Le développement de la perception de la parole : étude des conflits entre les modalités auditive et visuelle

Sophie Dupont
Laboratoire de phonétique, Département de linguistique et de didactique des langues, Université du Québec à Montréal.

Résumé

Plusieurs travaux ont montré que le décodage des voyelles et des consonnes par l'auditeur tire profit des caractéristiques de l'onde sonore, mais également des informations visuelles transmises par la position de la mâchoire et des lèvres. L'effet McGurk constitue un exemple de fusion audiovisuelle: un stimulus auditif [ba] superposé à un stimulus visuel conflictuel [ga] est perçu [da]. Si un grand nombre d'études ont été menées chez les adultes, peu ont tenté d'en décrire les conditions d'apparition. Les objectifs de cette étude consistent donc à décrire le rôle de la perception visuelle et auditive au cours du développement de la parole en français. Des enregistrements audio-visuels d'un locuteur adulte articulant les séquences [aba], [ada], [aga], [ava], [ibi], [idi], [igi], [ivi] ont été effectués. Le signal sonore a été numérisé et combiné au signal visuel de façon à générer des conflits (par exemple, un signal audio [aba] superposé à un signal visuel [ada]). 48 séquences ont ainsi été créées, incluant les cas de correspondances audio-visuelles et les conditions contrôle audio seule et visuelle seule. Ce corpus a été soumis à 8 enfants âgés de 4 à 5 ans et à 10 adultes lors d'un test de perception où les sujets devaient identifier la séquence syllabique perçue. Les taux d'identification de chaque séquence ont été reliés aux mesures articulatoires et acoustiques, afin d'évaluer statistiquement l'impact de ces dernières sur l'identification audio-visuelle. Les résultats confirment l'existence de l'illusion McGurk chez les enfants, mais suggèrent qu'il soit plus faible et sujet à la variabilité que chez les adultes. Finalement, les résultats sont reliés au rôle de la vision dans l'émergence du système phonologique chez les enfants.

Plusieurs travaux ont montré que la perception de la parole, et plus particulièrement le décodage des voyelles et des consonnes par l'auditeur, tire profit non seulement des caractéristiques de l'onde sonore produite par le locuteur, mais également des informations visuelles transmises par la position de sa mâchoire et de ses lèvres. On rencontre pourtant fréquemment des situations impliquant des interactions de communication unimodale réussies entre interlocuteurs ; par exemple, lorsqu'un individu parle au téléphone à un autre individu, seule la modalité auditive est utilisée, et lorsque deux personnes échangent des paroles alors qu'une vitre les sépare et les empêche de s'entendre, seule la modalité visuelle est utilisée. Des privations sensorielles telles que la surdité et la cécité constituent des contraintes physiologiques qui suppriment une des deux modalités de la parole audio-visuelle.

Si on aborde la parole du point de vue de l'interaction entre interlocuteurs «face à face», l'étude séparée de chacune des modalités, bien que fort pertinente à l'étude de plusieurs aspects de la parole, ne permet pas facilement de montrer l'apport de chacune des modalités et de leur intégration à la perception de la parole audio-visuelle. Toutefois, depuis qu'ont été publiées en 1976 les premières données relatives à l'effet McGurk, ce dernier constitue un paradigme de recherche privilégié pour l'étude de la perception de la parole audio-visuelle. En effet, l'effet McGurk a d'abord été abordé dans un article de McGurk et MacDonald (1976) dans lequel étaient présentés des résultats de recherche portant sur les percepts de stimuli audio-visuels conflictuels. La première manifestation de l'effet McGurk est survenue lors de la présentation à des adultes normaux d'un stimulus auditif /ba/ (désormais noté A /ba/) superposé à un stimulus visuel /ga/ (désormais noté V /ba/). Le plus souvent, ces derniers ont alors perçu [da], un percept qui constitue une fusion audio-visuelle qui ne relève ni exactement de la partie acoustique du stimulus ni exactement de sa partie visuelle; la nature des percepts étant acceptés comme faisant partie de l'effet McGurk s'est par la suite étendue à d'autres percepts, dans la mesure où ils différaient de la partie acoustique des stimuli conflictuels.

1. Cadre théorique

Si un grand nombre d'études ont été menées auprès d'adultes dans ce cadre de recherche, bien peu ont eu pour but de décrire les conditions d'apparition de l'effet McGurk au cours du développement de la parole. Nous vous présentons ici brièvement quelques études qui nous ont semblé pertinentes et qui portaient sur l'effet McGurk chez les bébés, chez les enfants et chez les adultes.

Rosenblum et col. (1997) ont étudié l'effet McGurk chez 10 bébés âgés de 5 mois et vivant en milieu anglophone ; ces sujets ne possèdaient alors qu'une très courte expérience linguistique et sensorielle. Les chercheurs ont mené leur étude dans le but de prouver que la représentation de la parole que possèdent les bébés est amodale et qu'ils procèdent à une association supramodale de l'input audio-visuel plutôt qu'à une association d'identité phonémique. Rosenblum et col. (1997) ont créé trois types de stimuli audio-visuels (non-conflictuel ; A /va/ - V /va/, conflictuels ; A /ba/ - V /va/ et A /da/ -V /va/) et trois stimuli audio seuls (A /va/, A /ba/ et A /da/), les parties audio étant synthétisées et les parties visuelles étant articulées par un homme. Ils ont cherché à déterminer si leurs sujets discriminaient les séquences audio-visuelles conflictuelles de la séquence non-conflictuelle. Les auteurs ont utilisé une technique d'habituation aux stimuli et ils ont ensuite mesuré les temps de fixation des stimuli par les sujets bébés. Ils ont trouvé une différence significative entre les temps de fixation du stimulus A /da/ -V /va/ et du stimulus A /va/ - V /va/, mais pas entre le stimulus A /ba/ - V /va/ et le stimulus A /va/ - V /va/. Cette dernière observation montrait que leurs sujets bébés manifestaient un effet McGurk, c'est-à-dire que leur percept du stimulus A /ba/ - V /va/ était probablement [va], un percept différent de la partie auditive du stimulus conflictuel et intégrant les informations auditives et visuelles en présence.

McGurk et MacDonald (1976) se sont intéressés à la robustesse de l'effet McGurk chez des enfants anglophones. Pour ce faire, ils ont testé 54 adultes (18-40 ans), 21 enfants d'âge préscolaire (3-4 ans) et 28 enfants d'âge scolaire (7-8 ans). Les sujets étaient soumis à une condition audio-visuelle (stimuli A /ba/ - V /ga/, A /ga/ - V /ba/, A /pa/ - V /ka/ et A /ka/ - V /pa/) et à une condition audio au cours desquelles ils devaient répéter ce qu'avait dit la locutrice. En classifiant les percepts des stimuli audio-visuels des sujets selon qu'ils étaient de catégorie audio, visuelle, fusion, combinaison ou autres, les chercheurs ont trouvé que les sujets adultes étaient davantage influencés par l'input visuel que ne l'étaient les deux groupes d'enfants. La description du phénomène faite par les auteurs montrait aussi que les percepts qui ne reflétaient qu'une seule modalité chez les enfants étaient davantage des percepts audio, alors que chez les adultes, on trouvait plus de percepts visuels.

MacDonald et McGurk (1978) ont cherché à expliquer plutôt qu'à décrire le phénomène d'intégration des informations auditives et visuelles conflictuelles en testant leur hypothèse *mode-lieu* (« manner-place hypothesis ») selon laquelle, en situation de communication entre interlocuteurs «face à face» dont l'audition est normale, le mode d'articulation des consonnes est mieux reconnu par l'oreille, alors que le lieu d'articulation est mieux reconnu par les yeux. 44 adultes anglophones (18-24 ans) ont passé un test de perception au cours duquel ils devaient dire à voix haute ce que la locutrice qu'ils voyaient à l'écran avait dit ; les stimuli étaient des superpositions conflictuelles des consonnes /p, b, t, d, k, g, m, n/ en contexte vocalique /a/. L'hypothèse *mode-lieu* s'est avérée vérifiée avec les percepts dont la partie auditive des stimuli était constituée de consonnes labiales et dont la partie

visuelle était constituée de consonnes non-labiales, mais pas avec les stimuli inverses. Ces résultats ont tout de même montré l'effet général de la vision dans la perception de la parole dans des interactions de type «face à face».

