UNIVERSITÉ DU QUÉBEC À MONTRÉAL

INTÉGRATION MULTISENSORIELLE DANS LES MÉCANISMES DE PERCEPTION ET DE PRODUCTION DE LA PAROLE CHEZ L'ENFANT ET L'ADULTE FRANCOPHONE.

THÈSE

PRÉSENTÉE

COMME EXIGENCE PARTIELLE

DU DOCTORAT EN LINGUISTIQUE

PAR

PAMÉLA TRUDEAU-FISETTE

DÉCEMBRE 2022

REMERCIEMENTS

Mes plus sincères remerciements vont à Lucie Ménard, ma directrice, avec qui j'ai le plaisir de travailler depuis le baccalauréat. Au cours de ces 9 dernières années, tu as su former la chercheuse en moi, et je t'en serai éternellement reconnaissante.

Tu m'as d'abord transmis ta passion pour la recherche. Au travers nos multiples projets, tu m'as insufflé une rigueur scientifique et m'as montré à sortir des sentiers battus. J'ai pu, grâce à toi, faire voyager nos idées aux quatre coins du monde et me tailler une place dans le milieu académique.

Tu m'as aussi ouvert les portes au merveilleux monde de l'enseignement. Je garde en tête la passion, le plaisir et la générosité avec lesquels tu te présentes devant tes étudiants. J'espère un jour être le huitième de la scientifique que tu es.

Au-delà de tout ça, tu m'as surtout donné l'exemple d'une direction inégalable. Il s'agit d'un travail laborieux, mais je crois pouvoir dire, sans peser mes mots, que je suis tombée sur la perle rare. Très chère Lucie, tu as su être la meilleure directrice pour moi.

Merci pour ta présence, ton temps, tes ressources, tes lectures (et tes relectures), ton intérêt, tes connaissances, ta vision, ta considération, mais surtout, merci pour ta confiance. Sans toi, jamais je n'aurais même pensé me rendre où je suis aujourd'hui. Je termine cette aventure avec une immense fierté et j'en garderai un excellent souvenir.

Ce travail, c'est aussi le tien.

I would also like to offer my heartfelt acknowledgment to David Ostry and Mark Tiede for taking the time to read this thesis. I would also like to recognize David's generosity in letting me use his research facilities for my data collection. I also wish to show my gratitude to Mark for his great patience, his infinite abilities to solve my MatLab problems (!), and for our discussions about travel. It is always a pleasure not to talk work with you :P

Je remercie aussi, Mme Marine Le Mene Guigoures d'avoir accepté de siéger sur mon jury, au pied levé, et par le fait même d'avoir sauvé ma soutenance. Merci pour votre lecture extraordinairement rapide, mais tout autant rigoureuse.

Je me dois aussi de remercier Pascal Perrier d'avoir accepté de m'accueillir «chez lui» à GISPA-Lab pour un séjour de recherche. Merci pour ta générosité et ta douceur. C'est une réelle joie de travailler auprès de toi. Je garde un plaisant souvenir de Grenoble et des amitiés formées (un gros clin d'œil pour toi J-F!)

Merci à mes amis du Laboratoire de phonétique de l'UQAM (dont certains remontent de loin!): Caro, Laureline, Lambert, Cristina. Un merci tout spécial à Camille, sans qui cette thèse n'aurait jamais vu le jour. Ta présence assidue à CHACUNE de mes collectes de données est un cadeau sans nom! Merci aussi à Christine, Dominique et Marie qui font partie de mon quotidien depuis 2012. Je suis bien chanceuse de vous avoir dans ma vie.

Merci à mes parents, à ma famille et à mes amis qui m'ont toujours encouragée dans cette aventure et qui ont su constamment me démontrer combien ils sont fiers de moi… Ça y est, c'est fini!

Merci finalement à Félix, l'homme de ma vie. Merci d'avoir écouté mes explications trop longues et détaillées sur «le résultat génial que je viens de trouver!» et de faire semblant d'y être hyper intéressé. Merci d'être resté patient (ou presque) quand je te disais (pour la 12e fois) que je ne trouvais plus ma plage de données. Merci de me

faire rire, de me faire confiance, de me pousser à être la meilleure version de moi-même et d'être le meilleur père pour notre fille. Je sais que tu aurais pu écrire cette thèse en 2-3 fins de semaine, merci donc aussi d'avoir respecté mon rythme ;) Sans blague, je ne peux m'imaginer de réaliser ce beau projet sans toi à mes côtés.

Un dernier petit mot pour mes filles Dalie et Zara (même si tu es arrivée à la toute fin de cette aventure). Le plus grand souhait que j'ai pour vous est que vous trouviez le moyen de vous accomplir. On passe notre vie à travailler, aussi bien trouver quelque chose de passionnant. Vous verrez, c'est merveilleux!

TABLE DES MATIÈRES

LISTE DES FIGURES

LISTE DES TABLEAUX

LISTE DES ABRÉVIATIONS, DES SIGLES ET DES ACRONYMES

IMS (MSI)	Intégration multisensorielle (Multisensory integration)
(GA)	(General Auditory approach)
(PACT)	(Perception-for-Action-Control Theory)
(VTE)	(Verbal Transformation Effect)
(AO)	(Auditory Only)
(SS)	(Skin Stretch)
(AV)	(Audiovisual)
(AVI)	(Audiovisual integration)

RÉSUMÉ

L'intégration multisensorielle fait référence à la capacité du cerveau à assimiler les signaux provenant de multiples modalités et joue un rôle déterminant dans le développement des habiletés de parole. Il est maintenant reconnu que les systèmes auditif et proprioceptif sont impliqués dans les procédés de production de la parole, et que les mécanismes de perception tiennent compte des informations auditives et visuelles; ces différents systèmes fonctionnant de façon complémentaire. Plus récemment, il a été démontré qu'un transfert des informations sensorielles s'opèrerait entre les systèmes de perception et de production, témoignant ainsi de l'étroite relation qu'ils entretiennent. Par exemple, la rétroaction proprioceptive serait aussi impliquée dans les mécanismes de perception. Mais qu'en est-il de l'enfant? À l'heure actuelle, les études menées chez la population en développement témoignent d'une énorme variabilité. Néanmoins, elles suggèrent que les enfants ne sont pas en mesure de traiter les informations multisensorielles comme les adultes. En ce qui concerne les principes de complémentarité et de transfert des informations sensorielles, à notre connaissance, aucune étude ne s'y est encore penchée. Au sein du présent travail, trois expérimentations ont été réalisées auprès de 65 locuteurs du français québécois (30 adultes et 25 enfants d'âge préscolaire). Via une étude de manipulation du retour auditif et une série de deux tests de perception bimodale (audiovisuel et audioproprioceptif), nous avons pu témoigner du rôle des informations auditives dans le développement du contrôle de la parole et observer les procédés d'intégration des informations auditive et proprioceptive, puis auditive et visuelle au sein des mécanismes de perception de la parole. Des analyses supplémentaires ont aussi été effectuées afin d'observer si les notions de complémentarités des systèmes et de transfert des informations sensorielles sont présentes chez l'enfant. À la lumière des analyses effectuées, il semble que les représentations sensorimotrices soient toujours en cours de maturation chez les individus de 5-6 ans, mettant ainsi un frein aux comportements de compensation auditive et aux habiletés d'intégration multimodale observés chez l'adulte. Il en ressort toutefois que les phénomènes de complémentarité des systèmes sensoriels et de transfert entre les mécanismes de perception et de production en sont facilités.

Mots clés : Intégration multimodale, contrôle moteur, perception de la parole, adultes, enfants, représentations sensorimotrices.

INTRODUCTION

Multimodalité et perception

Nous sommes quotidiennement confrontés à différents types d'indices sensoriels. Inconsciemment, ces stimuli (saveurs, odeurs, textures, etc.) sont combinés en un percept unique, riche en informations. L'intégration multisensorielle (IMS), également appelée intégration multimodale, fait référence à la capacité du cerveau à assimiler les signaux provenant de multiples modalités dans le but de tirer profit des informations de chaque sens et, ainsi, réduire l'ambiguïté et renforcer notre perception du monde (Molholm et al., 2002; Robert-Ribes et al., 1998; Stein et al., 1996; Stein & Meredith, 1993)

L'intégration multimodale occupe une place prépondérante dans la manière dont l'information est traitée en façonnant la manière dont les indices des multiples sources sont perçus. Si certains chercheurs suggèrent que le cerveau d'un nourrisson est doté de fonctionnalités multisensorielles dès la naissance (Bower et al., 1970; Streri & Gentaz, 2004), d'autres avancent plutôt que les habiletés d'IMS se développent au cours du développement, découlant des expériences sensorielles (Birch & Lefford, 1963; Burr & Gori, 2012; Yu et al., 2010). À ce titre, il a été démontré que nos différents systèmes sensoriels (gustatif, olfactif, visuel, etc.) se développent à des rythmes différents et de différentes manières. Par conséquent, les mécanismes d'IMS diffèreraient en fonction des systèmes sensoriels impliqués (Burr & Gori, 2012; Dionne-Dostie et al., 2015; Gori et al., 2008; Gougoux et al., 2004).

Bien qu'on ne sache toujours pas si l'IMS dépend d'une faculté innée, les données recueillies auprès des nourrissons, des enfants et des adultes suggèrent que les expériences sensorielles, tant unimodales que multimodales, et que la maturation du cerveau souscrivent la mise en place d'un traitement d'intégration efficace (Gori et al., 2008; Hillock et al., 2011; Krakauer et al., 2006; Nardini et al., 2008; Neil et al., 2006; Rentschler et al., 2004; Stein et al., 2014)

L'intégration multimodale est cruciale pour le développement de la parole. Lors de la période du babillage, les sons produits et perçus sont associés aux mouvements des articulateurs nécessaires à leur production et sont stockés en mémoire sous la forme d'un modèle sensorimoteur de production de la parole (pour une description des différents modèles de production de la parole, voir (Patri, J.-F., 2018)). Les liens sensori-moteurs associés à chaque phonème de la langue maternelle sont généralement acquis chez l'enfant dès l'âge de 4 ans. Cependant, ceux-ci se raffinent pendant l'enfance, la période de 4 à 6 ans étant une période charnière. Au plan perceptuel, la nature multimodale de la parole reste encore sous-estimée. Certes, les locuteurs démontrent souvent une dominance auditive (Hecht & Reiner, 2009; Lametti et al., 2012), mais de plus en plus d'études suggèrent que d'autres modalités sensorielles sont impliquées dans le traitement de la parole (Ito et al., 2009; Lametti et al., 2012; Perrier, 1995; Skipper et al., 2007; S. Tremblay et al., 2003).

L'objectif de la thèse

L'objectif de la présente thèse est d'approfondir les connaissances quant à la multimodalité de la parole en portant une attention particulière au développement des habiletés d'intégration multisorielle dans les processus de perception. Ainsi, trois expérimentations ont été réalisées afin de démontrer l'apport des informations auditives dans le contrôle de la parole, mais aussi de témoigner de l'intégration des informations audio-proprioceptives et audiovisuelles au sein de la perception des

mécanismes de perception de la parole chez l'enfant d'âge préscolaire et chez l'adulte francophone.

La structure de la thèse

Au premier chapitre, nous présenterons le contexte théorique sur lequel s'appuie le présent projet. Nous discuterons des théories de perception de la parole, de la relation entre les mécanismes de production et de perception de la parole et, finalement, nous présenterons les travaux ayant permis de mettre en lumière la nature multimodale de la parole.

Au chapitre 2, nous présenterons les résultats de notre première étude, soit celle discutant du rôle des informations auditives dans les stratégies de planification de la parole chez l'enfant d'âge préscolaire et chez l'adulte.

Au chapitre 3, nous discuterons de notre étude comparant les habiletés d'intégration des informations auditives et proprioceptives d'enfants de 5 à 6 ans et d'adultes francophones dans les procédés de perception de la parole. Cet article est publié dans la revue *Frontiers in Humain Neuroscience*[1].

Le chapitre 4 consistera en la présentation de notre troisième et dernière étude, soit celle portant sur le développement des habiletés d'intégration des informations

[1] Trudeau-Fisette, P., Ito T. et Ménard, L. (2019). Auditory and Somatosensory Interaction in Speech Perception in Children and Adults. In *Frontiers in Humain Neuroscience*. doi.org/10.3389/fnhum.2019.00344

auditive et visuelle dans les mécanismes de perception de la parole. Cet article est publié dans la revue *Frontiers in Psychology*[2].

Finalement, au chapitre 5, nous effectuerons un retour sur les trois études réalisées. Nous effectuerons également des analyses supplémentaires afin de démontrer de quelle manière les informations auditives, visuelles et proprioceptives fonctionnent de façon complémentaire dans le développement des habiletés de perception de la parole.

[2] Trudeau-Fisette, P., Arnaud L., & Ménard, L. (2022). Audiovisual interaction in speech perception in children and adults. In *Frontiers in Psychology*. doi.org/10.3389/fpsyg.2022.740271

CHAPITRE I

CADRE THÉORIQUE

Différentes approches théoriques cherchant à expliquer les mécanismes de perception de la parole sont décrites dans la littérature actuelle. Ayant le même but général, ces théories divergent tout de même quant à leurs principes fondamentaux. Elles s'opposent principalement quant à la nature de l'unité encodée par le locuteur-auditeur et quant aux processus spécifiques ou génériques responsables du traitement de la parole.

L'objectif de cette section est d'effectuer un bref retour sur la position des trois grandes familles de théories de perception de la parole. Ainsi, nous procéderons à une courte description des théories articulatoires et des théories acoustiques. Compte tenu du cadre théorique dans lequel se situe la présente thèse, les théories hybrides seront présentées de façon plus détaillée.

1.1 Les théories de perception de la parole

1.1.1 Les théories motrices

De façon générale, les théories motrices proposent que les informations essentielles à la perception de la parole soient présentes dans les configurations articulatoires. Ces théories s'appuient sur le fait que les paramètres acoustiques reliés aux cibles sonores sont trop variables pour que ceux-ci aient un quelconque rôle invariant.

Cette famille est composée de plusieurs théories, par exemple la *Analysis-by-Synthesis Theory of Speech Perception* (K. N. Stevens & Halle, 1967), la *Articulatory Phonology* (Browman & Goldstein, 1986) et la *Direct Realism Theory of Speech Perception* (DRT) (Fowler, 1986) et la *Motor Theory of Speech Perception* (Liberman et al., 1967), considérée comme la théorie fondatrice de cette approche.

L'hypothèse principale de ces théories est donc que les locuteurs perçoivent la parole via l'identification des positions du conduit vocal avec lesquelles elle est prononcée, plutôt que par les indices acoustiques. Tel qu'évoqué par Liberman: «*There is typically a lack of correspondence between acoustic cue and perceived phoneme, and in all these cases it appears that perception mirrors articulation more closely than sound*» (Liberman et al., 1967), p. 453). En d'autres mots, la transition entre la constriction du conduit vocal et le signal sonore serait sujette à trop de complexité, dont la principale cause est l'effet de coarticulation (Diehl et al., 2004).

Des observations classiques, comme *l'effet McGurk* (McGurk & MacDonald, 1976), une illusion perceptuelle découlant d'informations vues et entendues incongrues (nous y reviendrons plus tard), sont, pour les auteurs des théories motrices, une preuve considérable de l'apport des informations gestuelles dans le traitement de la parole. Ainsi, les théories motrices supposent que l'unité minimale de la parole soit de nature motrice.

1.1.2 Les théories auditives

D'autres maintiennent que l'information fondamentale à la perception du langage peut difficilement provenir des gestes articulatoires. La parole étant avant tout un procédé oral, les auteurs des théories acoustiques proposent plutôt que les informations nécessaires à la perception de la parole et, par le fait même, à la compréhension d'un message soient présentes dans le signal acoustique.

La *Quantum Theory* de (Stevens, S., 1972) et la *Auditory Enhancement Theory* de (Diehl & Kluender, 1989) font partie de la famille des théories acoustiques. De façon générale, les théories acoustiques partagent les principes de la *General Auditory approach* (GA) (Diehl et al., 2004).

La GA, ayant d'abord pris forme dans les années 80, tente de démontrer que le système auditif ne traite pas directement les informations acoustiques perçues, mais les transforme via des opérations de normalisation. Ainsi, l'identification des unités linguistiques peut s'effectuer de façon optimale et ce, peu importe les variables telles que l'âge, le sexe, le contexte ou encore le débit de parole pouvant influencer le signal acoustique (Lotto & Holt, 2015).

D'ailleurs, pour les auteurs en faveur de l'approche auditive, la perception de la parole est d'abord vue comme une habileté permettant d'identifier des patrons acoustiques afin de les encoder en unités significatives. Pour assurer une perception efficace, ces unités doivent donc être acoustiquement discriminables entre elles (Lotto & Holt, 2015; J. J. Ohala, 1996). Cette idée de l'importance du contraste auditif dans la perception est d'ailleurs reprise afin de démontrer la relation entre la distance maximale entre des voyelles et l'inventaire vocalique d'une langue (Liljencrants et al., 1972; Lindblom, 1986; Vallée et al., 1999). Pour ces auteurs, c'est ce qui permettrait d'expliquer pourquoi la majorité des systèmes à trois voyelles sont majoritairement composés des voyelles /i/, /a/ et /u/. Leurs patrons formantiques étant les plus distancés dans l'espace acoustique, ces voyelles sont les plus facilement reconnaissables auditivement.

1.1.3 Les théories hybrides

Finalement, les théories hybrides postulent que les informations nécessaires à la perception de la parole ne peuvent être uniquement motrices ou acoustiques. Ainsi, il semble que, lorsqu'ils sont pris individuellement, ces indices sensoriels sont

insuffisants à l'identification optimale des unités de parole. C'est d'ailleurs vers cette hypothèse que tendent la majorité des études récentes, qu'elles adoptent une approche comportementale, computationnelle ou neurologique (pour une revue, voir (Skipper et al., 2017)). De plus, les théories hybrides soulignent l'apport des informations multisensorielles au sein des habiletés de perception de la parole.

Afin de dresser un portait plus précis des caractéristiques de l'approche hybride, nous procéderons à la description de la *Perception-for-Action-Control Theory* (PACT) (Schwartz et al., 2012). Contrairement aux théoriciens prônant un traitement monosensoriel de la parole, les auteurs de la théorie de PACT défendent l'idée que les informations pertinentes à la perception de la parole sont présentes dans des unités perceptuo-motrices où les gestes (*articulatory knowledge*) sont modulés par la perception. Pour eux, c'est donc une unité perceptuo-motrice qui serait associée de façon invariante à un son.

En effet, la PACT tente de démontrer que, bien que les notions articulatoires soient à considérer dans la perception de la parole, celles-ci ne sont pas purement motrices, mais seraient en fait façonnées par la perception acoustique (Schwartz et al., 2007). Afin d'illustrer cette idée, les auteurs utilisent l'exemple du mouvement d'arrondissement des lèvres, comme celui impliqué dans le passage de la production du son [i] vers le son [y]. Plus précisément, ils expliquent que, malgré le fait que l'arrondissement survient de façon progressive, la perception du son produit demeure longtemps associée à la voyelle [i], puis devient abruptement associée à la voyelle [y] Tel que schématisé à la figure 1.1, le changement catégoriel entre les deux voyelles s'illustre par une courbe abrupte (en rouge), plutôt que progressive (en noir). Pour les auteurs, cela démontre qu'un geste articulatoire (ici, l'arrondissement) ne peut expliquer la perception d'un son d'un point de vue uniquement moteur et, alors, qu'il doit alors être considéré comme un paramètre articulo-acoustique dans lequel la perception module l'action (Schwartz et al., 2012).

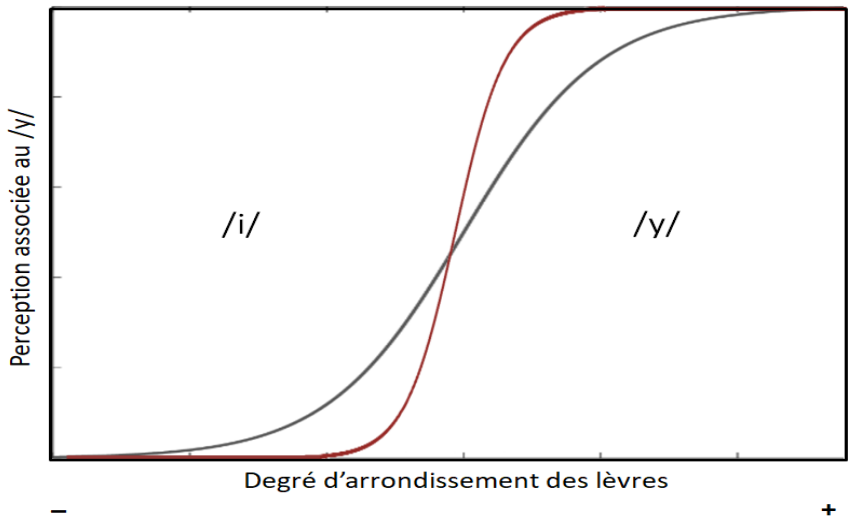

Figure 1.1 Schématisation du changement perceptuel abrupt entre les cibles [i] et [y]

Cet aspect perceptuo-moteur, ayant trouvé écho dans les récentes études de neuroimagerie et d'électrophysiologie, serait d'ailleurs impliqué tant dans les mécanismes de développement de la parole, où des prototypes moteurs sont confrontés à la perception catégorielle afin de développer un registre d'unités de parole, que dans son utilisation courante. Dans ce dernier cas, les prototypes moteurs sont alors comparés à la rétroaction sensorielle en temps réel permettant, par exemple, d'adapter la production en condition de parole bruitée ou encore de procéder à la segmentation efficace des unités linguistiques perçues. L'implication de l'unité perceptuo-motrice au sein de ces deux mécanismes, ceux-ci tenant une place majeure dans le cadre du présent projet doctoral, sera ici discutée.

1.1.3.1 Le développement de la parole

Afin de démontrer en quoi les informations impliquées dans le développement de la parole sont de nature perceptuo-motrice, les auteurs de la PACT font d'abord référence à la distribution de la hauteur des voyelles dans l'espace acoustique. Ceux-ci appuient leur hypothèse sur le principe d'*Adaptive Variability Theory* proposé par (Lindblom, 1990), voulant que la distribution des voyelles sur l'axe de F1

(correspondant, en termes articulatoires, au degré d'aperture de la cavité buccale) respecte le principe de distance suffisante à la perception d'un contraste entre les cibles acoustiques. Un exemple de répartitions possibles des voyelles avants, de trois locuteurs différents, est illustré à la figure 1.2. On y constate effectivement que, tant qu'elle est suffisante à la perception d'un contraste, la distance entre les voyelles [i], [e], [ɛ] et [a] (en noir sur la figure) est variable et peut différer d'un locuteur à l'autre.

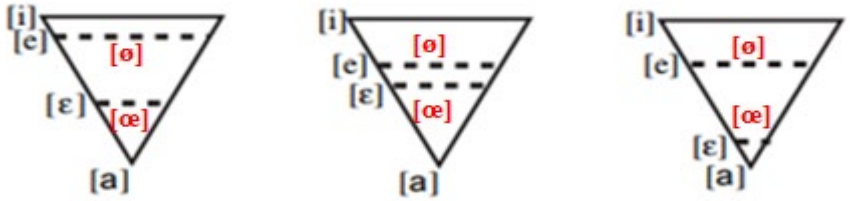

Figure 1.2 Exemples de la distribution de la hauteur des voyelles dans l'espace acoustique. Adapté de Schwartz et coll. (2012).

L'aspect perceptuo-moteur du développement des unités phonémiques s'observe alors dans la découverte de (Ménard & Schwartz, 2014) où les auteurs notent que, pour la distribution sur l'axe de F2 (faisant, cette fois-ci référence au lieu d'articulation), un transfert du contrôle acquis semble s'effectuer de manière à maintenir le contraste haut/bas des voyelles antérieures. De ce fait, tel qu'illustré par les cibles en rouge sur la figure 1.2, un locuteur pour qui les unités vocaliques [e] et [ɛ] sont très rapprochées, les unités [ø] et [œ] le seront aussi. Ce principe appelé *Maximal Use of Available Controls* permet ainsi de témoigner de l'apport des informations perceptuelles dans l'établissement des gestes articulatoires lors du développement de la parole (Schwartz et al., 2012).

Les auteurs de la PACT proposent également que les gestes articulatoires soient non seulement façonnés par la perception, mais qu'ils soient également définis en

fonction des valeurs perceptuelles acoustiques. À ce titre, ils discutent de l'apport des informations acoustico-perceptives dans l'organisation du système vocalique et reprennent l'hypothèse de Liberman voulant que la répartition des unités vocaliques soit expliquée par les configurations articulatoires extrêmes, en termes de hauteur, d'antéro-postériorité et d'arrondissement. Toutefois, selon cette conception, les unités vocaliques à retrouver dans un système à trois voyelles pourraient être tant [i a u] que [y a ɯ] (voir figure 1.3). Comment alors expliquer qu'aucun système vocalique n'est composé de ce dernier groupe de voyelles? Allant de pair avec la proposition de la GA, les auteurs de la PACT répondent que, si l'on considère les informations auditives (en plus des informations gestuelles), il est alors possible de constater que les cibles [i] et [u] sont davantage distancées que les cibles [y] et [ɯ] et concluent que les notions articulatoires sont sélectionnées en fonction de leurs valeurs perceptuo-acoustiques (Schwartz et al., 2007, 2012).

Compte tenu de l'importance accordée aux informations multisensorielles, cette approche théorique tient également compte des informations perceptuo-visuelles dans l'organisation des phonèmes. C'est ce qui expliquerait la forte présence des consonnes [m] et [n] dans les langues du monde. En effet, bien que le contraste entre ces dernières soit peu saillant acoustiquement, il s'avère visuellement très robuste (Schwartz et al., 2012).

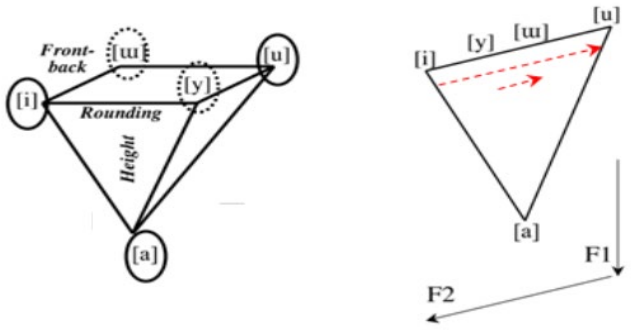

Figure 1.3 Schématisation de l'apport des informations acoustico-perceptives dans l'organisation des systèmes vocaliques. Adapté de Schwartz et coll. (2012).

Ainsi, la PACT démontre pleinement comment la distribution et l'organisation des phonèmes ne peuvent être expliquées que par des raisons purement acoustiques ou motrices et prouvent que, lors du développement des mécanismes de production et de perception de la parole, les gestes articulatoires sont en fait modulés par l'acoustique (Schwartz et al., 2007, 2012). Cette implication des unités perceptuo-motrices dans le développement de la parole est d'ailleurs expliquée par (Ménard, 2015) :

> «[…] According to PACT, speech goals correspond to multisensory perceptuo-motor units. In the course of speech development, perception and action are tightly linked, and speech perception necessarily involves procedural knowledge of the speech production mechanisms. Furthermore, perceptual mechanisms provide gestures with auditory, visual, and somatosensory templates that guide and maintain their development. » (p. 205)

1.1.3.2 La segmentation des unités de parole

Les informations perceptuo-motrices sont également mises à profit lors de la perception quotidienne de la parole, par exemple afin de faciliter la segmentation des unités de parole. Effectivement, alors qu'il était proposé depuis plusieurs années que les mécanismes permettant l'organisation du discours soient réalisés via des indices temporels, spectraux et fréquentiels (Bregman, 1990), la PACT défend plutôt que les indices utilisés lors de l'organisation du discours soient de nature perceptuo-motrice.

Pour illustrer l'interaction perception-articulation au sein du traitement de la parole, les auteurs présentent les travaux s'attardant à la *Verbal Transformation Effect* (VTE) (Warren, 1961; Warren & Gregory, 1958), un phénomène lors duquel la répétition en boucle d'un stimulus auditif engendre la perception multiple, mais stable d'un ou de plusieurs stimuli. Par exemple, l'écoute rapide et en boucle du mot «life» risque d'engendrer la perception du mot «fly», puis de revenir à celle de «life». Ce paradigme de *multistable perception* a d'ailleurs été largement étudié dans le domaine

de la perception visuelle, notamment avec l'illusion du cube de Necker (Necker, 1989).

Les travaux de (Pitt & Shoaf, 2001, 2002), s'attardant aux mécanismes qui sous-tendent la VTE, ont permis d'observer que ceux-ci s'opéraient via les étapes de segmentation, de considération lexico-sémantique et de filtrage auditif. Afin de comprendre davantage ce phénomène et dans le but d'y observer l'influence de la cohérence articulatoire, (Sato et al., 2006) proposent d'analyser la segmentation survenant lors de l'écoute de non-mots francophones comme «sep» et «pse». Selon leur hypothèse, comme les gestes nécessaires à l'articulation de la consonne [s] peuvent être anticipés lors de la production de la consonne [p], les participants de cette tâche perceptuelle auditive devraient majoritairement percevoir la cible «pse», celle-ci respectant davantage la synchronie articulatoire entre les gestes consonantiques. Les résultats de cette étude démontrent effectivement que le passage de la perception de «sep» vers «pse» est plus fréquent que celui de «pse» vers «sep»; le premier impliquant une resynchronisation articulatoire alors que le second nécessite une désynchronisation articulatoire (Sato et al., 2006; Schwartz et al., 2012). Cet impact de l'ordre des gestes articulatoires sur le résultat perceptuel témoigne ainsi pleinement du rapport important entre les mécanismes de perception et de production de la parole.

Peu après, (Sato et al., 2007) s'attardent aux procédés de segmentation de ces mêmes stimuli, mais cette fois-ci perçus en modalité audiovisuelle. En comparant la stabilité moyenne de la perception des stimuli «sep» ou «pse» présentés en boucle et en contexte de congruité ou d'incongruité audiovisuelle, les auteurs concluent que, lorsqu'un indice visuel est saillant (dans ce cas si, le geste articulatoire de [p]), le percept résultant sera de pair avec l'information visuelle (Sato et al., 2007). De ce fait, les auteurs de la PACT précisent que le poids respectif des paramètres perceptuo-moteur dépendra du type d'unités à percevoir (Schwartz et al., 2012).

Parallèlement aux données neurophysiologiques (Skipper et al., 2007, 2017), les expérimentations portant sur le développement de la parole et sur les procédés de segmentation du flux langagier dont nous avons discuté ont permis d'illustrer que les processus perceptuels mis à profit lors du traitement de la parole sont régis par des informations tant acoustiques qu'articulatoires et que ces indices perceptuo-moteurs multimodaux sont également impliqués au sein des mécanismes de production de la parole.

Vraisemblablement, bien qu'elles cherchent toutes à définir les procédés de perception de la parole, les différentes approches théoriques présentées proposent des hypothèses somme toute assez divergentes. Tel que nous l'avons vu, les théories articulatoires, acoustiques et hybrides s'opposent principalement quant à la nature de l'invariant. Toutefois, peu d'objections sont soulevées quant à la relation entre les mécanismes de perception et de production de la parole. Cette question sera plus amplement discutée dans la section qui suit.

1.2 La relation entre perception et production de la parole

Sans surprise, l'étroite relation entre les processus de perception et de production s'observe dans plusieurs domaines. La découverte des neurones miroirs (Di Pellegrino et al., 1992; Rizzolatti et al., 1996) témoigne d'ailleurs pleinement de l'existence neurocorticale d'un lien *perception-production*. Cette observation de neurones miroirs, d'abord réalisée chez les macaques, a permis de constater que certains neurones situés dans la région F5 du cortex pré moteur (zone normalement associée aux mouvements des mains) sont activés, non seulement lors de la réalisation d'un mouvement de saisie, mais aussi lors de l'observation de ce même mouvement.

Plus tard, des études rapportent que l'activation d'une région motrice lors de la perception d'un geste moteur s'observe aussi chez l'humain. Par exemple, (Fadiga, L. et al., 1995) montrent que, tout comme chez les macaques, les potentiels évoqués moteurs enregistrés lors de l'observation de mouvements de la main respectent les mêmes patrons d'activation que ceux recueillis lors de l'exécution de ces mouvements.

Dans le domaine des expérimentations langagières, (Fadiga, Luciano et al., 2002) démontrent ensuite qu'une activation plus importante de la région associée aux muscles de la langue est observée lors de l'écoute passive de mots nécessitant une importante activation de ce muscle, en comparaison avec l'écoute de mot impliquant davantage un mouvement labial. En ce sens, (D'Ausilio et al., 2009) montrent plus tard que l'identification de consonnes labiales et linguales produites dans le bruit est plus rapidement effectuée lorsqu'une stimulation magnétique transcrânienne est produite dans la région correspondant à la consonne à identifier.

Plusieurs études ont illustré la concomitance des procédés de perception et de production de la parole. À titre d'exemple, alors qu'ils mettent en relation les habiletés de discrimination de contrastes acoustiques et les habiletés de productions de cibles phonémiques, Perkell et al. (Perkell et al., 2004) montrent que les individus ayant un score de discrimination élevé produisent des contrastes acoustiques plus importants que ceux ayant un score de discrimination plus faible. Pour les auteurs, cela témoigne du fait que les gestes articulatoires sont organisés en fonction de l'espace auditif et démontre l'influence directe des capacités perceptuelles auditives sur les habiletés de production

(Shiller, Douglas M. et al., 2009) démontrent plus tard que l'apprentissage moteur découlant d'une manipulation acoustique peut générer un déplacement de la frontière perceptuelle entre deux phonèmes si ceux-ci sont affectés par la manipulation en

question. En effet, lors d'une tâche d'identification phonémique, les auteurs montrent que pour un même individu, l'emplacement de la frontière catégorielle entre les cibles /s/ et /ʃ/ diffère en fonction de si la tâche de perception a été réalisée avant ou après une tâche de manipulation de la fréquence centroïde (paramètre crucial pour la distinction de ces sibilantes). Des résultats similaires sont observés par (Lametti et al., 2014). Plus précisément, les auteurs concluent que l'adaptation à une manipulation du retour auditif, mais surtout l'apprentissage moteur qui s'en suit, mène à des changements perceptuels pour les cibles touchées par la manipulation acoustique.

(Nasir & Ostry, 2009) se sont aussi attardés aux conséquences d'un apprentissage moteur sur l'organisation acoustique des cibles de parole. Les auteurs ont constaté que l'adaptation motrice résultant de la compensation à une manipulation du retour somatosensoriel des gestes de la mâchoire lors de la production de la parole engendrait des changements dans la classification perceptive des sons de la parole, et ce, de façon proportionnelle au degré de compensation motrice. En effet, les résultats d'une tâche d'identification de cibles vocaliques tirées du continuum /æ - ɛ/, réalisée d'abord avant, puis après une tâche de production avec perturbation physique de la mâchoire, dénote un changement quant à la position de la frontière perceptuelle des cibles à identifier, laissant ainsi présumer qu'une certaine forme d'apprentissage perceptuel se serait opérée au travers de l'adaptation motrice.

Ainsi, l'influence du système moteur sur la perception de la parole illustre clairement de l'apport des informations proprioceptives (mouvements ressentis) sur cette habileté bien complexe. Un segment de parole ne serait donc pas seulement identifié en fonction des informations acoustiques, mais bien de façon multisensorielle. La section qui suit portera sur cet aspect multimodal de la parole; l'apport des informations auditives, mais surtout son interaction avec les informations proprioceptives et visuelles sera détaillé.

1.3 La multimodalité de la parole

1.3.1 Les informations auditives

Pour démontrer le rôle majeur de la rétroaction auditive sur les mécanismes de la parole, beaucoup d'études se sont penchées sur les conséquences de la surdité et ont montré que les personnes ayant une surdité congénitale ne portant pas d'appareils auditifs ont des difficultés à acquérir le langage oral (Lane et al., 2005; Oller & Eilers, 1988; Svirsky et al., 2004). Il a également été démontré que les personnes atteintes de surdité post-linguistique présentaient des signes de détérioration de la parole au niveau des paramètres segmentaux et suprasegmentaux, tels que la distance entre les voyelles acoustiques, la durée des voyelles, le rythme et l'intensité (Houde & Jordan, 1998; Lane et al., 2005; Lane & Wozniak Webster, 1998). En général, les études impliquant la population sourde oralisante soutiennent l'idée que la production de la parole dépend des informations sensorielles perçues par la rétroaction auditive.

