Intégration audiovisuelle dans le traitement du langage

Agathe Mestrallet, Pierre Martin

Mémoire de fin d'études en vue de l'obtention du
CERTIFICAT DE CAPACITÉ D'ORTHOPHONISTE

Par Pierre MARTIN et Agathe MESTRALLET

Centre de Formation Universitaire en Orthophonie de Marseille

INTÉGRATION AUDIOVISUELLE DANS LE TRAITEMENT DU LANGAGE

Sous la direction de Chotiga PATTAMADILOK et Marc SATO

Membres du jury
Chotiga PATTAMADILOK
Marc SATO
Florence TOUZE-LAVANDIER
Michel PITERMANN

Soutenu le 16/06/2021
à Aix-en-Provence

REMERCIEMENTS

Nous remercions sincèrement Chotiga Pattamadilok et Marc Sato pour leur encadrement tout au long de ce mémoire. Leurs conseils et encouragements ont été précieux, notamment pour nous immerger dans les lectures scientifiques liées au domaine que nous avons exploré. C'est grâce à leur accompagnement, leur bienveillance et leur disponibilité que nous avons pu mener ce mémoire à terme, et ainsi découvrir la recherche expérimentale.

Nous tenons également à adresser nos remerciements à Florence Touze-Lavandier et Michel Pitermann pour avoir accepté de constituer le jury de ce mémoire. Nous les remercions pour leur disponibilité, pour leur relecture attentive de notre mémoire ainsi que pour l'intérêt qu'ils ont porté à ce projet.

Nous adressons un remerciement particulier au Laboratoire Parole et Langage, qui nous a permis de mener à bien notre expérience en EEG.

Nous remercions également Mathilde Cans et Faustine Charignon, Marie-Sophie Villain--Bailly et Eléonore Clot, Amélie Brassart et Caroline Gruffy, qui ayant aussi fait un projet de mémoire au LPL encadré par Chotiga Pattamadilok, nous ont conforté dans notre envie de nous engager dans ce projet.

De plus, nous remercions les personnes ayant répondu présentes pour la participation à notre expérience, ainsi que celles qui ont répondu à notre pré-questionnaire de participation.

Nous tenons aussi à remercier toutes les personnes qui, de près ou de loin, nous ont soutenu dans la réalisation et l'écriture de ce mémoire. Il marque l'aboutissement de 5 années d'études riches en apprentissages au sein du Centre de Formation Universitaire en Orthophonie de Marseille.

Ce mémoire de recherche nous encourage ainsi à nous former de manière continue et à nous tenir constamment au courant des recherches en lien avec les domaines de l'orthophonie.

TABLE DES MATIÈRES

PARTIE THÉORIQUE

En 2020, la France est touchée par une pandémie due au COVID-19. Cette crise sanitaire mène alors à la mise en place de mesures spécifiques, et notamment le port d'un masque couvrant le nez et la bouche dans l'espace public. Dans le cadre d'une conversation, l'absence de perception visuelle des lèvres et de la bouche de l'interlocuteur ou de l'interlocutrice peut cependant entraver la compréhension du message linguistique perçu. Au-delà de la nécessité d'une telle mesure visant à freiner la progression du COVID-19, nous pouvons alors nous questionner sur l'influence possible du port d'un masque sur la communication verbale, à partir de cette période.

Ce projet de recherche s'inscrit, bien que de manière indirecte, dans cette perspective. Il a en effet pour objectif de mieux comprendre les mécanismes d'intégration des informations auditives et visuo-labiales, mais aussi auditives et visuo-orthographiques, lors du traitement du langage. Plus spécifiquement, nous nous intéresserons aux mécanismes neuronaux d'intégration mis en jeu lors de la perception simultanée de stimuli langagiers acoustiques et visuels vers l'émergence d'un percept « unifié ». Pour ce faire, nous étudierons les influences de ces types de stimuli visuels sur le traitement auditif de la parole et de là, l'apport et la contribution des informations véhiculées par chaque modalité dans la construction et l'émergence de ce percept.

Qu'elle se définisse par la perception des mouvements articulatoires de la parole d'une personne ou bien par la lecture d'un texte, l'information visuelle véhicule de nombreux indices facilitant l'accès au message linguistique. Ces derniers sont néanmoins bien différents. Premièrement, les temporalités d'acquisition au cours du développement divergent. La parole est l'expression physiologique du langage et s'acquiert de manière autonome dès le plus jeune âge, alors que l'apprentissage du code abstrait orthographique est plus tardif et nécessite un apprentissage intensif et supervisé, qui a généralement lieu à l'école. De plus, la nature de ces indices diffère dans leurs aspects dynamiques : la lecture labiale est appréhendée au travers de la cinématique des articulateurs de parole selon un prisme sensorimoteur, alors qu'un support verbal orthographique présente une forme figée et une relation arbitraire avec le signal acoustique de parole.

Du fait même de ces différences, l'étude conjointe des mécanismes d'intégration audio-visuo-labiaux et audio-visuo-orthographiques devrait permettre une meilleure compréhension de l'impact des processus perceptifs liés aux informations labiales et orthographiques dans la perception et la compréhension du langage.

1. La parole, expression physiologique du langage

Lors d'une situation de conversation orale, chaque auditeur ou auditrice s'appuiera sur plusieurs types d'informations pour appréhender le message linguistique qui lui est adressé, que nous réduisons souvent aux seules informations auditives. Ces informations auditives peuvent être de nature verbale, correspondant aux mots et phrases prononcés, et paraverbales, comme la prosodie ou l'intonation. A ces informations s'ajoutent néanmoins d'autres aspects, perceptibles visuellement, comme les mimiques et expressions du visage, ou encore les différentes configurations articulatoires observables. C'est l'ensemble de ces informations auditives, verbales et paraverbales, et visuelles qui nous permettront de segmenter le signal de parole en unités discrètes infra-lexicales, puis d'apparier ces unités avec les représentations lexicales vers une compréhension du message linguistique.