Massaro (1984) a mené une étude qui traitait de l'évaluation et de l'intégration de l'information dans la perception de la parole, dans le but d'étudier les aspects développementaux de la perception de la parole. Il a cherché à voir si ses 11 sujets enfants anglophones (4 à 6 ans) étaient aussi influencés par l'information visuelle que l'avaient été les sujets adultes de l'étude de Massaro et Cohen (1983). Pour ce faire, il a comparé les résultats de ses 11 sujets enfants à ceux de 11 adultes âgés entre 18 et 38 ans, qu'ils a soumis à une tâche de d'identification phonémique. La partie acoustique des stimuli audio-visuels auxquels ils étaient confrontés était synthétisée et consistait en cinq séquences d'un continuum de /ba/ à /da/ auxquelles étaient superposés successivement les mouvements articulatoires associés à /ba/, à /da/ et aucun mouvement articulatoire. Massaro (1984) a trouvé que les enfants présentaient la moitié moins d'influence visuelle que les adultes et que cette différence n'était pas de nature attentionnelle. Il a mentionné que les enfants semblaient sensibles à la correspondance des informations audio et visuelles, mais que l'information auditive avait une plus grande influence sur leur perception des catégories dans l'acquisition du langage.

À la lumière de ces études, on remarque que le développement de la perception audio-visuelle ne semble pas linéaire au cours du développement de la parole. Il nous a donc semblé pertinent de mener à nouveau une étude sur la perception de conflits audio-visuels auprès d'enfants d'âge préscolaire. La plupart des études ayant exploré la perception de stimuli conflictuels chez des sujets anglophones, nous nous sommes intéressés à des enfants locuteurs du français québécois. Notre objectif de recherche consistait donc à décrire le rôle de la perception visuelle et auditive au cours du développement de la parole en français. Pour ce faire, nous avons créé des stimuli conflictuels à l'aide de quatre consonnes dont les lieux d'articulation étaient différents (en ordre croissant de «visibilité : vélaire (/g/), dental (/d/), labio-dental (/v/) et bilabial (/b/)) mais dont le mode d'articulation demeurait le même (voisé) et nous avons comparé les percepts des sujets enfants à ceux des sujets adultes.

À cet égard, Robert-Ribes et col. (1998) ont identifié la complémentarité et la synergie entre l'audition et la vision comme étant deux facteurs qui influençaient l'efficacité de la perception de la parole audio-visuelle. La complémentarité est le pendant de l'hypothèse *mode-lieu* de McGurk et MacDonald (1978), dans la mesure où elle implique que le mode d'articulation est mieux transmis par le canal auditif et que le lieu d'articulation est mieux transmis par le canal visuel. La synergie est une propriété reliée au traitement de l'information et implique que la perception audio-visuelle est meilleure que la perception visuelle seule et que la perception auditive seule ; c'est d'ailleurs ce à quoi nous nous attendions dans le cadre de notre étude. Si Robert-Ribes et col. (1998) ont étudié la perception audio-visuelles des voyelles présentées avec du bruit et que nous nous intéressons plutôt à la perception audio-visuelles des consonnes sans bruit, la notion de visèmes qu'ils ont introduite est tout de même applicable à notre étude, les visèmes constituant les caractéristiques visuelles pertinentes à la reconnaissance des phonèmes.

2. Méthodologie

2.1. Enregistrement des stimuli

Des enregistrements audio-visuels d'un locuteur adulte du français québécois produisant des séquences bisyllabiques de type *voyelle$_i$-consonne-voyelle$_i$* ont été effectués à l'aide d'une caméra MiniDV Panasonic. Les images ont été numérisées à un taux de 25 images/seconde et à une fréquence d'échantillonnage du signal acoustique de 44100 Hz. Les voyelles /i/ et /a/ et les consonnes /b/, /d/, /g/ et /v/ ont servi à construire les séquences à l'étude et ont été enregistrées dans l'ordre suivant : [aba], [ada], [aga], [ava], [ibi], [idi], [igi] et [ivi]. Trois répétitions de ces huit séquences ont été produites par le locuteur dont le débit, le volume, l'intonation et l'intensité de la production ont été dirigés par les expérimentatrices afin qu'ils demeurent relativement constants. Un gros plan a été fait sur le visage du locuteur de telle sorte que seuls la partie inférieure de son visage (le bas de son nez, sa bouche, sa mâchoire) et le haut de son cou étaient visibles et occupaient alors environ les cinq sixièmes de l'image. Après la deuxième voyelle de chaque séquence, le locuteur revenait à une position neutre, caractérisée par les lèvres fermées, comme le montre la figure 1.

Figure 1. Image fixe du locuteur après avoir produit la séquence [aba]. Après chaque séquence, le locuteur revenait en position neutre d'une façon semblable, c'est-à-dire en refermant les lèvres.

Les données audiovisuelles ont été importées à l'aide du logiciel Imovie, puis les séquences bisyllabiques ont été segmentées à l'aide du logiciel Adobe Premiere Pro 7.0. Des trois répétitions produites de chacune des séquences, une seule occurrence a été conservée pour chacune d'elles, celle dont la qualité acoustique et la clarté des mouvements articulatoires ont été jugées les meilleures par l'expérimentatrice. Chaque séquence audio-visuelle avait une durée de 2833 ms environ, ce qui correspondait à 85 images.

2.2. Traitement des images et des signaux acoustiques

Les stimuli enregistrés ont ensuite été manipulés afin de créer les quatre conditions soumises aux sujets lors du test de perception décrit à la section 2.3. En condition bimodale, l'image et le son étaient présentés simultanément. Lorsque les signaux visuel et auditif d'une même séquence étaient superposés, les stimuli ont été qualifiés de 'non-conflictuels'. En revanche, nous réfèrerons aux cas où l'image d'une séquence donnée (par exemple [b]) est superposée au signal acoustique d'une séquence différente (par exemple [d]), par l'appellation stimuli 'conflictuels'. Deux conditions unimodales ont été créées : la condition unimodale 'visuelle',

correspondant à la présence seule de l'image, et la condition unimodale acoustique (ou unimodale 'audio'), correspondant à la présence seule du signal acoustique.

2.2.1. Condition unimodale acoustique

La création des stimuli de la condition unimodale acoustique a nécessité la manipulation des séquences audio-visuelles originales. En fait, le signal acoustique a été séparé du signal visuel à l'aide du logiciel Adobe Premiere Pro 7.0. Pour chacune des séquences bisyllabiques acoustiques conservées, l'image fixe d'un écran noir a été superposée. Ces manipulations ont donné lieu à huit stimuli 'audio' : [aba], [ada], [aga], [ava], [ibi], [idi], [igi] et [ivi].

2.2.2. Condition unimodale 'visuelle'

La création des stimuli 'visuels' n'a pas requis de manipulation des séquences audio-visuelles originales. Lors du test de perception, ce sont ces dernières qui ont été présentées aux sujets, à la seule différence que le volume des haut-parleurs a été placé à une valeur zéro, éliminant ainsi la partie acoustique des stimuli et ne donnant alors lieu qu'à la présentation des mouvements articulatoires effectués par le locuteur. La condition unimodale visuelle comptait donc huit stimuli 'visuels' : [aba], [ada], [aga], [ava], [ibi], [idi], [igi] et [ivi], où aucun son n'était émis.

2.2.3. Condition bimodale non conflictuelle

Les stimuli de la condition bimodale 'non-conflictuelle' consistaient en l'ensemble des séquences produites par le locuteur, sans qu'aucun traitement particulier ne leur soit appliqué. Huit stimuli 'non-conflictuels' étaient donc disponibles, comportant le son et l'image effectivement produits par le locuteur, soit [aba], [ada], [aga], [ava], [ibi], [idi], [igi] et [ivi].