Des études portant sur des populations au développement typique ont également permis de démontrer l'importance de la rétroaction auditive dans les habiletés de perception et de production de la parole. Dans ce contexte, le poids des informations auditives est examiné via les comportements compensatoires résultant de diverses manipulations sensorielles. La pertinence de ce type de paradigme expérimental a été établie il y a bien longtemps lors de travaux s'intéressant à la relation entre le contrôle moteur et les informations sensorielles. Parmi ceux-ci, on retrouve les travaux de (Held, 1965) démontrant qu'une manipulation du retour visuel pouvait engendrer un écart entre le mouvement planifié par le cerveau et celui réellement effectué lorsqu'il était demandé aux participants d'atteindre une cible. La figure 1.4 illustre le processus de compensation sensorielle découlant de l'altération de la rétroaction visuelle via des lunettes spécialisées, mais le phénomène observé est généralement le même, peu importe le type de manipulation. Avant l'application de la manipulation, une

concordance existe entre le geste planifié et le geste effectué (à gauche dans la figure 1.4). À la suite de l'introduction de la manipulation sensorielle, un écart entre *planification* et *exécution* survient (au centre). Puis, après une certaine période, un effet d'apprentissage est développé et une forme de correction motrice est produite afin de compenser la perturbation appliquée (à droite).

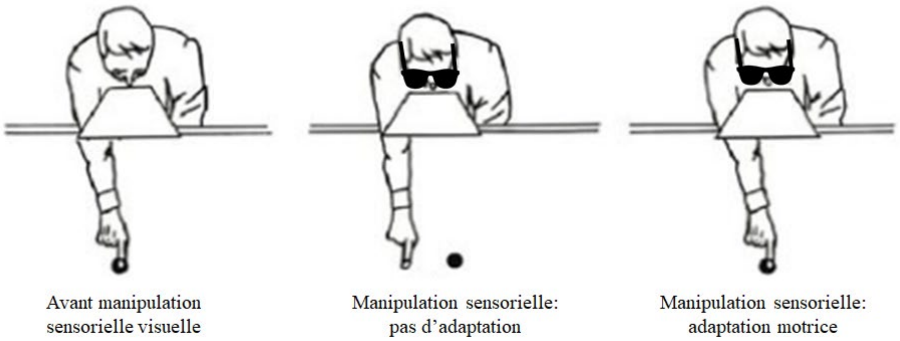

Figure 1.4 Représentation de la compensation motrice due à une manipulation du retour visuel (adaptée de (Pisella et al., 2006))

Ce procédé expérimental a par la suite été utilisé dans le domaine de la parole, soit par le biais d'une manipulation synthétique du retour auditif (Burnett et al., 1998; Jones & Munhall, 2000; MacDonald et al., 2010; Purcell & Munhall, 2006a). La modification acoustique peut être appliquée sur plusieurs paramètres, tels que le volume, le débit, la fréquence fondamentale ou encore sur les valeurs formantiques (voir schématisation à la figure 1.5).

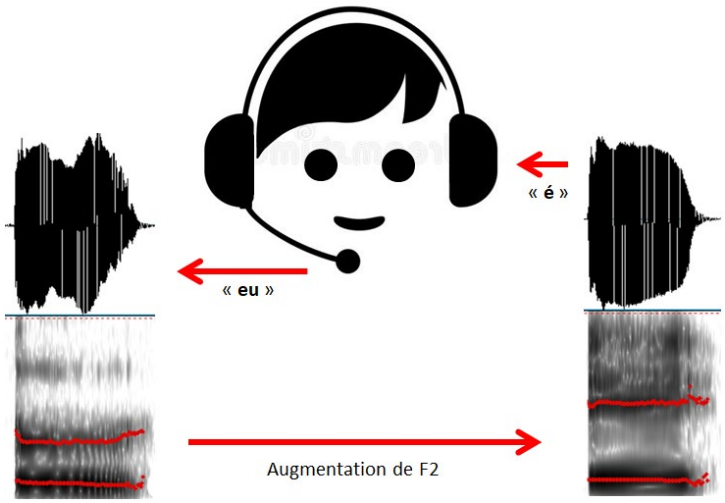

Figure 1.5 Schématisation d'une manipulation formantique.

Comme pour les tâches de manipulations visuelles, les tâches de manipulation de la rétroaction auditive démontrent que, chez l'adulte, un changement du retour auditif engendre une adaptation comportementale, conduisant ainsi à l'observation du rôle de la rétroaction auditive dans la planification des mouvements liés à la production de parole. Dans la section qui suit, nous effectuerons un retour sur quelques-unes des études ayant permis de témoigner du rôle des informations acoustiques dans le contrôle de la parole.

1.3.1.1 Le rôle de la rétroaction auditive dans le contrôle de la parole chez l'adulte

Les travaux portant sur la manipulation du retour auditif chez l'adulte montrent que celle-ci engendre rapidement une adaptation dans les productions (Burnett et al., 1998; Elman, 1981; Jones & Munhall, 2002; Max et al., 2003). Pour qu'un changement soit opéré, il est toutefois nécessaire que le locuteur soit confronté à plusieurs répétitions successives au cours desquelles il y a une incompatibilité entre le son émis et le son perçu (MacDonald et al., 2010). Effectivement, plusieurs auteurs soulèvent que les premières formes de compensation apparaissent peu avant la dixième production altérée (Max et al., 2003; Munhall et al., 2009; Neufeld, 2013). Par exemple, Purcell et

Munhall (2006a) observent que l'adaptation ne survient que lorsque le degré de la manipulation atteint un certain seuil. Malgré une importante variabilité intersujet, les auteurs ont établi que, pour une manipulation de 200 Hz, le seuil de compensation moyen, en fonction des deux directions de manipulation testées, est de 76 Hz. Globalement, ces résultats supposent qu'une certaine variabilité dans les productions vocaliques est acceptée par les locuteurs. Néanmoins, au-delà de cette zone tolérance, celle-ci étant variable d'un sujet à l'autre (nous y reviendrons), un comportement compensatoire est réalisé afin que les productions subséquentes s'alignent avec les buts perceptuels internes.

Que ce soit lors de manipulation de volume, de fréquence fondamentale ou de valeurs de F1 ou de F2, la majorité des études recensées rapportent que les participants adaptent généralement leurs productions dans la direction opposée à celle de la manipulation (Houde & Jordan, 1998; Mitsuya et al., 2013; Purcell & Munhall, 2006b, 2006a; Villacorta et al., 2007). En effet, afin de pallier la manipulation et d'entendre ce qu'ils croient produire, les participants émettent automatiquement des productions dont les paramètres acoustiques vont dans le sens contraire de celui de la manipulation.

Aussi, le degré de compensation semble être en lien avec l'ampleur de la modification acoustique reçue (MacDonald et al., 2010; Max et al., 2003; Purcell & Munhall, 2006b). En 2010, MacDonald et coll. manipulent simultanément, mais à différents degrés, les valeurs de F1 et de F2 des productions émises par les participants. Ils démontrent que la relation entre l'amplitude de la manipulation et le degré d'adaptation est linéaire pour les altérations de petites amplitudes seulement, soit de moins de 200 Hz pour F1 et de moins de 250 Hz pour F2. Il semblerait d'ailleurs qu'au-delà d'un certain point, l'effet de compensation peut stagner et même diminuer (MacDonald et al., 2010).

Le degré de compensation observé serait également en lien avec l'acuité auditive du locuteur. À ce titre, les sujets présentant de meilleures habiletés discriminatoires compenseraient davantage suite à une manipulation acoustique de F1 (Villacorta, 2006; Villacorta et al., 2007). Mitsuya et al. (Mitsuya et al., 2013) démontrent que l'organisation des cibles phonémiques dans le système vocalique peut également influencer l'ampleur de la compensation. Ils notent que comparativement aux participants anglophones, les sujets francophones compensent davantage à une manipulation lorsque les valeurs de F2 de la voyelle [ɛ] sont diminuées de 500 Hz. Les auteurs expliquent que les participants francophones seraient plus sensibles à cette modification acoustique en raison de la présence d'une cible vocalique supplémentaire dans leur répertoire phonémique.

Les études montrent que cet effet de compensation est assez robuste. Munhall et coll. (Munhall et al., 2009) ont d'ailleurs observé qu'il avait même lieu lorsqu'il était demandé aux participants de ne pas adapter leurs productions. Quant à lui, Villacorta (Villacorta, 2006) a observé que l'adaptation acoustique était présente même lorsque le retour auditif était masqué par du bruit. Néanmoins, peu importe le degré d'altération, les participants ne compensent que partiellement à la perturbation reçue (MacDonald et al., 2011; Munhall et al., 2009; Purcell & Munhall, 2006b, 2006a). Bien qu'une importante variation intersujet soit présente, les compensations moyennes relevées par les auteurs correspondent généralement à 25% de la manipulation appliquée (MacDonald et al., 2010, 2011; Munhall et al., 2009; Purcell & Munhall, 2006a).

Finalement, plusieurs auteurs notent un effet de persistance de la compensation (Jones & Munhall, 2002; Max et al., 2003; Neufeld, 2013; Purcell & Munhall, 2006b; Villacorta et al., 2007). En d'autres mots, les individus maintiendraient une certaine forme de compensation, même lorsque la manipulation acoustique est retirée. Il semblerait que cet effet de persistance ne soit pas influencé par la direction de la

manipulation ni par la durée d'exposition à une rétroaction auditive altérée. Encore une fois, ce phénomène s'observe autant lors de la manipulation de paramètres segmentaux que suprasegmentaux.

1.3.1.2 Le rôle de la rétroaction auditive dans le contrôle de la parole chez l'enfant

Compte tenu du rôle des informations auditives dans le développement des habiletés de parole, quelques auteurs ont aussi voulu observer si ces principes généraux étaient observables chez la population enfant. Les études de manipulation du retour acoustique chez l'enfant sont peu nombreuses, mais on peut en retirer que les jeunes enfants sont généralement en mesure d'adopter des comportements compensatoires se rapprochant de ceux des adultes et ce, que ce soit en réponse à une manipulation de la fréquence fondamentale (Liu et al., 2010; Scheerer et al., 2016), de la fréquence centroïde de certaines consonnes (Shiller et al., 2010) ou encore des valeurs de F1 et F2 de cibles vocaliques (Caudrelier et al., 2019b; Daliri et al., 2018; MacDonald et al., 2012; Shiller & Rochon, 2014). Ceci dit, la quasi-totalité des études stipule que les productions des enfants sont plus variables que celles des adultes. De plus, une étude récente montre que, bien que les adultes et les enfants agés entre 4 et 9 ans compensent de façon comparable à une manipulation de F2, les enfants compenseraient davantage en réponse à une manipulation de F1. Comme très peu d'auteurs s'y sont intéressés, il est d'ailleurs difficile de déterminer à partir de quand les comportements compensatoires des enfants rejoignent ceux des adultes. Néanmoins, MacDonald et coll. (MacDonald et al., 2012) avancent qu'aucune mesure compensatoire n'est employée par les enfants de 2 ans.

Des études employant d'autres paradigmes expérimentaux, comme ceux limitant certains gestes articulatoires (plaques occlusales, palais artificiel) ou modifiant la structure du conduit vocal (tube labial), ont aussi permis de comparer le rôle de la rétroaction auditive dans le contrôle des articulateurs de la parole chez l'enfant et l'adulte. En comparant les valeurs acoustiques et/ou articulatoires de productions

émises avant, pendant et après l'insertion du corps étranger, il est possible de rendre compte du degré de compensation réalisé afin de pallier son introduction. Cette capacité à trouver une configuration articulatoire alternative afin de produire la cible acoustique souhaitée témoigne, certes, de la finesse du contrôle des articulateurs de la parole, mais surtout du poids de la rétroaction auditive dans ce processus.

Contrairement aux expérimentations effectuées chez les adultes concluant que l'établissement de nouvelles stratégies articulatoires s'opère rapidement afin que les cibles acoustiques produites correspondent à celles souhaitées (Kelso et al., 1984; McFarland & Baum, 1995; Perrier, 1995; Savariaux et al., 1995), les études menées chez l'enfant conduisent globalement à des résultats opposés, quoique ceux-ci soient assez variables. Si certains concluent que les habiletés de compensations articulo-acoustiques sont acquises assez tôt (Baum & Katz, 1988), la majorité des études effectuées auprès des jeunes locuteurs démontrent qu'une expérience langagière importante et qu'une maturation du conduit vocal sont nécessaires à l'établissement de stratégies compensatoires efficaces (Gibson & McPhearson, 1980; Oller & MacNeilage, 1983). On en retient que les enfants n'ont pas suffisamment développé de représentations sensorimotrices nécessaires à l'élaboration de nouvelles stratégies articulo-acoustiques (Ménard, Perrier, et al., 2016).

Compte tenu de la diversité des protocoles expérimentaux et de la variabilité des résultats, il serait imprudent de tirer des conclusions fermes quant à l'apport des informations auditives dans le contrôle de la parole des jeunes locuteurs. Ceci dit, l'ensemble des études recensées suggère que l'utilisation des indices auditifs pour contrôler la production de la parole est liée à l'âge.

Globalement, bien qu'ils permettent de rendre compte de différents phénomènes, les travaux portant sur les locuteurs sourds et entendants et les études de manipulations synthétiques ou physiques ont su témoigner de leur fiabilité afin d'observer le rôle

des représentations acoustiques dans les mécanismes responsables de la parole. De plus, les travaux portant sur les habiletés compensatoires des locuteurs enfants et adultes permettent de rendre compte de l'importance de l'expérience du retour auditif dans la mise en place de liens perceptuo-moteurs précis et appuient l'hypothèse de l'apport des informations auditives dans l'établissement de liens articulo-acoustiques au cours du développement.

1.3.2 Les informations proprioceptives

Tel que mentionné plus tôt, plusieurs se sont intéressés à l'importante variabilité observée au sein des comportements compensatoires en réponse à une manipulation du retour auditif (Burnett et al., 1998; MacDonald et al., 2010; Munhall et al., 2009). Pour certains, l'absence ou incomplétude des compensations pourrait être expliquée par le fait que, pour ces individus, une telle perturbation acoustique n'engendrerait simplement pas d'incohérence majeure entre leurs attentes perceptuo-motrices et la rétroaction perçue. Cette explication laisse supposer que ces locuteurs auraient des représentations phonologiques plus larges, tolérant donc une variabilité intra-catégorielle plus importante (Guenther & Perkell, 2004; Purcell & Munhall, 2006a; Villacorta et al., 2007). D'autres avancent plutôt qu'un décalage entre la rétroaction acoustique et la rétroaction proprioceptive, soit la perception physique de nos propres mouvements, pourrait être identifié par le participant. Selon le poids qu'il accorde à chacune des rétroactions sensorielles, un individu pourrait se fier davantage à sa rétroaction proprioceptive afin de produire la cible souhaitée (Jones & Munhall, 2005; Katseff et al., 2012; MacDonald et al., 2010; Munhall et al., 2009; Purcell & Munhall, 2006b). L'apport des informations proprioceptives dans les mécanismes de perception de la parole sera plus amplement discuté dans la section qui suit.

Le nom proprioception, des termes latins *proprius* et *capio* signifiant respectivement propre et saisi, fait référence au sens de perception de la position des parties et des muscles du corps et de la force nécessaire à la réalisation d'un mouvement. Ces

informations proprioceptives sont décodées par le cerveau afin de guider l'interprétation du mouvement effectué (Stein, B.E. & Meredith, 1993). Par exemple, lors du soulèvement d'une boite, l'information transmise par nos muscles nous fournit des indices concernant le poids de la boite, nous permettant possiblement d'identifier les objets s'y trouvant ou encore d'en préciser la répartition. Au même titre, lors de la production d'un son, la position des articulateurs de la parole nous informe de l'objet produit. En ce sens, le ressentiment de protrusion des lèvres nous informe de la production d'un son nécessitant la fermeture presque complète de la cavité buccale et de l'allongement du conduit supra-glottique, comme pour la production de la cible /u/.

La reconnaissance du rôle de la rétroaction proprioceptive dans les mécanismes de production et de perception de la parole est relativement récente et découle principalement de travaux témoignant de l'influence des informations proprioceptives, aussi appelées kinesthésiques, dans les mécanismes de contrôle moteur (Ito & Ostry, 2010; Lametti et al., 2012). Par exemple, (Schürmann et al., 2004) ont démontré qu'un input de type vibrotactile sur la main pouvait influencer la perception de volume d'un son. La *parchment-skin illusion*, présentée par (Jousmäki & Hari, 1998), témoignant aussi de l'intégration des indices kinesthésiques et auditifs dans les habiletés de perception. Bien que la majorité des études aient démontré que les informations tactiles n'ont d'influence que si les participants sont conscients de la tâche ou s'ils ont suivi une forme d'entrainement, (Fowler & Dekle, 1991; Gick & Derrick, 2009) prouvent que des indices de types aérotactiles peuvent influencer la perception d'un son, même en l'absence des circonstances nommées ci-haut. Effectivement, à la suite d'une tâche d'identification syllabique, les auteurs démontrent que la perception du trait d'oralité des paires consonantiques /pa – ba/ et /ta – da/ est influencée par la présence simultanée d'une pression d'air sur le dos de la main ou dans le cou. Peu importe l'endroit où la pression d'air est ressentie ou encore le lieu d'articulation de la consonne, lorsque les syllabes sont accompagnées d'une

information aérotactile, elles sont davantage identifiées comme étant sourdes. Cette étude, ayant depuis été répliquée mainte fois, permet de démontrer l'apport des informations tactiles au sein des processus de la perception de la parole (Derrick et al., 2019; Gick & Derrick, 2009; Goldenberg et al., 2015).

L'implication des informations proprioceptives provenant des aires orofaciales diffère quelque peu de celles précédemment décrites. En effet, la rétroaction kinesthésique fait d'abord référence aux informations provenant de la position et du mouvement des muscles et des articulateurs (Proske & Gandevia, 2009). Certaines des régions orofaciales impliquées dans les mouvements de production de parole sont toutefois dépourvues de muscle; les indices proprioceptifs guidant nos habiletés de parole proviendraient donc aussi de microrécepteurs cutanés (Ito & Gomi, 2007; Ito & Ostry, 2010).

L'apport des indices proprioceptifs provenant des aires orofaciales dans les mécanismes de perception et de production de la parole demeure moins étudié. Néanmoins, depuis les dernières années, plusieurs auteurs ont voulu démontrer, via différents protocoles expérimentaux, l'apport de ces informations somatosensorielles dans les mécanismes de production et de perception de la parole. Une revue de ces études expérimentales sera présentée dans la section qui suit.

1.3.2.1 L'intégration audio-proprioceptive dans les tâches de parole chez l'adulte

L'un des premiers travaux s'attardant à l'implication du retour proprioceptif dans les mécanismes de production de la parole est celui de (Abbs et al., 1984). Lors de cette étude, les auteurs manipulent la lèvre inférieure des participants lors de productions vocaliques en y appliquant une légère force. Les auteurs démontrent que les locuteurs adultes compensent leurs déplacements articulatoires suite à la perception de la manipulation physique. Ils ajoutent que des gestes compensatoires s'observent tant en réponse à de petites ou de grandes perturbations et qu'ils peuvent être observés autant

sur l'articulateur obstrué que sur un articulateur voisin, dans ce cas-ci sur la lèvre supérieure.

(Feng et al., 2011) tentent également d'observer les compensations résultant d'une force externe lors de la production de mots CVC contenant les voyelles /ɛ/ et /æ/. Dans cette étude, la manipulation était constituée d'un déplacement vertical de la mâchoire et d'une manipulation du retour auditif. Les auteurs concluent que lorsque les manipulations sont compatibles, les participants compensent aux deux formes de perturbation. En contrepartie, lorsque les manipulations sont incompatibles, les sujets répondent à la manipulation acoustique seulement. Il semblerait donc que, lorsqu'elles sont en conflits, les informations auditives aient un poids plus important que les informations proprioceptives dans les habiletés de contrôle perceptuo-moteur.

Malheureusement, il est difficile de rendre compte de l'apport individuel de chacune des sources sensorielles dans les mécanismes de production de la parole. Une manipulation de type proprioceptive aura souvent un impact sur les informations acoustiques, laissant les causes d'une adaptation des productions à interpréter avec prudence (S. Tremblay et al., 2003).

De ce fait, dans le but de rendre compte du rôle exclusif des indices proprioceptifs dans les mécanismes de production de la parole, (Tremblay, S. et al., 2003) ont mis sur pied une tâche durant laquelle une manipulation de la position de la mâchoire était appliquée lors de la production de courtes syllabes chez des locuteurs adultes anglophones. Cette manipulation sensorielle, menant évidemment à une incohérence entre la planification motrice et le mouvement réalisé, mais n'engendrant pas de changement au niveau acoustique, permet donc de dissocier le rôle respectif des informations auditives des indices proprioceptifs. Les auteurs observent qu'en réponse à la force horizontale appliquée au niveau de leur mâchoire, les participants ont compensé leurs productions en déplaçant ce même articulateur dans le sens

opposé de la manipulation, soit en réalisant une protrusion de la mâchoire; soulignant le rôle distinct des informations proprioceptives dans les mécanismes de production de la parole. (Ito & Ostry, 2010) parviennent à des conclusions similaires alors qu'ils observent les comportements résultant d'une manipulation de la configuration labiale chez des sujets anglophones. Appliquant la manipulation proprioceptive (ici, un étirement des lèvres) quelques millisecondes avant la production de courtes syllabes, les auteurs remarquent que, bien que la manipulation n'engendre pas de conséquence au niveau de la perception acoustique, les participants compensent rapidement en produisant les syllabes suivantes avec une protrusion des lèvres plus importante.

Lametti et coll. (Lametti et al., 2012) ont aussi mené une étude permettant d'observer le rôle respectif des informations auditives et proprioceptives dans les mécanismes de production de la parole. Pour ce faire, les auteurs manipulaient, d'abord individuellement, puis simultanément, le retour acoustique (son entendu) et proprioceptif (position de la mâchoire) de participants anglophones. Les auteurs rapportent que plus la compensation était importante dans l'une des modalités, moins elle l'était dans la seconde. En plus de noter le rôle complémentaire des informations sensorielles, les auteurs soulignent l'importante variabilité inter sujet observée. En effet, peu importe la modalité dans laquelle une compensation de plus grande importance fut observée lors de la manipulation simultanée, cette dernière s'avérait aussi être celle engendrant une plus grande compensation lorsque les manipulations étaient appliquées de manière individuelle. (Nasir & Ostry, 2006) emploient un paradigme similaire et mentionnent que la perturbation de la mâchoire engendre un comportement compensatoire même si elle n'affecte pas le résultat acoustique. Contrairement aux observations de (Feng et al., 2011) proposant un apport plus important des informations auditives, les conclusions de (Nasir & Ostry, 2006) et de Lametti et coll., (Lametti et al., 2012) suggèrent plutôt que la rétroaction privilégiée par un participant afin de répondre à une manipulation sensorielle est variable.

En comparaison aux études de manipulations acoustiques, peu d'auteurs discutent de l'effet de persistance de la compensation à la suite d'une manipulation somatosensorielle (Feng et al., 2011). En effet, seuls (Tremblay, S. et al., 2003) et (Ito & Ostry, 2010) observent une persistance de la compensation motrice. (Tremblay, S. et al., 2003) précise toutefois que ce maintien de la compensation n'est observé que lors des conditions de parole oralisée et de parole chuchotée, mais pas en contexte de mouvements non liés à la parole. (Feng et al., 2011) et (Nasir & Ostry, 2006) indiquent quant à eux n'observer aucune forme de persistance. Ceci dit, il est important de préciser que ces études n'employaient ni le même type de manipulation ni la même force.

Globalement, ces études menées en contexte de production de parole ont permis de rendre compte de l'apport des informations somatosensorielles dans les mécanismes de contrôle articulo-acoustique, et ce, aux dépens des informations auditives. Sachant que les mécanismes de perception et de production sont étroitement liés, plusieurs se sont ensuite intéressés au rôle de la rétroaction proprioceptive dans les habiletés de perception.

(Hotting & Roder, 2004) ont démontré l'interdépendance des processus multisensoriels dans la perception d'un stimulus. Lors de cette étude, les participants, des adultes voyants et non voyants, avaient pour tâche d'identifier le nombre de stimuli tactiles qu'ils percevaient sur leur index (induit par une tige de métal). Ces stimuli tactiles pouvaient, ou non, être accompagnés de stimuli auditifs (courtes présentations de tons sinusoïdaux). Les participants devaient faire abstraction de ces inputs sonores et n'identifier que le nombre de percepts tactiles ressentis. Les auteurs notent que lorsqu'un stimulus tactile était accompagné de plus d'un stimulus auditif, les participants percevaient ce stimulus tactile simple comme étant multiple. De plus, ils précisent que ce phénomène s'est avéré moins important chez les locuteurs aveugles que chez les voyants. Bien que cette étude porte davantage sur l'intégration

des informations tactiles et auditives sur les habiletés de perception proprioceptive, elle illustre pleinement comment les différentes sources sensorielles sont étroitement reliées.

(Ito et al., 2009) ont brillamment su démontrer l'apport de ces informations sensorielles dans les mécanismes de perception de la parole. Les auteurs ont montré que la perception d'un mot était systématiquement influencée si la cible auditive à identifier était accompagnée d'une manipulation faciale imitant celle réalisée lors de la production d'un autre mot. Plus précisément, ils ont demandé à des individus, tous adultes et locuteurs natifs de l'anglais, d'identifier si la cible perçue était le mot *head* ou *had*. Lorsque les cibles acoustiques étaient précédées d'une manipulation orofaciale (soit celle rappelant le mouvement nécessaire à la production de la voyelle /ɛ/) le pourcentage d'identification de la cible comme *head* a significativement augmenté. En testant différentes directions de manipulation des muscles orofaciaux, les auteurs ont aussi démontré que cette influence n'était possible que si la manipulation physique appliquée rappelait celles requises en production de parole.

Les informations somatosensorielles semblent aussi impliquées dans le traitement de concepts perceptifs de plus haut niveau (Ogane et al., 2017). Dans une tâche similaire, des individus ont été invités à identifier si la cible acoustique perçue était "l'affiche" ou "la fiche". Les auteurs ont montré que la temporalité de la manipulation somatosensorielle (soit simultanément à la perception de la voyelle /a/ ou de la voyelle /i/) pouvait affecter la catégorisation de la cible lexicale.

1.3.2.2 L'intégration audio-proprioceptive dans les tâches de parole chez l'enfant

Malheureusement, bien que plusieurs modèles de production de la parole aient clairement illustré comment les informations sensorielles externes (telle l'acoustique) et internes (comme les informations proprioceptives et tactiles) interagissent afin de guider l'apprentissage entre un but acoustique et une configuration articulatoire, peu

d'études portent sur la reconnaissance des informations proprioceptives dans les habiletés langagières des jeunes enfants.

À notre connaissance, seulement deux études tentent d'observer le rôle des informations kinesthésiques des articulateurs de la parole dans les habiletés de perception chez les jeunes enfants.

Dans l'étude de (Yeung & Werker, 2013a) trois groupes de 32 sujets âgés entre 4 et 5 mois ont participé à une tâche de perception trimodale. Pour le groupe contrôle, une présentation simultanée de deux vidéos montrant l'articulation de la voyelle /i/ et de la voyelle /u/ étaient présentée accompagnée d'un stimulus sonore allant de pair avec l'un ou l'autre des visages parlant. Via les patrons de fixation observés, les auteurs rapportent que les jeunes participants sont davantage intéressés à la vidéo montrant l'articulation de la voyelle perçue auditivement.

Pour les deux groupes expérimentaux, les mêmes stimuli étaient présentés, mais cette fois-ci accompagnés de manipulations proprioceptives rappelant soit la production de la voyelle /i/ ou celle de la voyelle /u/. Les manipulations physiques étaient réalisées par l'insertion d'un doigt ou d'un jouet de dentition entre les lèvres des nourrissons. (Yeung & Werker, 2013b) rapportent que pour les participants pour lesquels la manipulation proprioceptive allait à l'encontre du stimulus auditif, les patrons de fixation collectés rappelaient ceux du groupe contrôle. En effet, les regards des participants étaient davantage dirigés vers la vidéo où le visage articulait la même voyelle que les sons entendus. En contrepartie lors des présentations où le mouvement des lèvres correspondait au stimulus auditif, les jeunes participants étaient moins portés à observer la vidéo dont l'articulation correspondait au son entendu, démontrant que les informations proprioceptives peuvent interférer dans le traitement audiovisuel de la parole.

Afin d'éliminer l'influence de l'expérience langagière dans la mise en relation des mécanismes de perception et de production de la parole (Bruderer et al., 2015) font passer un test de discrimination auditive à de jeunes bébés âgés de 6 mois. Le test était composé des phonèmes /d̪/ et /ɖ/, soit des consonnes de la langue Hindi se distinguant par leur caractéristique respectivement dentale et rétroflexe. Dans le but d'évaluer l'apport des informations proprioceptives lors de cette tâche, les auteurs ont utilisé des jouets de dentitions pour contrôler la position de la langue des jeunes participants. À l'aide d'un système d'imagerie par ultrason, ces derniers ont observé l'influence de la position des articulateurs de la parole dans la perception des sons. Cette influence n'a d'ailleurs été observée que lorsque l'articulateur qui aurait été impliqué dans la production du son entendu était manipulé. Les auteurs ont donc démontré que, malgré une expérience de production de parole quasi inexistante, les informations provenant de la configuration des articulateurs de la parole ont une influence sur la perception des sons, et ce, même s'ils proviennent d'une langue non familière.

Les études recensées ont permis d'établir que les informations de type proprioceptif, tels les mouvements de la langue, des lèvres et de la mâchoire ressentis ont un poids notable dans les mécanismes de production, mais aussi de perception de la parole. En fonction des quelques études portant sur le rôle de la rétroaction proprioceptive chez les locuteurs enfants, on suppose aussi que l'apport de ces informations sensorielles varie en fonction du stade développemental. Alors que des hypothèses similaires ont été confirmées dans des études s'intéressant au développement de l'interaction des habiletés sensorimotrices lors de tâches de tracé ou d'atteinte de cible (Contreras-Vidal et al., 2005; Kagerer & Clark, 2014) il s'avère nécessaire de poursuivre les recherches quant au développement des indices de nature proprioceptive dans les mécanismes de perception de la parole.

1.3.3 Les informations visuelles

Dans les sections précédentes, nous avons démontré que les indices de nature auditive et proprioceptive venaient à influencer la perception de la parole. Mais en réalité, seules les informations auditives et visuelles sont disponibles à l'interlocuteur. Effectivement, certains des articulateurs de la parole sont visibles lors de la production du langage et ces indices articulatoires auront, eux aussi un impact sur la perception des sons du langage. Dans la section qui suit, nous illustrerons de quelle façon les informations auditives et visuelles interagissent dans le traitement de la parole.

1.3.3.1 La complémentarité des informations auditives et visuelles

Comme le but principal de la production de la parole est de nous faire comprendre par notre interlocuteur, nous adaptons constamment nos stratégies de communication afin de présenter notre message sous sa forme la plus claire. Si nous discutons avec une personne ayant une origine linguistique différente de la nôtre, il nous sera spontané d'opter pour un débit de parole plus lent et de produire des mouvements de parole plus clairs en articulant de manière exagérée. Il s'agit d'un moyen de rendre notre discours plus intelligible (Rosenblum, 2008b; Rosenblum et al., 1997). Ainsi, l'apprenti locuteur pourra utiliser les informations visuelles fournies par les gestes articulatoires afin de mieux décoder notre message.

La relation qu'entretiennent les modalités auditive et visuelle dans la perception de la parole a beaucoup été étudiée et l'apport de chacune de ces sources est bien défini. Pour y parvenir, (Robert-Ribes et al., 1998) ont mis sur pied un test perceptuel lors duquel les participants devaient identifier des cibles vocaliques produites, en condition de parole bruitée. À la suite de cette étude, les auteurs ont proposé une échelle de robustesse des traits, cette dernière est illustrée à la figure 1.6. On y note que, pour le canal auditif, c'est le trait de hauteur qui est le plus résistant au bruit,

celui-ci suivi des traits d'antériorité et d'arrondissement. Pour ce qui est du canal visuel, c'est alors le trait d'arrondissement qui est le plus robuste, puis celui de hauteur et d'antériorité (Robert-Ribes et al., 1998).

Canal	**- robuste** ----------→ **+ robuste**		
Auditif	arrondissement	antériorité	hauteur
Visuel	antériorité	hauteur	arrondissement

Figure 1.6 Échelle de robustesse des traits, adaptée de Dupont (2006)

Les auteurs précisent que ces deux canaux perceptifs fonctionnent de manière complémentaire, soit que les contrastes les plus robustes dans une modalité seront les moins robustes dans l'autre. De plus, ils ajoutent que, bien que ces deux canaux assurent une certaine forme de perception de façon indépendante, l'utilisation simultanée des deux sources perceptuelles engendre une meilleure perception. D'ailleurs, les informations perçues visuellement permettent non seulement d'aider à l'identification des phonèmes, mais elles peuvent aussi permettre de récupérer de l'information de nature prosodique, surtout lorsque les informations auditives sont dégradées, comme en contexte de parole bruitée ou chuchotée (Dohen & Loevenbruck, 2009).

1.3.3.2 La perception audiovisuelle chez l'adulte

La contribution des indices visuels dans les mécanismes de perception de la parole a été répandue par l'étude de McGurk et MacDonald parue au milieu des années 80. Leur découverte, maintenant connue sous le nom d'*Effet McGurk*, rapporte les conséquences d'une illusion perceptuelle résultant de la combinaison d'indices auditifs et visuels incongrus, permettant de témoigner de l'apport incontestable des informations visuelles dans l'interprétation des percepts articulo-acoustique (McGurk & MacDonald, 1976).

Dans la version originale de l'étude, les participants avaient pour tâche d'identifier un input de nature audiovisuelle dont les informations acoustiques rapportent la production de la syllabe /ba/ et que les informations visuelles montrent l'articulation de la syllabe /ga/. De manière générale, les auteurs observent que cette source audiovisuelle est alors perçue /da/, soit une fusion des informations auditives et visuelles. Comme ces consonnes sont opposées quant à leur lieu d'articulation, le cerveau percevrait la consonne dont le lieu d'articulation y est intermédiaire, illustrant pleinement la relation de complémentarité des corrélats auditifs et visuels dans la perception de la parole (McGurk & MacDonald, 1976).

Depuis, plusieurs variantes de l'*Effet McGurk* ont été réalisées et l'utilisation du terme *fusion,* antérieurement employé pour définir le résultat, reflète mal la complexité du phénomène. Ainsi, la conclusion des diverses expérimentations employant le paradigme de présentations audiovisuelles incongrues montre que l'effet observé est plutôt variable. En effet, celui-ci peut, comme lors de l'étude de 1976, découler en une fusion des percepts auditifs et visuels, mais il peut également mener à une perception inversée d'une paire consonantique ou même à la perception du stimulus présenté visuellement (Saalasti et al., 2012). Afin de rendre compte de la variabilité des résultats, plusieurs auteurs -dont McGurk et MacDonald eux-mêmes- s'entendent pour définir l'*Effet McGurk* comme un phénomène ayant pour résultat de modifier la perception auditive d'un stimulus lorsqu'il est présenté avec un stimulus visuel incompatible (Tiippana et al., 2015).

Il a été démontré que l'illusion perceptuelle demeure malgré le fait que les sources perceptuelles soient dégradées, synthétiques ou encore que les composants visuels et auditifs soient produits par un homme et une femme (K. P. Green et al., 1991; Massaro et al., 1996; McGrath & Summerfield, 1985). De plus, l'intégration audiovisuelle est observée même lorsque les sujets sont informés du phénomène ou

encore lorsqu'ils ont pour consigne de ne porter attention à l'une seule des sources perceptuelles (McGurk & MacDonald, 1976; Rosenblum et al., 1997). Les résultats de ces diverses expérimentations ont donc permis d'établir que le rôle des informations visuelles dans la perception d'une cible audiovisuelle est très robuste.

Une récente étude a d'ailleurs souligné le rôle de la rétroaction visuelle de nos propres articulateurs lors de la production de la parole (Vidou et al., 2020). Il s'avère que lorsque les adultes au développement neurotypique reçoivent, en temps réel, via un avatar contrôlé par réalité virtuelle, un retour manipulé de leurs gestes de parole, ceux-ci compensent leurs productions en modifiant leurs tracés articulatoires. Les dimensions définies comme étant visuellement robustes sont d'ailleurs celles pour lesquelles des comportements adaptatifs plus importants ont été observés. Ainsi, en contexte où un retour optique est disponible, les représentations visuelles de nos propres articulateurs semblent, elles aussi, influencer le contrôle moteur de la parole.

1.3.3.3 La perception audiovisuelle chez l'enfant

D'un point de vue développemental, il a été établi que les jeunes enfants montrent des aptitudes d'imitation de mimiques faciales après seulement quelques jours de vie (Meltzoff & Moore, 1983) et qu'ils sont, dès l'âge de quatre mois, en mesure d'identifier si les informations visuelles et auditives perçues sont compatibles (Kuhl & Meltzoff, 1982, 1996; Legerstee, 1990; Patterson & Werker, 1999, 2003). Ceci dit, les conclusions des travaux portant sur les phénomènes d'intégration audiovisuelle chez les jeunes enfants sont assez variables.