La réception du message linguistique véhiculé par la parole est donc conditionnée par la perception de diverses informations, qu'elles soient auditives ou visuelles. Ce message peut être analysé par le prisme d'une approche **multimodale** et **multisensorielle** (Rosenblum, 2019). Notre perception de la parole est conditionnée par l'analyse de ce que nous entendons, c'est-à-dire par l'information auditive, ainsi que par l'information visuelle, qui nous permet d'obtenir divers indices essentiels à la bonne intégration du message.

Selon le Dictionnaire d'Orthophonie (Brin-Henry et al., 2011), la **parole** correspond à l'action de parler. Sa production se définit comme un acte physiologique volontaire, visant à produire un message grâce à la mise en jeu de divers procédés anatomiques, neurologiques et culturels.

La parole, et donc l'acte de parler, pourrait être définie comme l'agencement dynamique, complexe et volontaire de mouvements articulatoires produisant un signal acoustique de parole continue. Ce flux de parole peut être décomposé en unités sonores minimales distinctives appelées **phonèmes**. Leur production implique cependant des

phénomènes de coarticulation entraînant des modifications mutuelles articulatoires et acoustiques importantes. En effet, le passage d'un phonème à l'autre, dans un flux de parole continu, induit des changements spécifiques aux caractéristiques articulatoires des phonèmes concernés. Ces phénomènes répondent notamment à l'importance de minimiser, ou "lisser", l'effort articulatoire lors de la production de la parole et ont pour conséquence qu'un même phonème puisse posséder des propriétés articulatoires et donc acoustiques différentes selon les phonèmes qui le précèdent et le suivent. Face à ce phénomène de coarticulation, si la multitude des schémas articulatoires envisageables et la dynamique complexe de ces mouvements ne permettent pas de décrire avec exactitude la construction articulatoire d'une succession de sons, certains traits articulatoires peuvent tout de même être isolés. C'est cette base phonémique qui permettra à son tour de former des mots, puis des phrases, pour ainsi construire le sens du message prononcé.

a. Mouvements articulatoires et nature visémique de la parole

Pour mener à bien la production orale d'un message, il convient de mettre en jeu divers **processus physiologiques**.

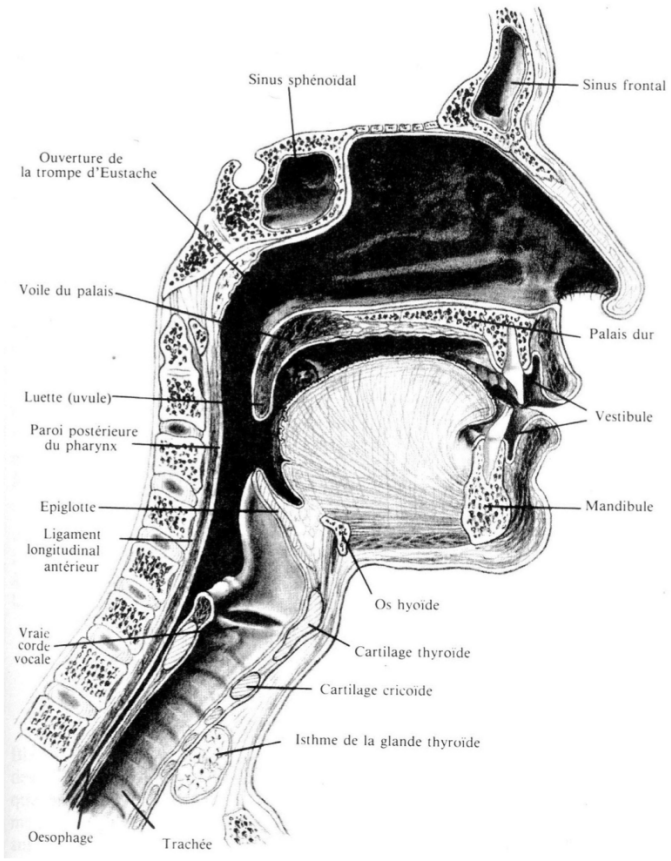


Figure 1. Anatomie du larynx et du pharynx (Bunch, 1982)

Dans un premier temps, les poumons produisent l'air nécessaire à la phonation. Cet air, après avoir emprunté la trachée, se rend dans le larynx et entraîne la vibration des plis vocaux - plus communément appelés "cordes vocales" - grâce à un phénomène de pression sous-glottique. L'air traverse ensuite divers articulateurs et résonateurs, situés pour la plupart au sein du pharynx et de la cavité buccale, qui influent sur la forme du tractus vocal et jouent un rôle de filtre. Parmi ceux-ci, on désigne notamment la langue, les joues, le voile du palais, les dents, etc. L'air sort ensuite par le nez et/ou la bouche, en produisant un son qui sera perçu par l'interlocuteur ou l'interlocutrice.

Les différentes configurations des articulateurs mentionnées plus haut permettent une description, bien que partielle, des caractéristiques acoustiques des sons et, de là, des phonèmes que nous produisons. Plusieurs aspects peuvent notamment être cités.

Les phonèmes consonantiques peuvent être caractérisés par la présence d'un voisement, phénomène dépendant de la mise en jeu, ou non, des plis vocaux. Nous observons également leur mode d'articulation, c'est-à-dire la manière dont l'air va être obstrué. Le point ou lieu d'articulation nous permet enfin d'appréhender où cet air sera obstrué. Les phonèmes vocaliques, quant à eux, sont caractérisés par leur degré d'ouverture et leur point d'articulation.