2.2.4. Condition bimodale conflictuelle

Les stimuli audio-visuels 'conflictuels' ont été construits de façon à générer des conflits consonantiques entre le signal acoustique et le signal visuel. Ces stimuli ont été créés à l'aide du logiciel Adobe Premiere Pro 7.0 qui, d'une part, a permis de séparer le signal acoustique du signal visuel de chacun des stimuli audio-visuels 'non-conflictuels' originaux, et d'autre part, de superposer successivement à chaque signal acoustique donné les trois signaux visuels dont la consonne différait de la sienne. Avant d'effectuer ces superpositions des signaux acoustiques et visuels, les temps d'ouverture de la mâchoire, de début de fermeture des lèvres, de fin de fermeture des lèvres et de fermeture de la mâchoire ont été notés à partir du signal visuel et pris en compte afin d'assurer une synchronisation optimale entre les signaux des deux modalités. Ces stimuli 'conflictuels' étaient au nombre de 24 : 4 consonnes 'audio' X 3 consonnes 'visuelles' X 2 voyelles. Le tableau 1 regroupe les combinaisons retenues pour la création des stimuli 'conflictuels'.

Tableau 1 Les 24 stimuli audio-visuels 'conflictuels' présentés aux sujets lors du test de perception.

Voyelle /i/		Voyelle /a/	
Signal visuel	Signal acoustique	Signal visuel	Signal acoustique
/ibi/	/idi/	/aba/	/ada/
/ibi/	/igi/	/aba/	/aga/
/ibi/	/ivi/	/aba/	/ava/
/idi/	/ibi/	/ada/	/aba/
/idi/	/igi/	/ada/	/aga/
/idi/	/ivi/	/ada/	/ava/
/igi/	/ibi/	/aga/	/aba/
/igi/	/idi/	/aga/	/ada/
/igi/	/ivi/	/aga/	/ava/
/ivi/	/ibi/	/ava/	/aba/
/ivi/	/idi/	/ava/	/ada/
/ivi/	/igi/	/ava/	/aga/

2.3. Test de perception

2.3.1. Sujets

8 enfants âgés entre 4 ans 3 mois et 5 ans 9 mois (moyenne de 4 ans 7 mois) ont participé à cette étude, avec le consentement écrit d'un parent. Ils étaient tous de sexe féminin et ont été recrutés au Centre de la petite enfance Frisson de Colline, à Montréal. À la fin du test de perception, les enfants ont reçu un certificat d'attestation de participation à leur nom en guise de cadeau. De plus, 10 étudiantes de premier et de deuxièmes cycles universitaires en linguistique de l'Université du Québec à Montréal, âgées entre 22 et 31 ans (moyenne de 25 ans), se sont prêtées à l'étude. Les sujets adultes n'ont reçu aucune compensation pour leur participation à l'étude. Le français était la langue maternelle de tous les sujets et ils avaient tous une vision normale (dans certains cas, corrigée par des lunettes ou des verres de contact) et une audition normale. Les sujets enfants ne connaissaient pas le but de l'expérience et n'ont pas semblé être conscients de la manipulation préalable des

stimuli qui leur étaient présentés. Au cours de l'expérience, trois sujets adultes ont manifesté qu'ils remarquaient que certains stimuli étaient conflictuels et qu'ils croyaient cette manipulation conçue pour perturber leur perception. Nous avons conservé les percepts fournis par ces sujets puisque McGurk et MacDonald (1976) ont mentionné que le fait qu'un individu soit conscient de la façon dont ont été construits les stimuli n'inhibait pas la manifestation de l'effet McGurk.

2.3.2. Déroulement du test

Pour les deux groupes de sujets, les stimuli 'conflictuels' et 'non-conflictuels' ont été regroupés en une seule catégorie de stimuli en condition bimodale. Trois conditions faisaient donc partie du test : les conditions unimodales visuelle et auditive, et la condition bimodale. Les conditions unimodales comportaient chacune 8 stimuli, alors que la condition bimodale comportait 32 stimuli. À l'intérieur de chacune des trois conditions, les stimuli étaient présentés une seule fois, en ordre aléatoire variable d'un sujet à l'autre. Les stimuli étaient présentés automatiquement à l'écran d'un ordinateur, séparés par une pause de 4000 ms, au cours de laquelle un écran noir sans signal acoustique était présenté.

2.3.2.1. Cas des sujet adultes

L'expérience auprès des adultes s'est déroulée à l'intérieur d'une seule séance d'une durée d'environ 15 minutes dans le laboratoire de phonétique de l'université à laquelle les sujets avaient été recrutés. La partie visuelle de chacun des 48 stimuli était présentée à l'aide du logiciel Lecteur Windows Media Player 9.0 sur un écran plat Philips $109B_4$ de grandeur 19 pouces. Les sujets étaient assis à environ 50 cm de l'écran de l'ordinauteur et ils devaient écouter la partie acoustique des stimuli à l'aide d'écouteurs Audio-Technica ATH-M20 à un volume qui leur était confortable.

La tâche des sujets adultes était d'être attentifs à ce qui était présenté sur l'écran et à ce qu'ils entendaient dans les écouteurs et de manifester la séquence qu'ils avaient perçue. Il leur avait été dit que la forme-type de ce qu'ils étaient susceptibles de percevoir était : *voyelle – une ou plusieurs consonnes – voyelle* et quelques exemples leur avaient été donnés verbalement («vous pourriez entendre des séquences telles que /aba/, /ada/, /abga/, etc»).

Bien que l'écran demeurait noir pendant 4 secondes afin de laisser suffisamment de temps aux sujets pour effectuer leur identification de la séquence audio-visuelle, ils étaient également libres d'appuyer sur «pause» afin d'avoir plus de temps pour écrire leur réponse avant que ne soit présenté le prochain stimulus. La tâche des cinq premiers sujets était d'écrire ce qu'ils avaient perçu sur une feuille prévue à cet effet, mais l'expérimentatrice ayant remarqué que les sujets n'utilisaient pas la fonctionnalité «pause» et qu'il était possible que l'attention des sujets ne soit alors pas optimale au début de chaque stimulus, la tâche des 5 sujets adultes suivants a été modifiée. Il leur était alors plutôt demandé de dire à voix haute ce qu'ils avaient perçu après chaque stimulus; c'est l'expérimentatrice elle-même qui notait leurs réponses, permettant aux sujets de ne se concentrer effectivement que sur qui était dit et sur ce qui était montré. Une liste aléatoire de stimuli était présentée aux sujets adultes pour chacune des conditions dans l'ordre suivant : stimuli audio-visuels ('non-conflictuels' et 'conflictuels' ensemble et en ordre aléatoire), stimuli 'audio' et stimuli 'visuels'.

2.3.2.2. Cas des sujets enfants

Le test de perception auprès des enfants s'est déroulé dans le cadre d'une seule séance d'une durée variable selon les besoins de l'enfant, mais d'une moyenne de 30 minutes, dans un local fermé à l'intérieur de la garderie à laquelle ils avaient été recrutés. La partie visuelle de chacun des 48 stimuli était présentée à l'aide du logiciel Lecteur Windows Media Player 9.0 sur un écran d'une grandeur de 15 pouces d'un ordinateur portable Toshiba modèle Satellite Pro. Les sujets enfants étaient assis à environ 40 centimètres de l'écran de l'ordinateur et ils devaient écouter la partie acoustique des stimuli présentée à l'aide de haut-parleurs Altec Lansing AVS200 à un volume qui leur était confortable.

L'expérimentatrice avait présenté la tâche aux sujets enfants comme étant un jeu au cours duquel ils étaient invités à être attentifs à ce qui était présenté sur l'écran et à ce qu'ils entendaient et à répéter à voix haute ce qu'ils croyaient que le locuteur venait de dire. Des exemples de la forme-type *voyelle - une ou plusieurs consonnes - voyelle* que pouvaient prendre les stimuli leur avaient aussi été donnés («tu entendras des choses telles que /aba/, /ada/, /abga/, etc»). L'expérimentatrice contrôlait la succession des stimuli en appuyant sur «pause» en chacun d'eux. C'est également elle qui écrivait sur la feuille réponse les percepts de l'enfants en les répétant au fur et à mesure, de façon à obtenir une forme d'accord inter-juges avec l'autre expérimentatrice. Toutes deux s'assuraient que l'enfant regardait bien l'écran lors de la présentation des stimuli 'visuels' seulement, audio-visuels 'conflictuels' et 'non-conflictuels'. Une liste aléatoire de stimuli était présentée aux sujets enfants pour chacune des conditions dans l'ordre suivant : stimuli *audio* seulement, stimuli *visuels* seulement et stimuli audio-visuels ('conflictuels' et 'non-conflictuels' ensemble et en ordre aléatoire).