Dans leur étude de 1976, McGurk et MacDonald observent que les deux groupes d'enfants testés (de 3 à 5 ans et de 7 à 8 ans) sont moins influencés par l'illusion audiovisuelle que les locuteurs adultes. (Massaro et al., 1986) rapportent aussi que les enfants âgés entre 4 et 10 ans sont moins influencés que les adultes par les informations visuelles lors de la présentation simultanée des stimuli /ba/ et /da/.

D'autres travaux avancent plutôt que les jeunes locuteurs intègreraient les informations audiovisuelles comme les adultes. Tel que souligné par (Kushnerenko et al., 2008), les nourrissons sont exposés à des mouvements de parole dès leur plus jeune âge. Les modalités auditive et visuelle devraient donc, toutes deux, avoir un poids considérable dans les habilités de développement de la parole et, par conséquent, être rapidement intégrées dans les représentations phonémiques. (Rosenblum et al., 1997) parviennent à cette conclusion à la suite d'une série de quatre expérimentations comparant, d'une part, les temps d'habituation et, d'autre part, les préférences de perception de stimuli audiovisuels congrus et incongrus d'enfants de 5 mois et d'adultes anglophones. Les auteurs mentionnent que les temps et les types de réactions aux stimuli incongrus des nouveau-nés rappellent ceux des adultes. Dans la même lignée, (Kushnerenko et al., 2008) avancent que, dès l'âge de 5 mois, les enfants sont en mesure d'anticiper le résultat auditif d'un stimulus perçu de façon visuelle seulement et que, par conséquent, ils auraient déjà formé des représentations neurologiques phonémiques multimodales.

1.3.3.4 Les impacts d'une privation visuelle

Les études s'attardant aux habiletés de perception et de production de la parole chez la population non voyante permettent également de noter l'importance des informations visuelles dans ces mécanismes. Bien que les personnes non voyantes démontrent des habiletés langagières similaires à celles des personnes voyantes, plusieurs études supportent l'idée d'une différence considérable entre les habiletés motrices et perceptuelles des locuteurs aveugles et voyants (Gougoux et al., 2004; Ménard et al., 2009; Niemeyer & Starlinger, 1981).

Ces différences au sein des habiletés perceptuelles s'expliqueraient par une compensation sensorielle, cette dernière n'étant toutefois observable que lorsque la cécité est survenue en bas âge (Kujala et al., 2000; Voss, 2019). Effectivement, différentes études s'attardant à la plasticité et à la réorganisation cérébrales appuient

l'hypothèse d'une compensation sensorielle chez les aveugles congénitaux; cette compensation étant particulièrement importante au niveau des informations auditives et tactiles (Fortin et al., 2007). Il s'avère que chez les aveugles congénitaux, les zones normalement réservées à la vision ne sont pas inopérantes, mais serviraient plutôt à traiter des informations provenant d'autres sources perceptuelles (Voss, 2019).

Une recension des études comparant les habiletés perceptuelles de locuteurs voyants et non voyants permet de conclure que les aveugles congénitaux performeraient mieux que les voyants lors de différentes tâches de perception telles que la discrimination de fréquences et de hauteur de la voix, la comparaison de courbes mélodiques et l'identification de phrases, de syllabes ou de voyelles présentées dans le bruit ou encore dans des tâches de localisation d'une source sonore (Arnaud et al., 2013; Gougoux et al., 2004; Lessard et al., 1998; Niemeyer & Starlinger, 1981). De plus, (Trudeau-Fisette et al., 2017) montrent que les aveugles congénitaux produisent des comportements compensatoires plus importants que les individus voyants en réponse à une manipulation de la rétroaction auditive. Les auteurs concluent que les locuteurs aveugles tolèreraient moins d'écart entre le feedback auditif réel et celui attendu que leurs pairs voyants.

Ces différences sur le plan de la perception auront des répercussions sur leurs habiletés de production de la parole, et celles-ci peuvent s'observer à un très bas âge. Lorsqu'elle survient à la naissance, une cécité occasionnait moins d'imitations des mouvements labiaux, une phase de babillage plus longue et un retard dans la production des premiers mots (pour une revue, voir (Leclerc, 2007; Ménard, 2013)). Chez les enfants d'âge scolaire, les impacts d'une cécité congénitale sur les habiletés langagières s'observent tant au plan acoustique qu'articulatoire. En effet, Ménard et coll. (à paraitre) montrent que des contrastes acoustiques et articulatoires plus importants entre les conditions de parole neutre et contrastive sont observés chez les

enfants voyants que chez les enfants aveugles. Des différences articulatoires sont aussi détectées chez les jeunes aveugles allemands (Göllesz, 1972).

Chez l'adulte, il a été démontré que les personnes aveugles congénitales produiraient des voyelles plus longues et moins contrastées que leurs pairs voyants (Ménard et al., 2013, 2014, 2017; Ménard, Trudeau-Fisette, et al., 2016; Trudeau-Fisette et al., 2013) et que les personnes voyantes et non voyantes utiliseraient des stratégies articulatoires différentes (Ménard et al., 2014, 2015; Ménard, Turgeon, et al., 2016). Les auteurs avancent que les locuteurs voyants utilisent davantage leurs lèvres (articulateurs visibles) alors que les locuteurs non-voyants compenseraient en usant principalement de leur langue (articulateur non visible). La rétroaction visuelle semble donc aussi avoir une incidence sur le contrôle des articulateurs de la parole et, par conséquent, sur les cibles acoustiques produites.

Considérant, d'une part, l'apport des informations auditives, visuelles et proprioceptives dans les mécanismes de production et de perception de la parole et, d'autre part, les conséquences de l'expérience sensorielle et langagière sur le contrôle de la parole, il s'avère pertinent de dresser un portrait complet des habiletés d'intégration multisensorielle dans le développement des mécanismes de perception de la parole.

L'objectif général du présent travail est ainsi d'approfondir les connaissances quant au développement des habiletés d'intégration multisorielle dans les processus de perception de la parole. Pour y parvenir, trois études complémentaires ont été effectuées chez des enfants d'âge préscolaire et chez des adultes francophones afin de rendre compte du rôle de la rétroaction auditive dans le contrôle de la parole, de témoigner de l'intégration des informations audioproprioceptives et, finalement, de démontrer l'interaction des informations audiovisuelles au sein de la perception des mécanismes de perception de la parole.

CHAPITRE II: ARTICLE 1

AUDITORY-MOTOR INTERACTION IN CHILDREN AND ADULTS'S SPEECH: ADAPTATIONS TO REAL-TIME AUDITORY FEEDBACK PERTURBATIONS

Paméla Trudeau-Fisette[1,2*], Camille Vidou[1,2], Lucie Ménard[1,2]

[1]Laboratoire de Phonétique, Université du Québec à Montréal, Montreal, Quebec, Canada
[2]Centre for Research on Brain, Language and Music, Montreal, Quebec, Canada

***Correspondence:**
Paméla Trudeau-Fisette
ptrudeaufisette@gmail.com

Keywords: **Formant shifting, compensatory responses, speech perception and production, adults, children**

Abstract

This study investigates the effects of development on the relationship between speech perception and production by examining compensatory responses to real-time perturbations of auditory feedback. Acoustic signals were recorded while preschoolers and adults speakers of Canadian French produced several utterances of the front rounded vowel /ø/ for which F2 was gradually shifted up to a maximum of 40%. The findings indicate that preschool-aged children show larger responses in speech productions when confronted with altered auditory feedback, whether they were compensating for or following the manipulated speech parameter. Results also indicate that children and adults use different strategies to adapt their acoustic output. We conclude that children rely more strongly on their auditory feedback system than adults and that their internal model of speech is not as robust as that of adults.

2.1 Background

Auditory feedback plays an important role in the development of speech mechanisms. During the babbling period, produced sounds and sounds in the environment are associated with the articulatory movements that are required for their production. In that way, developing children can store that auditory and proprioceptive information to build themselves a sensorimotor model of speech: a correspondence between articulatory movements and their acoustic and proprioceptive consequences (Guenther & Perkell, 2004; Guenther & Vladusich, 2012; Tourville & Guenther, 2011).

Acoustic feedback has a significant role in guiding this model during development in order to preserve speech fluency. Without articulatory adjustments, changes in the

shape, size and strength of speech articulators could have major effects on acoustic outputs (Callan et al., 2000; Guenther, 1994).

Once the model is mature, the feedforward system will control the articulators. Indeed, adult speakers no longer need to rely on the auditory feedback system, which would be too slow for real-time usage (Guenther & Perkell, 2004). In neurotypical adults, the auditory feedback system's role is rather to update the model and ensure sensorimotor adaptation in challenging speaking conditions (Guenther & Perkell, 2004; Jones & Munhall, 2002; Villacorta et al., 2007)

To demonstrate the major role of auditory feedback on speech mechanisms, some authors have addressed the consequences of deafness and have proved that individuals who are congenitally deaf and who do not wear hearing aids find it difficult to acquire oral language (Lane et al., 2005; Oller & Eilers, 1988; Svirsky et al., 2004). It has also been demonstrated that individuals with post-linguistic deafness showed evidence of speech deterioration in segmental and suprasegmental parameters, such as acoustic vowel distance, vowel duration, rhythm and intensity (Houde & Jordan, 1998; Lane et al., 2005; Lane & Wozniak Webster, 1998; Perkell et al., 2004). In general, studies of the deaf population support the idea that speech production depends on sensory information provided, mainly, by auditory feedback.

2.1.1 Auditory feedback experiments in adults

Studies using an altered auditory feedback paradigm also emphasize the importance of auditory feedback by showing that speakers change their subsequent productions to compensate for the altered feedback (Burnett et al., 1998; Houde & Jordan, 1998; Kawahara, 1998; Mitsuya et al., 2013; Purcell & Munhall, 2006a). Those compensatory behaviors are the outcome of a complex mechanism that includes, for instance, setting sensory goals, converting them into articulatory plans, predicting their sensory consequences, perceiving sensory feedback, comparing predictions to

feedback, and engages in inverse modeling to update motor commands (J. F. Patri et al., 2015).

Whether focusing on timing, volume, pitch or formant frequencies, authors have demonstrated that, when adult speakers hear their own feedback in which a parameter is shifted in real time, they adapt [3] their production by changing the modified parameter in the opposite direction to the perturbation (although they may also follow the perturbation) (Elman, 1981; Jones & Munhall, 2000, 2003, 2005; Larson et al., 2008; Purcell & Munhall, 2006b; Villacorta et al., 2007). Previous studies have also shown that the compensation level seems to be linked to the level of the alteration, but that it will remain incomplete (MacDonald et al., 2010, 2011; Purcell & Munhall, 2006b, 2006a). Finally, many studies have pointed out that compensation usually appears after several repetitions (MacDonald et al., 2010; Max et al., 2003; Munhall et al., 2009), which suggests that our production-perception model accepts some variability, but beyond that level, an adjustment of the production mechanisms occurs (Purcell & Munhall, 2006a).

2.1.2 Auditory feedback experiments in children

Despite the fundamental nature of auditory feedback for speech development, only a handful of studies have investigated the effect of sensory manipulation on speech motor control in children.

Liu et al. (Liu et al., 2010) compared speech compensation of school-aged children, young adults and elderly individuals in response to a real-time pitch manipulation on the vowel /u/. They found that, although all speakers compensated by shifting their

[3] In the current article, the terme adaptation adaptation is used to refer to adaptive behaviours in response to sensory manipulation, encompassing both compensatory and following behaviour concepts.

productions in the direction opposite to the manipulation, children's responses presented longer latency than those of the young adults or elderly participants. Scheerer et al. (Scheerer, Nichole E. et al., 2013) also investigated the developmental trajectory of the speech motor control system by exposing 100 speakers, aged 4 to 30 years old, to auditory manipulations of their F0. Here again, the compensatory responses of children and adults were found to be similar in magnitude. However, latency and variability were found to differ with age: younger children had longer latency responses and were more variable. In line with previous results, Scheerer et al. (Scheerer, N. E. et al., 2016) showed that toddlers were able to compensate for a real-time manipulation of fundamental frequency, suggesting that speakers as young as 2 years old are capable of using auditory information to plan upcoming speech motor commands.

Some authors have also investigated the effect of the manipulation of phonemic properties on speech adaptation. The results in this area are, however, less consistent. For example, Shiller et al. (Shiller, Douglas M. et al., 2010) chose to investigate the role of auditory feedback on speech control in 9- to 11-year-old children and adults by shifting the centroid frequency of the fricative consonant /s/, which resulted in acoustic feedback that was closer to the fricative /ʃ/. Although they found no effect of group on the magnitude of compensation, they noted greater variability in the children's productions. Moreover, when they compared the location of the categorical boundary between /s/ and /ʃ/, before and after the formant manipulation experiment, they found that the online manipulation led to a larger boundary shift (toward /ʃ/) for the adults. This indicates that, although children were able to compensate for the auditory manipulation, their perception-production links were not yet adult-like. That being said, Shiller et al. (Shiller, D. M. et al., 2013) later conducted an experiment in which they compared categorical boundaries and compensatory behaviors before and after perceptual training and found that relevant reinforced perceptual training leads to greater perceptual shifts and compensation in both children and adults. They

concluded that the acoustic correlates of phoneme categories are rapidly integrated into the process of speech motor planning.

MacDonald et al. (MacDonald et al., 2012) also reported that adults and preschool children (4 years old) showed similar compensation patterns in response to a simultaneous F1/F2 manipulation of the English vowel /ɛ/ (making it sound more like /a/). Those results are consistent with the idea that auditory feedback is important for monitoring and correcting speech production errors. However, they found that toddlers (2 years old) did not compensate for the formant shifts. Moreover, they noticed considerable variability in the younger speakers and reported that variability in production decreased with age, indicating that the use of auditory feedback to control speech production is age-related.

Shiller and Rochon (Shiller, Douglas M. & Rochon, 2014) investigated the relation between perceptual ability and motor learning and showed that 5- to 7-year-old children were capable of adapting to the altered auditory feedback of F1 and, consequently, that their perceptual abilities were sufficient for that kind of sensory-error-based learning. Still, the fact that adaptation behaviors improved with perceptual training suggests that their untrained perceptual ability somehow prevented them from making optimal speech motor adaptations.

Although they mainly investigated the role of stuttering in speech sensorimotor representations, Daliri et al. (Daliri et al., 2018) found that both control children (aged 7 to 11) and adults showed similar compensation patterns in response to F1/F2 perturbations. Similar findings were reported in a study of sensorimotor learning in preschool-aged and school-aged children and adults (Caudrelier et al., 2019b).

Finally, when they simultaneously shifted F1 and F2 of /i:/ in a study of Dutch children and adults, van Brenk and Terband (van Brenk & Terband, 2020) found

similar F2 compensations in both groups but larger F1 compensations in children (aged 4 to 9 years). Moreover, they reported that, unlike adults, children showed an F1 adaptation pattern as they maintained their compensation/adaptation behavior even when the auditory feedback was changed back to normal. In addition, the children showed greater token-to-token variability throughout the experiment. The authors concluded that existing linguistic representations may be less stable in children than adults.

While additional studies are required to gain a clearer understanding of the role of auditory feedback in the development of motor control of speech, past studies suggest that young children are generally as able to compensate for the manipulation of their auditory feedback as adults. Nevertheless, the substantial variability of their productions and their precocious learning behaviors suggest that their auditory-motor links have not yet matured fully. Furthermore, as the vast majority of studies only observe the acoustic impact on the manipulated parameter, it is difficult to say with certainty whether children and adults actually behave comparably. Indeed, there is evidence that altered auditory feedback can lead to compensatory behavior, although often less significant, affecting a parameter other than the one being manipulated (Mitsuya et al., 2013; Villacorta, 2006; Villacorta et al., 2007). Considering that the combination of these different acoustic parameters has been shown to be correlated with perceived phonetic features (Ménard et al., 2002; Syrdal & Gopal, 1986), which are precisely the manipulated objects in such a paradigm, it is essential to take into account the neighboring formants, not only independently but in relation which each other, in order to get a complete picture of the repercussions of a formant modification of acoustic feedback.

2.1.3 Perceptual correlate of rounding

In the case of a manipulation affecting the rounding characteristic of a vowel, a perceptual parameter such as F2' is most appropriate. F2', a weighted sum of F2, F3,

and F4, was introduced by Carlson et al. (Carlson et al., 1970) and was later identified by Traunmüller and Lacerda (Traunmüller & Lacerda, 1987) as a perceptual correlate of rounding for front vowels in Turkish and Swedish. F2' is based on the 3.5-Bark critical distance within which two formant peaks are perceived by the ear as one peak (Chistovich & Lublinskaya, 1979). (Schwartz et al., 1997) used this parameter, among others, to predict vowel system stability. It was also identified as an important perceptual correlate of rounding in French front vowels, as well as a significant gender normalizer (Mantakas, 1989). Simulating vocal tract morphology and acoustic output in seven growth stages (from 0 to 21 years) using the Maximal Vowel Space (Boe et al., 1989), Ménard et al. (Ménard et al., 2002) established that F2' is an invariant acoustic correlate of rounding throughout development, and that it is a more accurate perceptual correlate of rounding than F3–F2.

The goal of this article is to further explore the development of the speech motor control system by comparing preschool-aged children's and adults' adaptation strategies in response to auditory-feedback manipulations of formant frequencies. Complementing our previous studies (Trudeau-Fisette et al., 2019), we chose to focus our attention on the contrast in rounding between the phonemes /ø/ and /e/, a phonological feature that has rarely been investigated in acoustic manipulation studies.

First, we sought to investigate whether preschool-aged children and adults would adapt their productions to a similar degree when confronted with altered auditory feedback. Our hypothesis was that both children and adults would be able to compensate for the formant manipulation, but that children's productions would be more variable. We then wanted to observe whether those two age groups would employ comparable strategies to respond to this shifted feedback. Based on the premise that acoustic-motor links have not reached maturity in preschool children, we

believed that different articulatory strategies would be employed by the two experimental groups.

2.2 Methods

2.2.1 Participants

Sixty-five native speakers of Canadian French were recruited for this study. However, 6 potential participants were excluded due to equipment malfunction (1 adult), inability to perform the task (2 children) or poor data quality (3 children). In the end, 30 adults (aged 19–30, 13 females; mean = 26.2 years, SD = 4.0) and 29 preschoolers (aged 4–6, 15 females; mean = 5.2 years, SD = 0.7) were included in the analyses.

All participants were tested for their pure-tone detection threshold (DT) using an adaptive method (DT < 25 dB HL at 250, 500, 1,000, 2,000, 4,000 and 8,000 Hz) and had normal or corrected vision. They (or their parents) reported having no speech, language, psychological or neurologic disorder and gave written and informed consent to the experiment. The research protocol was approved by Université du Québec à Montréal's Institutional Review Board (no. 2015-05-4.2).

2.2.2 Experimental procedures

The task presented in this study was administered in a three-experiment session in which adults and children were asked to perform two bimodal perceptual tasks (one audiovisual, one audio-somatosensory) and one production task. This article discusses the results of the production task, in which real-time auditory perturbations were applied. The formant shifting task was always presented last, in order to prevent the formant manipulation from having an effect on the perceptual tests. Moreover, the experiments were separated by 15-minute breaks. The total duration of the session

was about 90 minutes, but the production task lasted no more than 10 minutes and was the same for children and adults.

Participants had to produce 70 utterances of the rounded vowel eu /ø/, as in the word eux ("them"). The vowel corresponded to the name of a cartoon character (although participants did not know that the name represented the vowel being tested). The task was to say the character's name each time it appeared on the screen. That way, we made sure that they were producing the vowel in the appropriate timeframe, and it made the task more appealing for the children. To make sure that all the participants would remember the character's name, a training session was administered prior to the experiment in which each participant was asked to name a series of different pictures, including Mr. Eu.

Before the experiment, each individual's formant values were collected in a short session in which they produced several instances of /ø/. Those values were then averaged and used to constrain formant measurements and avoid errors. Throughout the experiment, the produced vowel /ø/ was gradually shifted toward /e/, by increasing each individual's F2 by 40% using Audapter (Cai et al., 2008).

The experiment was divided into four phases (Figure 2.1). In the first phase (Baseline: 10 utterances), participants produced the vowel eu /ø/ ten times and received normal auditory feedback. In the second phase (Ramp: 30 utterances), participants received altered auditory feedback during which F2 was incrementally increased by 1.33% in each trial (the other formants were not shifted). Therefore, in the 40th trial, participants received auditory feedback in which F2 was raised by 40%. In the third phase (Hold: 15 utterances), the 500 Hz shift was simply maintained. In the 56th trial, the perturbation was abruptly removed. Thus, during the last phase (End: 15 utterances) participants received normal auditory feedback. Throughout the experiment, auditory feedback was amplified and mixed with white noise. Acoustic

signals were recorded with a high-quality Audio-Technica microphone (Omnidirectional condenser headworn microphone, model number BP892) and digitized at 44100 Hz using a Delta 1010 LT sound card.

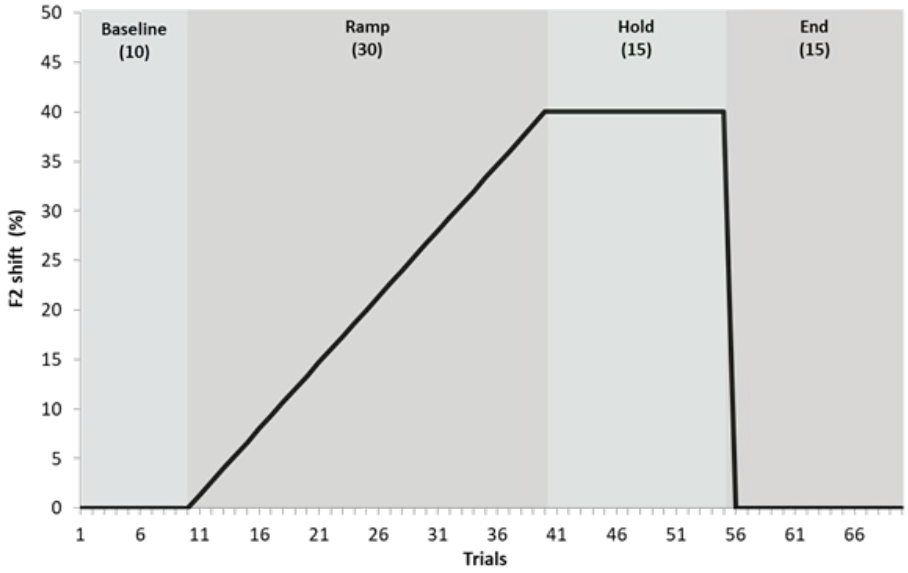

Figure 2.1. Schematic representation of the acoustic manipulation throughout the four experimental blocks.

2.2.3 Data analysis

Each of the 70 utterances of the vowel /ø/ was manually segmented with Praat software. Using a customized script, the first four formants were extracted at the vowel midpoint. Formant mislabeling was corrected manually.

The engineered acoustic manipulation created a contrast in regard to the rounding of the vowel (the produced rounded vowel /ø/ being perceived as its unrounded counterpart /e/). Because the perception of rounding is generally the result of the modification of multiple formants, it was difficult to identify one principal compensation pattern. Indeed, although our manipulation acted on the second formant, speakers reacted by adapting F2 and adjacent formants (a natural phenomenon

described by Vaissière (Vaissière, 2009)), leading to a multitude of compensation and following patterns. Therefore, although compensations to formant manipulations are usually measured by comparing the acoustic information on the first (baseline) and subsequent productions of the manipulated parameter (Jones & Munhall, 2000; MacDonald et al., 2012; Purcell & Munhall, 2006b), we opted for a method that took into account the perceptual consequences of a formant manipulation, rather than analyzing formants one by one.

F2' values were calculated for each produced vowel using the formula given by Schwartz et al. [45]: $F2' = ((c_2*F2) + (c_3*F3) + (c_4*F4))/(c_2+c_3+c_4)$. The values of c_2, c_3, and c_4 depend on the pattern of distance between F2, F3, and F4 (see Figure 2.2).

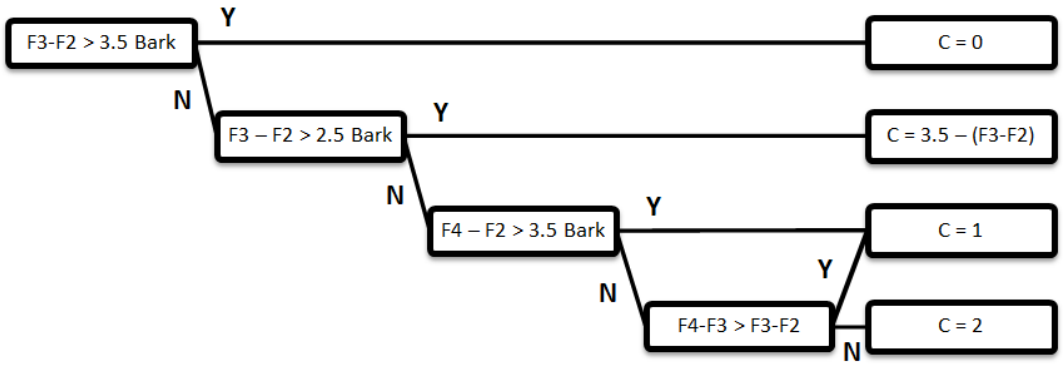

Figure 2.2. F2' computation flow chart (reproduced with permission from (Schwartz et al., 1997). If c = 0, then $c_2 = 1$, $c_3 = 0$, $c_4 = 0$ (f2' is set at F2; F3 and F4 are too distant to be taken into account); if $0 \leq c \leq 1$ then $c_2 = 1$, $c_3 = 0.5$, $c_4 = 0$ (F2' is set at center of gravity of F2 and F3; if c = 2 then $c_2 = 0$, $c_3 = 1$, $c_4 = 0.5$)F2' is set at the center of gravity of F3 and F4).

To determine whether a given participant was classified as a "compensator" or a "follower," the mean F2' value of each individual's baseline production was compared to their subsequent production in the ramp, hold, and end phases. These

ratios were then averaged in regard to the experimental blocks. If a subject presented a mean ratio lower than 1 for the Hold phase, he or she was categorized as a compensator. Otherwise, the subject was classified as a follower. Overall, 7 of the 30 adults and 8 of the 29 children were classified as "follower". As an example, Figure 2.3 shows the mean ratios for three participants. Individuals A06 and A10 were labeled as followers, while A07 was categorized as a compensator.

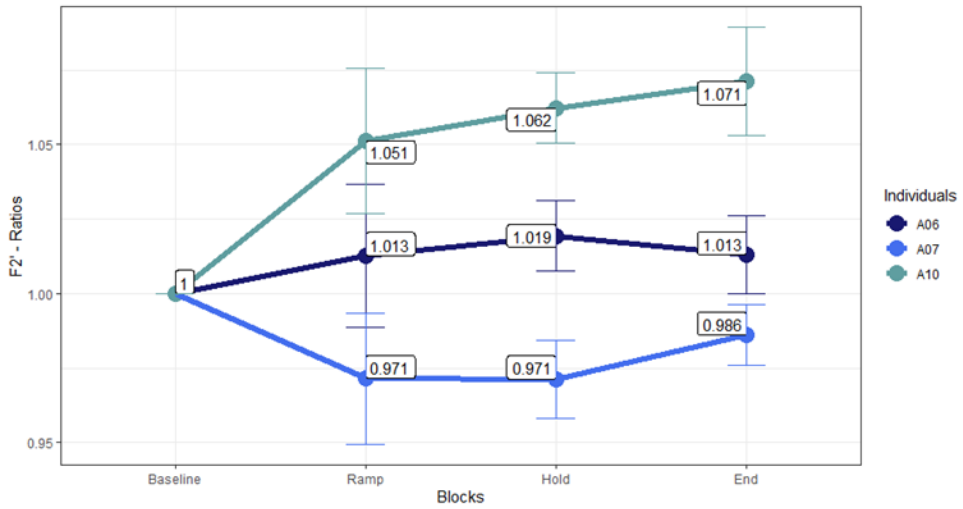

Figure 2.3. Mean F2' ratios (in Barks) for each experimental block. Data are presented across individuals. Error bars indicate standard deviations.

To evaluate acoustic adaptation across experimental phases, F2' values associated with every production (70 utterances collected from 59 participants) were fitted into a linear mixed-effects model (LMEM) (using the *lme4* package in R) in which the fixed factors were Group (Adults or Children), speaker Type (Compensators or Followers) and experimental Block (Baseline, Ramp, Hold, End) and the random factor was the individual participant. Post hoc analyses were performed using multiple comparisons (using *multcomp* package in R). Variability was investigated through standard deviations.

In order to evaluate whether individuals adapted their production within the different blocks, the first three and the last three trials of each block were identified and labeled as Initial and Final steps. Additional LMEM and post hoc tests were then performed where the variable Step was specified as a fixed factor instead of block. Again, standard deviations were used to investigate variability. LMEM were then conducted where Group (Adults or Children), Type (Compensators or Followers) and Step (Baseline: Initial/Final, Ramp: Initial/Final, Hold: Initial/Final, or End: Initial/Final) were assigned as fixed factors and the random factor was the individual participant.

Finally, to investigate whether both groups used similar strategies to modify the perceived rounding of their productions, we ran a multiple linear regression (MLR) using the *lme4* package in R. For both children and adults, an MLR was calculated to predict F2' prime values based on F2 and F3. (Because F4 was taken into account in the F2' computation of only 12 of the 4,130 total utterances, we chose not to include it in this subsequent analysis.)

2.3 Results

2.3.1 Acoustic adaptation – The perception of rounding

The mean F2' value of each block is shown in Figure 2.4. Data are averaged across speakers, within both speaker groups (Adults, Children) and presented in relation to speaker type (Compensators, Followers). For reference, the lower the value, the more rounded the vowel is perceived.

The LMEM revealed significant effects of Group ($\chi2(1) = 18.831$, $p < .001$) and Block ($\chi2(1) = 48.562$, $p < .001$). Those primary results indicate that, when all speakers are considered, children and adults did not produce comparable outputs; the different blocks led to changes in speech productions. Although no main effect of

Type was demonstrated, significant interactions of Type and Block ($\chi2(1) = 36.769$, $p < .001$), and of Group, Type and Block ($\chi2(4) = 37.948$, $p < .001$) were found.

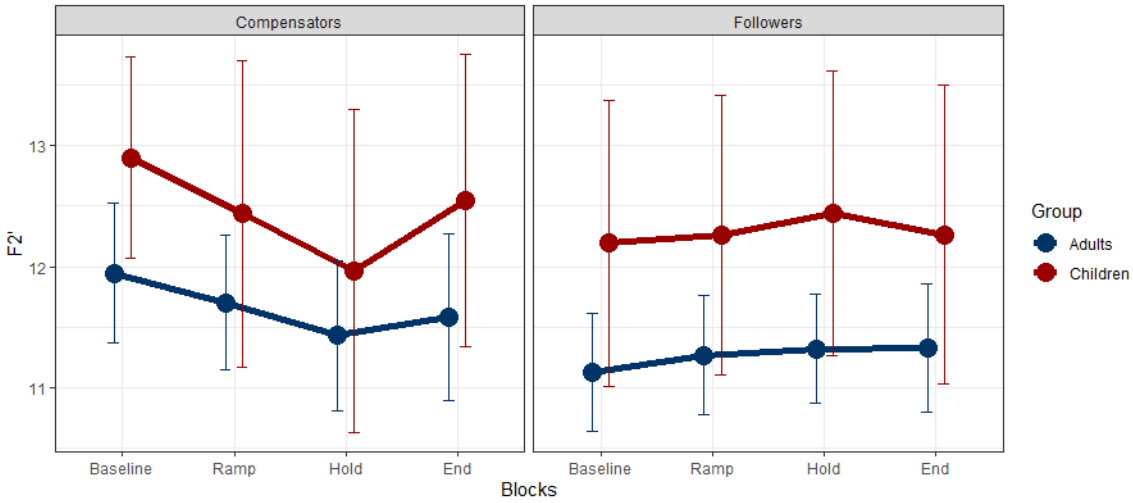

Figure 2.4. Mean F2' value (in Barks), for each experimental block. Data are presented across speaker groups and speaker types. Error bars indicate standard deviations.

Additional analyses of the Type/Block interaction (multiple comparisons) show that compensators produced larger and faster speech adaptations. Indeed, for speakers whose productions were changed in the opposite direction from the acoustic manipulation, every successive block led to a significant change in production (Baseline vs. Ramp: $z = –8.378$, $p < .001$; Ramp vs. Hold: $z = –10.949$, $p < .001$; Hold vs. End: $z = 8.874$, $p < .001$). For followers, a significant adjustment in production was observed only once in the Hold phase ($z = 3.270$, $p < .05$).

Moreover, the multiple comparisons we ran to tease out the three-way interaction of Group, Type and Block indicated that, while all compensators adapted their productions in the same way from one block to another (detailed z values for the differences between blocks, for each group, are presented in Table 2.1), only

children's productions changed significantly between the Hold and End blocks (z = 10.077, p < .001). No differences between blocks were found for followers.

Table 2.1

Summary of z values and significance levels of the differences between experimental blocks for children and adults.

Groups	Compensators					
	Baseline vs. Ramp	Baseline vs. Hold	Baseline vs. End	Ramp vs. Hold	Ramp vs. End	Hold vs. End
Adults	–4.30, $p < 0.001$	–8.61, $p < 0.001$	–6.10, $p < 0.001$	–6.11, $p < 0.001$		
Children	–7.79, $p < 0.001$	–14.37, $p < 0.001$	–5.38, $p < 0.001$	–9.59, $p < 0.001$		10.08, $p < 0.001$

Finally, the multiple comparisons also indicated that, for compensators, the Baseline (z = 4.554, p < .001), Ramp (z = 3.622, p < .05) and End phases (z = 4.703, p < .001) yielded different kinds of productions in children and adults. In contrast, for followers, only the Hold phase (z = 3.233, p < .05) led to significant differences between the two groups' productions. Those group differences can be seen in Figure 2.4. As expected, the LMEM showed that children performed more variably than adults. Only a main effect of Group was found ($\chi 2(1) = 25.499$, p < .001).

When Step was added as a fixed factor, the LMEM also indicated a main effect of Step ($\chi 2(1) = 46.651$, p < .001), an interaction of Type and Step ($\chi 2(1) = 32.965$, p < .001) and an interaction of Group, Type and Step ($\chi 2(4) = 38.651$, p < .001). The mean F2' value for each step is shown in Figure 2.5. Once again, data are averaged across speakers, within both speaker groups (Adults, Children) and presented for speaker type (Compensators, Followers).

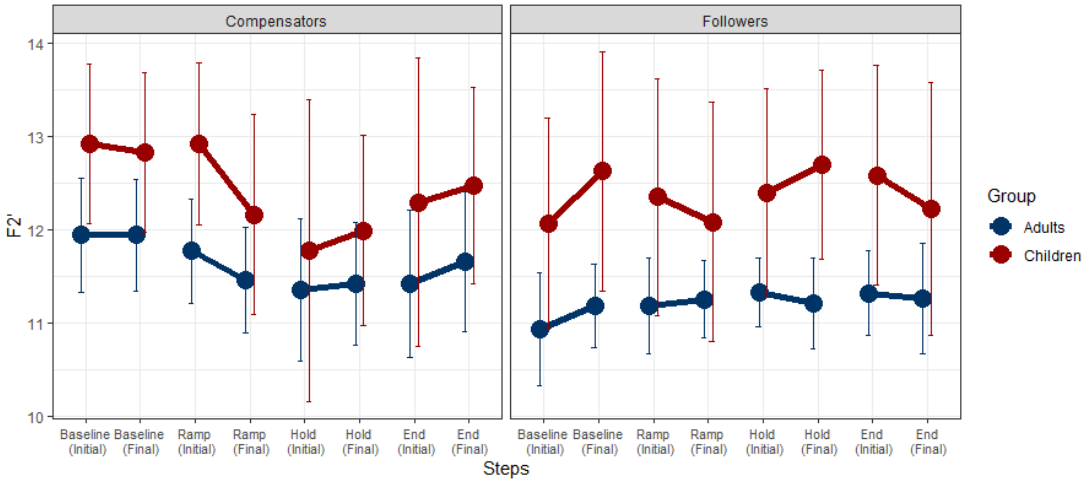

Figure 2.5. Mean F2' value (in Barks) for each step. Data are presented across speaker groups and speaker types. Error bars indicate standard deviations.

While both groups of compensators produced outputs that differed from those of the Baseline by the final step of the Ramp (Adults: $z = –5.089$, $p < .001$; Children: $z = –7.564$, $p < .001$), the multiple comparisons run to understand the three-way interaction of Group, Type and Step indicated that, within the Ramp phase, children compensated more ($z = –7.833$, $p < .001$) than adults ($z = –3.377$, $p < .05$), and that compensation continued in subsequent utterances as children's productions differed from the final step of the Ramp phase to the initial step of the Hold phase ($z = 3.650$, $p < .05$). In addition, F2' values for the final step of the End bloc, were significantly lower than those for the initial step of the baseline for both children ($z = –4.927$, $p < .001$) and adults ($z = –5.138$, $p < .001$). However, the absolute final step was different than the absolute initial one for child compensators only ($z = –4.113$, $p < .01$).