Par exemple, les phonèmes /p/ et /b/ peuvent être différenciés grâce au critère de voisement (i.e. la vibration, ou non, des cordes vocales) : /p/ est un phonème dit "sourd" (i.e. qui ne met pas en jeu la vibration des cordes vocales), à l'inverse du phonème /b/, qui est décrit comme un phonème "sonore". Le mode d'articulation se décline, quant à lui, sous diverses modalités. Lorsque l'air s'accumule suite à une occlusion du tractus vocal pour finalement être relâché, nous qualifions les phonèmes produits d'"occlusifs" (e.g., /p/, /b/, /t/, /d/, /k/, /g/). Lorsque la fermeture effectuée n'est pas complète et laisse donc passer l'air à travers un tractus vocal préalablement resserré, des frictions peuvent être observées. Ce phénomène nous permet la production de phonèmes "fricatifs" (e.g., /f/, /v/, /s/, /z/, /ch/, /ʒ/). Une mise en jeu du voile du palais, avec l'orientation de l'air vers les cavités nasales, nous permet de produire des phonèmes "nasaux" (e.g., /m/, /n/). Enfin, en parallèle de ces deux caractéristiques, nous pouvons également décrire le lieu d'articulation des phonèmes. Ce dernier nous permet notamment d'appréhender quelles structures anatomiques permettent les spécificités du phonème produit. Par exemple, les phonèmes /p/, /b/ et /m/ demandent une

action de fermeture des lèvres spécifique à l'expression de ces sons. Nous parlerons de phonèmes "bilabiaux". D'autres lieux d'articulation peuvent être cités : labiodental, dental, alvéolaire, palatal, uvulaire, etc.

Nous avons vu que la production des phonèmes se traduit par la mise en jeu spécifique de processus articulatoires. La perception visuelle de ces mouvements et configurations articulatoires représente un apport non négligeable dans certaines situations de perception de la parole. Néanmoins, certaines configurations articulatoires sont plus faciles à isoler visuellement que d'autres : nous parlerons de **saillance perceptive** visuelle (MacLeod et Summerfield, 1987, op. cit. in Fort, 2011). Il est par exemple impossible de percevoir visuellement le voisement d'un phonème (sauf dans le cadre d'un entraînement ou d'une situation de communication exceptionnelle), puisque celui-ci se traduit par la présence d'un mécanisme de vibration des cordes vocales, situées au sein du larynx et donc très peu ou non perceptible visuellement. De même, les phénomènes de nasalisation ne sont pas visuellement observables, ces derniers dépendant d'une action située au niveau du voile du palais. A l'inverse, certains lieux d'articulation sont plus facilement perceptibles, notamment lorsque ces derniers se situent au niveau des lèvres. Les groupes de phonèmes dits bilabiaux (e.g. /p/, /b/ et /m/) et labiodentaux (e.g. /t/, /d/) seront donc discriminés plus facilement. Pour les groupes de phonèmes vocaliques, le degré d'ouverture représente également un élément saillant.

C'est notamment dans l'optique de caractériser visuellement ces mouvements articulatoires qu'a été introduite la notion de **visème** (Fisher, 1968). De la contraction des mots "visuel" et "phonème", "la notion de visème renvoie à toute unité minimale de la parole qui est perçue comme visuellement distincte d'une autre unité." (Fisher, 1968, op. cit. in Fort, 2011), il s'agit donc de la représentation visuelle d'un phonème. Comme nous l'avons vu, certaines particularités comme le mouvement du voile du palais lors de la nasalisation d'un phonème ou le voisement ne sont pas perceptibles, alors que des mouvements labiaux le sont plus facilement. Pour illustrer ce concept, nous donnerons comme exemple les phonèmes /p/, /b/ et /m/. Ces derniers présentent la même configuration labiale, leurs différences de production se situant au niveau de phénomènes non perceptibles visuellement (voisement, nasalisation) : ils sont ainsi décrits par une même unité visémique. Les mots "pain", "bain" et "main" sont donc perçus visuellement de la même manière.

b. L'intégration audio-visuo-faciale

Après avoir présenté brièvement les mécanismes articulatoires mis en œuvre lors de la production de parole, nous nous attarderons sur l'apport des informations visuelles des mouvements articulatoires lors de la perception de la parole. Il a notamment été montré que l'information visuelle seule permettrait de discriminer 40% à 60% des phonèmes d'une langue et 10% à 20% des mots de la langue française (Schwartz, 2011, op. cit. in Fort, 2011).

Différentes études se sont intéressées à l'influence de la **perception du visage entier** comme indice visuel facilitant la compréhension de la parole. Dans un premier temps, il semble important de percevoir la différence entre "speechreading" (i.e. la lecture visuelle de la parole, avec la possibilité de voir le visage entier) et "lipreading" (i.e. lecture labiale) (Strelnikov, 2009). Dans ce paragraphe, nous parlerons de "speechreading". De manière générale, les performances en speechreading sont meilleures par comparaison avec une situation où seule la zone orale (i.e. les lèvres, soit en "lipreading") du visage est visible (Jordan et Thomas, 2011). Cependant, dans le cas de la perception d'un visage où la zone orale est cachée, l'identification visuelle reste possible. Bien que le masquage des lèvres réduise l'identification visuelle du message produit, les autres indices situés au niveau du visage nous permettent d'émettre des hypothèses quant à la configuration des articulateurs cachés, comme les lèvres (Jordan et Thomas, 2011). Ces données nous permettent donc d'appréhender l'importance de l'apport du visage entier dans notre compréhension du message : les indices visuels ne se résument pas aux seuls mouvements labiaux.

L'importance de ces éléments visuels prend tout son sens lorsque nous décrivons les stratégies mises en place par des **personnes malentendantes** lors de situations de communication orale. Les indices visuels permettent de pallier le manque d'informations auditives, pour ainsi permettre un meilleur accès au sens. De même, il a été démontré que, chez des personnes implantées, l'apport de l'information visuelle reste essentiel pour permettre aux personnes malentendantes de mieux percevoir le message (Strelnikov, 2009). D'autres expériences ont pu également montrer l'importance de la modalité visuelle en situation acoustique bruitée ou lors de l'apprentissage d'une seconde langue (Fort, 2011).

La perception audiovisuelle de la parole a également fait l'objet d'études explorant l'usage de stimuli incongruents. Une étude a notamment mis en évidence que, lorsqu'un

message auditif est présenté en même temps que des mouvements labiaux correspondant à un message différent, une “fusion perceptive” pouvait avoir lieu dans notre perception de cet ensemble syllabique : c’est l’**effet McGurk** (McGurk et MacDonald, 1976). Ainsi, lorsqu’un /ba/ auditif est présenté en même temps qu’un /ga/ visuel, le phonème /da/ pourra être perçu par la personne. Il est à noter que cet effet diffère d’un sujet à l’autre, ainsi qu’en fonction des stimuli présentés. Cette illusion perceptive nous permet d’appréhender l’influence qu’a l’information visuelle sur l’intégration d’un message, et ce même lorsque le stimulus auditif est parfaitement audible (McGurk et MacDonald, 1976 ; Rosenblum, 2019 ; Treille, 2017).

