastatter, M. P., McGuire, R. A., Kalinowski, J., & Stuart, A. (1997). Formant frequency characteristics of elderly speakers in contextual speech. *Folia Phoniatica et Logopaedica: Official Organ of the International Association of Logopedics and Phoniatics (IALP), 49*(1), 1-8.
- Reisberg, D., McLean, J., & Goldfield, A. (1987). Easy to hear but hard to understand: A lip-reading advantage with intact auditory stimuli. *Hearing by Eye: The Psychology of Lipreading*.
- Rouger, J. (2007). *Perception audiovisuelle de la parole chez le sourd postlingual implanté cochléaire et le sujet normo-entendant : étude longitudinale psychophysique et neurofonctionnelle* (Thèse de doctorat). Université de Toulouse III.
- Rouger, J., Lagleyre, S., Fraysse, B., Deneve, S., Deguine, O., & Barone, P. (2007). Evidence that cochlear-implanted deaf patients are better multisensory integrators. *Proceedings of the National Academy of Sciences of the United States of America, 104*(17), 7295-7300.
- Rouger, J., Lagleyre, S., Démonet, J.-F., Fraysse, B., Deguine, O., & Barone, P. (2012). Evolution of crossmodal reorganization of the voice area in cochlear-implanted deaf patients. *Human Brain Mapping, 33*(8), 1929-1940.
- Routhier, C., Bilodeau-Mercure, M., Lortie, C., Guitton, M., Sato, M., & Tremblay, P. (2013). Caractéristiques de la production de la parole au cours du vieillissement normal. *Affiche*

*présentée lors de la Journée scientifique du Réseau provincial de recherche en adaptation réadaptation (REPAR), Québec.*

- Russell, A., Penny, L., & Pemberton, C. (1995). Speaking fundamental frequency changes over time in women: a longitudinal study. *Journal of Speech and Hearing Research*, *38*(1), 101-109.
- Saltzman, E. (1986). Task Dynamic Coordination of the Speech Articulators: A preliminary model. *Experimental Brain Research Series*, *15*, 130 - 144.
- Sato, M., Tremblay, P., & Gracco, V. L. (2009). A mediating role of the premotor cortex in phoneme segmentation. *Brain and Language*, *111*(1), 1-7.
- Sato, M., Grabski, K., Garnier, M., Granjon, L., Schwartz, J.-L., & Nguyen, N. (2013). Converging toward a common speech code: imitative and perceptuo-motor recalibration processes in speech production. *Frontiers in Psychology*, *422*.
- Savariaux, C., Perrier, P., Orliaguet, J. P., & Schwartz, J. L. (1999). Compensation strategies for the perturbation of French [u] using a lip tube. II. Perceptual analysis. *The Journal of the Acoustical Society of America*, *106*(1), 381-393.
- Schroeder, M., R., Atal, B., & Hal, J. (1979). Objective measure of certain speech signal degradations based on masking properties of human auditory perception. *Frontiers of Speech Communication Research*, 217-229.
- Schwartz, J.-L., Boë, L.-J., Vallée, N., & Abry, C. (1997a). Major trends in vowel system inventories. *Journal of Phonetics*, *25*(3), 233-253.
- Schwartz, J.-L., Boë, L.-J., Vallée, N., & Abry, C. (1997b). The Dispersion-Focalization Theory of vowel systems. *Journal of Phonetics*, *25*(3), 255-286.
- Schwartz, J.-L., Abry, C., Boë, L.-J., & Cathiard, M.-A. (2002). Phonology in a theory of perception-for-action-control. *Phonetics, phonology and cognition*, (2002a), 255-280.
- Schwartz, J.-L., Basirat, A., Ménard, L., & Sato, M. (2012). The Perception-for-Action-Control Theory (PACT): A perceptuo-motor theory of speech perception. *Journal of Neurolinguistics*, *25*(5), 336-354.
- Sekiyama, K., Soshi, T., & Sakamoto, S. (2014). Enhanced audiovisual integration with aging in speech perception: a heightened McGurk effect in older adults. *Frontiers in Psychology*, *5*.
- Skinner, M. W., Fourakis, M. S., Holden, T. A., Holden, L. K., & Demorest, M. E. (1999). Identification of speech by cochlear implant recipients with the multipeak (MPEAK) and spectral peak (SPEAK) speech coding strategies II. Consonants. *Ear and Hearing*, *20*(6), 443-460.
- Skipper, J. I., van Wassenhove, V., Nusbaum, H. C., & Small, S. L. (2007). Hearing lips and seeing voices: how cortical areas supporting speech production mediate audiovisual speech perception. *Cerebral Cortex (New York, N.Y.: 1991)*, *17*(10), 2387-2399.
- Sommers, M. S., Tye-Murray, N., & Spehar, B. (2005). Auditory-visual speech perception and auditory-visual enhancement in normal-hearing younger and older adults. *Ear and Hearing*, *26*(3), 263-275.
- Stevens, K. N. (1960). Toward a Model for Speech Recognition. *The Journal of the Acoustical Society of America*, *32*(1), 47-55.