For the followers, only the children's productions varied over the course of the experiment. In fact, by the final trials of the Hold phase, F2' in children's utterances was significantly higher than in the initial trials of the Baseline phase ($z = 3.326$, p

< .05). Moreover, children's F2' value increased within the Baseline (z = –3.632, p < .05).

Once again, multiple comparisons also highlighted group differences. For compensators, initial and final steps of the Baseline, Ramp and End phases yielded different productions by children and adults. For followers, the final step of the Baseline, initial step of Ramp and final step of the End phase led to significant differences between the two groups' productions. Those group differences are noticeable in Figure 2.5. Complete z values for the differences between groups, for each step, are presented in Table 2.2.

Table 2.2

Summary of z values and significance levels of the differences between speaker groups for both compensators and followers.

		Compensators	Followers
		Children vs. Adults	
Baseline	Initial	4.664, $p < 0.001$	
	Final	4.406, $p < 0.001$	4.045, $p < 0.01$
Ramp	Initial	5.458, $p < 0.01$	3.506, $p < 0.05$
	Final	3.294, $p < 0.05$	
Hold	Initial		
	Final		3.792, $p < 0.01$
End	Initial	3.864, $p < 0.01$	
	Final	3.313, $p < 0.05$	

This time again, the LMEM showed that children performed more variably than adults. Only a main effect of Group was found to be significant ($\chi2(1) = 29.653$, $p < .001$).

2.3.2 Acoustic adaptation – The parameters of choice

MLR between F2' values and F2 and F3 (in Barks) are displayed in Figure 2.6. Data are averaged across speakers and presented for each speaker group (Children and Adults).

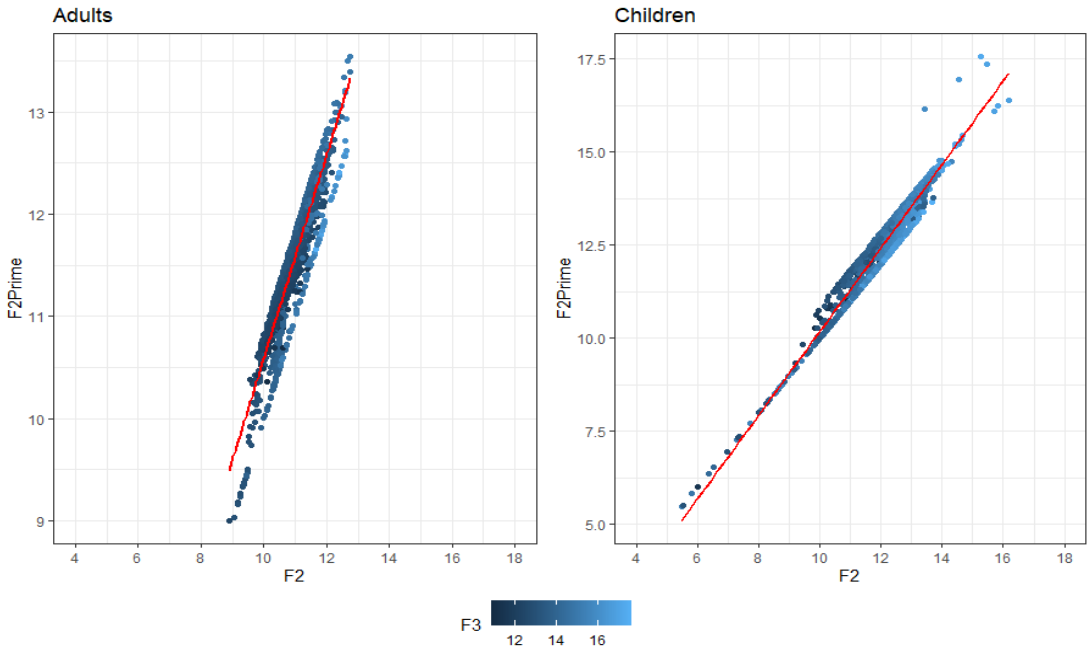

Figure 2.6. Linear regression between F2' values and F2 and F3 (in Barks). Data are presented across speaker groups. The red line indicates the linear regression between the three variables.

Multiple regression analyses were used to observe to what extent F2 and F3 significantly predicted both children's and adults' F2' output values.

The results of these regressions indicated that, for children, the two predictors explained 96% of the variance ($R2 = .961$, $F(3, 1761) = 1.43e+04$, $p < .001$). For school-aged speakers, F2 significantly predicted F2' ($\beta = 0.736$, $p < .001$), as did F3 ($\beta = –0.417$, $p < .001$) and the interaction between F2 and F3 ($\beta = 0.736$, $p < .001$).

In adults, the MLR indicated that these parameters explained 89% of the variance ($R2 = .889$, $F(3, 1922) = 5118$, $p < .001$). However, it was established that only F2 ($\beta = 1.267$, $p < .001$) and the interaction between F2 and F3 ($\beta = –0.506$, $p < .05$) significantly predicted the value of F2'. When a model was built with only the significant variables, it also explained 89% of the variance ($R2 = .889$, $F(3, 1923) = 7672$, $p < .001$). At this point, F2 significantly predicted F2' ($\beta = 1.126$, $p < .001$), as did the F2–F3 interaction ($\beta = –0.226$, $p < .001$).

2.4 Discussion

The goal of this article was to investigate the development of the speech motor control system by observing the strategies preschool-aged children and adults used to adapt to auditory feedback manipulations of formant frequencies.Our first objective was to find out whether children and adults would adapt their productions to a similar extent when dealing with altered auditory feedback. We then sought to determine if the two groups had used comparable strategies to adapt their productions and respond to the perceived shifted feedback.

We manipulated the auditory feedback of 30 adults (aged 19 to 30) and 29 young children (aged 4 to 6) by increasing the values of the second formants of their productions by up to 40%. The shifted feedback paradigm is known to generate a discrepancy between the sound produced by speakers and the sound they expect to hear in return. As mentioned by previous authors, the manipulation of any speech parameter can have an impact on other ones (Mitsuya et al., 2013; Villacorta, 2006;

Villacorta et al., 2007). In order to take that phenomenon into account and be able to objectively compare children's and adults' productions, we chose to investigate their compensatory behavior through F2', a perceptual correlate of front rounded vowels corresponding to a weighted sum of F2, F3, and, in some rare cases, F4.

The direction and strength of adaptation behaviors were assessed by comparing the two groups' mean F2' values, first during the four phases of the experiment and then during the first and last utterances of each phase. To assess whether children and adults used comparable articulatory strategies to change the perceived roundness of their productions, we related the F2' values of each participant's productions to those of their F2 and F3.

Previous authors who explored this matter generally found that, when faced with altered auditory feedback, school-aged and preschool-aged children showed similar adaptation patterns to adults (Daliri et al., 2018; Liu et al., 2010; MacDonald et al., 2012; Scheerer et al., 2013; Shiller et al., 2010). That said, there is a consensus that children's productions are much more variable than those of adults (MacDonald et al., 2012; Scheerer et al., 2013; Shiller et al., 2010; van Brenk & Terband, 2020).

The results of our first analyses indicated that, when faced with manipulated auditory feedback, both children and adults compensated either by producing outputs that were shifted in the opposite direction of the manipulation or by following it. Although similar adaptation patterns were observed when blocks' averaged F2' were considered, our step-based analyses indicate that children compensated more, thus contradicting our first hypothesis. Indeed, while all compensators produced utterances that differed from the Baseline by the end of the Ramp block, more substantial compensation occurred within this block for children. Thus, it appears that children felt the need to adapt their productions more significantly in order to respond to the perceived acoustic manipulation; they continued to compensate while they were responding to

the Hold block. Van Brenk and Terband (van Brenk & Terband, 2020) reported similar findings, although they observed greater compensation only in the Hold phase. Among speakers whose speech productions tracked the acoustic manipulation, only children displayed significant adaptations. Once again, this suggests that child followers reacted more strongly to the acoustic manipulation. The greatest adaptation occurred in the Hold phase, more specifically during the final utterances. This suggests that the vocalic representations of children aged 4 to 6 years may be less robust than those of adults and, therefore, that they seek to adapt their productions faster.

Although our results may seem to differ from those previously presented in the literature, where children and adults have generally been found to use comparable compensation strategies (Daliri et al., 2018; Liu et al., 2010; MacDonald et al., 2012; Scheerer et al., 2013; Shiller et al., 2010), it is important to note that, unlike the studies cited, we not only compared the mean values of each block but also chose to consider the first and last productions of each block independently.

In addition to the fact that children compensated more significantly and more abruptly than adults, we found that they also had difficulty returning to their baseline productions. The results of the step-based analyses indicate that, while the two groups produced utterances that were quite different from their Baseline utterances at the beginning of the End block, children's utterances remained different from their normal ones at the end of the End block. In other words, only adults managed to return to their natural productions. This observation is in line with the fact that, during infancy, children use feedback to construct their sensorimotor model of speech (Guenther & Perkell, 2004; Guenther & Vladusich, 2012; Tourville & Guenther, 2011). Since adults use auditory feedback mainly to maintain intelligibility in challenging contexts, it is not surprising that they were able to return to their natural production style quickly once the manipulation was removed.

In their investigation of the role of short-term changes in speech motor performance related to practice, Walsh et al. (Walsh et al., 2006) also found that children demonstrate greater plasticity in speech-motor learning, allowing them to acquire and store audio-articulatory mappings more quickly than adults. While Shiller et al. (Shiller, Douglas M. et al., 2010) reported that centroid manipulation led to a larger boundary shift for adults than for 9- to 11-year-olds, they later stated that 5- to 7-year-old children benefited greatly from relevant perceptual training, producing stronger responses to F1 perturbation. In our experiment, the fact that children's adaptation responses persisted even after the perturbation had been removed suggests that their internal speech model was affected by the previous productions, in which auditory and proprioceptive feedback were out of phase, and reflected the updating process that occurred through the feedback and feedforward systems (Mitsuya et al., 2013; van Brenk & Terband, 2020; Villacorta et al., 2007).

Surprisingly, a substantial increase in F2' values was also noted in the child followers' baseline production – a block where no feedback manipulation occurred. This can be interpreted as indicating that children's productions are not very stable. In fact, our results also showed that young speakers were more variable than adults, regardless of the direction of their adaptation patterns or whether blocks or steps were considered, which is in line with our second hypothesis. Those observations support previous findings of greater variability in children's productions (MacDonald et al., 2012; Scheerer et al., 2013; Shiller et al., 2010; van Brenk & Terband, 2020). As in van Brenk and Terband's study (van Brenk & Terband, 2020), the variability of the productions was no greater during the blocks where an acoustic manipulation occurred. The greater variability in children than adults can therefore be associated with the fragility of their speech representations, and not the additional cognitive load induced by sensorimotor integration.

Several group effects were also found. However, because we chose to analyze the actual output (and not ratios), it is important to be cautious when interpreting them. The group effects do not indicate that children and adults reacted differently to the shifted feedback. In fact, given that a group effect was also found during the first block, when no acoustic manipulation occurred, the evidence suggests instead that preschool-aged children produce rounded vowels that are generally perceived differently than those of adults. Since it is proposed that productions associated with a F2' value lower than 13 Barks are perceived as round (Ménard et al., 2002), we can assume that both speaker groups uttered vowels that were perceived as rounded, but that adults' productions were more so.

Following on our initial investigation, which found that children generally produced vowels that were perceptually more adapted to the acoustic manipulation than adults, the second line of analysis, confirming our third hypothesis, indicated that they also used different speech strategies to shape their acoustic output.

In this study, we gradually increased F2 values by a maximum of 40%. At that point, if no compensation occurred, the produced French vowel /ø/ sounded more like /e/, its unrounded counterpart. Considering the formant manipulation applied, it is consistent that a lowering of F2 was strongly associated with an increase in perceived rounding, in both speaker groups. Our results also showed that children increased F3 to intensify the perception of rounding (and vice versa). It is commonly known that projecting the lips creates a new cavity (the labial cavity), and thus a new resonance zone. Still, in general, the production of a rounded vowel is associated with a decrease in F3, which in turn causes a decrease in F2. However, in order to make their productions sound more like /ø/ to their own ears, children lowered F2 and increased F3 instead of lowering both F3 and F2. Moreover, the results of our MLR indicate that the interaction of F2 and F3, although significantly associated with F2', decreases as the perception of rounding increases for school-aged children.

Interestingly, we found the inverse correlation in adults, for whom the interaction of F2 and F3 increased as F2' decreased. In other words, in mature speakers, the relation between F2 and F3 is even stronger as the perception of rounding intensifies. This is actually consistent with articulatory-acoustic evidence highlighting the importance of the relationship between F2 and F3 in the perception of rounded vowels (Ménard et al., 2002; Schwartz et al., 1997; Vaissière, 2009).

2.5 Conclusion

The results of this study indicate that preschool-aged children show greater adaptation and greater variability in speech productions than adults when dealing with altered auditory feedback, whether they were compensating or following the manipulated speech parameter. This suggests that children relied more strongly on their auditory feedback system than adults, which could reflect the fact that their internal model of speech is not as robust as that of adults. Moreover, children and adults used different strategies to adapt their acoustic output; children's strategies were less typical.

When a discrepancy is perceived between the auditory and somatosensory systems, individuals may choose to rely most on one specific kind of sensory feedback (Katseff et al., 2012). Because children compensated or followed the acoustic manipulation more than adults and used atypical articulatory strategies, it is suggested that they rely less on somatosensory feedback than adults do. This is in line with a previous study of ours, which found that somatosensory information played a greater role in adults' speech than in children's, suggesting that the integration of auditory and somatosensory information (i.e., development of internal speech representations) evolves throughout the course of development (Trudeau-Fisette et al., 2019).

Au cours de ce chapitre, nous avons voulu comparer les stratégies de contrôle moteur de la parole chez les enfants d'âge préscolaire et les adultes francophones. Pour ce

faire, nous avons comparé les productions des deux groupes à l'étude alors que leur rétroaction auditive était manipulée; la voyelle /ø/ produite leur étant retournée sous la forme d'un /e/. Les résultats de notre étude montrent que les enfants d'âge préscolaire ont adapté leurs productions de façon plus importante que les adultes et que leurs productions étaient plus variables que celles des adultes. Nous avons aussi observé que les deux groupes de locuteurs utilisaient des stratégies différentes pour adapter leurs productions acoustiques, celles des enfants moins typiques. Nous avons conclu que les jeunes locuteurs se fiaient davantage à leur système auditif que les adultes dans le contrôle de parole et que, pour eux, les informations somatosensorielles semblent de moins grande importance. Ceci pourrait refléter le fait que leur modèle interne ne soit pas aussi robuste que celui des adultes.

Pour faire suite à cette première étude et dans le but de mieux définir la relation qu'ont les systèmes auditif et somatosensoriel au cours du développement, nous avons mis sur pied une expérimentation visant à observer le rôle des indices de type proprioceptifs dans la perception de la parole. Les mêmes individus ont participé à une tâche de perception dans laquelle des stimuli vocaliques étaient présentés, soit dans une condition unimodale auditive, ou dans une condition bimodale où l'entrée auditive était perçue simultanément à une entrée somatosensorielle (étirement des lèvres). Nos résultats indiquent que l'effet des informations somatosensorielles sur la catégorisation des sons est plus important chez les adultes que chez les enfants, suggérant que l'intégration des informations auditives et somatosensorielles évolue au cours du développement.

CHAPITRE III: ARTICLE 2

AUDITORY AND SOMATOSENSORY INTERACTION IN SPEECH PERCEPTION IN CHILDREN AND ADULTS[4]

Paméla Trudeau-Fisette[1,2*], Takayuki Ito[3,4], Lucie Ménard[1,2]

[1]Laboratoire de Phonétique, Université du Québec à Montréal, Montreal, Canada
[2]Centre for Research on Brain, Language and Music, Montreal, Quebec, Canada
[3]GIPSA-Lab, CNRS, Grenoble INP, Université Grenoble Alpes, Grenoble, France
[4]Haskins Laboratories, New Haven, CT, USA.

***Correspondence:**
Paméla Trudeau-Fisette
ptrudeaufisette@gmail.com

Keywords: **Multisensory integration, speech perception, auditory and somatosensory feedback, adults, children, categorization, maturation**

---

[4] Cet article est actuellement publié dans la revue *Frontiers in Human Neuroscience* doi.org/10.3389/fnhum.2019.00344

Abstract

Multisensory integration allows us to link sensory cues from multiple sources and plays a crucial role in speech development. However, it is not clear whether humans have an innate ability or whether repeated sensory input while the brain is maturing leads to efficient integration of sensory information in speech. We investigated the integration of auditory and somatosensory information in speech processing in a bimodal perceptual task in 15 young adults (age 19 to 30) and 14 children (age 5 to 6). The participants were asked to identify if the perceived target was the sound /e/ or /ø/. Half of the stimuli were presented under a unimodal condition with only auditory input. The other stimuli were presented under a bimodal condition with both auditory input and somatosensory input consisting of facial skin stretches provided by a robotic device, which mimics the articulation of the vowel /e/. The results indicate that the effect of somatosensory information on sound categorization was larger in adults than in children. This suggests that integration of auditory and somatosensory information evolves throughout the course of development.

3.1 Introduction

From our first day of life, we are confronted with multiple sensory inputs such as tastes, smells, and touches. Unconsciously, related inputs are combined into a single input with rich information. Multisensory integration (MSI), also called multimodal integration, is the ability of the brain to assimilate cues from multiple sensory modalities that allows us to benefit from the information from each sense to reduce perceptual ambiguity and ultimately reinforce our perception of the world (Molholm et al., 2002; Robert-Ribes et al., 1998; Stein et al., 1996; Stein & Meredith, 1993). MSI holds a prominent place in the way that information is processed, by shaping how inputs are perceived. This merging of various sensory inputs into common neurons was typically assumed to occur late in the perceptual process stream

(Dominic & Massaro, 1999) but recent studies in neurophysiology have even demonstrated that MSI can occur in the early stages of cortical processing, even in brain regions typically associated with lower-level processing of uni-sensory inputs (Foxe et al., 2002; Macaluso, 2000; Mercier et al., 2013; Mishra et al., 2007; Molholm et al., 2002; Raij et al., 2010).

While some researchers have suggested that an infant's brain is likely equipped with multisensorial functionality at birth (Bower et al., 1970; Streri & Gentaz, 2003, 2004), others have suggested that MSI likely develops over time as a result of experiences (Birch & Lefford, 1963; Burr & Gori, 2012; Yu et al., 2010). Several studies support the latter hypothesis. For example, studies have demonstrated that distinct sensory systems develop at different rates and in different ways, which suggests that several mechanisms are implicated in MSI depending on the type of interactions (Burr & Gori, 2012; Dionne-Dostie et al., 2015; Gori et al., 2008; Walker-Andrews, 1994). For example, researchers have reported that eye-hand coordination, a form of somatovisual interaction, can be observed in infants as young as a week old (Bower et al., 1970), and audiovisual association of phonetic information emerges around 2 months of age (Kuhl & Meltzoff, 1982; Patterson & Werker, 2003) but audiovisual integration in spatial localization behavior does not appear before 8 months of age (Neil et al., 2006).

Ultimately, although it is still unclear whether an innate system enables MSI in humans, data from infants, children, and adults suggest that unimodal and multimodal sensory experiences and brain maturation enables the establishment of efficient integration processing (Gori et al., 2008; Hillock et al., 2011; Krakauer et al., 2006; Nardini et al., 2008; Neil et al., 2006; Rentschler et al., 2004; Stein et al., 2014) and that multisensory tasks in school-aged and younger children are executed through unimodal dominance rather than integration abilities (Burr & Gori, 2012; Hatwell, 1987; McGurk & Power, 1980; McGurk & MacDonald, 1976; Misceo et al., 1999).

Moreover, according to the intersensory redundancy hypothesis, perception of multimodal information is only facilitated when information from various sources is redundant, and not when the information is conflicting (Bahrick & Lickliter, 2000, 2012).

Multimodal integration is crucial for speech development. According to the associative view, during infancy, the acoustic features of produced and perceived speech are associated with felt and seen articulatory movements required for their production (Kuhl, P. & Meltzoff, 1982; Patterson & Werker, 2003; Pons et al., 2009; Yeung & Werker, 2013b) Once acoustic information and proprioceptive feedback information are strongly linked together, this becomes part of an internal multimodal speech model (Guenther & Perkell, 2004; Guenther & Vladusich, 2012; Tourville & Guenther, 2011).

MSI can sometimes be overlooked in speech perception since speakers frequently have one dominant sensory modality (Hecht & Reiner, 2009; Lametti et al., 2012). However, even though audition is the dominant type of sensory information in speech perception, many researchers have suggested that other sensory modalities also play a role in speech processing (Ito et al., 2009; Lametti et al., 2012; Perrier, 1995; Skipper et al., 2007; S. Tremblay et al., 2003)The McGurk effect, a classic perceptual illusion resulting from incongruent simultaneous auditory and visual cues about consonants clearly demonstrates that information from multiple sensory channels are unconsciously integrated during speech processing (McGurk & MacDonald, 1976).

In the current study, we examined the integration of auditory and somatosensory interaction in speech perception. Previous research has suggested that to better understand how different types of sensory feedback interact in speech perception, we need to better understand how and when this becomes mature.

Hearing is one of the first sensory modalities to emerge in humans. While still in utero, babies can differentiate speech from non-speech and distinguish variability in speech length and intensity (for a review on auditory perception in the fetus, see(Lecanuet et al., 1995)). After birth, babies are very soon responsive to various rhythmic and intonation sounds (Demany et al., 1977) and can distinguish phonemic features such as voicing, manner, and place of articulation (Eimas et al., 1971). Specific perceptual aspects of one's first language, such as sensitivity to phonemes and phonotactic properties, are refined by the first year of life (Kuhl, 1991). Although auditory abilities become well established in the early years of life, anatomical changes and experiences will guide the development of auditory skills throughout childhood (Arabin, 2002; Turgeon, 2011)

Little is known about the development of oral somatosensory abilities in typically developing children. Yet, some authors have worked on the development of oral stereognosis in children and adults, where stereognosis is the ability to perceive and recognize the form of an object in the absence of visual and auditory information, by using tactile information. In oral stereognosis, the form of an object is recognized by exploring tactile information such as texture, size or spatial properties, in the oral cavity. This is usually evaluated by comparing the ability of children and adults to differentiate or identify small plastic objects in their mouths. Researchers have reported that oral sensory discrimination skills depend on age (Dette & Linke, 1982; Gisel & Schwob, 1988; McDonald & Aungst, 1967). McDonald and Aungst (McDonald & Aungst, 1967) showed that 6- to 8-year-old children correctly matched half of the presented forms; 17- to 31-year-old adolescents and adults had perfect scores; and scores declined significantly with age among the 52- to 89-year-olds. Dette and Linke (Dette & Linke, 1982) found similar results in 3- to 17-year-olds. The effect of age was also found in younger vs. older children. Kumin et al. (Kumin et al., 1984) showed that among 4- to 11-year-olds, the older children had significantly better oral stereognosis scores than younger children. Gisel and Schwob

(Gisel & Schwob, 1988) reported that 7- and 8-year-old children had better identification skills in an oral stereognosis experiment than 5- and 6-year-old children. Interestingly, only the 8-year-old children showed a learning effect, in that they got better scores as the experiment progressed.

To explain this age-related improvement in oral stereognosis, it was suggested that oral stereognosis maturity is achieved when the growth of the oral and facial structures is complete (Gisel & Schwob, 1988; McDonald & Aungst, 1967). This explanation is consistent with vocal tracctk growth data that shows that while major changes occur in the first three years of life (Vorperian et al., 1999), important growth of the pharyngeal region is observed between puberty and adulthood (Fitch & Giedd, 1999) and multidimensional maturity of the vocal tractk is not reached until adulthood (Burr & Gori, 2012).

A few recent studies have suggested that there is a link between auditory and somatosensory information in multimodal integration.

Lametti et al. (Lametti et al., 2012) proposed that sensory preferences in the specification of speech motor goals could mediate responses to real-time manipulations, which would explain the important variability in compensatory behavior to an auditory manipulation (MacDonald et al., 2010; Purcell & Munhall, 2006b; Villacorta et al., 2007). They point out that one's own auditory feedback is not the only reliable source of speech monitoring and, in line with the internal speech model theory, that somatosensory feedback would also be considered in speech motor control. In agreement with this concept, Katseff et al. (Katseff et al., 2012) suggested that partial compensation in auditory manipulation of real-time speech could be because both auditory and somatosensory feedback system monitor speech motor control and therefore, the two systems are competing when large sensory manipulation affects only one of the sensory channels.

A recent study of speech auditory feedback perturbations in blind and sighted speakers supports the latter explanation. It showed that typically developing adults, whose somatosensory goals are narrowed by vision were more likely to tolerate large discrepancies between the expected and produced auditory outcome, whereas blind speakers, whose auditory goals had primacy over somatosensory ones, tolerated larger discrepancies between their expected and produced somatosensory feedback. In this sense, blind speakers were more inclined to adopt unusual articulatory positions to minimize divergences of their auditory goals (Trudeau-Fisette et al., 2017).

Researchers have also suggested that acoustic and somatosensory cues are integrated. As far as we know, Von Schiller (cited in (Jousmäki & Hari, 1998; Krueger, 1970)) was the first one to report that sound could modulate touch. Indeed, although he was mainly focused on the interaction between auditory and visual cues, he showed in his 1932's paper that auditory stimuli, such as tones and noise bursts, could influence an object's physical perception. Since then, studies have shown how manipulations of acoustic frequencies or even changes in their prevalence can influence the tactile perception of objects, events, and skin deformation such as their perceived smoothness, occurrence, or magnitude (Guest et al., 2002; Hotting & Roder, 2004; Ito & Ostry, 2012; Jousmäki & Hari, 1998; Krueger, 1970). Multimodal integration was stronger when both perceptual sources were presented simultaneously (Guest et al., 2002; Jousmäki & Hari, 1998).

This interaction between auditory and tactile channels is also found in the opposite direction, in that somatosensory inputs can influence the perception of sounds. For example, Schürmann et al. (Schürmann et al., 2004) showed that vibrotactile cues can influence the perception of sound loudness. Later, Gick and Derrick (Gick & Derrick, 2009) demonstrated that aerotactile inputs could modulate the perception of a consonant's oral property.

Somatosensory information coming from orofacial areas is somewhat different from those typically intended. Kinesthetic feedback usually refers to information retrieved from position, movement, and receptors in muscles and articulators (Proske & Gandevia, 2009). However, some of the orofacial regions involved in speech production movement are devoid of muscle proprioceptors. Therefore, the somatosensory information guiding our perception and production abilities likely also come from cutaneous mechanoreceptors (Ito & Gomi, 2007; Ito & Ostry, 2010; Johansson et al., 1988).

Although many studies have reported on the role of somatosensory information derived from orofacial movement in speech production (Feng et al., 2011; Ito & Ostry, 2010; Lametti et al., 2012; Nasir & Ostry, 2006; S. Tremblay et al., 2003), few studies have reported its role in speech perception.

Researchers recently investigated the contribution of somatosensory information on speech perception mechanisms. Ito et al., (Ito et al., 2009) designed a bimodal perceptual task experiment where they asked participants to identify if the perceived target was the word "head" or "had." When the acoustic targets (all members of the "head/had" continuum) were perceived simultaneously to a skin manipulation recalling the oral articulatory gestures implicated in the production of the vowel /ɛ/, the identification rate of the target "head" was significantly improved. The researchers also tested different directions of the orofacial muscle manipulation and established that the observed effect was only found if the physical manipulation reflected a movement required in speech production (Ito et al., 2009).

Somatosensory information appears to even be involved in the processing of higher-level perceptual concepts (Ogane et al., 2017). In a similar perceptual task, participants were asked to identify if the perceived acoustic target was "l’affiche" (the poster) or "la fiche" (the form). The authors showed that the appropriate temporal

positions of somatosensory skin manipulation in the stimulus word, simulating somatosensory inputs concerning the hyperarticulation of either the vowel /a/ or the vowel /i/, could affect the categorization of the lexical target.

Although further study would reinforce these findings, these experiments highlight the fact that the perception of linguistic inputs can be influenced by the manipulation of cutaneous receptors involved in speech motion (Ito et al., 2009, 2014; Ito & Ostry, 2010) and furthermore, attest of a strong link between auditory and somatosensory channels within the multimodal aspect of speech perception in adults.

The fact that sounds discrimination if facilitated when included in the infants' babbling register vihman 1996 is surely part of the growing body of evidence that demonstrates how somatosensory information that is derived from speech movement also influences speech perception in young speakers (Bruderer et al., 2015; Depaolis et al., 2011; Werker, 2018). However, to our knowledge, only two studies have investigated how somatosensory feedback is involved in speech perception abilities in children (Bruderer et al., 2015; Yeung & Werker, 2013b). In both studies, the researchers manipulated oral somatosensory feedback by constraining tongue or lip movement, thus forcing the adoption of a precise articulatory position. Although multisensory integration continues to evolve until late childhood (Ross et al., 2011), these two experiments in toddlers shed light on how this phenomenon emerges.

In their 2013 paper, Yeung and Werker, reported that when 4- and 5-month-old infants were confronted with incongruent auditory and labial somatosensory cues, they were more likely to fix the visual demonstration corresponding to the vowel perceived through the auditory channel. In contrast, congruent auditory and somatosensory cues, did not call for the need to add a corresponding visual representation of the perceived vowel.

Also using a looking-time procedure, Bruderer et al. (Bruderer et al., 2015) focused on the role of language experience on the integration of somatosensory information. They found that the ability of 6-month-old infants to discriminate between the non-native dental /d̪/ and the retroflex /ɖ/ Hindi consonant was influenced by the insertion of a teething toy. When the toddlers' tongue movements were restrained, they showed no evidence of phonetic contrast discrimination of tongue tip position. As shown by Ito et al. (Ito et al., 2009), the effect of somatosensory cues was only observed if the perturbed articulator would have been involved in the production of the sound that was heard.

While these two studies mainly focused on perceptual discrimination rather than categorical representation of speech, they suggest that proprioceptive information resulting from static articulatory perturbation plays an important role in speech perception mechanisms in toddlers, and that the phenomenon of multimodal integration in the perception-production speech model starts early in life. The authors suggested that, even at a very young age, babies can recognize that information can come from multiple sources and they react differently when the sensory sources are compatible. However, it is still unknown when children begin to integrate various sensory sources to treat them as a single sensory source.

In the current study, we aimed to investigate how dynamic somatosensory information from orofacial cutaneous receptors is integrated in speech processing in children compared to adults. Based on previous research, we hypothesized that (1) when somatosensory inputs are presented simultaneously with auditory inputs, this affects their phonemic categorization; (2) auditory and somatosensory integration is stronger in adults than in children; and (3) multisensory integration is facilitated when both types of sensory feedback are consistent.

3.2 Materials and Methods

3.2.1 Participants

We recruited 15 young adults (aged 19 to 30), including 8 females. We also recruited 21 children (aged 4 to 6) and after excluding 7 children due to equipment malfunction (1), non-completion (2), or inability to understand the task (4), this left 14 children (aged 5 to 6) including 10 females, for the data analysis. Five to six-year-old is a particularly interesting age window since children master all phonemes of their native language. However, they have not yet entered the fluent reading stage, during which explicit teaching of reading has been shown to alter multimodal perceptual (Horlyck et al., 2012).

All participants were native speakers of Canadian French and were tested for pure-tone detection threshold using an adaptive method (DT < 25 dB HL at 250, 500, 1000, 2000, 4000, and 8000 Hz). None of the participants reported having speech or language impairments. The research protocol was approved by the Université du Québec à Montréal's Institutional Review Board (no 2015-05-4.2) and all participants (or the children's parents) gave written informed consent. The number of participants was limited due to the age of the children and the length of the task (3 different tasks were executed on the same day).

3.2.2 Experimental procedure

As in the task used by Ito et al. (Ito et al., 2009) the participants were asked to identify the vowel they perceived and were asked to choose between /e/ and /ø/. Based on (Ménard et al., 2004) the auditory stimulus consisted of 10 members of a synthesized /e–ø/ continuum generated using the Maeda model (see Table 3.1). This continuum was created such that the first four formants were equally distributed from those corresponding to the natural endpoint tokens of /e/ and /ø/. To ensure that the

children understood the difference between the two vocalic choices, the vowel /e/ was represented by an image of a fairy (/e/ as in fée) and the vowel /ø/ was represented by an image of a fire (/ø/ as in feu). Since we wanted to minimize large head movements during the experiment, the children were asked to point out the image corresponding to their answers. Both images were placed in front of them at shoulder level, three feet away from each other on the horizontal plane. The adults were able to use the keyboard without looking at it and they used the right and left arrows to indicate their responses.

Table 3.1. Formant and bandwidth values of the synthesized stimuli used in the perceptual task.

	Formants					Bandwiths				
	F1	F2	F3	F4	F5	B1	B2	B3	B4	B5
stim1	364	1922	2509	3550	4000	48	55	60	50	100
stim2	364	1892	2469	3500	4000	48	55	60	50	100
stim3	364	1862	2429	3450	4000	48	55	60	50	100
stim4	364	1832	2389	3400	4000	48	55	60	50	100
stim5	364	1802	2349	3350	4000	48	55	60	50	100
stim6	364	1772	2309	3300	4000	48	55	60	50	100
stim7	364	1742	2269	3250	4000	48	55	60	50	100
stim8	364	1712	2229	3200	4000	48	55	60	50	100
stim9	364	1682	2189	3150	4000	48	55	60	50	100
stim10	364	1652	2149	3100	4000	48	55	60	50	100

Figure 3.1 shows the experimental set-up for the facial skin stretch perturbations. The participants were seated with their backs to a Phantom 1.0 device (SensAble Technologies) and they wore headphones (Sennheiser HD 380 pro). This small unit, composed of a robotic arm to which a wire is attached, allows for minor lateral skin manipulation at the side of the mouth, where small plastic tabs (2 mm x 3 mm), located on the ends of the wire, were placed with double-sided tape. The robotic arm was programed to ensure that when a 4 Newton flexion force was administered it led to a 10- to 15-mm lateral skin stretch.

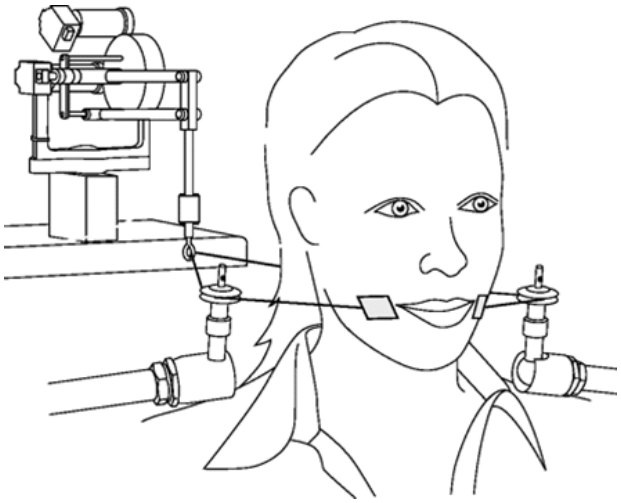

Figure. 3.1 Experimental set up for facial skin stretch perturbations (reproduced with permission from Ito & Ostry, 2010).

When this facial skin stretch is applied at lateral to the oral angle in the backward direction as shown in the figure, it mimics the articulation associated with the production of the unrounded vowel /e/. Therefore, auditory and somatosensory feedback were either congruent (with /e/-like auditory inputs) or incongruent (with /ø/-like auditory inputs). As stated early, cutaneous receptors found in the within the labial area provides speech related kinesthetic information (Ito & Gomi, 2007). Since the skin manipulation was programed to be perceived at the same time as the auditory stimuli, it was possible to investigate the contribution of the somatosensory system to the perceptual processing of the speech targets.

The auditory stimuli were presented in 20 blocks of 10 trials each. Within each block, all members of the 10-step continuum were presented, in a random order. For half of the trials, only the auditory stimulus was presented (unimodal condition). For the other half of the trials, a facial skin manipulation was also applied (bimodal condition). Alternate blocks of unimodal and bimodal conditions were presented to

the participants. In total, 200 perceptual judgments were collected, 100 in the auditory-only condition and 100 in the combined auditory and skin-stretch condition.