Une des caractéristiques de l’intégration audio-visuelle est son **aspect prédictif**. En effet, notre cerveau possède la capacité d’anticiper et d’émettre des hypothèses sur la suite des informations produites. Par exemple, en perception de la parole, il a été admis que l’information auditive succédait l’information des lèvres, avec une différence d’environ 200ms (Van Wassenhove et al., 2007). Certaines situations non verbales nous permettent d’appréhender cette notion d’intégration prédictive. Par exemple, visualiser un applaudissement induit l’arrivée du signal acoustique à mesure que les mains se rapprochent. A l’inverse, l’action de déchirer une feuille ne permet pas cette prédiction, et ce puisque l’arrivée de l’information auditive se fait en même temps que le début de l’action (Stekelenburg et Vroomen, 2007).

En résumé, les relations entre informations visuelles et acoustiques lors de la perception et l’intégration audiovisuelle sont complexes, qu’il s’agisse de la correspondance entre configuration anatomique du conduit vocal et propriétés acoustiques du signal, de la saillance et possible ambiguïté visuelle, de la prise en compte de ces informations visuelles en fonction de l’environnement et de la qualité du signal acoustique, ou encore de la dynamique et précédence variable de ces informations visuelles sur les informations acoustiques. Pour les appréhender au mieux, il convient de prendre en compte leurs natures prédictives et multisensorielles, les divers aspects qui régissent notre compréhension d’un message, ainsi que l’importance et la complémentarité des flux auditifs et visuels.

2. Le langage écrit, système graphique du langage

Le langage écrit, quant à lui, est apparu phylogénétiquement après le langage oral. L'invention d'un système d'écriture permet alors un nouvel accès à de nombreuses connaissances, et ainsi de stocker des informations jusqu'ici conservées par la tradition orale. En français, son apprentissage se fait sur la base d'un code graphique alphabétique, enseigné à l'école et au cours du développement de l'enfant. On peut le désigner comme le deuxième code visuel (i.e. après le code naturel visémique), à la fois artificiel, non naturel et sans aspect dynamique. Malgré un processus d'apprentissage intensif, il devient quasiment automatique une fois maîtrisé. Nous étudierons donc dans cette partie les caractéristiques qui décrivent son apprentissage et son effet sur le langage oral. D'après le Dictionnaire d'Orthophonie (Brin-Henry et al., 2011), le **langage écrit** "recouvre à la fois la compréhension (i.e. la lecture) et la production ou l'expression (i.e. l'écriture), d'un système codé en signes graphiques permettant, sur tout support possible, la transmission d'informations et la communication entre individus d'une même communauté linguistique ayant reçu un enseignement dans ce domaine".

a. De la littératie au langage parlé

On nomme « **littératie** » l'acquisition de l'écriture et de la lecture. La littératie permet l'acquisition de nouvelles connaissances, au travers de la lecture, et l'augmentation de la conservation de ces connaissances grâce à un stockage extérieur. L'effet principal de la littératie est alors de créer une nouvelle interface grâce à laquelle on peut accéder, à partir de la vision, à des représentations qui sont normalement propres au langage parlé (Kolinsky et al., 2014).

Un graphème correspond à la plus petite unité d'un système d'écriture, soit une lettre ou un groupe de lettres transcrivant un phonème. Ces associations entre graphèmes et phonèmes sont la base de la lecture dans les systèmes d'écriture alphabétique. Bien que la littératie soit un phénomène culturel artificiel, la plupart des gens n'ont pas de difficultés à acquérir des compétences dans ce domaine. De manière intéressante, le plus faible niveau d'alphabétisation des personnes ayant une surdité peut amener à se demander si les mécanismes liés au langage oral interviennent dans l'apprentissage de la lecture (Perfetti et Sandak, 2000, op. cit. in Van Atteveldt et al, 2004). D'autre part, l'impact de l'acquisition de la lecture sur la nature des représentations phonologiques a été étudié en comparant trois populations distinctes : une première composée d'adultes qui n'ont jamais appris à lire

(illettrés), une seconde composée d'adultes lettrés, et une dernière composée d'adultes qui ont appris à lire à l'âge adulte. Les résultats nous montrent qu'il y a eu une activation des représentations phonologiques à partir du code écrit pour les personnes sachant lire et que les représentations phonologiques seraient devenues plus fines et plus stables grâce à la littératie (Kolinsky et al., 2014).

D'après une étude de Kolinsky et al. (2012), les processus impliqués dans le traitement auditif de la parole pourraient être affectés par des connaissances dépendantes de l'alphabétisation. Elles ont pu montrer l'implication des représentations orthographiques dans la reconnaissance des mots parlés par la comparaison de populations de personnes alphabétisées, ou anciennement analphabètes, à des populations de personnes non alphabétisées. Cette problématique a également été explorée via l'impact de la connaissance orthographique sur le traitement de la parole chez des personnes alphabétisées. Par exemple, dans une étude de Morais et al. (Morais et al., 1979), deux tâches ont été proposées à des adultes portugais analphabètes ou anciens analphabètes. Ces tâches consistaient en la suppression, ou en l'ajout, d'un phonème au début d'une courte expression verbale. Les adultes analphabètes étaient incapables d'accomplir ces tâches, et près de la moitié d'entre eux n'ont pu donner une réponse correcte. A l'inverse, les adultes anciennement analphabètes obtenaient une moyenne raisonnable, puisque près de la moitié d'entre eux ont réussi à chaque essai. Les résultats nous montrent ainsi que lors de tâches métaphonologiques (e.g. la suppression du premier phonème d'un mot), les personnes non alphabétisées présentaient plus de difficultés que les personnes alphabétisées.