2.4. Traitement des données

Pour chacune des trois conditions de présentation des stimuli, des critères de classification des percepts des sujets ont été établis. Cette étude portant sur les conflits entre les modalités auditive et visuelle ayant cours dans la perception des consonnes des séquences *voyelle$_i$-consonne-voyelle$_i$* et les deux voyelles /i/ et /a/ n'ayant été utilisées qu'afin de pouvoir créer un plus grand nombre de stimuli et d'empêcher l'habituation des sujets à la forme des séquences lors du test de perception, nous n'avons pas analysé les données perceptives pour chacune des voyelles séparément.

Les résultats obtenus proviennent de moyennes effectuées à partir du nombre de percepts d'un groupe de sujets donné à une condition donnée pour un critère de classification donné pour une consonne ou une combinaison de consonnes donnée. Par la suite, les résultats obtenus par les sujets enfants à un critère de classification donné ont été statistiquement comparés aux résultats obtenus par les sujets adultes au même critère. Le logiciel Statistica 6.1 a été utilisé pour faire les analyses de variance (*one-way ANOVA*) entre les résultats obtenus par les deux groupes de sujets. Des tests *t* de Student ont été utilisés pour mesurer la différence de perception à l'intérieur même d'un groupe de sujets, mais pour des conditions différentes.

3. Présentation et discussion des résultats

3.1 Conditions contrôle unimodales

L'identification phonémique des stimuli unimodaux acoustiques et visuels constituait deux conditions contrôle. En effet, l'analyse des résultats obtenus par les sujets adultes et enfants aux stimuli 'audio' a permis de juger de la qualité acoustique des séquences produites par le locuteur. D'autre part, elle a servi à évaluer la compétence des sujets pour une tâche d'identification phonémique de stimuli acoustiques. De la même façon, l'analyse des percepts des sujets des stimuli 'visuels' a révélé si les mouvements articulatoires du locuteur étaient suffisamment visibles, s'ils étaient typiques des séquences cibles et si les sujets étaient habiles à une tâche d'identification phonémique de stimuli visuels. Aussi, les résultats obtenus dans ces conditions unimodales cherchaient à mesurer la robustesse «naturelle» auditive et visuelle de certains stimuli.

3.1.1. Percepts des stimuli 'audio' unimodaux

Seuls les percepts dont le lieu et la mode d'articulation étaient identiques à ceux des stimuli 'audio' présentés ont été jugés comme étant des percepts *audio* et qualifiés de *bonnes réponses*. Le tableau 2 montre la classification des percepts fournis par les sujets dans les catégories *bonnes réponses* et *autres*, cette dernière catégorie comprenant les percepts qui ne répondaient pas aux critères de correspondance de mode et de lieu d'articulation des percepts avec ceux des stimuli.

Tableau 2. Classification des percepts des sujets des stimuli acoustiques présentés.

Stimuli acoustiques	Percepts	
	Audio (*bonnes réponses*)	*Autres*
/b/	[b]	-
/d/	[d]	-
/g/	[g]	[d]
/v/	[v]	[b] [bv]

La figure 2 présente la moyenne de *bonnes réponses* fournies lors de l'identification des stimuli unimodaux 'audio' par les sujets enfants et adultes. La proportion de *bonnes réponses* fournies par les enfants n'est pas significativement différente de celle des adultes.

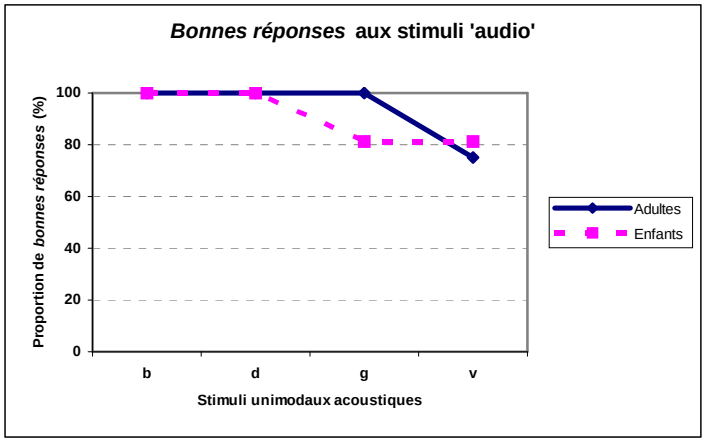

Figure 2. Proportions de *bonnes réponses*, c'est-à-dire de percepts dont le lieu et le mode d'articulation correspondaient à ceux des stimuli acoustiques présentés.

Si l'identification de /b/ et de /d/ est parfaite pour les deux groupes d'âge dans les deux contextes vocaliques, en contexte /a/, 37,5% des sujets enfants ont perçu la consonne /g/ comme étant /d/. Cette tendance n'ayant nullement été observée en contexte /i/ chez les enfants et en aucun cas chez les adultes, cela peut suggérer que le stimulus unimodal 'audio' /aga/ n'était pas ambigu, mais que ce sont plutôt les enfants qui ont été moins habiles à faire la distinction entre les réalisations acoustiques des lieux d'articulation alvéolaire (/d/) et vélaire (/g/) dans ce contexte ou qu'ils ont été moins attentifs à la tâche d'identification lors de la présentation de ce stimulus.

La fricative labio-dentale du stimulus /ava/ semblait quant à elle réellement plus difficile à identifier. En effet, 37,50% des enfants l'ont perçue comme étant plutôt la bilabiale [b]; 30,00% des adultes l'ont aussi perçu ainsi ou comme une combinaison des lieux d'articulation labio-dental et bilabial, [bv]. 20,00% des adultes ont aussi ainsi perçu /bv/ dans le contexte vocalique /i/. Nous proposons que la durée un peu longue de la friction et le fort accent mis sur la deuxième voyelle puissent être à l'origine de cette légère ambiguïté.

Par ailleurs, ces résultats montrent que tous les auditeurs étaient compétents pour effectuer une tâche d'identification phonémique de stimuli 'audio' et que tous les stimuli étaient acoustiquement valables. Nous verrons dans les prochaines sections si la potentielle ambiguïté acoustique de /v/ a eu une influence sur les percepts des stimuli bimodaux.

3.1.2. Percepts des stimuli 'visuels' seuls

Les percepts des stimuli 'visuels' seuls qui ont été considérés comme de *bonnes réponses* sont ceux dont le lieu d'articulation correspondait à celui des stimuli; la mode d'articulation et la distinction entre certains lieux d'articulation (par exemple, vélaire versus uvulaire) n'étant pas visibles, cela a parfois permis d'admettre plusieurs percepts pour un même stimulus. Le tableau 3 fait état de notre classification des percepts des sujets adultes et enfants et présente aussi les réponses non-acceptées (*autres*).

Tableau 3. Classification des percepts des sujets des stimuli 'visuels' présentés.

Stimuli visuels	*Visuels* (*bonnes réponses*)	*Autres*
/b/	[b] [p]	[k] [gb] [v]
/d/	[d] [t]	[v] [b] [g]
/g/	[g] [r] [j]	[b] [d] [v]
/v/	[v] [f]	[b] [d] [g] [j] [bv]

La figure 3 présente la moyenne de *bonnes réponses* fournies par les deux groupes de sujets comme identification des stimuli 'visuels'. La proportion de *bonnes réponses* fournies par les sujets enfants est significativement inférieure à celle des adultes ($F(1,6)=27,960$, $p<0,05$).

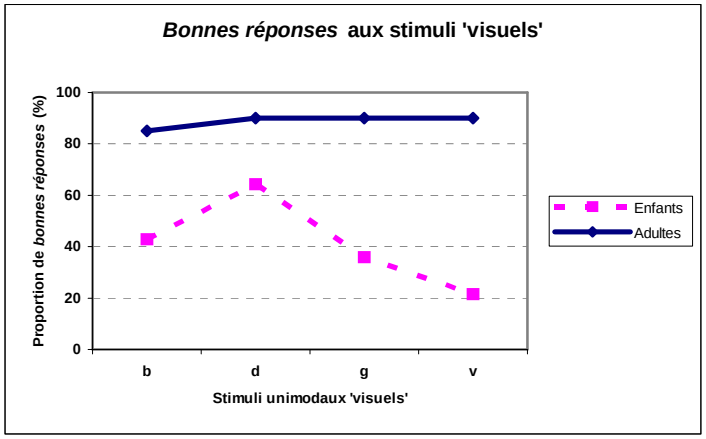

Figure 3. Proportions de *bonnes réponses*, c'est-à-dire de percepts dont le lieu d'articulation correspond aux stimuli 'visuels' présentés.