3.2.3 Data analysis

For each participant, stimulus, and condition, we calculated the percentage of /e/ responses. The experiment was closely monitored, and the responses in trials where a short pause was requested by the participant were excluded from the analysis. In doing so, we sought to eliminate categorical judgments for which the participants were no longer in a position to properly respond to the task (fewer than 1.1 % and 0.2 % of all responses were excluded for children and adults, respectively). These perceptual scores were then fitted onto a logistic regression model (Probit model) to obtain psychometric functions from which the labeling slopes and 50% crossover boundaries were computed. The value of the slope corresponds to the sharpness of the categorization (the lower the value, the more distinct the categorization), while the boundary value indicates the location of the categorical boundary between the two vowel targets (the higher the value, the more toward /ø/ the frontier). Using the lme4 package in R, we carried out a linear mixed-effects model (Baayen et al., 2008) for both the steepness of the slopes and the category boundaries in which group (adult or children) and condition (unimodal or bimodal) were specified as fixed factors and individual participant was defined as a random factor.

Each given answer (5800 perceptual judgments collected from 29 participants) was fitted into a linear mixed-effects model where fixed factors included stimuli (the 10-step continuum), group (adult or children), and condition (unimodal or bimodal), and the random factor was the individual participant. The mean categorization of the first and last two stimuli were also compared. Once again, the averages of the given answers (116 mean perceptual judgments collected from 29 participants) were fitted into a linear mixed-effects model where the fixed variables included stimuli (head stimuli or tail stimuli), group (adult or children), and condition (unimodal or bimodal)

and where the random variable was the individual participant. Finally, independent t-tests were carried out in order to compare variability in responses between both experimental groups and conditions. In both cases, Kolmogorov-Smirnov tests indicated that categorizations followed a normal distribution.

3.2.4 Results

The overall percentage of /e/ responses for each stimulus is shown in Figure 3.2. The data were averaged across speakers, within both groups. Figure 3.3 displays the values for the labeling slope (distinctiveness of the vowels' categorization) and 50% crossover boundary (location of the categorical frontier) averaged across experimental conditions and groups. As can be seen in both figures, regardless of the experimental condition, the children had greater variations in overall responses compared to the adults, which was confirmed in an independent t-test ($t(38)=2.792$, $p<0.01$)($t(28)=-5.503$, $p<0.001$).

3.2.5 Psychometric functions

3.2.5.1 Labeling slope results

The linear mixed-effects model revealed a significant main effect of group on the steepness of the slope ($\chi2(1)=23.549$, $p<0.001$), indicating that there was more categorical perception in adults than in children (see Figure 3.2, black lines and Figure 3.3, left-hand part of the graph).

Although no effect of condition as a main effect was observed ($\chi2(1)=3.618$, $p>0.05$), a significant interaction between group and condition was found ($\chi2(1)=4.956$, $p<0.05$). Post-hoc analysis revealed that in the bimodal condition the slope of the labeling function was more abrupt for the adults ($z=-3.153$, $p<0.01$) but not for the children, suggesting that the skin stretch condition led to a more categorical identification of the stimuli in adults only.

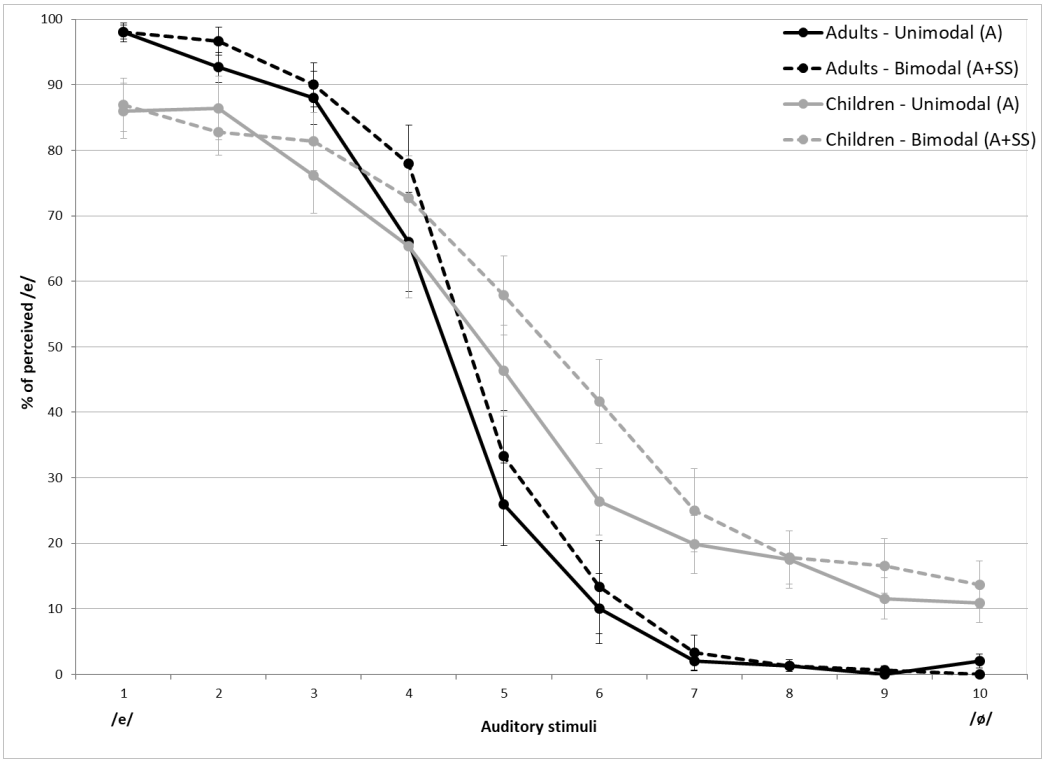

Figure 3.2. Percent identification of the vowel [e] for stimuli on the [e–ø] continuum, in both experimental conditions, for both groups. Error bars indicate standard errors.

3.2.5.2 The 50% crossover boundary results

A linear mixed-effects model analysis carried out on the 50% crossover boundaries revealed a single main effect of condition ($\chi 2(1)=9.245$, $p<0.01$). For both groups, the skin stretch perturbation led to a displacement of the 50% crossover boundary. In the bimodal condition (A+SS), the boundary was located closer to /ø/ than in the unimodal condition (A). This result is consistent with the expected effect of the skin stretch perturbation; more stimuli were perceived as /e/ than /ø/. No effect of group, as a main effect or with condition was found. The results are presented in Figure 3.2 and in Figure 3.3, in the right-hand part of the graphs.

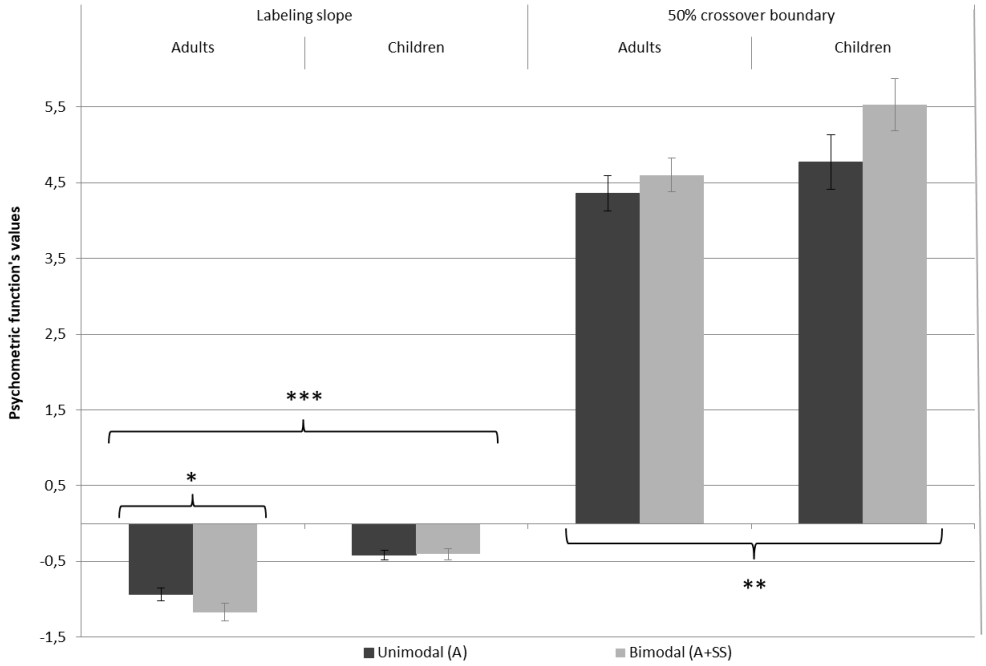

Figure 3.3: Psychometric functions of labeling slope and 50% crossover boundary, in both experimental conditions, for both groups. Error bars indicate standard errors.

3.2.6 Categorical judgments

A linear mixed-effects model analysis performed on the categorical judgments revealed that in addition to the expected main effect of stimuli ($\chi2(1)=3652.4$, $p<0.001$), there were significant effects of group ($\chi2(1)=4.586$, $p<0.05$) and condition ($\chi2(1)=15.736$, $p<0.001$), suggesting that children and adults did not categorize the stimuli in a similar manner and that both experimental conditions prompted different categorization. Moreover, a significant interaction of group and stimuli ($\chi2(1144.52$, $p<0.001$) revealed that irrespectively of the experimental condition, some auditory stimuli were categorized differently by the two groups.

Post-hoc tests revealed that whether a skin stretch manipulation was applied or not, stimulus 7 (A $z=-3.795$, $p<0.1$ / A+SS $z=-4.648$, $p<0.01$), 8 (A $z=-3.445$, $p<0.5$ /

A+SS $z=-3.544$, $p<0.1$) and 9 (A $z=-3.179$, $p<0.5$ / A+SS $z=-4.347$, $p<0.01$) were more systematically identified as /ø/ by the adults than by the children. While no other two-way interactions were found, a significant three-way interaction of group, condition, and stimuli was observed ($\chi2(4)=117.26$, $p<0.001$) suggesting that, for some specific stimuli, the skin stretch condition affected the perceptual judgment of both groups in a different manner.

First, it was found that the skin stretch manipulation had a greater effect on stimulus 6, in children only ($z=-3.251$, $p<0.5$). For this group, the skin stretch condition caused a 15.8% increase of /e/ labeling on stimulus 6. For the adults, the addition of somatosensory cues only led to a 3.3% increase in /e/ categorization.

Although less expected, the skin stretch manipulation also led to some perceptual changes at the endpoint of the auditory continuum. As shown in Figure 3.2, stimulus 2 ($z=3.053$, $p<0.5$) and stimulus10 ($z=-3.734$, $p<0.1$) were labeled differently by the two groups, but only in the bimodal condition. In fact, stimulus 2 (an /e/-like stimulus) was more likely to be identified as an /e/ by the adults in the experimental condition. In contrast, children were less inclined to label it so. As for stimulus 10 (an /ø/-like stimulus), the addition of somatosensory inputs decreased the correct identification rate in children only. In adults, although it barely affected their categorical judgments, the skin stretch manipulation mimicking the articulatory gestures of the vowel /e/ resulted in an increase of /ø/ labeling, as if it had a reverse effect.

Last, a comparison of mean categorizations of the first and last two stimuli revealed a main effect of stimuli ($\chi2(1)=313.52$, $p<0.001$) and a significant interaction of group and stimuli ($\chi2(1)=36.260$, $p<0.001$). More importantly, it also revealed a 3-way interaction of group, condition, and stimuli ($\chi2(4)=37.474$, $p<0.001$). Post-hoc tests indicated that those endpoint stimuli of the continuum were categorized differently by the two groups, but only when a skin stretch manipulation was applied. In agreement

with previous results, in the skin stretch condition, children labeled more /e/-like stimuli as /ø/ ($z=3.434$, $p<0.5$), and more /ø/-like stimuli as /e/ ($z=-4.139$, $p<0.01$).

3.3 Discussion

This study aimed to investigate how auditory and somatosensory information is integrated in speech processing by school-aged children and adults, by testing three hypotheses.

As hypothesized, the overall perceptual categorization of the auditory stimuli was affected by the addition of somatosensory manipulations. The results for psychometric functions and categorical judgments revealed that auditory stimuli perceived simultaneously with skin stretch manipulations were labeled differently than when they were perceived on their own. Sounds were more perceived as /e/ when they were accompanied by the proprioceptive modification.

The second hypothesis that auditory and somatosensory integration would be greater in adults than in children was also confirmed. As shown in Figures 3.2 and 3.3, orofacial manipulation affected the position of the 50% crossover boundary of both groups; when backward skin stretches were perceived simultaneously with the auditory stimulus, it increased its probability of being identified as an /e/. This impact of skin stretch manipulation on the value corresponding to the 50th percentile was also reported in Ito et al.'s experiment (Ito et al., 2009). However, bimodal presentation of auditory and somatosensory inputs affected the steepness of the slope in adults only. Figure 3.2 also shows that adult participants were more likely to label /e/-like stimuli as /e/ in the bimodal condition. Since negligible changes were observed for /ø/-like stimuli, it led to a more categorical boundary between the two acoustic vocalic targets. This difference in the integration patterns between children

and adults suggests that linkage of specific somatosensory inputs with a corresponding speech sound evolves with age.

The third hypothesis that MSI would be stronger when auditory and somatosensory information was congruent was confirmed in adults but not in children. Only adults' perception was facilitated when both sensory information was consistent. In children, a decrease in the correct identification rate resulted from the bimodal presentation when auditory and proprioceptive inputs were compatible. Moreover, while adults seemed to not be affected by the /e/-like skin stretches when auditory stimuli were alongside the prototypical /ø/ vocalic sound (see Figure 3.2), children's categorization was influenced even when sensory channels were clearly contrasting, as if the bimodal presentation of vocalic targets blurred the children categorization abilities. Moreover, thought somatosensory information mostly affected specific stimuli in adult, it's effect in children was further distributed along the auditory continuum. These last observations support our second hypothesis that multisensory integration is strongly defined in adults.

As many have suggested, MSI continues to develop during childhood (Dionne-Dostie et al., 2015; Ross et al., 2011). The fact that young children are influenced by somatosensory inputs in a different manner then adults could therefore be due to their underdeveloped MSI abilities. Related findings have been reported for audiovisual integration (Desjardins et al., 1997; Massaro, 1984; McGurk & MacDonald, 1976). It has also been demonstrated that the influence of visual articulators in audition is weaker in school-aged children than in adults.

In agreement with the concept that MSI continues to develop during childhood, the differences observed between the two groups of perceivers could also be explained by the fact that different sensory systems develop at different rates and in different ways. In that sense, it has also been found that school-aged children were not only less

likely to perceive a perceptual illusion resulting from incongruent auditory and visual inputs, but they also had poorer results in the identification of unimodal visual targets (Massaro, 1984).

Studies of the development of somatosensory abilities also support this concept. As established earlier, oral sensory acuity continues to mature until adolescence (Dette & Linke, 1982; Holst-Wolf et al., 2016; McDonald & Aungst, 1967). The young participants who were 5 to 6 years of age in the current study may have had underdeveloped proprioceptive systems, which may have caused their less clearly defined categorization of bimodal presentations.

It is generally accepted that auditory discrimination is poorer and more variable in children than in adults (Buss et al., 2009; Macpherson & Akeroyd, 2014), and children's lower psychometric scores are often related to poorer attention (Moore et al., 2008).

MSI requires sustained attention, and researchers have suggested that poor psychometric scores in children might be related to an attentional bias between the recruited senses in children versus adults (Alsius et al., 2005; Barutchu et al., 2009; Spence & McDonald, 2004). For example, Barutchu et al. (Barutchu et al., 2009) observed a decline in multisensory facilitation when auditory inputs were presented with a reduced signal-to-noise ratio. They suggested that the increased level of difficulty in performing the audiovisual detection task under high noise condition may be responsible for the degraded integrative processes.

If this attention bias might explain some of the between-group performance differences found when /e/-like somatosensory inputs were presented with /ø/-like auditory inputs (high level of difficulty), it would not justify differences between children and adults when the auditory and somatosensory channels agreed. The

children showed decreased multisensory ability when both sensory inputs were compatible. Since difficulty level was reduced when multiple sensory sources were compatible, we should only have observed confusion in the children's categorization when auditory and somatosensory information was incongruent. According to the intersensory redundancy hypothesis, MSI should be improved when information from multiple sources is redundant. Indeed, Bahrick and Lickliter (Bahrick & Lickliter, 2000) suggested that concordance of multiple signals would guide attention and even help learning (Barutchu et al., 2010). In the current study, this multisensory facilitation was only found in the adult participants.

This latter observation and the fact that no significant differences in variability were found across experimental conditions makes it difficult to link the dissimilar patterns of MSI found between the two groups to an attentional bias in children. However, finding a greater variability in MSI in children in both conditions, combined with their distinct psychometric and categorical scores provides support for the concept that perceptual systems in school-aged children are not yet fully shaped, which prevents them from attaining adult-like categorization scores.

As speech processing is multisensory and 5- to 6-year-olds have already experienced it, it is not surprising that some differences, even typical MSI ones, were found between the two experimental conditions in children. Since even very young children recognize that various speech sensory feedback can be compatible—or not (Bruderer et al., 2015; Patterson & Werker, 2003; Werker, 2018; Yeung & Werker, 2013b) the different behavioral patterns observed in this study suggest that some form of multimodal processing exists in school-aged children, but complete maturation of the sensory systems is needed to achieve adult-like MSI.

3.4 Conclusion

When somatosensory input was added to auditory stimuli, it affected the categorization of stimuli at the edge of the categorical boundary for both children and adults. However, while the oral skin stretch manipulation had a defining effect on phonemic categories in adults, it seemed to have a blurring effect in children, particularly on the prototypical auditory stimuli. Overall, our results suggest that since adults have fully developed sensory channels and more experiences in MSI, they have stronger auditory and somatosensory integration than children.

Although longitudinal observations are not possible, two supplementary experiments in these participants have been conducted to further investigate how multisensory integration takes place in speech processing in school aged children and adults, two supplementary experiments in these participants has been conducted. These investigate the role of visual and auditory feedback in speech processing.

Au cours du présent chapitre, nous avons établi que le rôle des informations somatosensorielles sur la catégorisation des sons s'opérait de façon plus importante chez les adultes que chez les enfants de 5 à 6 ans, suggérant ainsi que l'intégration des informations auditives et proprioceptives évoluerait au cours du développement. Cette étude est venue appuyer les résultats de notre premier article dans lequel nous avions conclu que les informations auditives étaient, pour les individus encore en développement, celles ayant le plus de poids dans le contrôle de la parole. À la lumière de ce deuxième travail, il semble que les informations acoustiques soient également celles qui gouvernent les habiletés de catégoration chez les jeunes enfants, les informations liées au système moteur paraissant encore avoir un rôle de second plan.

Néanmoins, les indices associés au système moteur s'observent aussi via les informations visuelles. En effet, en contexte de conversation en face à face, un interlocuteur verra les gestes articulatoires de son interlocuteur, comme le mouvement des lèvres et de la mandibule, ce qui facilite la perception des cibles (Dohen & Loevenbruck, 2009; Robert-Ribes et al., 1998). Afin de dresser un portrait complet du développement des habiletés d'intégration multimodale, nous avons réalisé une troisième étude dont l'objectif était de rendre compte du développement du phénomène d'intégration audiovisuelle de la parole. Il s'agit d'une thématique somme toute bien définie chez l'adulte, mais donc les conclusions sont encore très variables chez l'enfant. Ainsi, des tests de perception audio et audiovisuelle ont été réalisés auprès des mêmes individus. De façon générale, nos résultats montrent que les enfants d'âge préscolaire sont en mesure d'intégrer des signaux auditifs et visuels comme les adultes, mais que des différences individuelles substantielles sont présentes à ce stade. Nous proposons que le traitement bimodal audiovisuel ne soit optimal que chez les enfants dont les systèmes sensoriels ont atteint maturité.

CHAPITRE IV: ARTICLE 3

VISUAL INFLUENCE ON AUDITORY PERCEPTION OF VOWELS BY FRENCH-SPEAKING CHILDREN AND ADULTS[5]

Paméla Trudeau-Fisette[1,2*], Laureline Arnaud[2,3], Lucie Ménard[1,2]

[1]Laboratoire de Phonétique, Université du Québec à Montréal, Montreal, Quebec, Canada
[2]Centre for Research on Brain, Language and Music, Montreal, Quebec, Canada
[3] Integrated Program in Neuroscience, McGill University, Montreal, QC, Canada

***Correspondence:**
Paméla Trudeau-Fisette
ptrudeaufisette@gmail.com

Keywords: **speech perception, adults, children, sensorimotor maturation, audiovisual interaction**

---

[5] Cet article est actuellement publié dans la revue *Frontiers in Psychologie. doi.org/10.3389/fpsyg.2022.740271*

Abstract

Audiovisual interaction in speech perception is well defined in adults. Despite the large body of evidence suggesting that children are also sensitive to visual input, very few empirical studies have been conducted. To further investigate whether visual inputs influence auditory perception of phonemes in preschoolers in the same way as in adults, we conducted an audiovisual identification test. The auditory stimuli (/e/-/ø/ continuum) were presented either in an auditory condition only or simultaneously with a visual presentation of the articulation of the vowel /e/ or /ø/. The results suggest that, although all participants experienced visual influence on auditory perception, substantial individual differences exist in the 5- to 6-year-old group. While additional work is required to confirm this hypothesis, we suggest that auditory and visual systems are developing at that age and that multisensory phonological categorization of the rounding contrast took place only in children whose sensory systems and sensorimotor representations were mature.

4.1 Introduction

In neurotypical individuals, face-to-face communication is multisensory (Rosenblum, 2008a, 2008b). In addition to body movements and facial expressions, effective communication relies heavily on multisensory information, such as auditory, visual, and proprioceptive cues ((Stein et al., 2014; Stein & Stanford, 2008). Multisensory processing is crucial for efficient perception, as it optimizes brain functions and reduces perceptual ambiguity (Gori, 2015; Stein et al., 2014). From a developmental perspective, several studies provide evidence that children do not display adult-like multisensory processing until late childhood(Barutchu et al., 2009; Burr & Gori, 2012; Gori et al., 2008). First, since sensory systems are not mature at birth, but evolve and are calibrated throughout childhood (Burr & Gori, 2012), multisensory processing constantly adapts to different kinds of inputs (Birch & Lefford, 1963; Yu et al., 2010).

Second, the brain areas shown to be involved in multisensory processing are not operational at birth but develop with experience (Stein et al., 2014). Despite its crucial role in speech perception and its continuing refinement in the first years of life, very little is known about the development of multisensory processing in the specific area of speech.

This study is part of a larger project investigating the development of such perceptual processes in school-aged children. In a recent paper (Trudeau-Fisette et al., 2019), we investigated a specific case of multisensory processing, namely, the interaction between auditory and somatosensory input during vowel perception in children and adults. More specifically, 10 synthesized vowels equally stepped on a continuum from /e/ (as in “fée” fairy) to /ø/ (as in “feu” fire) were presented in the auditory modality to francophone adults and children ranging in age from 4 to 6 years old. The participants’ task was to categorize the sounds they perceived as either /e/ or /ø/. In some trials, a facial skin stretch applied by a mechanical device on the participant’s cheeks was synchronized with the audio signal, mimicking the articulation normally associated with the production of the vowel /e/. The data showed that the effects of somatosensory feedback on auditory vowel categorization were reduced in children compared to adults. Our results thus suggest that preschool-aged children do not combine auditory and somatosensory information in the same way as adults do. In the current follow-up paper, we focus on the interaction between the auditory and visual sensory modalities in the perception of the same vowel continuum in francophone children and adults.

4.1.1 The Development of Audiovisual Interaction in Speech Perception

Substantial work has shown evidence of early sensitivity to auditory and visual interaction in speech perception. For instance, it has been shown that babies show facial mimicking skills after only a few days of life (Meltzoff & Moore, 1983) and

that they are able, after a couple of months, to recognize whether or not information they receive through auditory and visual channels is compatible (Burnham & Dodd, 2004; Dodd, 1979; Kuhl & Meltzoff, 1982; Legerstee, 1990; Patterson & Werker, 1999). Prelinguistic infants are also sensitive to the famous McGurk effect, whereby an auditory stimulus /ba/ dubbed on a visual stimulus/ga/triggers the perception of/da/ (McGurk & MacDonald, 1976). However, in their original work, McGurk and MacDonald (1976) observed that anglophone children (aged 3–4 and 7–8) were generally less subject to audiovisual illusions than adults (see also Massaro et al., 1986). Dupont et al. (2005) also reported a reduced influence of visual input on audiovisual consonant perception in French-speaking children aged 4 and 5 years compared to adults. In incongruent stimuli (were the visual signal corresponded to a different phoneme than the auditory signal), children relied more frequently on the auditory signal only than adults did. Several papers have since confirmed that, during the first decade of life, children do not integrate auditory and visual cues as much as adults do (Burr & Gori, 2012; Desjardins & Werker, 2004; Hockley & Polka, 1994; Knowland et al., 2014; C. Tremblay et al., 2007).

Another common manifestation of the interaction between auditory and visual cues in speech perception is multisensory enhancement, whereby identification scores are greater in the audiovisual condition than in either the auditory or the visual condition (see Stein et al., 2014 for a discussion of this process in MSI). This audiovisual enhancement (or audiovisual gain) is found in quiet conditions (Robert-Ribes et al., 1998) as well as in contexts where the auditory signal is degraded (Dohen & Loevenbruck, 2009; Grant & Seitz, 2000). Indeed, in noisy environments, visual information on the speech articulators helps shape the overall perception of speech signals by recovering part of the information that is lost from the auditory channel. Similarly to the pattern found for the McGurk effect, Ross et al. (2011) found that children (aged 5–14 years old) had a smaller gain in correct identification scores than adults when exposed to audiovisual signals compared to auditory signals in low

signal-to-noise ratios (SNRs; see also Barutchu et al., 2010). Along the same lines, Wightman et al. (2006) showed that, until 9 years of age, visual information is not used to recover masked auditory signals.

To summarize, past findings indicate that, while prelinguistic infants are able to distinguish bimodal from unimodal speech stimuli and show a preference for compatible information, children do not attach as much weight to visual sensory cues as adults do in incongruent audiovisual conditions or degraded auditory conditions. While the literature indicates that MSI processes in general require sensory experiences and brain maturation (Gori et al., 2008; Hillock et al., 2011; Stein et al., 2014), it is suggested that some form of audiovisual interaction exists early in life

4.1.2 The Case of the Rounding Contrast

Perceptual facilitation of audiovisual information is largely due to the fact that information recovered by the ear and eye is complementary: while auditory cues mainly convey voicing and manner, visual inputs transmit information about place of articulation and rounding (Macleod & Summerfield, 1987; Ménard, 2015; Peelle & Sommers, 2015; Robert-Ribes et al., 1998). In a study of vowel perception in French in various sensory conditions (auditory alone, visual alone, and audiovisual) at different noise levels, Robert-Ribes et al. (Robert-Ribes et al., 1998) proposed robustness scales for vowel features (the higher the correct identification score in noisy conditions, the greater the robustness). In the audio channel, height is the most robust feature, followed by place of articulation, which in turn is more robust than rounding. In the visual channel, rounding is the most robust feature, followed by height, while place of articulation is the least robust feature.

Because of its visual saliency, the contrast between rounded vowels and unrounded vowels is ideal for studying audiovisual interactions in speech perception, particularly

in languages like French, Dutch, or Swedish, in which this feature is phonologically relevant. In French, (Lisker & Rossi, 1992) instructed experts to rate vowel rounding based on an auditory presentation of vowels, sometimes accompanied by unmatched visually articulated vowels. Despite clear acoustic signals denoting rounded vowels, most listeners were influenced by the visually presented unrounded vowels (although the extent of the interaction varied across participants). Furthermore, (Traunmuïler & Niklas Öhrström, 2007) presented nonsense syllables containing unrounded vowels (/gig/ and /geg/) and rounded vowels (/gyg/ and /gøg/) to adult Swedish participants. Their task was to identify the syllable they perceived. In some trials, the auditory and visual parts of the stimuli corresponded to similar rounding values (congruent stimuli) while in others, the two modalities denoted different values (incongruent stimuli). Responses to the incongruent stimuli suggested that visually indexed rounding is heavily weighted in listeners' perception: a stimulus in which the auditory part is unrounded and the visual part is rounded is generally perceived as rounded. In a later paper, Valkenier et al. (2012) combined visually articulated Dutch vowels contrasting in terms of height and rounding to congruent and incongruent auditory signals of the vowels, mixed with noise. Native Dutch adult listeners were instructed to identify the vowel they perceived. The results showed an audiovisual facilitation effect, as correct identification was enhanced by congruent audiovisual presentation. On the contrary, incongruent presentation degraded correct identification.

Although the influence of visual cues on the auditory perception of the rounding feature in languages like French has been established, nothing is known about its development in children. In this paper, we report on an experiment carried out to investigate whether processing of auditory and visual information occurs in preschool-aged children in the same way as it does in adults. As in our previous study of auditory and somatosensory interaction (Trudeau-Fisette et al., 2019), we focused on the perception of the unrounded/rounded vowel pair /e/-/ø/ in Canadian French.

Based on the previous work on the developmental time course of MSI processes presented in Section “The Development of Audiovisual Interaction in Speech Perception”, it is expected that the effect of vision on the perception of rounding contrasts will be reduced in schoolaged children compared to adults.

4.2 Materials and Methods

4.2.1 Participants

Thirty young adults and 30 children were recruited. After excluding three adults and seven children due to equipment malfunction (two adults and two children), non-completion (three children), or inability to perform the task (one adult and two children), we were left with 27 adults (aged 19–30; mean=26.3, 13 females) and 23 children (aged 5–6; mean=5.6, 15 females). At that age, the phonemic categories under study (/e/ and /ø/) are well mastered (Ménard et al., 2007)

All participants were native speakers of Canadian French and were tested for pure-tone detection threshold using an adaptive method (DT < 25dB HL at 250, 500, 1,000, 2,000, 4,000, and 8,000Hz). Every participant (or their parents) reported that they had never had speech, language, neurological, or psychological disorders. They also reported having normal or corrected-to-normal vision. Every participant (or their parents) gave written informed consent to participate in the experiment. The research protocol was approved by Université du Québec à Montréal’s Institutional Review Board (no. 2015-05-4.2).

4.2.2 Experimental Procedures

The experiment consisted of an audiovisual identification test. The auditory stimuli corresponded to 10 equally stepped synthesized five-formant vowels on the /e–ø/ continuum. The vowels were synthesized using the Maeda model (Maeda, 1990).

First, prototypical formant and bandwidth values for French /e/ and /ø/ were determined for the model (Ménard et al., 2004). For each of the vowels on the continuum, the values of the first and fifth formants were fixed (F1=364Hz and F5=4,000Hz), while the second, third, and fourth formant values (in Hertz) were interpolated from the values of the two endpoints /e/ and /ø/ (see Table 4.1). Formant bandwidth values were as follows: B1=48Hz, B2=55Hz, B3=60Hz, B4=50Hz, and B5=100Hz. Each stimulus lasted 600 milliseconds and had a mean fundamental frequency of 130Hz.

Table 4.1 Values of the second, third, and fourth formants (in Hz) of the synthesized stimuli used in the perceptual task.

Stimulus number	F2	F3	F4
1	1,922	2,509	3,550
2	1,892	2,469	3,500
3	1,862	2,429	3,450
4	1,832	2,389	3,400
5	1,802	2,349	3,350
6	1,772	2,309	3,300
7	1,742	2,269	3,250
8	1,712	2,229	3,200
9	1,682	2,189	3,150
10	1,652	2,149	3,100

For all stimuli, F1=364Hz, F5=4,000Hz, B1=48Hz, B2=55Hz, B3=60Hz, B4=50Hz, and B5=100Hz. Stimulus 1 is the prototypical /e/ and stimulus 10 is the prototypical /ø/.

In our experiment, the identification scores in various audiovisual conditions involving different visual components mixed with the same auditory stimuli will be compared. The auditory modality will be considered the dominant modality. Following a method used previously in speech perception (Ito et al., 2009; Trudeau-Fisette et al., 2019), the synthesized vowels were presented either in an auditory-only (AO) condition or simultaneously with a visual presentation of the prototypical articulation of the vowel /e/ or /ø/ (see Figure 4.1 for a schematic description of the audiovisual conditions). The visual signals were obtained from an adult French Canadian male speaker producing the two vowels /e/ and /ø/. The use of prototypical

visual stimuli has been shown to be efficient in evaluating audiovisual speech perception in a young population (Kuhl & Meltzoff, 1982; Patterson & Werker, 2003; Yeung & Werker, 2013b). Several repetitions of the vowels were obtained with instructions to start and end with a neutral position. The best occurrence of each vowel was selected by the experimenter. In one condition (AV /e/), each of the 10 synthesized auditory stimuli of the /e/-/ø/ continuum was manually mapped into the muted visual articulation of /e/. In the other condition (AV /ø/), each auditory stimulus was combined with the visual articulation of /ø/. Participants were asked to identify the vowel they perceived and were forced to choose between /e/ and /ø/. To ensure that the children understood the difference between the two vowel choices, the vowel /e/ was represented by an image of a fairy (/e/ as in fée) and the vowel /ø/ was characterized by an image of a fire (/ø/ as in feu). Adults were asked to use the right and left arrows of the computer keyboard to indicate their responses. Children pointed to the image corresponding to their answers (placed right in front of them on the top left and right corners of the laptop) and the experimenter selected the corresponding keyboard key. A practice round was conducted with each participant to ensure that they understood the task.

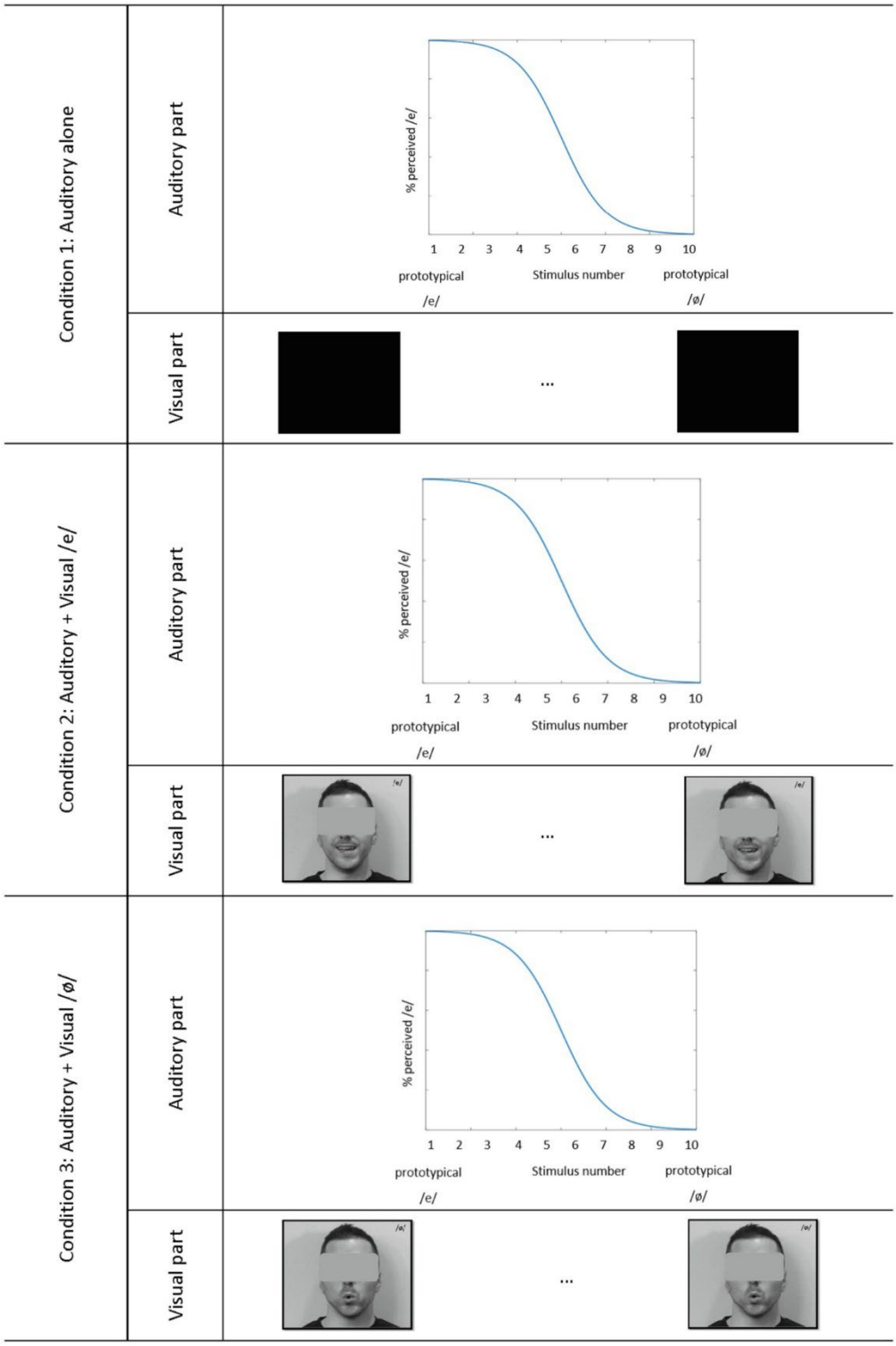

Figure 4.1. Schematic representation of the audiovisual stimuli in the three conditions

The stimuli were presented in 18 blocks of 10 trials each. Within each block, all members of the 10-step continuum were presented in a randomized fashion. For one-third of the trials, only the auditory stimulus was presented (Auditory Only). For the other two-thirds of the trials, the visual articulation corresponding to the vowels /e/ and /ø/ was also presented randomly. Alternate blocks of unimodal and bimodal conditions (Audiovisual /e/ condition and Audiovisual /ø/ condition altogether) were presented to the participants. In total, 180 perceptual judgments were collected, 60 in the auditory-only condition, and 120 in the combined audiovisual conditions.