b. L'intégration audio-visuo-orthographique

L'intégration audio-visuo-orthographique, c'est-à-dire les mécanismes d'intégration mis en jeu lors de la présentation simultanée d'un son de parole et d'un stimulus visuel orthographique, a également fait l'objet d'investigations. C'est notamment le cas d'une étude qui a cherché à établir un possible équivalent orthographique à l'effet McGurk (Stekelenburg et al., 2018). Neuf stimulations auditives se situant sur un continuum défini par les syllabes /aba/ et /ada/ ont été présentées, associées ou non à un stimulus visuel orthographique. La personne devait alors choisir si elle entendait /aba/ ou /ada/. ll a été montré que la présentation simultanée d'un stimulus orthographique avec une production auditive ambiguë se situant sur le continuum /aba/-/ada/ cité plus haut entraîne l'émergence d'une illusion perceptive, bien que celle-ci soit moindre comparée à l'effet McGurk. En effet, les résultats

de cette étude montrent une légère déviation de la courbe de reconnaissance audio-visuo-orthographique comparé à la courbe de reconnaissance en modalité auditive seule. Ainsi, les stimuli acoustiques les plus ambigus situés sur le continuum /aba/-/ada/ sont plus facilement assimilés à la transcription orthographique proposée. Un texte écrit, tout comme son équivalent visémique, peut donc induire un changement dans la perception des sons de la parole (Stekelenburg et al., 2018).

Nous l'avons vu, lorsqu'une personne lit, elle peut associer les lettres aux sons de la parole. Cela permet d'émettre l'hypothèse que les lettres et les sons n'agissent pas de manière isolée (Stekelenburg et al., 2018). L'indiçage visuel orthographique a une influence sur notre perception du message acoustique. On retrouve alors une similitude avec l'indiçage visémique, pourtant sans dynamique temporelle ici. Une étude en imagerie par résonnance magnétique fonctionnelle (IRMf) portant sur l'intégration des lettres et des sons de la parole (Van Atteveldt et al., 2004) a révélé un traitement spécifique dans le cortex temporal, au niveau du sillon temporal supérieur (STS) et du gyrus temporal supérieur (STG) considérés comme des régions hétéro-modales. Dans cette étude, des sons de la parole et des lettres ont été présentés de façon unimodale (auditive ou visuelle seule) et bimodale (auditive et visuelle). Les stimuli audiovisuels pouvaient être congruents ou incongruents. L'activité de régions du cortex auditif impliquées dans le traitement du son de la parole, situées dans le gyrus de Hesch et le planum temporale, a été influencée par la congruence des lettres et des sons présentés simultanément. Comme ces régions n'ont pas répondu aux lettres seules, cette influence pourrait être vue comme une modulation par rétroaction provenant des régions STS/STG hétéro-modales où l'intégration a eu lieu. De manière importante, ces données suggèrent que l'intégration des lettres et des sons de la parole recrute des mécanismes neuronaux similaires à ceux classiquement observés lors de l'intégration des informations auditives et visuo-faciales (Calvert et al., 2000).

Du fait des similitudes observées entre les modalités audiovisuelles faciales et orthographiques (i.e. l'indiçage visuel, la possibilité d'illusion perceptive, les liens étroits avec le code phonologique) et de leurs différences (i.e. l'aspect figé du langage écrit en comparaison de l'ambiguïté d'une configuration articulatoire et des aspects prédictifs de la parole, le caractère artificiel du système orthographique), il convient dès lors d'approfondir les mécanismes sous-jacents à ces deux types d'intégration, pour ainsi comprendre et comparer leurs fonctionnements respectifs.

3. Apport de l'électroencéphalographie dans l'étude de l'intégration audiovisuelle

Nous avons vu, principalement par le biais d'études comportementales, l'importance des mécanismes d'intégration lors de la perception simultanée de signaux acoustiques et visuels langagiers. De nombreuses études en neuroimagerie se sont également intéressées à ces processus d'intégration audiovisuelle par l'utilisation de techniques diverses comme l'électro-encéphalographie (EEG), la magnéto-encéphalographie (MEG), ou encore l'IRMf. On peut noter que l'intégration audiovisuelle faciale a été nettement plus explorée que l'intégration audiovisuelle orthographique. Dans le cadre de notre projet de recherche, nous dresserons ici un état des lieux des études effectuées en électroencéphalographie.

L'**électroencéphalographie** est une méthode d'exploration cérébrale de l'activité électrique du cerveau mesurée à l'aide d'électrodes placées au niveau du cuir chevelu. C'est une méthode non invasive et indolore qui permet d'observer les activations cérébrales en temps réel. Par cette technique, la modification de l'activité électrique du système nerveux lors de la présentation d'un stimulus auditif ou visuel est concomitante à l'apparition de potentiels évoqués. Ces derniers peuvent être décrits par différents paramètres comme leur amplitude ou leur latence, paramètres dépendant des processus perceptifs et cognitifs mis en œuvre ainsi que de leur temporalité (Pinto et Sato, 2016).

Dans ce manuscrit, nous nous intéressons aux potentiels évoqués auditifs (PEAs) qui traduisent l'activité électrique des populations neuronales du cortex auditif, et donc aux PEs associés au traitement de l'information auditive. Lors d'une stimulation auditive, différents PEAs sont observés, et nous nous intéresserons ici aux marqueurs électrophysiologiques connus N1/P2, qui sont très étudiés dans la littérature :

- Le **N1**, une composante d'amplitude négative qui apparaît 100 ms après le début d'un signal auditif
- Le **P2**, une composante d'amplitude positive qui apparaît 200 ms après le début d'un signal auditif

Ce complexe **N1/P2** présente des marqueurs typiques de l'intégration audiovisuelle (Treille, 2017). Il traduit notamment les processus de décodage inhérents au signal acoustique (i.e. de parole ou non), et s'acquiert par le biais d'électrodes placées au niveau fronto-central. Pour parler d'intégration, il faut que l'on puisse observer une différence entre l'amplitude de la modalité audiovisuelle (AV) et la somme de l'amplitude des modalités auditive (A) et visuelle (V) seules. C'est ce que l'on appelle le modèle additif, décrit aussi comme suit : AV

≠ A + V (Baart 2016, op. cit. in Pinto, 2019). Une illustration d'une intégration audiovisuelle est présentée dans la figure 2.