Il peut d'abord sembler étonnant que les stimuli /d/ soient les mieux identifiés (90,00% de *bonnes réponses* chez les adultes et 64,29% chez les enfants) et que les stimuli /b/ le soient moins bien (85,00% chez les adultes et 42,86% chez les enfants), étant donné que le lieu d'articulation bilabial (/b/) est plus visible que le lieu

d'articulation dental (/d/). Nous avons toutefois relevé que chez les sujets enfants, parmi toutes les réponses exclues des *bonnes réponses*, c'est le percept [d] qui était le plus souvent fourni. La dentale représente en effet entre 45,45% et 75,00% des réponses *autres* fournies pour les stimuli /b/, /g/ et /v/; les sujets 2 et 8 n'ont d'ailleurs fourni presque exclusivement que des percepts /d/. Cela suggère donc que chez les sujets enfants, le percept /d/ constituait peut-être un percept « par défaut » pour chaque perception incertaine, parce que fréquent dans le vocabulaire de locuteurs du français québécois de cet âge, et qu'il n'était alors par véritablement mieux reconnu que les stimuli dont le lieu d'articulation est plus visible.

Aussi, la condition unimodale 'visuelle' constituait sans doute la condition du test de perception la plus difficile à comprendre pour les sujets enfants, bien qu'elle leur ait été bien expliquée. En effet, le fait d'être confrontés à des stimuli que visuels ne constitue pas une situation courante dans les jeux ni même dans les activités éducatives d'enfants de cet âge. D'ailleurs, les sujets 7 et 8 n'ont pas fourni de réponses à tous les stimuli 'visuels' présentés. Mais globalement, les enfants ayant fourni de *bonnes réponses* entre 21,43% et 64,29% dans cette condition unimodale 'visuelle', nous pouvons dire qu'ils étaient sensibles aux mouvements articulatoires du locuteur et qu'ils étaient aptes à effectuer une tâche d'identification phonémique à partir de stimuli 'visuels'.

D'autre part, on remarque que la proportion de *bonnes réponses* dans la perception des quatre consonnes à l'étude par les adultes est demeurée presque constante pour chacune d'elles (entre 85,00% et 90,00%). Ces résultats étant tous très élevés, un certain *effet plafond* a pu s'opérer et empêcher de mettre en relief l'échelle de visibilité des lieux d'articulation des quatre consonnes présentées (en ordre croissant de visibilité des mouvements articulatoires, tel que mentionné précédemment: /g/, /d/, /v/, /b/). Toutefois, ces résultats montrent que les adultes étaient très sensibles aux mouvements articulatoires du locuteur et qu'ils étaient compétents pour effectuer une tâche d'identification phonémique à partir de stimuli 'visuels'.

3.2. Conditions test bimodales

La présentation des deux types de stimuli bimodaux 'non-conflictuels' et 'conflictuels' constituait des conditions test. L'analyse des résultats obtenus par les sujets enfants et adultes aux stimuli 'non-conflictuels' a fourni des indicateurs de la validité acoustique et articulatoire des enregistrements. Elle a également permis de s'assurer de la capacité des sujets à effectuer une tâche d'identification phonémique en condition bimodale, mais surtout d'évaluer l'apport de l'information visuelle à la perception des consonnes à l'étude. La condition bimodale 'conflictuelle' a quant à elle permis d'évaluer l'intégration des informations acoustiques et visuelles et mis en lumière le rôle de l'information visuelle dans la perception des consonnes. Les résultats obtenus par les sujets enfants dans chacune des conditions ont été comparés à ceux des sujets adultes afin de voir si les percepts différaientt selon l'âge et si oui, dans quelle mesure ils se distinguaient.

3.2.1. Percepts des stimuli 'non-conflictuels'

Les percepts dont le lieu et le mode d'articulation étaient identiques à ceux des stimuli 'non-conflictuels' présentés ont été jugés comme étant de *bonnes*

réponses. Le tableau 4 montre les percepts *bonnes réponses* et les percepts *autres*, qui ne répondaient pas à ces deux critères.

Tableau 4. Classification des percepts des sujets des stimuli audio-visuels 'non-conflictuels' présentés.

Stimuli		Percepts	
Partie acoustique	Partie visuelle	*Bonnes réponses*	*Autres*
/b/	/b/	[b]	[bv] [v]
/d/	/d/	[d]	[gb]
/g/	/g/	[g]	[d] [dg]
/v/	/v/	[g]	[b] [bv]

Les figures 4 et 5 présentent respectivement la proportion de *bonnes réponses* fournies pour les stimuli 'non-conflictuels' (lignes vertes pointillées) par les sujets enfants et adultes. Sont également reproduits les résultats obtenus aux conditions unimodales 'audio' (ligne orange) et 'visuelle' (ligne bleue) afin de voir les différences de perception entre ces différentes conditions qui regroupent tous les mêmes stimuli.

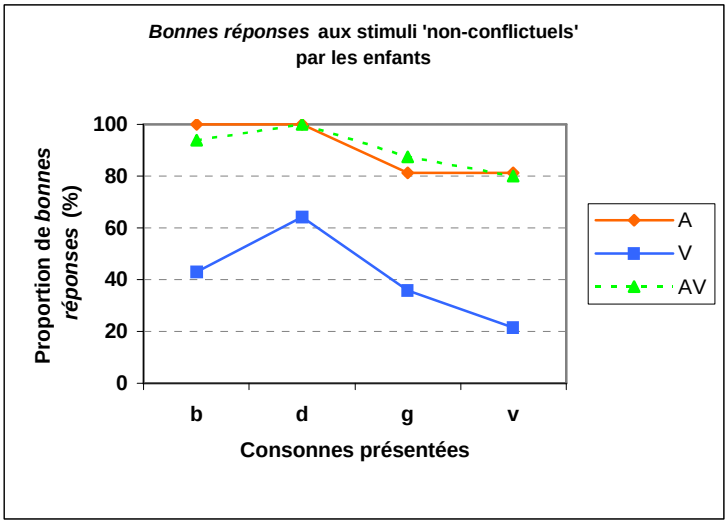

Figure 4. Proportion de *bonnes réponses* fournies par les enfants aux stimuli 'non-conflictuels'.

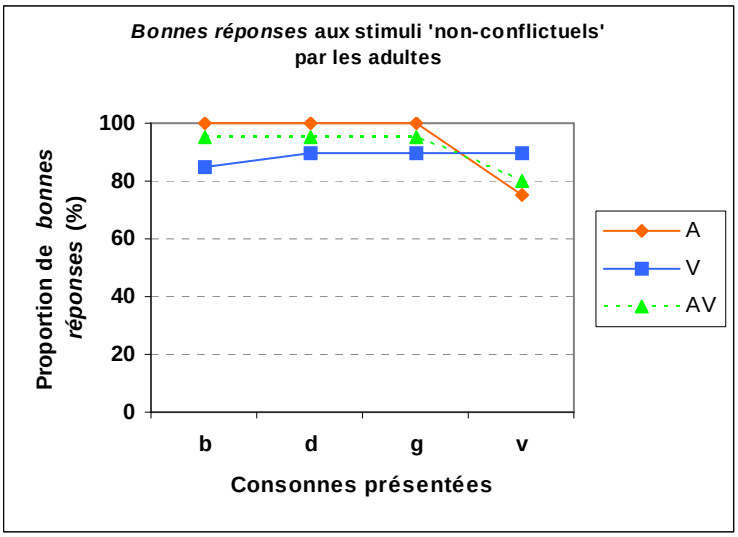

Figure 5. Proportion de *bonnes réponses* fournies par les adultes aux stimuli 'non-conflictuels'.

Regardons tout d'abord la différence entre la proportion de *bonnes réponses* fournies par les enfants et par les adultes (ligne verte pointillée de chaque graphique) aux stimuli 'non-conflictuels'. Il n'y pas de différence significative entre ces deux groupes d'âge pour cette condition. Cela suggère donc que les enfants sont aussi compétents que les adultes pour effectuer une tâche d'identification phonémique de stimuli bimodaux 'non-conflictuels'.