4.2.3 Data Analysis

For each participant, stimulus, and condition, we calculated the percentage of /e/ responses. Although reaction time (RT) could not be analyzed due to differences in the experimental procedures between the two groups, it was used to exclude responses where RT was ±2 standard deviations from the blocks' mean RT. In doing so, we sought to eliminate categorical judgments for which the participants were no longer in a position to properly respond to the task (less than 2.2 and 1.1% of all responses were excluded for children and adults, respectively). Excluded responses were fairly distributed across speakers and conditions; no more than 3% of the responses were discarded for each individual. While perceptual categorization of speech targets is often analyzed through psychometric functions (e.g., 50% crossover boundaries and labeling slopes), responses collected from children, for whom the categorical boundary was rarely crossed in the audiovisual conditions, prevent us from using these paradigms to describe the obtained results. Therefore, each answer given (9,000 perceptual judgments collected from 50 participants) was fitted into a linear mixed-effects model (LMEM; using the lme4 package in R) in which the fixed factors were stimulus (the 10-step continuum), group (adult or children), and condition [Auditory Only (AO), Audiovisual /e/ (AV /e/), or Audiovisual /ø/ (AV /ø/)], and the random factor was the individual participant. Post-hoc analyses were

performed using multiple comparisons (using the multcomp package in R). Values of p were corrected using the Bonferroni method. It should be specified that, since we were interested in the influence of vision on the auditory phonemic categorization of the rounding contrast, only individuals with a typical psychometric curve in the AO condition (those for whom the endpoint stimuli belonged to the two different phonological categories) were included in the analyses. As mentioned previously, two children were excluded because of their "inability to perform the task."

4.3 Results

4.3.1 Mean Perceptual Scores across Conditions and Groups

The mean percentage of /e/ responses for each stimulus is shown in Figure 4.2. Data are averaged across speakers, within each group (Adults, Children) and experimental condition (AO, AV/e/, AV/ø/).

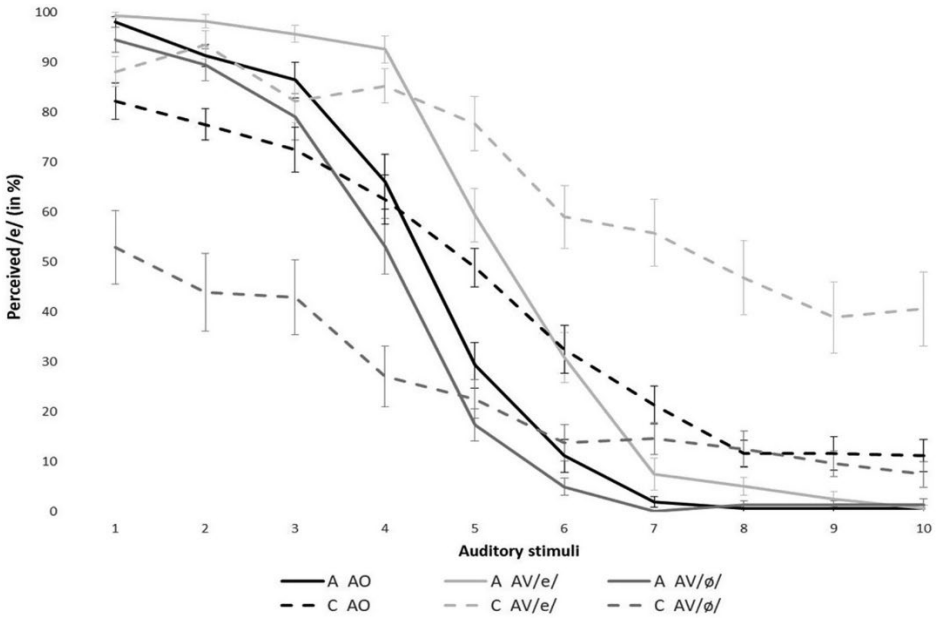

Figure 4.2 Mean percentage identification of the vowel [e] for stimuli on the [e–ø] continuum across speaker groups (adults and Children) and experimental conditions. Error bars indicate standard errors.

In addition to the expected main effect of stimulus ($\chi2(1) = 4360$, $p < .001$), LMEM revealed significant effects of group ($\chi2(1) = 4.88$, $p < .05$) and condition ($\chi2(2) = 772.14$, $p < .001$). Overall, this means that, regardless of the experimental condition, children and adults labeled the perceived stimuli differently and that, for both groups, the three experimental conditions led to different categorization scores.

Significant interactions between group and stimulus ($\chi^2(1) = 482.92$, $p < .001$), group and condition ($\chi^2(2) = 197.21$, $p < .001$) and condition and stimulus ($\chi^2(2) = 51.73$, $p < .001$) were also revealed. In the AO condition, children and adults had significantly different identification scores. For the prototypical stimuli, children had lower identification scores than adults for /e/-like stimuli (stimuli 1–3). Conversely, for /ø/-like stimuli (stimuli 8–10), children perceived /e/ more than adults (thus, less /ø/). The perception of several ambiguous stimuli (stimuli 5–7) also yielded significantly higher identification scores for /e/ in children than adults, suggesting a less categorical shape of the perception function. More importantly, a significant three-way interaction of group, condition, and stimulus ($\chi^2(7) = 785.80$, $p < .001$) was found, revealing that, for various stimuli, audiovisual presentation affected the perceptual categorization of the speech target differently for children and adults.

Post hoc tests indicated that, in children, responses corresponding to the perception of stimuli 1 to 6 (/e/-like to ambiguous) presented with the visual articulation of the vowel /ø/ (dashed dark gray line) were significantly more associated with the label /ø/ than when perceived only auditorily (1: $z = –.815$, $p < .001$; 2: $z = –9.256$, $p < .001$; 3: $z = –7.242$, $p < .001$; 4: $z = –7.746$, $p < .001$; 5: $z = –5.348$, $p < .001$; 6: $z = –3.875$, $p < .01$). Likewise, categorization of stimuli 4 to 10 (ambiguous to /ø/-like) presented under the bimodal /e/ condition (dashed light gray line) was more associated with the vowel /e/ than in the AO condition (4: $z = 4.932$, $p < .001$; 5: $z = 5.888$, $p < .001$; 6: $z = 5.539$, $p < .001$; 7: $z = 8.321$, $p < .001$; 8: $z = 10.225$, $p < .001$; 9: $z = 8.347$, $p < .01$; 10: $z = 9.887$, $p < .01$). In adults, only the visual /e/ condition (solid light gray line)

on ambiguous auditory targets (stimuli 4, 5 and 6) led to a significant change in categorization (4: $z = 6.439$, $p < .001$; 5: $z = 6.259$, $p < .001$; 6: $z = 4.452$, $p < .001$).

To better evaluate the effect of visual stimuli on children's auditory perception, we divided the young participants into two groups based on their categorization pattern in the AV conditions. Since we are interested in the perception of the rounding contrast, we divided the child participants based on the presence or absence of a categorical distinction in the AV conditions. Figure 4.3 displays the overall percentage of /e/ responses for each stimulus, according to the three groups: Adults, Children 1 (C1), and Children 2 (C2). C1 consisted of 11 (mean age: 5.8) children whose categorical slopes crossed the 50% boundary in both AV conditions. C2 was composed of 12 children (mean age: 5.4) for whom the 50% boundary was not crossed in at least one of the AV conditions. For 5 of those 12 individuals, the 50% boundary was not crossed in any of the AV conditions. Interestingly, the 50% boundary was crossed in the AV /e/ condition for only six of the remaining children while it was crossed in the AV /ø/ condition for only a single child. As Figure 4.3 shows, individuals from all groups had a typical psychometric curve in the AO condition. The data presented in Figure 4.3 reveal that, rather than performing at the chance level, children adopted two completely different behavioral patterns. The first group of children, C1 (middle graph in Figure 4.3) displayed a clear categorical distinction between /e/ and /ø/ in the three experimental conditions. However, the C2 group of children (right-hand graph in Figure 4.3) did not display a response pattern in the two AV conditions consistent with the categorical perception of /e/ and /ø/: unlike the children in the C1 group, those in the C2 group made responses dominated by visual input (either /e/ or /ø/, depending on the condition). To account for the differences between these groups, LMEM analyses were computed where the fixed factors were stimulus (the 10-step continuum), group (adult, C1, or C2), and condition [AO, AV /e/, or AV /ø/], and the random factor was the individual participant. While no main effect of group was detected [$\chi 2$ (1)=3.049, p>0.05],

significant interactions between group and stimulus [$\chi 2$ (1)=161.27, $p < 0.001$], group and condition [$\chi 2$ (2)=44.75, $p<0.001$], and group, stimulus, and condition were identified [$\chi 2$ (7)=267.14, $p<0.001$].

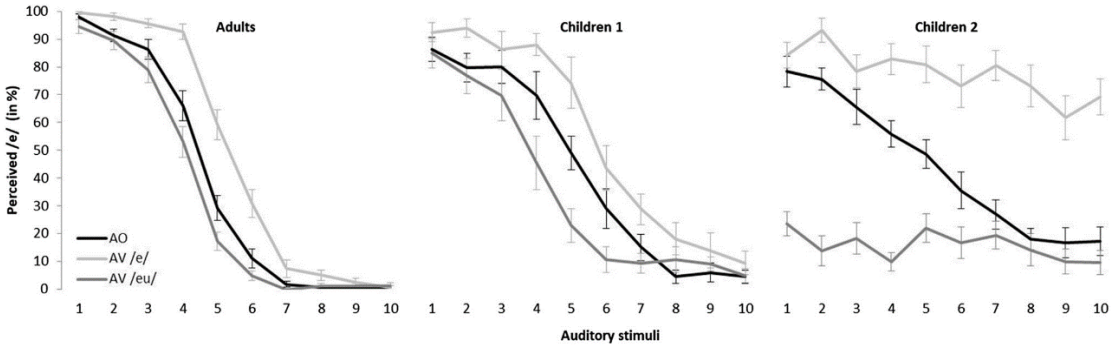

Figure 4.3. Mean percentage identification of the vowel [e] for stimuli on the [e] – [ø] continuum across speaker groups and experimental conditions. Error bars indicate standard errors.

4.3.2 Visual Gain on Categorization Scores

In order to better investigate the weight of experimental condition in the three-way interaction displayed in Figure 4.3, the difference between the identification scores in the two AV conditions relative to the AO condition was computed. This difference, corresponding to the visual gain, is shown in Figure 4.4 for the three groups. The difference in categorization between the AV/e/ and AO conditions is shown in the left-hand panel, while the difference between the AV/ø/ and AO conditions is displayed in the right-hand panel. For the sake of clarity, significance is shown only with asterisks, but detailed z values for group differences are presented in Table 4.2.

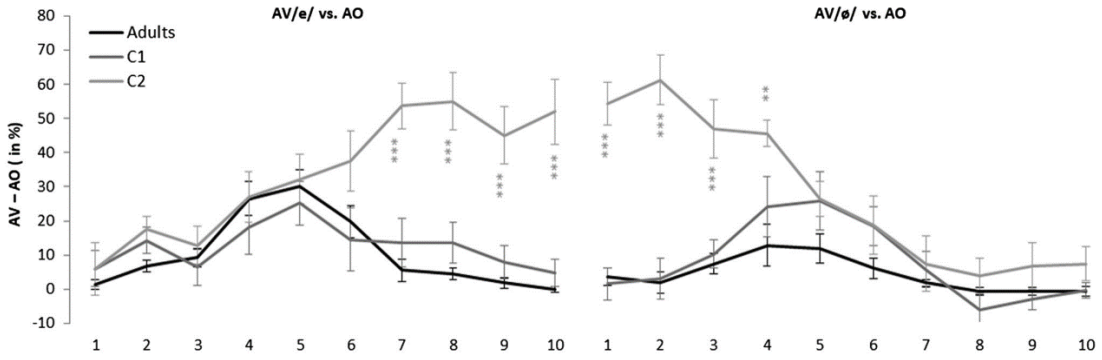

Figure 4.4. Mean visual influence on the categorization of auditory stimuli on the [e] – [ø] continuum across speaker groups and experimental conditions. Error bars indicate standard errors. **p < 0.01 and ***p < 0.001.

Table 4.2. Summary of z values and significance levels of visual influence on the categorization of stimuli 1–10 according to audiovisual condition (cases where no significant difference is found are denoted by the symbol "–").

Stimuli		AV/e/ vs. AO			AV/ø/ vs. AO		
		Adults vs. C1	Adults vs. C2	C1 vs. C2	Adults vs. C1	Adults vs. C2	C1 vs. C2
/e/-like	1	–	–	–	–	9.392, $p < 0.001$	–8.584, $p < 0.001$
	2	–	–	–	–	10.984, $p < 0.001$	–9.18, $p < 0.001$
	3	–	–	–	–	6.950, $p < 0.001$	–5.398, $p < 0.001$
Ambiguous	4	–	–	–	–	4.289, $p < 0.01$	–
	5	–	–	–	–	–	–
	6	–	–	–	–	–	–
	7	–	6.699, $p < 0.001$	–5.253, $p < 0.001$	–	–	–
/ø/–like	8	–	9.018, $p < 0.001$	–6.582, $p < 0.001$	–	–	–
	9	–	8.519, $p < 0.001$	–6.092, $p < 0.001$	–	–	–
	10	–	10.321, $p < 0.001$	–7.327, $p < 0.001$	–	–	–

Overall, Figure 4.4 and Table 4.2 show that Adults, and C1's categorization patterns are quite similar. In fact, no differences were found between the Adult and C1 groups, regardless of AV condition and auditory stimulus. Although C2 classifications form a completely different pattern, it seems that visual inputs affect all three groups in a similar way for stimuli 1–6 (under AV /e/ condition, left-hand panel) and stimuli 5–10 (under AV /ø/ condition, right-hand panel). However, Figure 4.4 clearly shows

that substantial group differences were found and that those differences reflect the congruity of auditory and visual information.

When ambiguous auditory stimuli were perceived, all three groups were affected by visual information. Indeed, as Figure 4.4 shows, large differences in categorization were observed between the AV conditions and the AO condition for stimuli 4–7. Setting aside group C2's particular classification pattern, it can be seen that, compared to C1 (dark gray lines), for whom both AV conditions led to considerable changes, the visual cues barely affected adults' classifications of ambiguous auditory stimuli in the AV/ø/ condition (black line, right-hand panel). As a matter of fact, unlike both groups of children, adults were influenced only by seeing the articulatory movement of the vowel /e/. This pattern is clearly illustrated in Figure 4.3 (left-hand panel) and Table 4.3, where z values corresponding to the differences in categorization between the two AV conditions and the AO condition are presented, this time according to the three experimental groups. The effect of visual information on the categorization of ambiguous stimuli is highlighted in light gray

Table 4.3. Summary of z values and significance levels of visual influence on the categorization of stimuli 1–10.

Stimuli		Adults		Children 1 (C1)		Children 2 (C2)	
		AV/e/ vs. AO	AV/ø/ vs. AO	AV/e/ vs. AO	AV/ø/ vs. AO	AV/e/ vs. AO	AV/ø/ vs. AO
/e/–like	1	–	–	–	–	–	–11.584, $p < 0.001$
	2	–	–	–	–	–	–12.932, $p < 0.001$
	3	–	–	–	–	–	–8,629, $p < 0.001$
Ambiguous	4	6.518, $p < 0.001$	–	–	–3.758, $p < 0.01$	4.205, $p < 0.01$	–7.272, $p < 0.001$
	5	6.571, $p < 0.001$	–	3.547, $p < 0.05$	–3.698, $p < 0.05$	4.803, $p < 0.001$	–3.893, $p < 0.05$
	6	4.471, $p < 0.001$	–	–	–	5.739, $p < 0.001$	–
	7	–	–	–	–	9.490, $p < 0.001$	–
/ø/–like	8	–	–	–	–	11.877, $p < 0.001$	–
	9	–	–	–	–	10.391, $p < 0.001$	–
	10	–	–	–	–	13.210, $p < 0.001$	–

Data are presented in terms of experimental groups (cases where no significant difference is found are denoted by the symbol "–").

While both child groups were affected by both AV conditions when they perceived ambiguous auditory stimuli (stimuli 4–7), Table 4.3 shows that, compared to C1, C2's responses were more influenced by the visual inputs. Furthermore, for C2, visual influence tended to become more pronounced as a function of the distance from the auditory stimulus. Interestingly, this phenomenon also extends to the endpoint stimuli.

Post-hoc tests revealed that C2's responses were also affected when incongruent visual inputs were presented with closeto-endpoint auditory stimuli (highlighted in dark gray in Table 4.3). As Figure 4.3 (right-hand panel) reveals, when /e/-like auditory stimuli (1–3) were presented with the /ø/ visual articulation (dark gray line), children in C2 identified most of the targets as /ø/. Likewise, when /ø/−like auditory stimuli (8–10) were perceived simultaneously with the visual articulation of the vowel /e/ (light gray line), C2 children generally disregarded auditory cues and based their decisions primarily on the visual input. Thus, in contexts where acoustic and visual information were incongruent, the contribution of visual articulatory gestures played a leading role for children in C2 (see Figure 4.4) such that no phonological change was observed over the course of the auditory continuum, regardless of the clarity of the auditory stimulus.

To further explore the dissimilarities between the two groups of children, additional LMEMs were run and confirmed that behavioral differences between the two groups were not due to sex [$\chi 2$ (1)=0.039, p>0.5] or age [$\chi 2$ (1)=0.002, p>0.5]. Moreover, no differences between C1 and C2 were found in the AO condition.

Finally, independent t-tests were performed on the mean standard error of every stimulus, in each of the three experimental conditions. None of the conditions led to significant differences in variability between C1 and C2 responses [AO: t(12.411)=−0.321, p < 0.05]; AV/e/: t(18)=−0.762, p<0.05]; AV/ø/: t(10.854)=0.667,

$p<0.05$]. Moreover, when comparing variability of responses to endpoint stimuli, no significant differences were detected between the two groups of children, regardless of whether the auditory and visual information were compatible [$t(21)=-0.956$, $p>0.05$] or not [$t(21)=-1.264$, $p<0.05$]. Greater variability among C2 children could have been an indication of an attentional bias.

While no variability differences were found between the two child groups, the data show that child participants were generally more variable than adults in both congruent [Adults vs. C1: $t(11.194) =-42.990$, $p < 0.05$, Adults vs. C2: $t(12.406) =-4.423$, $p < 0.001$] and incongruent AV conditions [Adults vs. C1: $t(36) =-39.742$, $p < 0.01$, Adults vs. C2: $t(37) =-5.242$, $p < 0.001$].

4.4 Discussion

The goal of this study was to investigate whether the audiovisual perception of speech differs between preschool-aged children and adults. We chose to investigate the rounding feature, which is the most visually salient, and focused on the perception of the French vowel pair /e/ and /ø/ in 27 adults and 23 preschoolaged children. Our results suggest that the visual influence on auditory perception may occur early in development, but its effect on phonological categorization differs in children and adults. This mechanism likely requires mature sensory systems and sensorimotor representations.

4.4.1 Audiovisual Interaction in Perception

It has been shown that perception is usually facilitated by congruent auditory and visual presentation. Indeed, according to the intersensory redundancy hypothesis, concordance of multiple signals guides attention, reduces RT, and, ultimately,

disambiguates perceptual processing (Bahrick & Lickliter, 2000, 2012). Although, no major impact of visual information was found in our data when auditory and visual information were compatible (probably due to a ceiling effect), Figure 4.3 suggests that endpoint stimuli (stable ones) were classified either similarly or better with congruent visual cues.

Bimodal presentation also facilitates perception when one of the sensory sources is damaged or unstable (Barutchu et al., 2010; Ross et al., 2011). In that case, additional cues help recover the weaker signal. In our study, a similar outcome was observed when optimal visual inputs were perceived simultaneously with ambiguous auditory stimuli. Indeed, although the three experimental groups were not affected to the same extent by the different visual inputs, all participants' categorizations of ambiguous auditory percepts were influenced by the visible articulatory movements. This could mean that the auditory and visual sensory modalities interact in the perception of speech targets in children and adults, either when both sensory sources are compatible or when one of them (in our case, auditory) is imprecise.

Yet a disparity in our results is observed when participants perceived auditory and visual information sources that were incompatible. As shown in Figure 4.3, when adults (left-hand panel) and children in C1 (middle panel) perceived /e/-like stimuli (1–3) combined with the visual input for /ø/ (dark gray lines), the overall percentage of perceived /e/ barely decreased. However, a massive change in overall categorization was observed in C2 (right-hand panel). The same effect was found when /ø/-like auditory stimuli (8–10) were presented with visual articulatory movements for /e/ (light gray lines): trivial changes were observed in adults and C1 while a substantial increase in the perception of /e/ was found in C2.

As Burr and Gori (2012) suggested, the first 8 years of life are critical for brain plasticity. During that period, experiences and comparisons are used to calibrate our

senses in order to benefit from them. According to cross-modal calibration theory, before optimal integration abilities are achieved, the most robust sense prevails over the others (Burr & Gori, 2012). In neurotypical individuals, once brain maturation has occurred, cross-calibration gives way to multisensory facilitation. At that point, when confronted with hard-to-define inputs (ambiguous, noisy, and poor quality), the sturdiest sensory modality has more weight in the perceptual process (Burr & Gori, 2012). In our study, the results of bimodal perception of ambiguous auditory targets in adults and C1 children, for whom overall categorization was influenced by visible speech movements, are in line with this hypothesis. Although the classification patterns of the adult and C1 groups are similar, the fact that the responses of children in C1 were more variable than the adults' may indicate that their cross-modal calibration is still being refined.

Consequently, the children from the C2 group, who based their decisions on visual cues when unrelated bimodal information was presented, may have done so because they used vision as the calibrator in incongruent speech processing. As Burr and Gori (2012) mentioned, the calibrator is not necessarily the most precise sense but the most robust one. Since the vowel targets used in this study contrasted in terms of rounding, and this feature is the most visually salient one in French (Robert-Ribes et al., 1998), it is possible that this particular group of children, who had not yet achieved adult-like MSI, chose to rely on this contextually stronger sensory signal. The results of Kushnerenko et al.'s (2008) study could also be interpreted in that way. They found that 5 month olds processed "auditory /ga/ visual /ba/" as a mismatched signal but not "auditory /ba/ visual /ga/." Because the consonant [b] is produced in a more visually salient way than [g], it could be that infants were more influenced by the visual information when it contained prominent cues.

Of course, multisensory perceptual tasks generally require increased attention. Contrary to the facilitation effect resulting from congruent signals, the increased

difficulty and need for sustained attention is often cited as the reasons for children's poorer scores in atypical multisensory conditions, such as reduced SNR (Alsius et al., 2005; Barutchu et al., 2009; Spence & McDonald, 2004). If greater variability had been identified in C2's responses than in C1's in the incongruent AV presentation (/e/-like auditory + /ø/ visual or /ø/-like auditory + /e/ visual), divergent categorizations between the two groups of children could be explained (at least partially) by an attention bias. However, our comparison of the variability of responses indicated that both groups of children showed similar patterns of variance, in both congruent and incongruent conditions.

Hence, taking into account the relative weight of visual and auditory information for phonemic identification, the crosscalibration hypothesis seems to offer an interesting explanation of why some of the children based their decisions on the visual input, even though audition is considered to be the dominant type of sensory information in speech perception mechanisms. Importantly, the fact that no adult acted similarly to the C2 individuals and that no differences in variability were found between the two groups of children lead us to believe that the differences in categorization are due to a developmental variable, rather than an individual one.

4.4.2 Sensorimotor Maturation

The group difference in terms of audiovisual perception could, however, also reflect the maturation of sensorimotor representations of speech (Caudrelier et al., 2019). Indeed, children with more mature articulatory speech production patterns for certain articulators would attribute more weight to these articulators at the perceptual level.

The link between motor experience, its sensory consequences, and children's still developing feedforward models, while less often applied to the domain of speech development, is well established in the field of motor control, particularly with regard

to arm movements. Children's gestures are known to be slower, less precise, and more inconsistent than adults', due to their lack of sensory experience (Jansen-Osmann et al., 2002; Lambert & Bard, 2005). A recent study of speech motor control maturity in preschoolaged francophone children also acknowledges the role of underspecified sensorimotor maps in explaining inaccurate and unstable predictions of speech motor commands (Barbier et al., 2020). The authors conclude that the onset of maturation and sensorimotor development occurs at around 4 years old.

In our experiment, this could mean that unlike children from the C1 group, children from the C2 group still have incomplete sensorimotor maps. Thus, they might have welldefined articulatory patterns for the lips, but not yet for, say, the tongue. Consequently, they would assign more weight to perceptual information concerning aperture, lip rounding, and lip stretching which is transmitted more prominently through the visual channel. This is in line with the hypothesis that immature sensorimotor representations can interfere with multisensory processing in speech. We believe that this hypothesis is not incompatible with the cross-modal calibration hypothesis and that it may even underlie it.

4.4.3 Limitations of the Study

A potential variable that was not taken into account in this study is between-subject variability in terms of lip-reading skills. Given that this competence evolves and is refined throughout childhood, a precise measurement of lip-reading skills could have provided a more detailed account of the children's perceptual responses. Furthermore, although we chose to use prototypical visual presentations of the /e/ and /ø/ vowels in combination with prototypical and ambiguous auditory stimuli, the use of ambiguous visual information (intermediate degrees of rounding between /e/ and /ø/) in combination with the auditory stimuli could help refine our analyses of children's perceptual responses. Indeed, in cases of visually ambiguous stimuli, reliance on

auditory cues might be enhanced in children as it is in adults. Follow-up studies are currently underway to further investigate these issues.

4.5 Conclusion

This study investigated whether the visual modality influences the auditory perception of the French rounded vowels /e/ and /ø/ in preschool-aged children in the same way as in adults. Our results suggest two distinct patterns of response for children. In the first group, visual influence on auditory perception is similar to (or greater than) that of adults, while in the second group, responses are dominated by the visual content of the stimuli. Thus, substantial individual differences in audiovisual perception still exist at that stage. We suggest that auditory and visual speech perception skills are still developing around that age and that multisensory processing took place only for children whose sensory systems and sensorimotor representations were mature.

CHAPITRE V

DISCUSSION

L'intégration multisensorielle, aussi appelée intégration multimodale, fait référence à la capacité du cerveau à assimiler les signaux provenant de multiples modalités (Molholm et al., 2002; Robert-Ribes et al., 1998; Stein et al., 1996; Stein & Meredith, 1993) et joue un rôle déterminant dans le développement des habiletés de parole (Ito et al., 2009; Lametti et al., 2012; Perrier, 1995; Skipper et al., 2007; S. Tremblay et al., 2003).

La nature multimodale de la parole est reconnue dans les processus de production chez l'adulte et plusieurs modèles de contrôle de la parole accordent une place prédominante aux informations auditives et proprioceptives (voir (Patri, J.-F., 2018)). Au plan perceptuel, il est aussi reconnu que le locuteur adulte exploite les informations auditives et visuelles pour identifier un segment de parole (McGurk & MacDonald, 1976; Robert-Ribes et al., 1998; Rosenblum, 2008a).

Plus récemment, des travaux ont démontré que les informations de type proprioceptif seraient aussi impliquées dans les procédés de perception de la parole (Ito et al., 2009; Yeung & Werker, 2013a), et même que les informations visuelles joueraient un rôle dans les mécanismes de production (Sams et al., 2005; Vidou et al., 2020) suggérant ainsi un transfert des informations liées au système moteur entre les mécanismes de perception et de production.

Bien que les informations auditives occupent généralement une place prédominante dans les habiletés de parole, on sait maintenant que le poids accordé à chacune des rétroactions sensorielles varie d'un locuteur à l'autre et que les systèmes sensoriels fonctionnent, dans une certaine mesure, de façon complémentaire (Katseff et al., 2012; Lametti et al., 2012). Ainsi, un locuteur pour lequel les informations relatives à l'emplacement ressenti des articulateurs (informations proprioceptives) ont une très grande importance dans le contrôle de la parole, accordera moins de poids aux informations auditives.

La multimodalité de la parole est aussi présente chez l'enfant. Lors de la période du babillage, il mettra en relation les informations entendues et ressenties relatives aux sons de sa ou ses langues premières dans le but de former le modèle de contrôle de parole que l'on retrouve chez l'adulte. Alors que ces liens sensori-moteurs sont généralement acquis chez l'enfant dès l'âge de 4 ans, ils continueront à se raffiner pendant l'enfance. Bien que leurs résultats soient plutôt variables, les quelques études de compensation à des manipulations acoustiques ou proprioceptives menées chez l'enfant concluent généralement que ces derniers n'ont pas suffisamment développé de représentations sensorimotrices nécessaires à l'élaboration de nouvelles stratégies articulo-acoustiques ((Gibson & McPhearson, 1980; Oller, D. K. & MacNeilage, 1983) Ménard, Perrier, et al., 2016). Au plan perceptuel, les résultats des études menées chez l'enfant sont aussi très variables et proviennent majoritairement du domaine de la perception audiovisuelle. À ce jour, il demeure donc imprudent de tirer des conclusions sur la complémentarité des systèmes sensoriels et sur le transfert entre les mécanismes de perception et de production de la parole chez l'enfant.

L'objectif de la présente thèse était de rendre compte de la multimodalité de la parole en portant une attention particulière au développement des habiletés d'intégration dans les processus de perception de la parole. Trois expérimentations ont été réalisées auprès d'adultes et d'enfants francophones. Individuellement, elles ont permis 1) de

témoigner du rôle des informations auditives dans le développement du contrôle de la parole, 2) d'observer les procédés d'intégration des informations auditive et somatosensorielle chez l'enfant et l'adulte et 3) de rendre compte de l'évolution des procédés d'intégration des informations auditive et visuelle au sein des mécanismes de perception de la parole.

Les résultats de notre première étude ont montré qu'en réponse à une manipulation de la rétroaction auditive, les enfants d'âge préscolaire ont davantage adapté leurs productions que les adultes. De plus, les enfants et les adultes ont utilisé des stratégies différentes dans le but de répondre à la manipulation perçue, celles adoptées par les adultes étant plus typiques. En ce sens, nous avons conclu que les jeunes individus se fieraient davantage à leur rétroaction auditive pour guider leurs productions et que, pour eux, les informations proprioceptives auraient un rôle de second plan. Pour ces jeunes locuteurs, il semble donc que les liens sensorimoteurs ne soient pas complètement robustes. Le modèle interne de contrôle de la parole serait toujours en cours de développement.

Notre seconde étude portant sur les habiletés d'intégration audioproprioceptives est venue corroborer ces premiers résultats. Nous avons observé que l'impact des informations proprioceptives sur la perception avait été plus important chez les locuteurs adultes que chez les jeunes enfants. Il semble donc que les jeunes individus aient de la difficulté à intégrer les indices de type proprioceptifs dans les tâches de perception de la parole. Chez l'adulte, cet ajout sensoriel a eu pour effet, au contraire, de faciliter la catégorisation des cibles perçues; la frontière perceptuelle entre les cibles perçues étant plus nette en condition audioproprioceptive.

Finalement, dans notre troisième étude, nous avons conclu que les enfants de 5 à 6 ans étaient en mesure de traiter les informations audiovisuelles de façon comparable aux adultes, mais, qu'à ce stade développemental, des différences individuelles

substantielles existaient encore. Nous avons proposé que l'intégration audiovisuelle optimale n'ait eu lieu que pour les enfants dont les systèmes sensoriels et les représentations sensorimotrices étaient bien construits.

La richesse du présent projet réside toutefois dans le fait que les trois expérimentations ont été réalisées auprès des mêmes locuteurs. Ainsi, des analyses supplémentaires ont pu être réalisées afin de déterminer si des liens existaient entre nos trois études.

Considérant le principe de complémentarité des systèmes sensoriels, nous avons d'abord effectué des analyses de corrélation afin d'observer si le poids accordé aux buts auditifs était corrélé à celui accordé aux informations proprioceptives, puis dans un deuxième temps, aux informations visuelles. Afin d'observer si un transfert des informations liées au système moteur s'observerait chez les individus rencontrés, nous avons ensuite cherché à voir si les enfants du groupe C2, pour qui la rétroaction visuelle était particulièrement importante, ont aussi accordé plus de poids aux informations de type proprioceptif. Nous avons donc procédé à une réanalyse des données de la tâche de catégorisation audioproprioceptive en tenant compte, cette fois-ci, des trois groupes formés dans notre étude de perception audiovisuelle.

Pour témoigner du poids des informations acoustiques, nous avons considéré la valeur du ratio moyen de F2' entre la phase de maintien (Hold) et la phase de base (Baseline). Ainsi, un individu pour qui la valeur du ratio moyen est en deçà de 1 aura compensé à la manipulation acoustique. Un individu dont le ratio moyen est au-delà de 1 aura, quant à lui, suivi la manipulation acoustique. Aussi, plus la valeur s'éloigne de 1, plus l'adaptation est importante, peu importe sa direction. Nous avons vu, dans notre première étude, que les enfants avaient davantage adapté leurs productions en réponse à la manipulation acoustique perçue, stipulant par le fait même qu'ils se fiaient davantage à la rétroaction auditive dans les mécanismes de contrôle de la

parole. L'idée, ici, est plutôt de diviser les individus en fonction de l'importance du but auditif dans les mécanismes de parole. À ce titre, les grands compensateurs sont ceux pour qui le but auditif prime dans le contrôle de la parole puisqu'ils ont adapté leurs productions avec l'objectif d'entendre ce qu'ils avaient prévu produire. Les suiveurs (followers), eux, n'ont pas cherché à atteindre la cible vocalique, et s'en sont même éloignés. Ainsi, même s'ils se sont fiés à leur rétroaction auditive pour produire les cibles subséquentes, celle-ci ne prime pas dans le contrôle de la parole.

Pour témoigner du rôle des informations proprioceptives, nous avons calculé la différence moyenne entre la catégorisation de la cible médiane dans les conditions audioproprioceptive et auditive. De ce fait, un individu pour qui la différence entre les deux conditions est très faible aura été peu influencé par les informations somatosensorielles. Au contraire, un individu pour qui la différence entre la catégorisation en contexte auditif et audioproprioceptif est très grande aura accordé un rôle important aux informations de nature proprioceptive.

Le même principe a été utilisé pour mesurer le poids des informations visuelles. Dans un esprit d'uniformité des variables, nous n'avons considéré que la condition audiovisuelle /e/. Ainsi les variables témoignant du poids des informations proprioceptives et visuelles sont propres aux indices référents à la voyelle /e/.

Pour notre première méta-analyse, nous avons mis en relation le rôle du but auditif et des informations proprioceptives afin d'observer si les individus rencontrés montrent une complémentarité de leurs systèmes auditif et proprioceptif. Cette corrélation vient bonifier les analyses de notre deuxième article où l'on discutait plutôt les habiletés d'intégration entre les systèmes auditif et proprioceptif. Les résultats de cette analyse sont présentés à la figure 5.1.