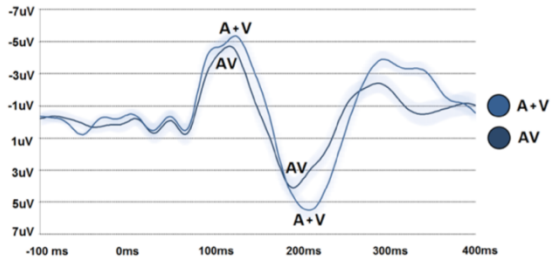


Figure 2. PEAs moyens sur les électrodes fronto-centrales liés aux signaux EEG AV et A+V observés dans l'étude de Pinto et al. (2019). Les auteurs ont observé des différences entre A+V et AV pour les composantes N1, à environ 100 ms, et P2, à environ 200 ms, illustrant ainsi une intégration observée grâce au modèle additif.

L'amplitude des PEAs pour les composantes N1 et P2 est fonction de la taille de la population neuronale activée par une stimulation auditive. Un premier résultat issu des études en EEG (e.g. Klucharev, Möttönen et Sams, 2003 ; Van Wassenhove et al., 2005) portant sur l'intégration audio-visuo-faciale est celui de de la réduction d'amplitude des PEAs lors de la présentation d'un stimulus audiovisuel congruent, en comparaison avec la présentation d'un stimulus auditif seul ou de la somme des modalités A et V seules (AV < A et AV < A + V ; Klucharev, Möttönen et Sams, 2003 ; Stekelenburg et Vroomen, 2007 ; Treille, 2017). Cette réduction de l'amplitude pourrait notamment être expliquée par des phénomènes de prédiction, induisant un changement au niveau du cortex auditif. Les informations visuelles prédictives inhérentes au signal acoustique à venir et envoyées au cortex auditif impliqueraient une première sélection neuronale avant le traitement auditif propre à la stimulation acoustique. Du fait de cette sélection, une moindre population neuronale serait alors activée par comparaison à la modalité auditive seule, expliquant la réduction d'amplitude observée (Klucharev, Möttönen et Sams, 2003). Face à cette hypothèse, une étude de Van Wassenhove et collègues a cependant montré que cette réduction d'amplitude était indépendante de la congruence des signaux auditifs et visuels (Van Wassenhove et al., 2005). Au sein de cette expérience, les stimuli étaient composés d'éléments visuels faciaux et auditifs correspondant à l'articulation et au signal auditif des syllabes /pa/, /ta/ et /ka/. Ces derniers étaient présentés de manière congruente et incongruente, de manière à créer un effet McGurk dans ce dernier cas. La réduction d'amplitude observée pour la modalité audiovisuelle quelle que soit la congruence ou non des stimuli auditifs et visuels suggère

l'existence de deux mécanismes distincts prédictifs : temporel et phonétique. Les prédictions temporelles issues des informations visuelles permettraient une anticipation de l'onset acoustique du stimulus auditif, les prédictions visuelles phonétiques permettraient, quant à elles, une anticipation du phonème perçu. Le fait que cette réduction d'amplitude des composantes N1 et P2 des PEAs soit indépendante de la congruence des signaux auditifs et visuels suggère ainsi l'implication de prédictions visuelles temporelles, mais pas celle de prédictions visuelles phonétiques. Une étude appuie cette hypothèse par le prisme de trois expériences différentes (Stekelenburg et Vroomen, 2007). Dans une première expérience, des événements audiovisuels liés à des actions vocales (syllabes) et non vocales (applaudissements) étaient confrontés à des stimuli auditifs ou visuels seuls. Une même réduction d'amplitude a été observée pour la modalité AV par comparaison à la somme des modalités A et V quelle que soit le type de stimulus (vocal ou non-vocal). Ce résultat suggère ainsi que le phénomène de réduction d'amplitude des PEAs lors de l'intégration des informations auditives et visuelles n'est pas spécifique à la parole. La deuxième expérience explorait l'éventualité d'une modification des PEAs en cas d'incongruence des stimuli. Les stimuli étaient les mêmes que dans la première expérience. Le même effet a été retrouvé pour les deux modalités, attestant d'une absence d'influence de la congruence phonétique sur les PEAs. Enfin, une troisième expérience a permis de démontrer qu'en cas d'absence de mouvement visuel anticipatif (comme lorsque l'on déchire une feuille par exemple, où l'onset acoustique est synchrone avec l'onset visuel), aucune réduction d'amplitude des PEAs n'était observée. Une autre étude a également montré l'importance des prédictions visuelles temporelles en modifiant ou non l'onset visuel de stimuli audiovisuels (Vroomen et Stekelenburg, 2010). Une réduction d'amplitude des PEAs a été retrouvée lorsque les mouvements visuels permettaient une anticipation de l'onset acoustique. En revanche, lorsque le mouvement visuel était présenté de manière asynchrone, trop tôt ou trop tard, cette réduction d'amplitude était absente. Une étude de Stekelenburg et collègues (Stekelenburg et al., 2018) s'est également intéressée à l'intégration audiovisuelle faciale, ainsi qu'orthographique. Lors de cette étude, les stimuli auditifs correspondaient à des syllabes issues d'un continuum /aba/-/ada/, les stimuli visuels correspondaient soit à l'articulation des syllabes /aba/ ou /ada/, soit à leurs transcriptions orthographiques. Au niveau comportemental, un effet d'illusion perceptive a pu être observé lors de la perception audio-visuo-orthographique de stimuli incongruents, bien que moindre que celui observé lors de la perception audio-visuo-faciale. Néanmoins, si les résultats EEG ont montré une influence des

signaux visuels faciaux sur le traitement auditif, une telle influence n'a pas été observée dans le cas des stimuli audio-visuo-orthographiques.