Attardons-nous maintenant aux différences de perception entre les conditions unimodales 'audio' et 'visuelle' et la condition bimodale 'non-conflictuelle' à l'intérieur même des groupes d'âge afin d'évaluer l'apport de chacune des modalités à la perception en condition audio-visuelle. Un test *t* de Student a révélé que chez les sujets enfants, la proportion de *bonnes réponses* est signicativement plus grande dans la condition unimodale 'audio' que dans le condition unimodale 'visuelle' ($t=4{,}745554$ $p<0{,}01$), alors que chez les adultes, il n'y a pas de différence significative entre ces deux conditions. La différence observée chez les enfants entre les percepts des stimuli 'audio' et 'visuels' était prévisible, étant donné leur difficulté plus importante à percevoir les stimuli 'visuels' seuls.

En ce qui a trait à la différence entre les résultats obtenus en condition unimodale 'visuelle' seule et en condition bimodale 'non-conflictuelle', les enfants ont fourni significativement plus de *bonnes réponses* aux stimuli audio-visuels 'non-conflictuels' qu'aux stimuli unimodaux 'audio', tel qu'attendu ($t=-4{,}97240$, $p<0{,}005$). Cette différence suggère que l'apport d'information acoustique à l'information visuelle permet une meilleure perception audio-visuelle des consonnes étudiées.

Aucune différence significative n'a été trouvée entre les percepts *bonnes réponses* à des stimuli unimodaux 'audio' et à des stimuli bimodaux 'non-conflictuels', ni chez les enfants ni chez les adultes. En effet, chez les sujets enfants, seule la vélaire /g/ semble avoir été mieux reconnue en condition bimodale qu'en condition unimodale, la dentale /d/ et la labio-dentale /v/ ayant été presque également reconnues dans les 2 conditions et la bilabiale /b/ ayant même été légèrement mieux reconnue en condition unimodale 'audio'. Ces résultats sont

étonnants, étant donné que le lieu d'articulation vélaire est moins visible que le lieu bilabial.

Chez les adultes, les 3 occlusives /b/, /d/ et /g/ ont été légèrement mieux reconnues en condition unimodale 'audio' qu'en condition bimodale 'non-conflictuelle', ne montrant pas du tout l'apport de l'information visuelle sur la perception bimodale. Aucune différence significative n'a été trouvée chez les sujets adultes entre la proportion de *bonnes réponses* fournies entre les conditions unimodales 'audio' et 'visuelle', ni entre chacune d'elles avec la condition bimodale 'non-conflictuelle'. En fait, les adultes obtenant des résultats près de la note parfaite dans les trois conditions, nous pouvons à nouveau postuler que l'*effet plafond* observé nous empêche de discerner une potentielle différence de performance entre les différentes conditions; ce sont plutôt les résultats obtenus en condition bimodale 'conflictuelle' qui le permettront.

Bref, si les résultats en condition bimodale 'non-conflictuelle' chez les enfants ne sont pas révélateurs d'une performance accrue par rapport à la condition unimodale 'audio', les résultats obtenus à la condition unimodale 'visuelle' suffisent à montrer que les enfants peuvent identifier des consonnes à la seule présentation des mouvements articulatoires qui leur sont associés et qu'ils peuvent traiter cette information visuelle. Les percepts des stimuli bimodaux 'conflictuels' présentés à la prochaine section permettront également de montrer l'importance de l'information visuelle dans la perception de la parole chez les deux groupes de sujets.

3.2.2. Percepts des stimuli 'conflictuels'

Rappelons-nous que les stimuli bimodaux 'conflictuels' étaient formés d'un signal acoustique d'une certaine séquence $voyelle_i$-$consonne_j$-$voyelle_i$ où la consonne était /b/, /d/, /g/ ou /v/ et d'un signal visuel $voyelle_i$-$consonne_k$-$voyelle_i$ dont la consonne différait de celle de la partie acoustique. Nous avons classifié les percepts auxquels ces stimuli 'conflictuels' ont donné lieu en 5 types : *audio*, *visuels*, *combinaisons*, *fusions* et *autres*. Dans le tableau 5 se trouvent des exemples de percepts de chaque type, tirés de notre corpus. Nous verrons alors clairement les conflits qui ont cours entre les modalités visuelle et auditive lors de la perception de la parole.

Tableau 5. Exemples de la classification des percepts des sujets des stimuli audio-visuels 'conflictuels' présentés.

Stimuli		Percepts					
				Combinaisons			
Partie acoustique	Partie visuelle	*Audio*	*Visuels*	*strictes*	*étendues*	*Fusions*	*Autres*
/b/	/d/	[b]	[d]	-	-	-	[t] [vg]
/b/	/g/	[b]	[g]	-	-	[d] [θ]	-
/g/	/b/	[g]	[b]	[bg] [gb]	-	-	-
/g/	/v/	[g]	-	[vg]	[bg]	[d]	[f]
/v/	/d/	[v]	[d]	-	[bv] [lv] [zd]	[th]	[t]

Les percepts audio sont ceux dont le lieu et le mode d'articulation correspondaient à ceux de la partie acoustique des stimuli bimodaux 'conflictuels'. Par exemple, pour un stimulus A /b/ - V /g/, un percept [b] constitue un percept *audio*.

Les percepts qui ont été placés dans la catégorie *visuels* sont ceux dont le lieu et le mode d'articulation correspondaient à ceux de la partie visuelle des stimuli bimodaux 'conflictuels' présentés. Par exemple, pour un stimulus A /v/ - V /g/, un percept [g] constituait un percept *visuel*.

Jusqu'à présent, les percepts consonantiques dont nous avons parlé n'étaient formés que d'un seul phonème. Or, les percepts *combinaisons* sont ceux qui comportent deux phonèmes. Au départ, notre classification des percepts *combinaisons* était stricte et n'admettait que les percepts constitués de la consonne de la partie auditive et de la consonne de la partie visuelle, peu importe l'ordre dans lequel ils étaient identifiés. Par exemple, pour un stimulus A /g/ - V /b/, les percepts [bg] et [gb] ont été classés comme étant des *combinaisons strictes*. Nous avons toutefois décidé d'élargir notre définition de *combinaisons* de façon à pouvoir inclure les percepts biphonémiques dont seul un des phonèmes était directement issu de la partie acoustique ou de la partie visuelle de stimuli 'conflictuels'. Par exemple, pour un

stimulus A /v/ - V /d/, les percepts [bv], [lv] et [zd] ont été classés comme étant des *combinaisons étendues*.

Lorsque les sujets ont identifié les stimuli 'conflictuels' à l'aide d'une seule consonne et que cette consonne n'était ni celle de la partie acoustique des stimuli ni celle de la partie visuelle, mais plutôt une consonne «intermédiaire», une «fusion» de l'information fournie par les 2 modalités, nous avons classé ces percepts comme étant des *fusions*. Nous avons par exemple retrouvé des percepts *fusions* [d] lors de la présentation d'un stimulus A /b/ - V /g/.

Finalement, toutes les réponses qui ont été fournies par les sujets et qui ne répondaient pas aux critères des 4 autres catégories mentionnées précédemment ont été placées dans la catégorie *autres*. Par exemple, un percept [vg] à la suite de la présentation d'un stimulus A /b/ - V /d/ a été placé dans la catégorie *autres*.

3.2.2.1. Percepts audio

Les percepts *audio* montrent l'absence de perception conflictuelle entre les modalités auditive et visuelle ou la prépondérance de la modalité auditive. La figure 6 illustre la moyenne d'identification des stimuli 'conflictuels' présentés comme étant des percepts *audio*.