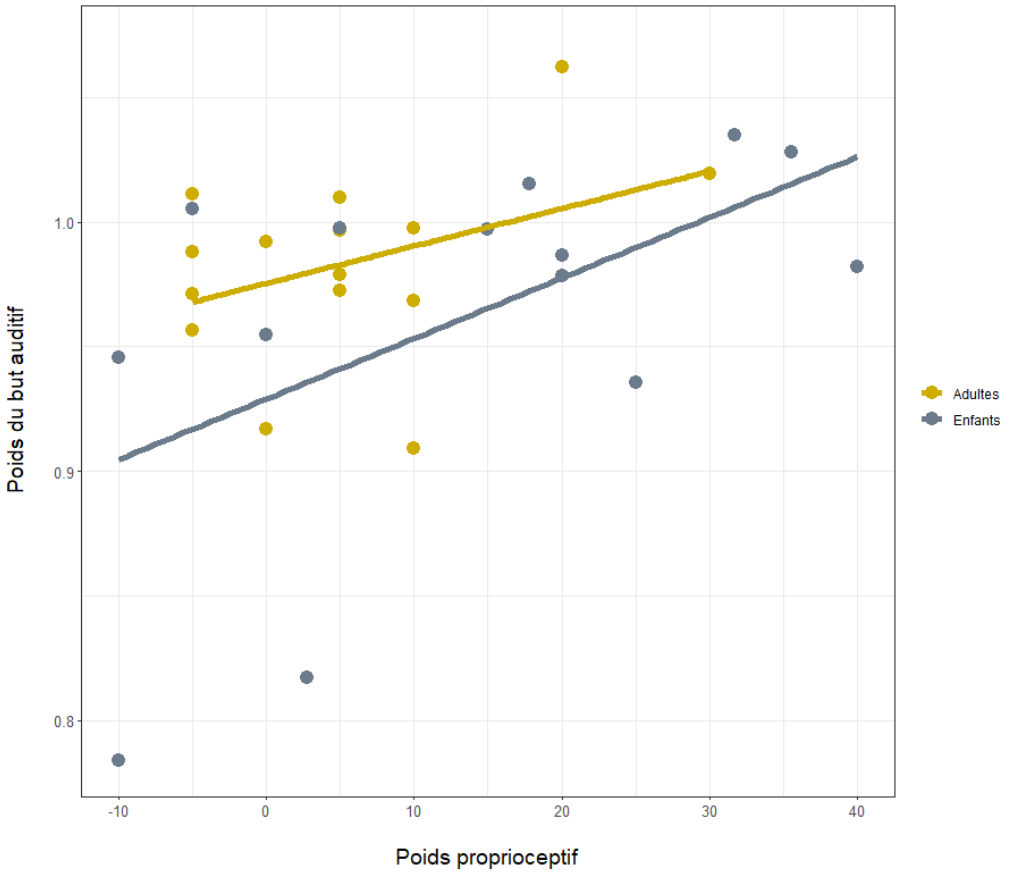

Figure 5.1 Corrélation entre les variables de "Poids du but auditif" et de "poids proprioceptif". Les données sont présentées en fonction des groupes Adultes et Enfants.

Une analyse de corrélation entre ces deux variables a démontré que plus le but auditif prime dans le contrôle de la parole (plus la valeur est en deçà du 1), moins les informations proprioceptives auront un rôle déterminant ($R = 0.38$, $p < .05$). Ainsi, un individu ayant grandement compensé lors de la tâche de manipulation auditive n'aura été que peu ou pas affecté par les indices provenant de la manipulation proprioceptive lors de la tache de perception bimodale. Au contraire, les individus qui n'ont pas cherché à atteindre le but acoustique dans la tâche de manipulation auditive et qui ont ainsi failli à compenser leurs productions ou encore qui ont suivi la manipulation

perçue sont ceux pour qui l'étirement ressenti des lèvres a eu beaucoup d'influence sur la catégorisation des cibles vocaliques.

La variable de groupe s'est aussi avérée déterminante au sein de la corrélation. En effet, une corrélation positive entre les variables de "Poids du but auditif" et de "poids proprioceptif" a aussi été relevée chez les enfants ($R = 0.54$, $p < .05$). Cette observation va de pair avec les résultats de nos deux premières études, lors desquelles nous avons conclu que les enfants accorderaient plus d'importance aux informations auditives, mais moins aux informations proprioceptives.

Dans notre deuxième méta-analyse, nous avons mis en relation le poids du but auditif et celui des informations visuelles afin, encore une fois, de déterminer si une complémentarité des systèmes auditif et visuel est présente chez les individus rencontrés. Cette analyse se voit être un complément à celles effectuées dans notre troisième article où l'on discutait des habiletés d'intégration des informations audiovisuelles. Les résultats de cette seconde corrélation sont présentés à la figure 5.2.

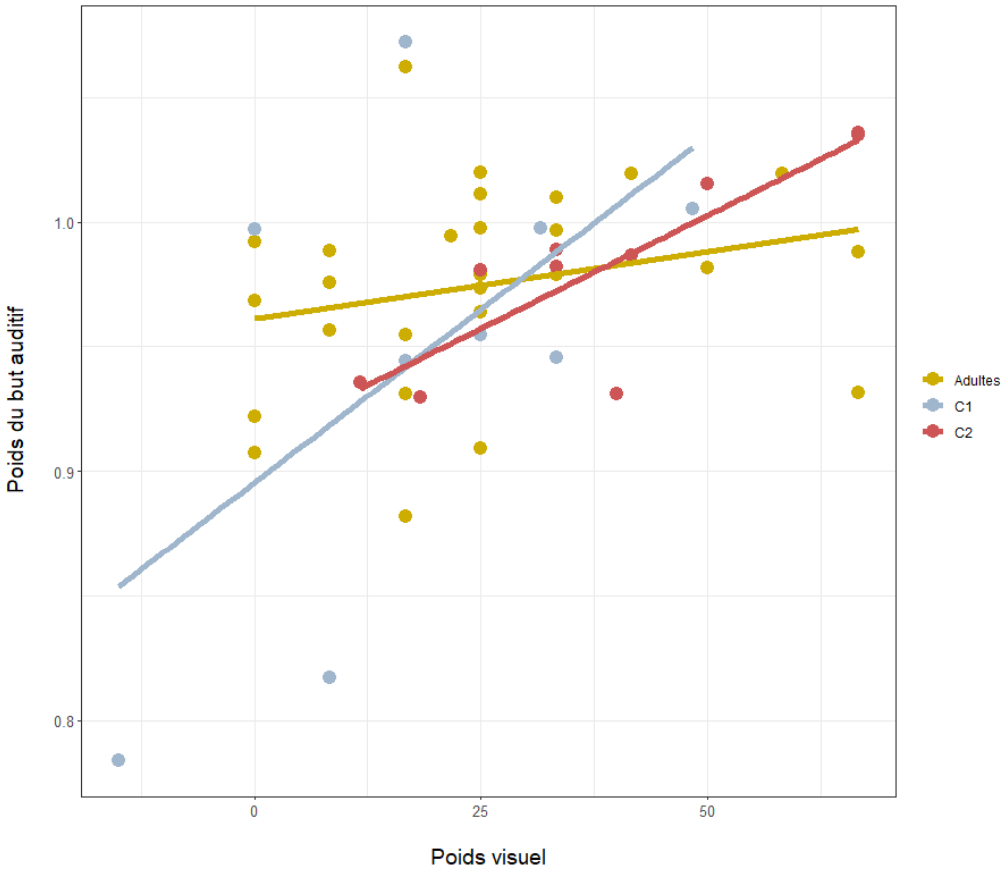

Figure 5.2 Corrélation entre les variables de "Poids auditif" et de "poids visuel". Les données sont présentées en fonction des groupes Adultes, C1 et C2.

Une analyse de corrélation entre ces deux variables a démontré que plus le but acoustique prime dans le contrôle de la parole, moins les informations visuelles auront un rôle déterminant ($R = 0.45$, $p < .01$). En d'autres mots, un individu ayant beaucoup compensé lors de la tâche de manipulation auditive n'aura été que peu affecté par les indices visuels lors de la tache de perception audiovisuelle. Au contraire, les individus pour qui la vision des articulateurs de la parole a beaucoup influencé la catégorisation des sons sont ceux qui n'ont pas cherché à atteindre le but acoustique dans la tâche de manipulation auditive et qui ont failli à compenser leurs productions.

Il est intéressant de noter que cette corrélation est d'autant plus présente chez les jeunes enfants formant le groupe C2 ($R = 0.84$, $p < .01$). À titre de rappel, nous avons proposé que le fait que certains enfants (C2) s'appuient grandement sur les informations visuelles reflétait le fait que, pour eux, les représentations sensorimotrices ne sont pas pleinement formées et qu'ils accorderaient ainsi plus de poids aux informations en lien avec les représentations motrices les plus matures comme, dans ce cas, celles reliées aux lèvres. Nous avons émis l'hypothèse que ce principe de maturation des représentations sensorimotrices puisse même sous-tendre celui de la calibration intermodale (cross-modal calibration). Les représentations motrices liées aux lèvres étant davantage construites, les indices sensoriels y étant associés agiraient à titre de calibrateur dans les processus de perception de la parole.

Les enfants formant le groupe C2 étant ceux pour qui la complémentarité était la plus importante, il semble donc que la maturation des représentations motrices et la calibration des indices sensoriels aient aussi un impact sur le phénomène de complémentarité des systèmes; un rôle plus important serait donné aux représentations les plus matures, soit aux calibrateurs.

Finalement, nous nous sommes intéressés au lien entre le développement des habiletés d'intégration audiovisuelle et audioproprioceptive. Partant du principe que les systèmes visuel et proprioceptif traitent, tous deux, des informations liées au système moteur, nous avons effectué une dernière méta-analyse dont l'objectif était d'observer si ces deux procédés d'intégration bimodale fonctionnent de façon analogue.

Dans notre tâche de perception audiovisuelle, des différences considérables ont été observées au niveau de la catégorisation des cibles vocaliques, particulièrement au sein du groupe d'enfants, menant à la création de deux groupes d'enfants distincts. En ce sens, C1 était des enfants dont les pentes catégorielles traversaient la frontière des

50 % et dont les schémas de classification ressemblaient ainsi à ceux des adultes et C2 était composé des enfants pour lesquels la frontière à 50 % n'était pas franchie dans au moins une des conditions audiovisuelles (AV /e/ ou AV/ø/). Afin d'observer si les enfants des groupes C1 et C2 ont aussi eu des comportements divergents lors de la tâche de perception audio-somato, nous avons choisi de refaire les analyses de catégorisation en considérant, cette fois-ci, les trois groupes utilisés dans la tâche de perception audiovisuelle. Ainsi, chaque réponse obtenue lors de la tâche de perception audio-somato a été intégrée dans un modèle linéaire à effets mixtes où les facteurs fixes comprenaient les stimuli (les mêmes que pour la tâche audiovisuelle, tirés du continuum e-eu), le groupe (Adultes, C1, C2) et la condition (unimodale ou bimodale), et le facteur aléatoire était le participant. Les résultats de cette nouvelle analyse sont illustrés à la figure 5.3.

L'interaction des variables de groupe et de condition s'est avérée significative ($\chi^2(2)=11.071$, $p<0.01$). Des tests posthoc ont revelé que la catégorisation des cibles vocaliques entre les conditions unimodale auditive et audioproprioceptive était significativement différente pour le groupe C2 seulement ($z=3.231$, $p<0.05$). Ainsi, comme lors de la tâche de perception audiovisuelle, les adultes et les enfants formant le groupe C1 ont intégré les informations proprioceptives de façon comparable. L'apport des informations proprioceptives était, par contre, plus important pour le groupe C2.

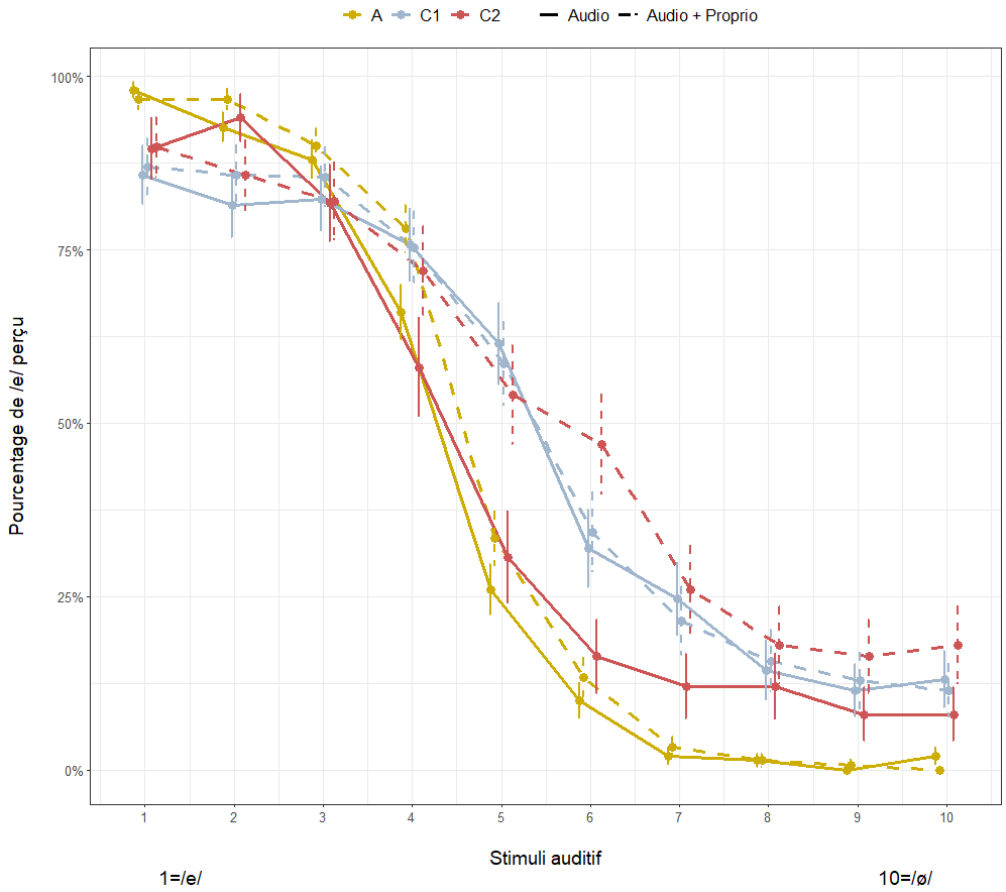

Figure 5.3 Pourcentage d'identification de la voyelle [e] pour les stimuli du continuum [e-ø], dans les deux conditions expérimentales, pour les groupes Adultes, C1 et C2. Les barres d'erreur indiquent les erreurs types.

Encore une fois, le principe de calibration intermodale offre une explication intéressante à ce phénomène. Tel que proposé lors de notre étude sur l'intégration des informations audiovisuelles, et réitéré lors de notre seconde méta-analyse portant sur la complémentarité des systèmes auditif et visuel, pour les enfants formant C2, les représentations sensorimotrices liées aux articulateurs visibles servent de calibrateurs dans les mécanismes de perception de la parole. Ainsi, qu'elles soient perçues via le système visuel (vision des articulateurs de parole) ou le système somatosensoriel (mouvement des articulateurs ressentis), les informations sensorielles renvoyant aux

représentations les plus développées ont eu un rôle plus important dans la catégorisation des cibles phonémiques pour les jeunes locuteurs du groupe C2.

À la lumière des analyses effectuées, il semble que les représentations sensorimotrices soient toujours en cours de maturation chez les individus de 5-6 ans, mettant ainsi un frein aux comportements de compensation auditive et aux habiletés d'intégration multimodale observés chez l'adulte. Il en ressort toutefois que les phénomènes de complémentarité des systèmes sensoriels et de transfert entre les mécanismes de perception et de production en sont facilités.

Il apparait que les représentations motrices progressant le plus rapidement sont celles faisant référence aux lèvres, probablement dû au fait que ce sont des articulateurs visibles. Il est d'ailleurs démontré que l'organisation et le contrôle des lèvres et de la mandibule se développeraient très tôt dans les premières années de vies, bien qu'ils continueraient à se raffiner après l'âge de 6 ans (J. R. Green et al., 2000, 2002). Effectivement, les représentations sensorimotrices des articulateurs visibles sont alimentées, non seulement par les systèmes auditif et proprioceptif, mais aussi par le système visuel, hâtant donc leurs maturations. Le fait que les consonnes labiales soient acquises plus tôt que les palatales ou les vélaires, et ce, peu importe la langue ambiante (Mcleod & Crowe, 2018) et que les locuteurs non-voyants produisent des déplacements des lèvres moins importants que les voyants (Ménard et al, 2013, 2014) laisse à penser que les mécanismes de perception et de production de la parole pourraient aussi être guidés par des buts visuels.

Tel que mentionné plus tôt, lors des premiers mois de vie, les bébés enregistrent les sons perçus et les mouvements des articulateurs qui y sont associés pour former un modèle interne de parole. Ceci dit, les phonèmes visibles ont l'avantage sur les autres. Bien que des travaux supplémentaires devraient être réalisés afin de confirmer cette hypothèse, il pourrait être intéressant d'intégrer le système visuel dans les modèles de

développement de la parole, permettant ainsi de mieux conceptualiser la trimodalité des mécanismes de production de la parole, encore reconnus de manière bimodale et de mieux refléter le transfert entre les systèmes moteurs et perceptuels, particulièrement chez l'enfant.

CONCLUSION

L'objectif de cette thèse était de dresser un portait complet de la nature multimodale de la parole chez l'enfant et l'adulte francophone.

Dans un premier temps, nous avons mis en évidence le rôle des informations auditives dans le contrôle de la parole, puis examiné les procédés d'intégration audioproprioceptive et audiovisuelle dans les mécanismes de perception. Dans un second temps, nous avons mis en relation les résultats précédemment obtenus afin de rendre compte de la complémentarité des systèmes sensoriels et du transfert des représentations somatosensorielles de parole.

Nous concluons que la période entre 5 et 6 est cruciale dans le développement des représentations sensorimotrices. Il apparait qu'à ce stade, les représentations n'ayant pas bénéficié du système visuel sont encore immatures, ce qui mettrait un frein aux comportements compensatoires et aux habiletés d'intégration multimodale qu'il est possible d'observer chez les locuteurs adultes. En contrepartie, cette immaturité des représentations sensorimotrices intensifierait les phénomènes de complémentarité des systèmes sensoriels et de transfert entre les mécanismes de perception et de production de la parole.

BIBLIOGRAPHIE

Abbs, J. H., Gracco, V. L., & Cole, K. J. (1984). Control of multimovement coordination: sensorimotor mechanisms in speech motor programming. *Journal of Motor Behavior*, *16*(2), 195–231. https://doi.org/10.1080/00222895.1984.10735318

Alsius, A., Navarra, J., Campbell, R., & Soto-Faraco, S. (2005). Audiovisual Integration of Speech Falters under High Attention Demands. *Current Biology*, *15*(9), 839–843. https://doi.org/10.1016/j.cub.2005.03.046

Arabin, B. (2002). Music during pregnancy. *Ultrasound in Obstetrics and Gynecology*, *20*(5), 425–430. https://doi.org/10.1046/j.1469-0705.2002.00844.x

Arnaud, L., Sato, M., Ménard, L., & Gracco, V. L. (2013). Repetition Suppression for Speech Processing in the Associative Occipital and Parietal Cortex of Congenitally Blind Adults. *PLoS ONE*, *8*(5), e64553. https://doi.org/10.1371/journal.pone.0064553

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Bahrick, L. E., & Lickliter, R. (2000). Intersensory Redundancy Guides Attentional Selectivity and Perceptual Learning in Infancy. *Developmental Psychology*, *36*(2), 190–201. https://doi.org/0.1037//0012-1649.36.2. 190

Bahrick, L. E., & Lickliter, R. (2012). The role of intersensory redundancy in early perceptual, cognitive, and social development. *Multisensory Development*, 183–205. https://doi.org/10.1093/acprof:oso/9780199586059.003.0008

Barbier, G., Perrier, P., Payan, Y., Tiede, M. K., Gerber, S., Perkell, J. S., & Ménard, L. (2020). What anticipatory coarticulation in children tells us about speech motor control maturity. *PLOS ONE*, *15*(4). https://doi.org/10.1371/journal.pone.0231484

Barutchu, A., Crewther, D. P., & Crewther, S. G. (2009). The race that precedes coactivation: development of multisensory facilitation in children. *Developmental Science*, *12*(3), 464–473. https://doi.org/10.1111/j.1467-7687.2008.00782.x

Barutchu, A., Danaher, J., Crewther, S. G., Innes-Brown, H., Shivdasani, M. N., & Paolini, A. G. (2010). Audiovisual integration in noise by children and adults. *Journal of Experimental Child Psychology*, *105*(1–2), 38–50. https://doi.org/10.1016/j.jecp.2009.08.005

Baum, S. R., & Katz, W. F. (1988). Acoustic analysis of compensatory articulation in children. *Journal of the Acoustical Society of America*, *84*(5), 1662–1668.

Birch, H. G., & Lefford, a. (1963). Intersensory Development in Children. *Monographs of the Society for Research in Child Development*, *28*(5), 1–47. https://doi.org/10.2307/1165681

Boe, L.-J., Perrier, P., Guérin, B., & Schwartz, J.-L. (1989). Maximal vowel space. *European Conference on Speech Communication and Technology (Eurospeech)*, 281–284.

Bower, T. G., Broughton, J. M., & Moore, M. K. (1970). The coordination of visual and tactual input in infants. *Perception & Psychophysics*, *8*, 51–53.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. The MIT Press. https://psycnet.apa.org/record/1990-98046-000

Browman, C. P., & Goldstein, L. M. (1986). Towards an articulatory phonology. *Phonology Yearbook*, *3*, 219–252. https://doi.org/10.1017/s0952675700000658

Bruderer, A. G., Danielson, D. K., Kandhadai, P., & Werker, J. F. (2015). Sensorimotor influences on speech perception in infancy. *Proceedings of the National Academy of Sciences*, *112*(44), 13531–13536. https://doi.org/10.1073/pnas.1508631112

Burnett, T. A., Freedland, M. B., Larson, C. R., & Hain, T. C. (1998). Voice F0 responses to manipulations in pitch feedback. *The Journal of the Acoustical Society of America*, *103*(6), 3153–3161. https://doi.org/10.1121/1.423073

Burnham, D., & Dodd, B. (2004). Auditory-visual speech integration by prelinguistic infants: Perception of an emergent consonant in the McGurk effect. *Developmental Psychobiology*, *45*(4), 204–220. https://doi.org/10.1002/dev.20032

Burr, D., & Gori, M. (2012). Multisensory Integration Develops Late in Humans. In M. M. Murray & M. T. Wallace (Eds.), *The Neural Bases of Multisensory Processes* (p. 810). CRC Press. www.ncbi.nlm.nih.gov/books/NBK92864/

Buss, E., Hall, J. W., & Grose, J. H. (2009). Psychometric functions for pure tone intensity discrimination: Slope differences in school-aged children and adults. *Citation: The Journal of the Acoustical Society of America*, *125*, 1050. https://doi.org/10.1121/1.3050273

Cai, S., Boucek, M., Ghosh, S., Guenther, F., & Perkell, J. (2008). A system for online dynamic perturbation of formant frequencies and results from perturbation of the Mandarin triphthong /iau/. *Proceedings of the 8th Intl. Seminar on Speech Production*, 65–68.

Callan, D. E., Kent, R. D., Guenther, F. H., & Vorperian, H. K. (2000). An Auditory-Feedback-Based Neural Network Model of Speech Production That Is Robust to Developmental Changes in the Size and Shape of the Articulatory System. *Journal of Speech, Language, and Hearing Research*, *43*(3), 721–736. https://doi.org/10.1044/jslhr.4303.721

Carlson, R., Granström, B., & Fant, B. (1970). *Some studies concerning perception of isolated vowels*.

Caudrelier, T., Ménard, L., Perrier, P., Schwartz, J. L., Gerber, S., Vidou, C., & Rochet-Capellan, A. (2019a). Transfer of sensorimotor learning reveals phoneme representations in preliterate children. *Cognition*, *192*. https://doi.org/10.1016/j.cognition.2019.05.010

Caudrelier, T., Ménard, L., Perrier, P., Schwartz, J.-L., Gerber, S., Vidou, C., & Rochet-Capellan, A. (2019b). Transfer of sensorimotor learning reveals phoneme representations in preliterate children. *Cognition*, *192*, 103973. https://doi.org/10.1016/j.cognition.2019.05.010

Chistovich, L. A., & Lublinskaya, V. V. (1979). The "center of gravity" effect in vowel spectra and critical distance between the formants: Psychoacoustical study of the perception of vowel-like stimuli. *Hearing Research*, *1*(3), 185–195. https://doi.org/10.1016/0378-5955(79)90012-1

Contreras-Vidal, J. L., Bo, J., Boudreau, J. P., & Clark, J. E. (2005). Development of visuomotor representations for hand movement in young children. *Experimental Brain Research*, *162*(2), 155–164. https://doi.org/10.1007/s00221-004-2123-7

Daliri, A., Wieland, E. A., Cai, S., Guenther, F. H., & Chang, S. E. (2018). Auditory-motor adaptation is reduced in adults who stutter but not in children who stutter. *Developmental Science*, *21*(2). https://doi.org/10.1111/desc.12521

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. https://doi.org/10.1016/j.cub.2009.01.017

Demany, L., McKenzie, B., & Vurpillot, E. (1977). Rhythm perception in early infancy. *Nature*, *266*(5604), 718–719.

Depaolis, R. A., Vihman, M. M., & Keren-Portnoy, T. (2011). *Do production patterns influence the processing of speech in prelinguistic infants?* https://doi.org/10.1016/j.infbeh.2011.06.005

Derrick, D., Madappallimattam, J., & Theys, C. (2019). Aero-tactile integration during speech perception: Effect of response and stimulus characteristics on syllable identification. *The Journal of the Acoustical Society of America*, *146*(3), 1605–1614. https://doi.org/10.1121/1.5125131

Desjardins, R. N., Rogers, J., & Werker, J. F. (1997). An Exploration of Why Preschoolers Perform Differently Than Do Adults in Audiovisual Speech Perception Tasks. *Journal of Experimental Child Psychology*, *66*(1), 85–110. https://doi.org/10.1006/jecp.1997.2379

Desjardins, R. N., & Werker, J. F. (2004). Is the integration of heard and seen speech mandatory for infants? *Developmental Psychobiology*, *45*(4), 187–203. https://doi.org/10.1002/dev.20033

Dette, M., & Linke, P. G. (1982). The development of oral and manual stereognosis in children from 3 to 10 years old (Die Entwicklung der oralen und manuellen Stereognose bei Kindem im Altervon 3 bis 10 Jahren). *Stomatologie*, *32*, 269–274.

Di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Experimental Brain Research*, *91*(1), 176–180. https://doi.org/10.1007/BF00230027

Diehl, R. L., & Kluender, K. R. (1989). On the Objects of Speech Perception. *Ecological Psychology*, *1*(2), 121–144. https://doi.org/10.1207/s15326969eco0102_2

Diehl, R. L., Lotto, A. J., & Holt, L. L. (2004). Speech Perception. *Annual Review of Psychology*, *55*(1), 149–179. https://doi.org/10.1146/annurev.psych.55.090902.142028

Dionne-Dostie, E., Paquette, N., Lassonde, M., & Gallagher, A. (2015). Multisensory Integration and Child Neurodevelopment. *Brain Sciences*, *5*(1), 32–57. https://doi.org/10.3390/brainsci5010032

Dodd, B. (1979). Lip reading in infants: Attention to speech presented in- and out-of-synchrony. *Cognitive Psychology*, *11*(4), 478–484. https://doi.org/https://doi.org/10.1016/0010-0285(79)90021-5

Dohen, M., & Loevenbruck, H. (2009). Interaction of audition and vision for the perception of prosodic contrastive focus. *Language and Speech*, *52*(0), 177–206. https://doi.org/10.1177/0023830909103166

Dominic, I., & Massaro, W. (1999). Speechreading: Illusion or window into pattern recognition. *Trends in Cognitive Sciences*, *3*(8), 310–317. https://doi.org/10.1016/S1364-6613(99)01360-1

Dupont Jérôme Aubin Lucie Ménard, S. (2005). A study of the McGurk effect in 4 and 5-year-old French Canadian children. In *Papers in Linguistics* (Vol. 40).

Eimas, P. D., Siqueland, Einar, R., Jusczyk, Peter., & Vigorito, J. (1971). Speech perception in Infants. *Science*, *171*(3968), 303–306.

Elman, J. L. (1981). Effects of frequency-shifted feedback on the pitch of vocal productions. *The Journal of the Acoustical Society of America*, *70*(1), 45–50. https://doi.org/10.1121/1.386580

Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: A TMS study. *European Journal of Neuroscience*, *15*(2), 399–402. https://doi.org/10.1046/j.0953-816x.2001.01874.x

Fadiga, L., Fogassi, L., Pavesi, G., & Rizzolatti, G. (1995). Motor facilitation during action observation: A magnetic stimulation study. *Journal of Neurophysiology*, *73*(6), 2608–2611. https://doi.org/10.1152/jn.1995.73.6.2608

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. https://doi.org/10.1152/jn.00638.2010

Fitch, W. T., & Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *The Journal of the Acoustical Society of America*, *106*(3), 1511–1522. https://doi.org/10.1121/1.427148

Fortin, M., Voss, P., Lassonde, M., & Lepore, F. (2007). Sensory loss and brain reorganization. In *Medecine/Sciences* (Vol. 23, Issue 11, pp. 917–922). Editions EDK. https://doi.org/10.1051/medsci/20072311917

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct–realist perspective. *Journal of Phonetics*, *14*(1), 3–28. https://doi.org/10.1016/s0095-4470(19)30607-2

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. https://doi.org/10.1037/0096-1523.17.3.816

Foxe, J. J., Wylie, G. R., Martinez, A., Schroeder, C. E., Javitt, D. C., Guilfoyle, D., Ritter, W., & Murray, M. M. (2002). Auditory-somatosensory multisensory processing in auditory association cortex: an fMRI study. *Journal of Neurophysiology*, *88*(1), 540–543. https://doi.org/DOI 10.1152/jn.00694.2001

Gibson, A., & McPhearson, L. (1980). Production of bite-block vowels by children. *PERLUS 2*, 26–43.

Gick, B., & Derrick, D. (2009). Aero-tactile integration in speech perception. *Nature*, *462*(7272), 502–504. https://doi.org/10.1038/nature08572

Gisel, E. G., & Schwob, H. (1988). Oral Form Discrimination in Normal 5-to 8-Year-Old Children: An Adjunct to an Eating Assessment. *The Occupational Therapy Journal of Research*, *8*(4), 195–209. https://doi.org/https://doi.org/10.1177/153944928800800401

Goldenberg, D., Tiede, M. K., & Whalen, D. H. (2015). Aero-tactile influence on speech perception of voicing continua. *International Congress of Phonetic Sciences*.

Göllesz, V. (1972). Über die Lippen Artikulation der von Geburt an Blinden. *Interdisciplinary Speech Research: Proceedings of the Symposium*.

Gori, M. (2015). Multisensory Integration and Calibration in Children and Adults with and without Sensory and Motor Disabilities. *Multisensory Research*, *28*(1–2), 71–99. https://doi.org/10.1163/22134808-00002478

Gori, M., del Viva, M., Sandini, G., & Burr, D. C. (2008). Young Children Do Not Integrate Visual and Haptic Form Information. *Current Biology*, *18*(9), 694–698. https://doi.org/10.1016/j.cub.2008.04.036

Gougoux, F., Lepore, F., Lassonde, M., Voss, P., Zatorre, J. R., & Belin, P. (2004). Pitch discrimination in the early blind. *Nature*, *430*, 309–310.

Grant, K. W., & Seitz, P.-F. (2000). The use of visible speech cues for improving auditory detection of spoken sentences. *Citation: The Journal of the Acoustical Society of America*, *108*, 1197. https://doi.org/10.1121/1.1288668

Green, J. R., Moore, C. A., Higashikawa, M., & Steeve, R. W. (2000). The Physiologic Development of Speech Motor Control: Lip and Jaw Coordination. *Journal of Speech, Language & Hearing Research*, *43*, 239–255.

Green, J. R., Moore, C. A., & Railly, K. J. (2002). The Sequential Development of Jaw and Lip Control for Speech. *Journal of Speech, Language & Hearing Research*, *45*.

Green, K. P., Kuhl, P., Meltzoff, A. N., & Stevens, E. B. (1991). Integrating speech information across talkers, gender, and sensory modality: Female faces and male voices in the McGurk effect A par. In *Perception &: Psychophysics* (Vol. 50, Issue 6).

Guenther, F. H. (1994). A neural network model of speech acquisition and motor equivalent speech production. *Biological Cybernetics*, *72*(1), 43–53. https://doi.org/10.1007/BF00206237

Guenther, F. H., & Perkell, J. S. (2004). A neural model of speech production and its application to studies of the role of auditory feedback in speech. *Speech Motor Control in Normal and Disordered Speech*, *02*, 29–49.

Guenther, F. H., & Vladusich, T. (2012). A neural theory of speech acquisition and production. *Journal of Neurolinguistics*, *25*(5), 408–422. https://doi.org/10.1016/j.jneuroling.2009.08.006

Guest, S., Catmur, C., Lloyd, D., & Spence, C. (2002). Audiotactile interactions in roughness perception. *Experimental Brain Research*, *146*(2), 161–171. https://doi.org/10.1007/s00221-002-1164-z

Hatwell, Y. (1987). Motor an Cognitive Functions of the Hand in Infancy and Chilhood. *International Jornal of Behavioral Development*, *10*(4), 509–526.

Hecht, D., & Reiner, M. (2009). Sensory dominance in combinations of audio, visual and haptic stimuli. *Experimental Brain Research*, *193*(2), 307–314. https://doi.org/10.1007/s00221-008-1626-z

Held, R. (1965). Plasticity in sensory-motor systems. *Scientific America*, *123*(5), 84–94.

Hillock, A. R., Powers, A. R., & Wallace, M. T. (2011). Binding of sights and sounds: Age-related changes in multisensory temporal processing. *Neuropsychologia*, *49*(3), 461–467. https://doi.org/10.1111/j.1743-6109.2008.01122.x.Endothelial

Hockley, N. S., & Polka, L. (1994). A developmental study of audiovisual speech perception using the McGurk paradigm. *The Journal of the Acoustical Society of America*, *96*(5), 3309–3309. https://doi.org/10.1121/1.410782

Holst-Wolf, J. M., Yeh, I.-L., Konczak, J., Seidler, R. D., Cheyne, D. O., & Farne, A. (2016). Development of Proprioceptive Acuity in Typically Developing Children: Normative Data on Forearm Position Sense. *Frontiers in Human Neuroscience*, *10*(August), 1–8. https://doi.org/10.3389/fnhum.2016.00436

Horlyck, S., Reid, A., & Burnham, D. (2012). The relationship between learning to read and language-specific speech perception: Maturation versus experience. *Scientific Studies of Reading*, *16*(3), 218–239.

Hotting, K., & Roder, B. (2004). Hearing Cheats Touch, but Less in Congenitally Blind Than in Sighted Individuals. *Psychological Science*, *15*(1), 60–64. https://doi.org/10.1111/j.0963-7214.2004.01501010.x

Houde, J. F., & Jordan, M. I. (1998). Sensorimotor Adaptation in Speech Production. *Science*, *279*(5354), 1213–1216. https://doi.org/10.1126/science.279.5354.1213

Ito, T., & Gomi, H. (2007). Cutaneous mechanoreceptors contribute to the generation of a cortical reflex in speech. *NeuroReport*, *18*(9), 907–910. https://doi.org/10.1097/WNR.0b013e32810f2dfb

Ito, T., Gracco, V. L., & Ostry, D. J. (2014). Temporal factors affecting somatosensory-auditory interactions in speech processing. *Frontiers in Psychology*, *5*(OCT), 1–10. https://doi.org/10.3389/fpsyg.2014.01198

Ito, T., & Ostry, D. J. (2010). Somatosensory contribution to motor learning due to facial skin deformation. *Journal of Neurophysiology*, *104*(3), 1230–1238. https://doi.org/10.1152/jn.00199.2010

Ito, T., & Ostry, D. J. (2012). Speech sounds alter facial skin sensation. *Journal of Neurophysiology*, *107*(1), 442–447. https://doi.org/10.1152/jn.00029.2011

Ito, T., Tiede, M., & Ostry, D. J. (2009). Somatosensory function in speech perception. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(4), 1245–1248. https://doi.org/10.1073/pnas.0810063106

Jansen-Osmann, P., Richter, S., Konczak, J., & Kalveram, K.-T. (2002). Force adaptation transfers to untrained workspace regions in children. *Exp Brain Res*, *143*, 212–220. https://doi.org/10.1007/s00221-001-0982-8

Johansson, R. S., Trulsson, M., Olssor, K. A., & Abbs, J. H. (1988). Mechanoreceptive afferent activity in the infraorbital nerve in man during speech and chewing movements. *Experimental Brain Research*, *72*, 209–214. https://doi.org/10.1007/BF00248519

Jones, J. A., & Munhall, K. G. (2000). Perceptual calibration of F0 production: Evidence from feedback perturbation. *The Journal of the Acoustical Society of America*, *108*(3), 1246. https://doi.org/10.1121/1.1288414

Jones, J. A., & Munhall, K. G. (2002). The role of auditory feeback during phonation: Studies of Mandarin tone production. *Journal of Phonetics*, *30*(3), 303–320. https://doi.org/10.1006/jpho.2001.0160

Jones, J. A., & Munhall, K. G. (2003). Learning to produce speech with an altered vocal tract: The role of auditory feedback. *The Journal of the Acoustical Society of America*, *113*(1), 532–543. https://doi.org/10.1121/1.1529670

Jones, J. A., & Munhall, K. G. (2005). Remapping auditory-motor representations in voice production. *Current Biology*, *15*(19), 1768–1772. https://doi.org/10.1016/j.cub.2005.08.063

Jousmäki, V., & Hari, R. (1998). Parchment-skin illusion: sound-biased touch. *Current Biology*, *8*(6), R190. https://doi.org/10.1016/S0960-9822(98)70120-4

Kagerer, F. A., & Clark, J. E. (2014). Development of interactions between sensorimotor representations in school-aged children. *Human Movement Science*, *34*(1), 164–177. https://doi.org/10.1016/j.humov.2014.02.001

Katseff, S., Houde, J. F., & Johnson, K. (2012). Partial Compensation for Altered Auditory Feedback: A Tradeoff with Somatosensory Feedback? *Language and Speech*, *55*(2), 295–308. https://doi.org/10.1177/0023830911417802

Kawahara, H. (1998). Hearing voice: Transformed auditory feedback effects on voice pitch control. In *Computational auditory scene analysis.* (pp. 335–349).