Comme nous l'avons mentionné, la **latence** des composantes N1 et P2 des PEAs nous permet également de mieux appréhender les mécanismes neuronaux mis en œuvre lors de l'intégration audiovisuelle. Elle correspond au délai d'apparition des PEAs après la présentation du stimulus, qu'il soit auditif ou audiovisuel. Par extension, nous pouvons donc observer la vitesse de traitement des processus mis en œuvre grâce à l'étude du délai d'apparition des PEAs. Les études EEG portant sur l'intégration audio-visuo-faciale montrent que la latence des PEAs est atténuée lors de la présentation d'un stimulus AV congruent en comparaison avec la présentation d'un stimulus A seul ou la somme des modalités A et V seules (Van Wassenhove et al., 2005 ; Stekelenburg et Vroomen, 2007 ; Treille, 2017). De manière importante, cet effet de facilitation temporelle dépend des prédictions visuelles phonétiques : plus le lieu d'articulation est visible, plus la réduction de latence est importante (Van Wassenhove et al., 2005). Ce résultat est également renforcé par le fait qu'aucune réduction de latence n'est observée dans le cas de stimuli audiovisuels incongruents (Van Wassenhove et al., 2005).

OBJECTIFS DE CETTE ÉTUDE

Les aspects écrits et oraux du langage s'appréhendent selon des approches développementales, sociétales et comportementales. Dans notre expérience, nous avons comme objectif de confronter deux types d'intégrations : les intégrations audio-visuo-orthographique et audio-visuo-labiale (i.e. l'intégration audio-visuo-faciale réduite à la zone orale, voir ci-après). En effet, ces deux modalités nous permettent d'accéder au langage par différents prismes. Il nous semble alors pertinent d'évaluer les différences et les similitudes qui peuvent exister au sujet de la nature de ces deux types d'intégration.

Néanmoins, certaines différences peuvent être observées et ce notamment grâce aux études mentionnées dans ce manuscrit. Premièrement, les études qui abordent l'intégration audiovisuelle faciale l'ont fait par le biais de stimuli visuels dynamiques, en respectant les aspects prédictifs de la parole. Par conséquent, des extraits de parole dynamique étaient présentés. Dans notre expérience, l'intégration audiovisuelle faciale est appréhendée d'une manière différente. Les images de configurations articulatoires sont des photos de la bouche (i.e. des images figées de visèmes) et perdent ainsi les aspects dynamique et prédictif de la parole. Du fait de l'utilisation de photos de la bouche et non du visage entier du locuteur, nous utiliserons préférentiellement le terme d'**intégration audio-visuo-labiale** pour décrire notre étude. Nous avons fait ce choix de présentation d'images figées pour rapprocher la parole du code orthographique. Cela nous permet notamment de cibler les deux types d'informations visuelles en termes de contenu phonétique.

De même, nous avons décidé de confronter les stimuli auditifs et visuels dans des conditions congruentes et incongruentes. Il est alors possible d'observer si une intégration audiovisuelle est possible dès lors que des informations visuelles et auditives sont présentées, et ce peu importe leur éventuelle congruence, ou si le lien de congruence entre stimuli auditifs et visuels joue un rôle sur les mécanismes mis en jeu. Cette observation de l'influence de la congruence nous permet donc de juger le rôle de l'information sur le contenu phonétique, et sur l'intégration éventuellement observée.

Enfin, nous confrontons des tâches de décision lexicale et de décision phonémique. Nous pourrons aussi évaluer l'influence de l'attention du sujet sur l'intégration. Ainsi, il sera

possible de déduire si le fait que la personne porte son attention sur le premier phonème (i.e. dans la tâche de décision phonémique) ou sur le contenu lexical (i.e. dans la tâche de décision lexicale) affecte l'intégration.

Pour notre étude, nous observerons l'amplitude et la latence du complexe N1/P2 lors de deux tâches précises, phonémiques et lexicales. Lors de précédentes études sur l'intégration audiovisuelle faciale (Klucharev, Möttönen et Sams, 2003 ; Stekelenburg et Vroomen, 2007 ; Vroomen et Stekelenburg, 2010), il a été observé une diminution de l'amplitude et une accélération des PEAs au niveau du complexe N1/P2. Cette modification d'amplitude serait alors indépendante de la congruence du contenu des signaux auditifs et visuels (Van Wassenhove et al., 2005 ; Stekelenburg et Vroomen, 2007). De manière importante, alors que la réduction d'amplitude des PEAs peut être reliée à des mécanismes prédictifs temporels, la réduction de latence dépend également des mécanismes prédictifs phonétiques. Les prédictions temporelles, dans le cas d'une présentation de situation audiovisuelle de parole, seraient néanmoins nécessaires à l'observation de cette réduction d'amplitude (Vroomen et Stekelenburg, 2010). Ainsi, l'usage d'images figées de configurations labiales nous laisse supposer la potentielle disparition de la réduction d'amplitude envisagée précédemment.

Nous souhaitons confronter ces deux types d'intégration et explorer leurs différences et similitudes. L'indiçage visuel permettrait de réduire le nombre de possibilités acoustiques, et donc de réduire les réponses neuronales auditives (Treille, 2017). Qu'ils soient orthographiques ou faciaux, certaines études comportementales ont montré que ces indices ont une influence sur la perception qu'aura la personne du phonème produit. Il conviendra donc de chercher à savoir si l'activité électrique du cerveau nous permet de mieux comprendre les mécanismes sous-jacents, en explorant les bases sur lesquels ils s'appuient pour dresser leurs points communs. Nous confrontons donc des situations d'intégrations audio-visuo-labiales et audio-visuo-orthographiques, lorsque les informations auditives et visuelles sont congruentes ou incongruentes. Pour ce faire, les modalités audio-visuo-labiales et audio-visuo-orthographiques seront comparées réciproquement à la modalité auditive seule, afin d'extraire un éventuel effet d'intégration. Ainsi, par le biais de la comparaison de ces différences de traitement en fonction des deux modalités et d'un éventuel effet de congruence, nous souhaitons explorer l'influence du contenu phonétique sur les mécanismes d'intégration audiovisuelle.