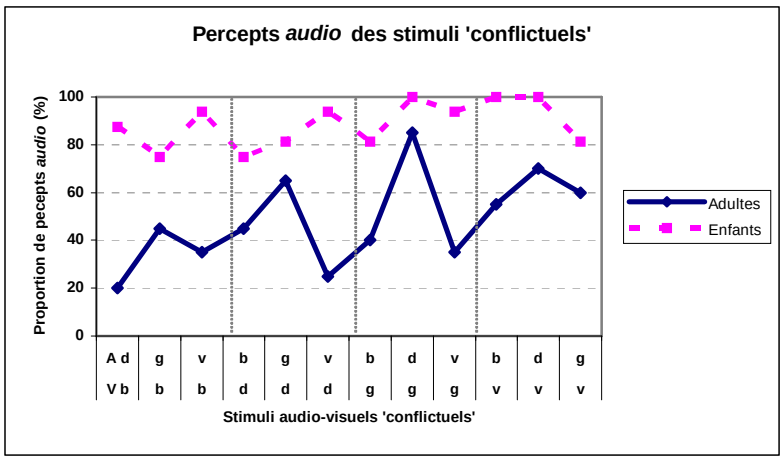

Figure 6. Moyenne des identifications des stimuli bimodaux 'conflictuels' comme étant des percepts *audio*, c'est-à-dire des percepts qui ne reflètent le lieu et le mode d'articulation que de la partie acoustique des stimuli présentés.

La proportion de percepts *audio* est significativement plus élevée chez les sujets enfants ($F(1,22)=42,098$, $p<0,001$) que chez les sujets adultes. Ces résultats vont dans le même sens que ceux observés dans les conditions unimodales chez les enfants, dans la mesure où ces derniers semblaient davantage sensibles à l'information acoustique. Nous pouvons identifier une tendance dans la distribution des percepts *audio* chez les enfants : les stimuli dont les lieux d'articulation sont les plus rapprochés donnent lieu à davantage de percepts *audio*. Cette tendance se manifeste dans les stimuli impliquant les conflits de lieux d'articulation dental - vélaire, bilabial - labio-dental et dental - labio-dental, tels que les stimuli A /d/ -V /g/, A /v/ - V /b/, A /b/ - V /v/ et A /v/ - V /d/ et A /d/ - V /v/. Cette tendance montre que les enfants sont plus sensibles à l'information acoustique là où le moins

d'information visuelle conflictuelle est fournie, tel qu'attendu. Il est à noter que même si les adultes manifestent moins de percepts *audio* que les enfants, la même tendance prévaut pour les mêmes lieux d'articulation et impliquent des stimuli tels que A/g/ - V /d/, A /g/ - V /d/ et A /d/ - V /v/.

Ces résultats montrent donc que les sujets enfants sont plus sensibles à l'information acoustique, mais nous verrons dans les prochaines sections qu'ils manifestent tout de même d'autres formes de percepts faisant intervenir le traitement de l'information visuelle dans neuf des douze contextes conflictuels.

3.2.2.2. Percepts visuels

Les percepts 'visuels' montrent la prépondérance de la modalité visuelle sur la modalité auditive. La figure 7 illustre la moyenne d'identification des stimuli 'conflictuels' présentés comme étant des percepts *visuels*.

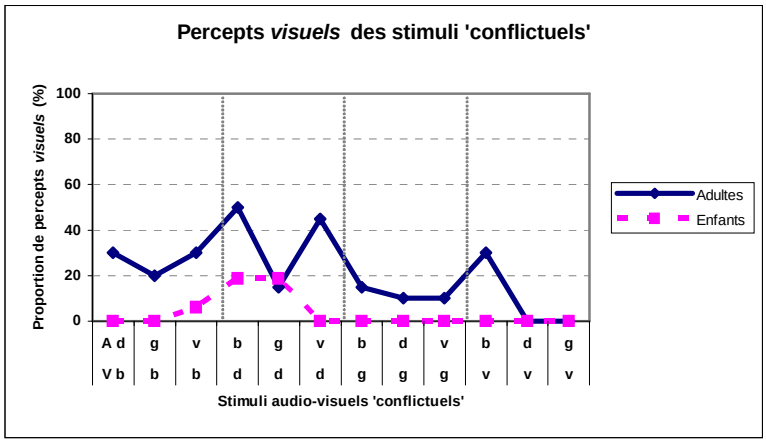

Figure 7. Moyenne des identifications des stimuli bimodaux 'conflictuels' comme étant des percepts *visuels*, c'est-à-dire des percepts qui ne reflètent le lieu et le mode d'articulation que de la partie visuelle des stimuli présentés.

La proportion de percepts *visuels* est significativement plus petite chez les enfants que chez les adultes ($F(1,22)=11,895$, $p<0,01$). Chez les enfants, on retrouve tout de même trois contextes dans lesquels leurs percepts ne représentaient que les caractéristiques de la partie visuelle des stimuli. Chez les adultes, ce sont les stimuli dont la partie visuelle impliquait la bilabiale /b/ et la dentale /d/ qui ont généré les plus grandes proportions de percepts *visuels*.

La présence de percepts *visuels* marque une importante influence de la modalité visuelle dans la perception de la parole; les stimuli étaient présentés dans une pièce silencieuse où peu de bruits environnants pouvaient venir distraire les sujets dans leur perception de la partie acoustique des stimuli. Pourtant, dans leur intégration des informations acoustiques et visuelles, les enfants et les adultes se sont appuyés sur le lieu d'articulation vu plutôt qu'entendu dans plusieurs contextes, les adultes plus fréquemment que les enfants.

3.2.2.3. Percepts combinaisons

Les percepts *combinaisons* marquent de façon explicite l'influence simultanée des informations fournies par les modalités visuelle et auditive. La figure 8

illustre la moyenne d'identification des stimuli 'conflictuels' présentés comme étant des percepts *combinaisons* par les sujets enfants et adultes.

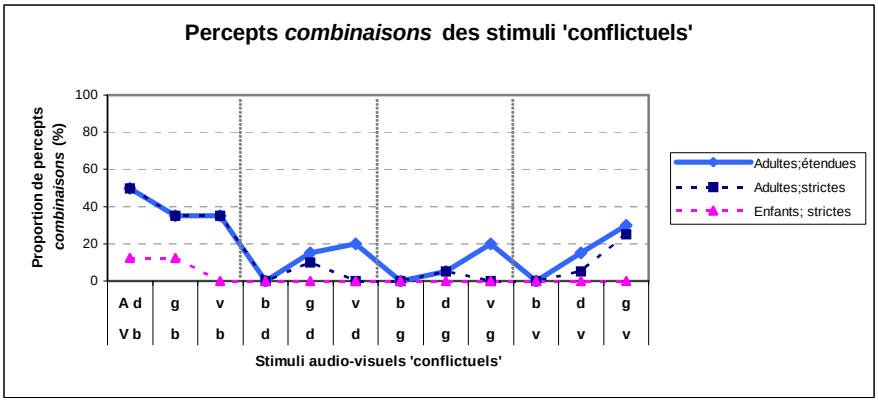

Figure 8. Moyenne des identifications des stimuli bimodaux 'conflictuels' comme étant des percepts *combinaisons*. Sont représentés les percepts *combinaisons strictes*, c'est-à-dire les percepts formés des consonnes composant les parties acoustique et visuelle des stimuli 'conflictuels', et les percepts *combinaisons étendues*, c'est-à-dire les percepts formés d'au moins un des deux phonèmes des parties acoustique et visuelle des stimuli 'conflictuels'.

La proportion de percepts *combinaisons* est significativement plus petite chez les enfants que chez les adultes ($F(1,22)=11,579$, $p<0,01$). On peut remarquer que les enfants ne présentent que des *combinaisons strictes*, ce qui fait que nous avons comparé le nombre de combinaisons strictes des enfants avec le nombre de *combinaisons étendues* des adultes. Les enfants présentent des percepts *combinaisons* dans deux contextes : A /d/ - V /b/ et A /g/ - V /b/. On peut remarquer que ce sont dans ces mêmes contextes que les adultes en présentent le plus, ce qui suggère que les mouvements articulatoires de la bilabiale /b/ induisent plus de percepts *combinaisons* lorsque combinée avec une contre-partie acoustique dont le lieu d'articulation est moins visible (dental /d/ et vélaire /g/).

Ces résultats montrent que les enfants sont sensibles à l'information visuelle présente dans les stimuli 'conflictuels', mais qu'ils le sont moins que les adultes.

3.2.2.3. Percepts fusions

Les sujets étant des locuteurs natifs du français et l'inventaire des consonnes du français étant limité, tous les stimuli 'conflictuels' n'étaient pas susceptibles de donner lieu à un percept *fusion*. Il est à noter que le percept fusion constitue le percept le plus «typique» de l'effet McGurk. La figure 9 illustre la moyenne d'identification des stimuli 'conflictuels' présentés comme étant des percepts *fusions* par les sujets enfants et adultes.