- Stevens, K., & Halle, M. (1967). Remarks on analysis by synthesis and distinctive features. *Models for the Perception of Speech and Visual Form.*, 88-102.
- Stevens, K. N. (1972). The quantal nature of speech: Evidence from articulatory-acoustic data. In *Human communication: A unified view* (p. 51-66). EE David et PB Denes editors.
- Stevens, K. N., & Blumstein, S. E. (1978). Invariant cues for place of articulation in stop consonants. *The Journal of the Acoustical Society of America*, 64(5), 1358-1368.
- Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics*, 17, 3-46.
- Stevens, K. N., & Keyser, S. J. (2010). Quantal theory, enhancement and overlap. *Journal of Phonetics*, 38(1), 10-19.
- Stevenson, R. A., Zemtsov, R. K., & Wallace, M. T. (2012). Individual differences in the multisensory temporal binding window predict susceptibility to audiovisual illusions. *Journal of Experimental Psychology. Human Perception and Performance*, 38(6), 1517-1529.
- Stuart, A., Kalinowski, J., Rastatter, M. P., & Lynch, K. (2002). Effect of delayed auditory feedback on normal speakers at two speech rates. *The Journal of the Acoustical Society of America*, 111(5 Pt 1), 2237-2241.
- Sumby, W. H., & Pollack, I. (1954). Visual Contribution to Speech Intelligibility in Noise. *The Journal of the Acoustical Society of America*, 26(2), 212-215.
- Torre, P., & Barlow, J. A. (2009). Age-related changes in acoustic characteristics of adult speech. *Journal of communication disorders*, 42(5), 324-33.
- Treille, A., Cordeboeuf, C., Vilain, C., & Sato, M. (2014). Haptic and visual information speed up the neural processing of auditory speech in live dyadic interactions. *Neuropsychologia*, 57, 71-77.
- Treille, A., Vilain, C., & Sato, M. (2014). The sound of your lips: electrophysiological cross-modal interactions during hand-to-face and face-to-face speech perception. *Frontiers in Psychology*, 5.
- Tremblay, P., Dick, A. S., & Small, S. L. (2013). Functional and structural aging of the speech sensorimotor neural system: functional magnetic resonance imaging evidence. *Neurobiology of Aging*, 34(8), 1935-1951.
- Turgeon, C., Prémont, A., Trudeau-Fisette, P., & Ménard, L. (2015). Exploring production-perception relationships in normal hearing and cochlear implant adults: a lip-tube perturbation study. *Clinical Linguistics and Phonetics*, 29(5), 378-400.
- Tyler, R. S., Fryauf-Bertschy, H., Kelsay, D. M., Gantz, B. J., Woodworth, G. P., & Parkinson, A. (1997). Speech perception by prelingually deaf children using cochlear implants. *Otolaryngology-Head and Neck Surgery: Official Journal of American Academy of Otolaryngology-Head and Neck Surgery*, 117(3 Pt 1), 180-187.
- Välimaa, T. T., Määttä, T. K., Löppönen, H. J., & Sorri, M. J. (2002a). Phoneme recognition and confusions with multichannel cochlear implants: consonants. *Journal of Speech, Language, and Hearing Research: JSLHR*, 45(5), 1055-1069.
- Välimaa, T. T., Määttä, T. K., Löppönen, H. J., & Sorri, M. J. (2002b). Phoneme recognition and confusions with multichannel cochlear implants: vowels. *Journal of Speech, Language, and Hearing Research: JSLHR*, 45(5), 1039-1054.

- van Wieringen, A., & Wouters, J. (1999). Natural vowel and consonant recognition by Laura cochlear implantees. *Ear and Hearing, 20*(2), 89-103.
- Vick, J., Lane, H., Perkell, J., Matthies, M., Gould, J., & Zandipour, M. (2000). Speech perception, production, and intelligibility improvements in vowel-pair contrasts among adults who receive cochlear implants. *The Journal of the Acoustical Society of America, 108*(5), 2600-2601.
- Wilson, S. M., Saygin, A. P., Sereno, M. I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience, 7*(7), 701-702.
- Wilson, S. M., & Iacoboni, M. (2006). Neural responses to non-native phonemes varying in producibility: evidence for the sensorimotor nature of speech perception. *NeuroImage, 33*(1), 316-325.
- Winkler, R., Ma, L., & Perrier, P. (2011). A model of optimal speech production planning integrating dynamical constraints to achieve appropriate articulatory timing. In D. J. O. S.R. Baum, V.L. Gracco L. Ménard (éd.), *9th International Seminar on Speech Production (ISSP 2011)* (p. 235-236). Montréal, Canada.
- Xu, L., Zhou, N., Chen, X., Li, Y., Schultz, H. M., Zhao, X., & Han, D. (2009). Vocal singing by prelingually-deafened children with cochlear implants. *Hearing research, 255*(1-2), 129-134.
- Zekveld, A. A., Heslenfeld, D. J., Festen, J. M., & Schoonhoven, R. (2006). Top-down and bottom-up processes in speech comprehension. *NeuroImage, 32*(4), 1826-1836.