Kelso, J. A. S., Tuller, B., Vatikiotis-Bateson, E., & Fowler, C. A. (1984). Functionally specific articulatory cooperation following jaw perturbations during speech: evidence for coordinative structures. *Journal of Experimental Psychology. Human Perception and Performance*, *10*(6), 812–832. https://doi.org/10.1037/h0090451

Knowland, V. C. P., Mercure, E., Karmiloff-Smith, A., Dick, F., & Thomas, M. S. C. (2014). Audio-visual speech perception: a developmental ERP investigation. *Developmental Science*, *17*(1), 110–124. https://doi.org/10.1111/desc.12098

Krakauer, J. W., Mazzoni, P., Ghazizadeh, A., Ravindran, R., & Shadmehr, R. (2006). Generalization of motor learning depends on the history of prior action. *PLoS Biology*, *4*(10), 1798–1808. https://doi.org/10.1371/journal.pbio.0040316

Krueger, L. E. (1970). David Katz's Der Aufbau der Tastwelt (The world of touch): A synopsis. *Perception & Psychophysics*, *7*(6), 337–341. https://doi.org/10.3758/BF03208659

Kuhl, P. (1991). Human adults and human infants show a perceptual magnet effect for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, *50*(2), 93–107.

Kuhl, P., & Meltzoff, A. (1982). The Bimodal Perception of Speech in Infancy. *Science*, *218*(4577), 1138–1141.

Kuhl, P., & Meltzoff, A. N. (1996). Infant vocalizations in response to speech: Vocal imitation and developmental change. *The Journal of the Acoustical Society of America*, *100*(4), 2425–2438. https://doi.org/10.1121/1.417951

Kujala, T., Alho, K., & Näätänen, R. (2000). Cross-modal reorganization of human cortical functions. In *Trends in Neurosciences* (Vol. 23, Issue 3, pp. 115–120). https://doi.org/10.1016/S0166-2236(99)01504-0

Kumin, L. B., Saltysiak, E. B., Bell, K., Forget, K., Goodman, M. S., Goytisolo, M., Padden, J. V., Schroeter, N. C., & Thomas, S. (1984). Relationships of Oral Stereognostic Ability to Age and Sex of Children. *Perceptual and Motor Skills*, *59*(1), 123–126. https://doi.org/10.2466/pms.1984.59.1.123

Kushnerenko, E., Teinonen, T., Volein, A., & Csibra, G. (2008). Electrophysiological evidence of illusory audiovisual speech percept in human infants. *Proceedings of the National Academy of Sciences*, *105*(32), 11442–11445. https://doi.org/10.1073/pnas.0804275105

Lambert, J., & Bard, C. (2005). Acquisition of visuomanual skills and improvement of information processing capacities in 6- to 10-year-old children performing a 2D pointing task. *Neuroscience Letters*, *377*(1), 1–6. https://doi.org/10.1016/j.neulet.2004.11.058

Lametti, D. R., Nasir, S. M., & Ostry, D. J. (2012). Sensory Preference in Speech Production Revealed by Simultaneous Alteration of Auditory and Somatosensory Feedback. *Journal of Neuroscience*, *32*(27), 9351–9358. https://doi.org/10.1523/JNEUROSCI.0404-12.2012

Lametti, D. R., Rochet-Capellan, A., Neufeld, E., Shiller, D. M., & Ostry, D. J. (2014). Plasticity in the human speech motor system drives changes in speech perception. *Journal of Neuroscience*, *34*(31), 10339–10346. https://doi.org/10.1523/JNEUROSCI.0108-14.2014

Lane, H., Denny, M., Guenther, F. H., Matthies, M. L., Menard, L., Perkell, J. S., Stockmann, E., Tiede, M., Vick, J., & Zandipour, M. (2005). Effects of bite blocks and hearing status on vowel production. *The Journal of the Acoustical Society of America*, *118*(3), 1636–1646. https://doi.org/10.1121/1.2001527

Lane, H., & Wozniak Webster, J. (1998). *Speech deterioration in postlingually deafened adults*.

Larson, C. R., Altman, K. W., Liu, H., & Hain, T. C. (2008). Interactions between auditory and somatosensory feedback for voice F 0 control. *Experimental Brain Research*, *187*(4), 613–621. https://doi.org/10.1007/s00221-008-1330-z

Lecanuet, J.-P., Granier-Deferre, C., & Busnel, M.-C. (1995). Human fetal auditory perception. In J.-P. Lecanuet, W. P. Fifer, N. A. Krasnegor, & W. P. Smotherman (Eds.), *Fetal development : a psychobiological perspective* (p. 512). Lawrence Erlbaum Associates, Inc.

Leclerc, A. (2007). *Le rôle de la vision dans la production de la parole*. Université du Québec à Montréal.

Legerstee, M. (1990). Infants use multimodal information to imitate speech sounds. *Infant Behavior and Development*, *13*(3), 343–354. https://doi.org/10.1016/0163-6383(90)90039-B

Lessard, N., Paré, M., Lepore, F., & Lassonde, M. (1998). Early-blind human subjects localize sound sources better than sighted subjects. *Nature*, *395*(6699), 278–280. https://doi.org/10.1038/26228

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*(6), 431–461. https://doi.org/10.1037/h0020279

Liljencrants, J., Lindblom, B., & Lindblom, B. (1972). Numerical Simulation of Vowel Quality Systems: The Role of Perceptual Contrast. *Language*, *48*(4), 839. https://doi.org/10.2307/411991

Lindblom, B. (1986). Phonetic Universals in Vowel Systems. In J. Ohala & J. Jaeger (Eds.), *Experimental phonology* (pp. 13–44). Academic Press.

Lindblom, B. (1990). Explaining Phonetic Variation: A Sketch of the H&H Theory. In *Speech Production and Speech Modelling* (pp. 403–439). Springer Netherlands. https://doi.org/10.1007/978-94-009-2037-8_16

Lisker, L., & Rossi, M. (1992). Auditory and Visual Cueing of the [±Rounded] Feature Of Vowels. *Http://Dx.Doi.Org/10.1177/002383099203500402*, *35*(4), 391–417. https://doi.org/10.1177/002383099203500402

Liu, H., Russo, N. M., & Larson, C. R. (2010). Age-related differences in vocal responses to pitch feedback perturbations: A preliminary study. *The Journal of the Acoustical Society of America*, *127*(2), 1042–1046. https://doi.org/10.1121/1.3273880

Lotto, A. J., & Holt, L. L. (2015). Speech Perception: The View from the Auditory System. In *Neurobiology of Language* (pp. 185–194). Elsevier Inc. https://doi.org/10.1016/B978-0-12-407794-2.00016-X

Macaluso, E. (2000). Modulation of Human Visual Cortex by Crossmodal Spatial Attention. *Science (New York, NY)*, *289*(5482), 1206–1208. https://doi.org/10.1126/science.289.5482.1206

MacDonald, E. N., Goldberg, R., & Munhall, K. G. (2010). Compensations in response to real-time formant perturbations of different magnitudes. *The Journal of the Acoustical Society of America*, *127*(2), 1059–1068. https://doi.org/10.1121/1.3278606

MacDonald, E. N., Johnson, E. K., Forsythe, J., Plante, P., & Munhall, K. G. (2012). Children's development of self-regulation in speech production. *Current Biology*, *22*(2), 113–117. https://doi.org/10.1016/j.cub.2011.11.052

MacDonald, E. N., Purcell, D. W., & Munhall, K. G. (2011). Probing the independence of formant control using altered auditory feedback. *The Journal of the Acoustical Society of America*, *129*(2), 955–965. https://doi.org/10.1121/1.3531932

Macleod, A., & Summerfield, Q. (1987). Quantifying the contribution of vision to speech perception in noise. *British Journal of Audiology*, *21*(2), 131–141. https://doi.org/10.3109/03005368709077786

Macpherson, A., & Akeroyd, M. A. (2014). Variations in the slope of the psychometric functions for speech intelligibility: A systematic survey. *Trends in Hearing*, *18*. https://doi.org/10.1177/2331216514537722

Maeda, S. (1990). Compensatory Articulation During Speech: Evidence from the Analysis and Synthesis of Vocal-Tract Shapes Using an Articulatory Model. *Speech Production and Speech Modelling*, 131–149. https://doi.org/10.1007/978-94-009-2037-8_6

Mantakas, M. (1989). *Application du second formant effectif F'2 à l'étude de l'opposition d'arrondissement des voyelles antérieures du francais*. Institut National Polythechnique de Grenoble.

Massaro, D. W. (1984). Children's Perception of Visual and Auditory Speech. *Child Development*, *55*(5), 1777–1788.

Massaro, D. W., Cohen, M. M., & Smeele, P. M. T. (1996). Perception of asynchronous and conflicting visual and auditory speech. *The Journal of the Acoustical Society of America*, *100*(3), 1777–1786. https://doi.org/10.1121/1.417342

Massaro, D. W., Thomson, L. A., Barron, B., & Laren, E. (1986). Developmental Changes in Visual and Auditory Contributions to Speech Perception. *Journal of Experimental Child Psychology*, *41*, 93–113.

Max, L., Wauacet, M. E., & Vincene, I. (2003). Sensorimotor adaptation to auditory perturbations during speech : Acoustic and kinematic experiments. *Proceedings of the 15th International Congress of Phonetic Sciences*, 1053–1056.

McDonald, E. T., & Aungst, L. F. (1967). Studies in oral sensorimotor function. In C. C. Thomas (Ed.), *Symposium on Oral Sensation and Perception* (pp. 202–220).

McFarland, D. H., & Baum, S. R. (1995). Incomplete compensation to articulatory perturbation. *Journal of the Acoustical Society of America*, *97*(3), 1865–1873.

McGrath, M., & Summerfield, Q. (1985). Intermodal timing relations and audio-visual speech recognition by normal-hearing adults. *Journal of the Acoustical Society of America*, *77*(2), 678–685. https://doi.org/10.1121/1.392336

McGurk, H., & Power, R. P. (1980). Intermodal coordination in young children: Vision and touch. *Developmental Psychology*, *16*(6), 679–680. https://doi.org/10.1037/0012-1649.16.6.679

McGurk, & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746–748. https://doi.org/10.1038/264746a0

Mcleod, S., & Crowe, K. (2018). Children's Consonant Acquisition in 27 Languages: A Cross-Linguistic Review. *American Journal of Speech-Language Pathology*, *27*(4), 1546–1571.

Meltzoff, A., & Moore, M. (1983). Newborn infants imitate adult facial gestures. *Child Development*, *54*, 702–709.

Ménard, L. (2013). Sensorimotor constraints and the organization of sound patterns. In C. Lefebvre, B. Comrie, & H. Cohen (Eds.), *New Perspectives on the Origins of Language, eds C. Lefebvre, B. Comrie and H. Cohen (Amsterdam: John Benjamins)*. John Benjamins Pubicating Compagny.

Ménard, L. (2015). Multimodal Speech Production. In *The Handbook of Speech Production* (pp. 200–221). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118584156.ch10

Ménard, L., Côté, D., & Trudeau-Fisette, P. (2017). Maintaining distinctiveness at increased speaking rates: A comparison between congenitally blind and sighted speakers. *Folia Phoniatrica et Logopaedica*, *68*(5), 232–238. https://doi.org/10.1159/000470905

Ménard, L., Dupont, S., Baum, S. R., & Aubin, J. (2009). Production and perception of French vowels by congenitally blind adults and sighted adults. *The Journal of the Acoustical Society of America*, *126*(3), 1406–1414. https://doi.org/10.1121/1.3158930

Ménard, L., Leclerc, A., & Tiede, M. (2014). Articulatory and acoustic correlates of contrastive focus in congenitally blind adults and sighted adults. *Journal of Speech, Language, and Hearing Research*, *57*(3), 793–804.

Ménard, L., Leclerc, A., Tiede, M., Prémont, A., Turgeon, C., Trudeau-Fisette, P., & Côté, D. (2013). Correlates of contrastive focus in congenitally blind adults and sighted adults. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*.

Ménard, L., Perrier, P., & Aubin, J. (2016). Compensation for a lip-tube perturbation in 4-year-olds: Articulatory, acoustic, and perceptual data analyzed in comparison with adults. In *The Journal of the Acoustical Society of America* (Vol. 139, Issue 5, pp. 2514–2531). https://doi.org/10.1121/1.4945718

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. https://doi.org/10.3813/AAA.918747

Ménard, L., Schwartz, J.-L., & Boë, L.-J. (2004). Role of Vocal Tract Morphology in Speech Development. *Journal of Speech Language and Hearing Research*, *47*(5), 1059. https://doi.org/10.1044/1092-4388(2004/079)

Ménard, L., Schwartz, J.-L., Boë, L.-J., & Aubin, J. (2007). Articulatory-acoustic relationships during vocal tract growth for French vowels: Analysis of real data and simulations with an articulatory model. *Journal of Phonetics*, *35*, 1–19. https://doi.org/10.1016/j.wocn.2006.01.003

Ménard, L., Schwartz, J.-L., Boë, L.-J., Kandel, S., & Vallée, N. (2002). Auditory normalization of French vowels synthesized by an articulatory model simulating growth from birth to adulthood. *The Journal of the Acoustical Society of America*, *111*(4), 1892–1905. https://doi.org/10.1121/1.1459467

Ménard, L., Trudeau-Fisette, P., Côté, D., & Turgeon, C. (2016). Speaking clearly for the blind: Acoustic and articulatory correlates of speaking conditions in sighted and congenitally blind speakers. *PLoS ONE*, *11*(9). https://doi.org/10.1371/journal.pone.0160088

Ménard, L., Turgeon, C., Trudeau-fisette, P., & Bellavance-courtemanche, M. (2015). *Effects of blindness on production – perception relationships : Compensation strategies for a lip- tube perturbation of the French [u]. 9206*(September), 227–248. https://doi.org/10.3109/02699206.2015.1079247

Ménard, L., Turgeon, C., Trudeau-Fisette, P., & Bellavance-Courtemanche, M. (2016). Effects of blindness on production–perception relationships: Compensation strategies for a lip-tube perturbation of the French [u]. *Clinical Linguistics & Phonetics*, *30*(3–5), 227–248. https://doi.org/10.3109/02699206.2015.1079247

Mercier, M. R., Foxe, J. J., Fiebelkorn, I. C., Butler, J. S., Schwartz, T. H., & Molholm, S. (2013). Auditory-driven phase reset in visual cortex: Human electrocorticography reveals mechanisms of early multisensory integration. *NeuroImage*, *79*, 19–29. https://doi.org/10.1016/j.neuroimage.2013.04.060

Misceo, G. F., Hershberger, W. a, & Mancini, R. L. (1999). Haptic estimates of discordant visual-haptic size vary developmentally. *Perception & Psychophysics*, *61*(4), 608–614. https://doi.org/10.3758/BF03205533

Mishra, J., Martinez, A., Sejnowski, T. J., & Hillyard, S. A. (2007). Early Cross-Modal Interactions in Auditory and Visual Cortex Underlie a Sound-Induced Visual Illusion. *Journal of Neuroscience*, *27*(15), 4120–4131. https://doi.org/10.1523/JNEUROSCI.4912-06.2007

Mitsuya, T., Samson, F., Ménard, L., & Munhall, K. G. (2013). Language dependent vowel representation in speech production. *The Journal of the Acoustical Society of America*, *133*(5), 2993–3003. https://doi.org/10.1121/1.4795786

Molholm, S., Ritter, W., Murray, M. M., Javitt, D. C., Schroeder, C. E., & Foxe, J. J. (2002). Multisensory auditory – visual interactions during early sensory processing in humans : a high-density electrical mapping study. *Cognitive Brain Research*, *14*, 115–128.

Moore, D. R., Ferguson, M. A., Halliday, L. F., & Riley, A. (2008). Frequency discrimination in children: Perception, learning and attention. *Hearing Research*, *238*, 147–154.

Munhall, K. G., MacDonald, E. N., Byrne, S. K., & Johnsrude, I. (2009). Talkers alter vowel production in response to real-time formant perturbation even when instructed not to compensate. *The Journal of the Acoustical Society of America*, *125*(1), 384–390. https://doi.org/10.1121/1.3035829

Nardini, M., Jones, P., Bedford, R., & Braddick, O. (2008). Development of Cue Integration in Human Navigation. *Current Biology*, *18*(9), 689–693. https://doi.org/10.1016/j.cub.2008.04.021

Nasir, S. M., & Ostry, D. J. (2006). Somatosensory Precision in Speech Production. *Current Biology*, *16*(19), 1918–1923. https://doi.org/10.1016/j.cub.2006.07.069

Nasir, S. M., & Ostry, D. J. (2009). Auditory plasticity and speech motor learning. *Proceedings of the National Academy of Sciences*, *106*(48), 20470–20475. https://doi.org/10.1073/pnas.0907032106

Necker, L. A. (1989). Observations on some remarkable optical phaenomena seen in Switzerland; and on an optical phaenomenon which occurs on viewing a figure of a crystal or geometrical solid. *London and Edinburgh Philosophical Magazine and Journal of Science*, *1*(5), 329–337.

Neil, P. A., Chee-Ruiter, C., Scheier, C., Lewkowicz, D. J., & Shimojo, S. (2006). Development of multisensory spatial integration and perception in humans. *Developmental Science*, *9*(5), 454–464. https://doi.org/10.1111/j.1467-7687.2006.00512.x

Neufeld, C. (2013). *Multimodal targets in speech production: Acoustic, articulatory and dynamic evidence from formant perturbation*. University of Toronto.

Niemeyer, W., & Starlinger, I. (1981). Do the blind hear better? investigations on auditory processing in congenital or early acquired blindness II. Central functions. *International Journal of Audiology*, *20*(6), 510–515. https://doi.org/10.3109/00206098109072719

Ogane, R., Schwartz, J.-L., & Ito, T. (2017). Somatosensory information affects word segmentation and perception of lexical information. *Society for the Neurobiology of Language*.

Ohala, J. J. (1996). Speech perception is hearing sounds, not tongues. *The Journal of the Acoustical Society of America*, *99*(3), 1718–1725. https://doi.org/10.1121/1.414696

Oller, D. K., & Eilers, R. E. (1988). The Role of Audition in Infant Babbling. *Child Development*, *59*(2), 441. https://doi.org/10.2307/1130323

Oller, D. K., & MacNeilage, P. F. (1983). Development of Speech Production: Perspectives from Natural and Perturbed Speech. In M. P.F. (Ed.), *The Production of Speech*. Springer.

Patri, J. F., Diard, J., & Perrier, P. (2015). Optimal speech motor control and token-to-token variability: a Bayesian modeling approach. *Biological Cybernetics*, *109*(6), 611–626. https://doi.org/10.1007/s00422-015-0664-4

Patri, J.-F. (2018). *Université Grenoble Alpes*. Université Grenoble Alpes.

Patterson, M. L., & Werker, J. F. (1999). Matching phonetic information in lips and voice is robust in 4.5-month-old infants. *Infant Behavior and Development*, *22*(2), 237–247. https://doi.org/10.1016/S0163-6383(99)00003-X

Patterson, M. L., & Werker, J. F. (2003). Two-month-old infants match phonetic information in lips and voice. *Developmental Science*, *6*(2), 191–196. https://doi.org/10.1111/1467-7687.00271

Peelle, J. E., & Sommers, M. S. (2015). Prediction and constraint in audiovisual speech perception. *Cortex*, *68*(5), 169–181. https://doi.org/10.1016/j.cortex.2015.03.006

Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Stockmann, E., Tiede, M., & Zandipour, M. (2004). The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts. *The Journal of the Acoustical Society of America*, *116*(4), 2338–2344. https://doi.org/10.1121/1.1787524

Perrier, P. (1995). Control and representations in speech production. *ZAS Papers in Linguistics, 40*, 109–132.

Pisella, L., Rode, G., Farnè, A., Tilikete, C., & Rossetti, Y. (2006). Prism adaptation in the rehabilitation of patients with visuo-spatial cognitive disorders. In *Current Opinion in Neurology* (Vol. 19, Issue 6, pp. 534–542). https://doi.org/10.1097/WCO.0b013e328010924b

Pitt, M. A., & Shoaf, L. (2001). The source of a lexical bias in the verbal transformatiion effect. *Language and Cognitive Processes*, *16*(5–6), 715–721. https://doi.org/10.1080/01690960143000056

Pitt, M. A., & Shoaf, L. (2002). *Linking Verbal Transformations to Their Causes*. https://doi.org/10.1037/0096-1523.28.1.150

Pons, F., Lewkowicz, D. J., Soto-Faraco, S., & Sebastian-Galles, N. (2009). Narrowing of intersensory speech perception in infancy. *Proceedings of the National Academy of Sciences*, *106*(26), 10598–10602. https://doi.org/10.1073/pnas.0904134106

Proske, U., & Gandevia, S. C. (2009). The kinaesthetic senses. *Journal of Physiology*, *587*(17), 4139–4146. https://doi.org/10.1113/jphysiol.2009.175372

Purcell, D. W., & Munhall, K. G. (2006a). Adaptive control of vowel formant frequency: Evidence from real-time formant manipulation. *The Journal of the Acoustical Society of America*, *120*(2), 966–977. https://doi.org/10.1121/1.2217714

Purcell, D. W., & Munhall, K. G. (2006b). Compensation following real-time manipulation of formants in isolated vowels. *The Journal of the Acoustical Society of America*, *119*(4), 2288–2297. https://doi.org/10.1121/1.2173514

Raij, T., Ahveninen, J., Lin, F. H., Witzel, T., Jääskeläinen, I. P., Letham, B., Israeli, E., Sahyoun, C., Vasios, C., Stufflebeam, S., Hämäläinen, M., & Belliveau, J. W. (2010). Onset timing of cross-sensory activations and multisensory interactions in auditory and visual sensory cortices. *European Journal of Neuroscience*, *31*(10), 1772–1782. https://doi.org/10.1111/j.1460-9568.2010.07213.x

Rentschler, I., Jüttner, M., Osman, E., Müller, A., & Caelli, T. (2004). Development of configural 3D object recognition. *Behavioural Brain Research*, *149*(1), 107–111. https://doi.org/10.1016/S0166-4328(03)00194-3

Rizzolatti, G., Fadiga, L., Matelli, M., Bettinardi, V., Paulesu, E., Perani, D., & Fazio, F. (1996). Localization of grasp representations in humans by PET: 1. Observation versus execution. *Experimental Brain Research*, *111*(2), 246–252. https://doi.org/10.1007/BF00227301

Robert-Ribes, J., Schwartz, J.-L., Lallouache, T., & Escudier, P. (1998). Complementarity and synergy in bimodal speech: Auditory, visual, and audio-visual identification of French oral vowels in noise. *The Journal of the Acoustical Society of America*, *103*(6), 3677–3689. https://doi.org/10.1121/1.423069

Rosenblum, L. D. (2008a). Primacy of Multimodal Speech Perception. *The Handbook of Speech Perception*, 51–78. https://doi.org/10.1002/9780470757024.ch3

Rosenblum, L. D. (2008b). Speech Perception as a Multimodal Phenomenon. *Current Directions in Psychological Science*, *17*(6), 405–409. https://doi.org/10.1111/j.1467-8721.2008.00615.x

Rosenblum, L. D., Schmuckler, M. A., & Johnson, J. A. (1997). The McGurk effect in infants. *Perception & Psychophysics*, *59*(3), 347–357. https://doi.org/10.3758/BF03211902

Ross, L. A., Molholm, S., Blanco, D., Gomez-Ramirez, M., Saint-amour, D., & Foxe, J. J. (2011). The development of multisensory speech perception continues into the late childhood years. *European Journal of Neuroscience*, *33*(12), 2329–2337. https://doi.org/10.1111/j.1460-9568.2011.07685.x.The

Saalasti, S., Kätsyri, J., Tiippana, K., Laine-Hernandez, M., Von Wendt, L., & Sams, M. (2012). Audiovisual speech perception and eye gaze behavior of adults with asperger syndrome. *Journal of Autism and Developmental Disorders*, *42*(8), 1606–1615. https://doi.org/10.1007/s10803-011-1400-0

Sams, M., Möttönen, R., & Sihvonen, T. (2005). Seeing and hearing others and oneself talk. *Cognitive Brain Research*, *23*(2–3), 429–435. https://doi.org/10.1016/j.cogbrainres.2004.11.006

Sato, M., Schwartz, J.-L., Abry, C., Cathiard, M. A., & Lœvenbruck, H. (2006). Multistable syllables as enacted percepts: A source of an asymmetric bias in the verbal transformation effect. *Perception and Psychophysics*, *68*(3), 458–474. https://doi.org/10.3758/BF03193690

Sato, M., Vallée, N., Schwartz, J.-L., & Rousset, I. (2007). A perceptual correlate of the labial-coronal effect. *Journal of Speech, Language, and Hearing Research*, *50*(6), 1466–1480. https://doi.org/10.1044/1092-4388(2007/101)

Savariaux, C., Perrier, P., & Orliaguet, J. P. (1995). Compensation strategies for the perturbation of the rounded vowel [u] using a lip tube: A study of the control space in speech production. *Journal of the Acoustical Society of America*, *98*(5), 2428-2442.

Scheerer, N. E., Jacobson, D. S., & Jones, J. A. (2016). Sensorimotor learning in children and adults: Exposure to frequency-altered auditory feedback during speech production. *Neuroscience*, *314*, 106–115. https://doi.org/10.1016/j.neuroscience.2015.11.037

Scheerer, N. E., Liu, H., & Jones, J. A. (2013). The developmental trajectory of vocal and event-related potential responses to frequency-altered auditory feedback. *European Journal of Neuroscience*, *38*(8), 3189–3200. https://doi.org/10.1111/ejn.12301

Schürmann, M., Caetano, G., Jousmäki, V., & Hari, R. (2004). Hands help hearing: Facilitatory audiotactile interaction at low sound-intensity levels. *The Journal of the Acoustical Society of America*, *115*(2), 830–832. https://doi.org/10.1121/1.1639909

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. https://doi.org/10.1016/j.jneuroling.2009.12.004

Schwartz, J.-L., Boë, L.-J., & Abry, C. (2007). Linking the Dispersion-Focalization Theory (DFT) and the Maximum Utilization of the Available Distinctive Features (MUAF) principle in a Perception-for-Action-Control Theory (PACT). In M. J. Solé, P. Beddor, & M. Ohala (Eds.), *Experimental approaches to phonology* (pp. 104–124). Oxford University Press.

Schwartz, J.-L., Boë, L.-J., Vallé, N., & Abry, C. (1997). The Dispersion-Focalization Theory of vowel systems. In *Journal of Phonetics* (Vol. 25).

Shiller, D. M., Gracco, V. L., & Rvachew, S. (2010). Auditory-motor learning during speech production in 9- 11-year-old children. *PLoS ONE*, *5*(9). https://doi.org/10.1371/journal.pone.0012975

Shiller, D. M., Lametti, D. R., & Ostry, D. J. (2013). Auditory plasticity and sensorimotor learning in speech production. *Proceedings of Meetings on Acoustics*, *19*, 1–6. https://doi.org/10.1121/1.4799848

Shiller, D. M., & Rochon, M.-L. (2014). Auditory-perceptual learning improves speech motor adaptation in children. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(4), 1308–1315. https://doi.org/10.1037/a0036660

Shiller, D. M., Sato, M., Gracco, V. L., & Baum, S. R. (2009). Perceptual recalibration of speech sounds following speech motor learning. *The Journal of the Acoustical Society of America*, *125*(2), 1103–1113. https://doi.org/10.1121/1.3058638

Skipper, J. I., Devlin, J. T., & Lametti, D. R. (2017). The hearing ear is always found close to the speaking tongue: Review of the role of the motor system in speech

perception. *Brain and Language*, *164*, 77–105. https://doi.org/10.1016/j.bandl.2016.10.004

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*, *17*(10), 2387–2399. https://doi.org/10.1093/cercor/bhl147

Spence, C., & McDonald, J. (2004). The Cross-Modal Consequences of the Exogenous Spatial Orienting of Attention. In G. A. Calvert & C. Spence (Eds.), *The handbook of multisensory processes* (B. E. Stein, pp. 3–25). https://psycnet.apa.org/record/2004-17469-001

Stein, B. E., London, N., Wilkinson, L. K., & Price, D. D. (1996). Enhancement of Perceived Visual Intensity by Auditory Stimuli: A Psychophysical Analysis. *Journal of Cognitive Neuroscience*, *8*(6), 497–506. https://doi.org/10.1162/jocn.1996.8.6.497

Stein, B. E., & Meredith, M. A. (1993). *The Merging of the Senses*. MIT Press.

Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: current issues from the perspective of the single neuron. *Nature Reviews Neuroscience 2008 9:4*, *9*(4), 255–266. https://doi.org/10.1038/nrn2331

Stein, B. E., Stanford, T. R., & Rowland, B. A. (2014). Development of multisensory integration from the perspective of the individual neuron. *Nature Reviews Neuroscience*, *15*(8), 520–535.

Stevens, K. N., & Halle, M. (1967). Remarks on analysis by synthesis and distinctive features. In *Models for the perception of speech and visual form* (MIT Press, pp. 88–102). Wathen-Dunn, W.

Stevens, S. (1972). A Neural Quantum in Sensory Discrimination. *Science*, *177*, 749–762.

Streri, A., & Gentaz, E. (2003). Cross-modal recognition of shape from hand to eyes and handedness in human newborns. *Somatosensory & Motor Research*, *20*(01), 13–18. https://doi.org/10.1080/0899022031000083799

Streri, A., & Gentaz, E. (2004). Cross-modal recognition of shape from hand to eyes and handedness in human newborns. *Neuropsychologia*, *42*(10), 1365–1369. https://doi.org/10.1016/j.neuropsychologia.2004.02.012

Svirsky, M. A., Teoh, S. W., & Neuburger, H. (2004). Development of language & speech perception in congenitcally, profound deaf children.pdf. In *Audiology Neurootology* (Vol. 9, pp. 224–233).

Syrdal, A. K., & Gopal, H. S. (1986). A perceptual model of vowel recognition based on the auditory representation of american english vowels. *Journal of the Acoustical Society of America*, *79*(4), 1086–1100. https://doi.org/10.1121/1.393381

Tiippana, K., Möttönen, R., & Schwartz, J.-L. (2015). Multisensory and sensorimotor interactions in speech perception. *Frontiers in Psychology*, *6*(MAR), 2014–2016. https://doi.org/10.3389/fpsyg.2015.00458

Tourville, J. A., & Guenther, F. H. (2011). The DIVA model: A neural theory of speech acquisition and production. *Language and Cognitive Processes*, *26*(7), 952–981. https://doi.org/10.1080/01690960903498424

Traunmuïler, H., & Niklas Öhrström. (2007). Audiovisual perception of openness and lip rounding in front vowels. *Journal of Phonetics*, *35*, 244–258. https://doi.org/10.1016/j.wocn.2006.03.002

Traunmüller, H., & Lacerda, F. (1987). Perceptual relativity in identification of two-formant vowels. *Speech Communication*, *6*, 143–157.

Tremblay, C., Champoux, F., Voss, P., Bacon, B. A., Lepore, F., & Théoret, H. (2007). Speech and non-speech audio-visual illusions: A developmental study. *PLoS ONE*, *2*(8). https://doi.org/10.1371/journal.pone.0000742

Tremblay, S., Shiller, D. M., & Ostry, D. J. (2003). Somatosensory basis of speech production. *Nature*, *423*(6942), 866–869. https://doi.org/10.1038/nature01710

Trudeau-Fisette, P., Ito, T., & Ménard, L. (2019). Auditory and Somatosensory Interaction in Speech Perception in Children and Adults. *Frontiers in Human Neuroscience*, *13*. https://doi.org/10.3389/fnhum.2019.00344

Trudeau-Fisette, P., Tiede, M., & Ménard, L. (2017). Compensations to auditory feedback perturbations in congenitally blind and sighted speakers: Acoustic and articulatory data. *PLoS ONE*, *12*(7). https://doi.org/10.1371/journal.pone.0180300

Trudeau-Fisette, P., Turgeon, C., & Côté, D. (2013). Vowel production in sighted adults and blind adults: A study of speech adaptation strategies in high-intensity

background noise. *Proceedings of Meetings on Acoustics*, *19*. https://doi.org/10.1121/1.4799432

Turgeon, C. (2011). *Mesure du développement de la capacité de discrimination auditive et visuelle chez des personnes malentendantes porteuses d'un implant cochléaire*.

Vaissière, J. (2009). *Articulatory Modeling and the Definition of Acoustic- -Perceptual Targets for Reference Vowels*. *2*, 22–33.

Valkenier, B., Duyne, J. Y., Andringa, T. C., & Başkenta, D. (2012). Audiovisual Perception of Congruent and Incongruent Dutch Front Vowels. *Journal of Speech, Language, and Hearing Research*, *55*(6), 1788–1801. https://doi.org/10.1044/1092-4388(2012/11-0227)

Vallée, N., Schwartz, J.-L., & Escudier, P. (1999). Phase spaces of vowel systems : A typology in the light of the Dispersion-Focalisation Theory (DFT). *Proc. of the XIVth International Congress of Phonetic Sciences*, 333–336.

van Brenk, F., & Terband, H. (2020). Compensatory and adaptive responses to real-time formant shifts in adults and children. *The Journal of the Acoustical Society of America*, *147*(4), 2261–2270. https://doi.org/10.1121/10.0001018

Vidou, C., Uribe, C., Boukhalfi, T., Labbé, D., & Ménard, L. (2020). Compensatory responses to real-time perturbation of visual feedback during vowel production. *International Seminar on Speech Production*.

Villacorta, V. M. (2006). *Sensorimotor adaptation to perturbations of vowel acoustics and is relation to perception*. Massachusetts Institute of Technology.

Villacorta, V. M., Perkell, J. S., & Guenther, F. H. (2007). Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception. *The Journal of the Acoustical Society of America*, *122*(4), 2306–2319. https://doi.org/10.1121/1.2773966

Vorperian, H. K., Kent, R. D., Gentry, L. R., & Yandell, B. S. (1999). Magnetic resonance imaging procedures to study the concurrent anatomic development of vocal tract structures: preliminary results. *International Journal of Pediatric Otorhinolaryngology*, *49*(3), 197–206. https://doi.org/10.1016/S0165-5876(99)00208-6

Voss, P. (2019). Brain (re)organization following visual loss. *Wiley Interdisciplinary Reviews: Cognitive Science*, *10*(1), e1468. https://doi.org/10.1002/wcs.1468

Walker-Andrews, A. (1994). Taxonomy for intermodal relations. In D. J. Lewkowicz & R. Lickliter (Eds.), *The Development of Intersensory Perception: Comparative Perspectives* (pp. 39–56).

Walsh, B., Smith, A., & Weber-Fox, C. (2006). Short-Term Plasticity in Children's Speech Motor Systems. *Developmental Psychobiology*, *48*(8), 660–674. https://doi.org/10.1002/dev.20185

Warren, R. M. (1961). Illusory changes of distinct speech upon repetition: The verbal transformation effect. *British Journal of Psychology*, *52*(3), 249–258. https://doi.org/10.1111/j.2044-8295.1961.tb00787.x

Warren, R. M., & Gregory, R. L. (1958). An auditory analogue of the visual reversible figure. *The American Journal of Psychology*, *71*(3), 612–613. https://doi.org/10.2307/1420267

Werker, J. F. (2018). Perceptual beginnings to language acquisition. *Applied Psycholinguistics*, *39*, 703–728. https://doi.org/10.1017/S0142716418000152

Wightman, F., Kistler, D., & Brungart, D. (2006). Informational masking of speech in children: Auditory-visual integration. *The Journal of the Acoustical Society of America*, *119*(6), 3940–3949. https://doi.org/10.1121/1.2195121

Yeung, H. H., & Werker, J. F. (2013a). Lip Movements Affect Infants' Audiovisual Speech Perception. *Psychological Science*, *24*(5), 603–612. https://doi.org/10.1177/0956797612458802

Yeung, H. H., & Werker, J. F. (2013b). Lip Movements Affect Infants' Audiovisual Speech Perception. *Psychological Science*, *24*(5), 603–612. https://doi.org/10.1177/0956797612458802

Yu, L., Rowland, B. A., & Stein, B. E. (2010). Initiating the Development of Multisensory Integration by Manipulating Sensory Experience. *Journal of Neuroscience*, *30*(14), 4904–4913. https://doi.org/10.1523/JNEUROSCI.5575-09.2010