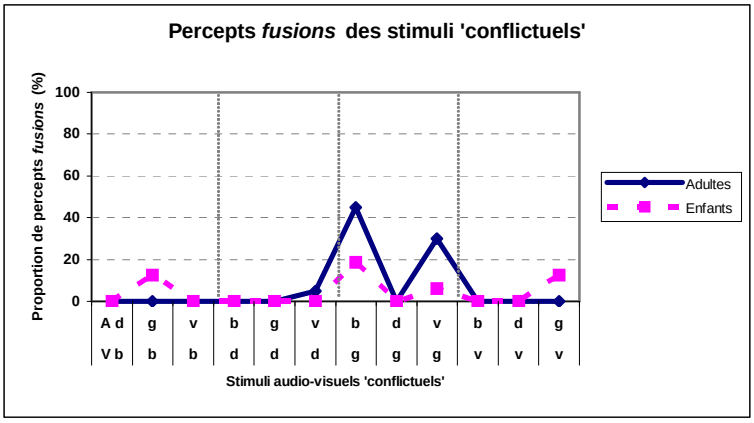

Figure 9. Moyenne des identifications des stimuli bimodaux 'conflictuels' comme étant des percepts *fusions*, c'est-à-dire les percepts formés d'une seule consonne, qui n'est ni celle de la partie acoustique des stimuli, ni celle de la partie visuelle, mais plutôt une consonne intermédiaire à celles-ci.

La proportion de percepts *fusions* chez les enfants n'est pas significativement différente de celle des adultes. Il est toutefois intéressant de remarquer que les enfants présentent des percepts *fusions* dans quatre contextes alors que les adultes n'en présentent que dans deux contextes. Les stimuli A /b/ -V /g/ semblent être ceux qui donnent le plus souvent lieu à des percepts *fusions* /d/ ou /θ/. Ces résultats montrent que les enfants en les adultes sont sensibles à l'information visuelle et que dans certains contextes, ils l'intègrent à l'information acoustique fournie et forme des percepts *fusions*.

3.2.2.4. Percepts autres

La figure 10 illustre la moyenne d'identification des stimuli 'conflictuels' présentés comme étant des percepts *autres* par les sujets enfants et adultes.

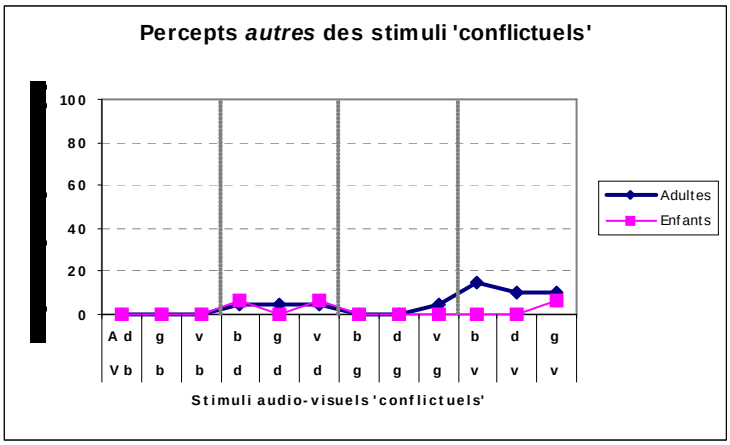

Figure 10. Moyenne des identifications des stimuli bimodaux 'conflictuels' comme étant des percepts *autres*, c'est-à-dire tous les percepts qui ne correspondaient pas autres critères des catégories *visuels*, *audio*, *combinaisons* et *fusions*.

La proportion de percepts *autres* chez les enfants n'est pas significativement différente de celle des adultes. Certains de ces percepts ne s'expliquent pas facilement par les indices acoustiques et visuels des stimuli (par exemple, un stimulus

A /b/ - V /d/ perçu comme /vg/) et peuvent constituer des percepts «aléatoires», à défaut de pouvoir fournir un percept précis et certain de la part des sujets. Par contre, l'interprétation d'autres percepts *autres* peut faire intervenir l'information visuelle; chez les adultes, 36,36% des percepts *autres* sont formés de la partie visuelle des stimuli 'conflictuels', mais dévoisée. Par exemple, un stimulus A /g/ - V /v/ donne lieu à un percept /f/.

Il est donc intéressant de noter que les enfants n'ont pas fourni significativement plus de percepts *autres* que les adultes. En effet, on aurait pu croire qu'ils fourniraient plus de réponses aléatoires dû à la difficulté de la tâche ou à un manque d'attention.

4. Discussion générale

Plusieurs différences entre les percepts de nos sujets d'âge préscolaire et de nos sujets adultes se sont avérées significatives. Globalement, les sujets enfants se sont avérés plus sensibles à l'information acoustique. D'ailleurs, la plus petite proportion de leurs percepts *bonnes réponses* aux stimuli unimodaux 'visuels' et la plus grande proportion de leur percepts *audio* aux stimuli bimodaux 'conflictuels' rejoignent le postulat de Massaro (1984) selon lequel l'information auditive a une plus grande influence sur leur perception des catégories phonétiques dans l'acquisition du langage chez les enfants.

De plus, nos sujets adultes ayant manifesté beaucoup de percepts *visuels* en condition unimodale 'conflictuelle' et nos sujets enfants ayant manifesté beaucoup de percepts *audio,* ils reflètent une tendance qu'avaient notée McGurk et MacDonald (1976) selon laquelle lorsque le percept d'un stimulus conflictuel ne représentait qu'une seule modalité, les percepts des adultes provenaient plus souvent de la modalité visuelle et ceux des enfants provenaient plus souvent de la modalité auditive.

Cependant, si nos sujets enfants ont été davantage influencés par l'information acoustique fournie, ils ont tout de même manifesté l'effet McGurk, mais cette manifestation a été plus faible et plus sujette à la variabilité que chez les adultes. Ces résultats sont évocateurs du rôle de la vision dans le développement du système phonologique.

Ces résultats obtenus auprès de sujets francophones vont dans le même sens que ceux obtenus par Massaro (1984). Mis en parallèle avec les résultats de Rosenblum (1997) montrant un effet McGurk fort chez des bébés âgés de 5 mois pour certaines séquences audio-visuelles conflictuelles, nos résultats suggèrent que la pente du développement de la perception de la parole bimodale ne soit pas toujours ascendante. Nous proposons le postulat selon lequel le développement des mécanismes de perception de la parole audio-visuelle progresse en forme de U (*U shape*), c'est-à-dire que très jeunes, les bébés intègrent les informations auditives et visuelles présentées, puis au cours de l'enfance, une «surcharge» cognitive étant causée par le développement rapide de nombreuses habiletés, l'efficacité d'autres habiletés déjà acquises diminue, expliquant les résultats obtenus par nos sujets d'âge préscolaire, et parvient à un niveau optimal une fois ce développement cognitif achevé, soit vers l'âge adulte.

Il serait intéressant de poursuivre des recherches dans ce sens auprès de plus de sujets dont l'étendue des âges est plus grande, de façon à pouvoir cerner à quel moment est atteint le plateau d'intégration optimale des informations auditives et

visuelles et à pouvoir corréler cette période à l'atteinte d'habiletés cognitives particulières.

Références

MacDonald, J. M., H. (1978). "Visual influences on speech perception process." Perception and Psychophysics 24: 253-257.

Massaro, D. W., Cohen, M. M. (1983). "Evaluation and Integration of Visual and Auditory Information in Speech Perception." Journal of Experimental Psychology: Human Perception and Performance 9 (5): 753-771.

Massaro, D. W. (1984). "Children's Perception of Visual and Auditory Speech." Child Development 55: 1777-1888.

McGurk H., M., J. (1976). "Hearing lips and seeing voices." Nature 264: 746-748.

Robert-Ribes, J., Schwartz, J-L, Tallouache, T., Escudier, P. (1998). "Complementary and synergy in bimodal speech: Auditory, visual, and audio-visual identification of French oral vowels in noise." Journal of Acoustical Society of America 103 (6): 3677-3689.

Rosenblum, L. D., Schmuckler, M.A., & Johnson, J.A. (1997). "The McGurk effect in infants." Perception & Psychophysics 59(3): 347-357.