MÉTHODOLOGIE

Note de lecture : pour la rédaction des parties « Méthodologie », « Présentation des résultats » et « Discussion des résultats », nous faisons des choix d'écriture à l'image de l'importance que nous accordons au langage et aux inégalités qui s'y immiscent. Nous utilisons ainsi un maximum de mots épicènes et l'accord à la majorité. Notre étude comportant plus de participantes, l'accord pluriel est donc effectué au féminin.

1. Population

a. Description

20 adultes saines (17 femmes et 3 hommes) ont participé à l'étude menée dans ce projet de mémoire. Elles étaient âgées de 18 à 28 ans, avec une moyenne d'âge de 22 ans (±3 ET)[1].

Toutes les participantes étaient droitières selon le Standard Handness Inventory (Oldfled, 1971) avec un score moyen de latéralité manuelle de 86% (±16 ET). Leurs niveaux d'études allaient de bac à bac +5, avec une moyenne de niveau bac +2 (±2 ET).

b. Critères d'inclusion et d'exclusion

Toutes les personnes ayant participé à l'étude étaient de langue maternelle française, droitières, présentaient une vision normale ou corrigée et étaient âgées de 18 à 30 ans, critères d'inclusion définis préalablement. Elles ne présentaient pas d'antécédents ni troubles actuels de la parole, du langage, de l'audition, de problèmes d'origine neurologique, psychiatrique et/ou neuropsychologique, critères d'exclusion définis préalablement.

c. Considérations éthiques

Le protocole a été établi en accord avec les standards éthiques définis par la Déclaration d'Helsinki de 1964. Toutes les personnes ayant participé à l'étude ont présenté au préalable et par écrit leur consentement après avoir été informées des différents aspects inhérents à notre expérimentation. Elles ont été dédommagées à hauteur de 10 euros par heure. L'expérience nécessitait 2 heures de participation.

[1] La population initiale testée était composée de 29 participantes. Par la suite, les 8 premières participantes ont été exclues suite à une modification du protocole expérimental. Les données d'une autre participante n'ont pu être analysées du fait de problèmes techniques durant l'acquisition du signal EEG.

2. Matériel

a. Stimuli auditifs

Les stimuli auditifs utilisés lors de cette expérience ont été enregistrés dans une chambre sourde au Laboratoire Parole et Langage d'Aix-en-Provence, par un locuteur masculin au moyen d'un microphone situé approximativement à 25 cm de sa bouche (fréquence d'enregistrement de 48kHz). Ces stimuli étaient composés de productions sonores correspondant à un mot ou un non-mot bisyllabique. Lors de l'enregistrement, chaque mot ou non-mot était répété deux fois. La liste de ces stimuli est présentée en annexe de ce mémoire.

Les stimuli enregistrés correspondant à ce que nous appellerons les essais "NoGo" (voir ci-dessous la description du protocole expérimental) étaient au nombre de 144, soit 8 groupes de 18 mots. Chaque groupe était composé de 18 mots commençant par le même phonème. Il y avait ainsi un groupe pour chacun des phonèmes suivants : /a/, /i/, /y/, /o/, /p/, /t/, /s/ et /f/. D'un point de vue orthographique, les mots choisis commencent par les graphèmes [a], [i], [u], [o], [p], [t], [f]. Ils étaient composés en moyenne de 4,61 phonèmes (ET : 0,96) et 5,94 lettres (ds : 1,22 lettres). Leur fréquence lexicale selon la base de données Lexique (New et al., 2004) était de 3,11 en langage parlé (ET : 5,85) et de 5,69 en langage écrit (ET : 4,77).

Les stimuli correspondant aux essais "Go" de l'épreuve de décision phonémique se composaient de 12 mots commençant par le phonème /ʒ/ (orthographié [g], par exemple « gencive », « gésier »). Ils étaient composés en moyenne de 4,5 phonèmes et 6,25 lettres. Les stimuli correspondant aux essais "Go" de l'épreuve de décision lexicale étaient composés de 12 non-mots et partageaient les mêmes phonèmes initiaux que les mots (par exemple « fékase », « pifare »). Ils étaient composés en moyenne de 4,25 phonèmes et de 5,33 lettres.

La segmentation acoustique des stimuli a été effectuée manuellement via le logiciel PRAAT (Boersma and Weenink, 2019 ; version 6.0.33). Via le spectrogramme, l'onset de chaque stimulus a été défini 5 ms avant chaque attaque initiale consonantique ou vocalique au point de passage zéro de l'onde sonore. L'offset correspondait à une durée totale du stimulus d'une seconde. Comme mentionné plus haut, 2 enregistrements ont été effectués pour chaque mot. Le plus saillant de ces enregistrements était conservé lors de cette étape. Pour l'ensemble des stimuli, la valeur moyenne du pic d'intensité était de 77dB (±2 ET) et la

valeur moyenne de la fréquence fondamentale de 171Hz (±18 ET). Du fait d'une faible dispersion des valeurs d'intensité entre les stimuli, l'intensité de ces derniers n'a pas été normalisée.

b. Stimuli visuels

Suite à l'enregistrement des stimuli auditifs, les stimuli visuels labiaux ont été enregistrés par le même locuteur à l'aide d'une caméra numérique haute définition centrée sur son visage vu de face (fréquence d'enregistrement de 30 images/seconde, résolution de 1920x1080 pixels). Ces stimuli correspondaient aux visèmes décrivant les phonèmes /a/, /i/, /o/, /p/, /t/, /ʒ/ et /f/. Les stimuli visuels correspondant aux phonèmes /s/ et /y/ n'ont pas été conservés, de par leurs proximités visémiques respectives avec les phonèmes /t/ et /o/. Pour chaque stimulus, une image correspondant à l'onset du phonème et recentrée sur la bouche du locuteur a été créée via le logiciel AdobePremiere.

3. Protocole expérimental

L'expérience s'est déroulée dans une pièce close, isolée phoniquement et faiblement éclairée. Le sujet était assis face à un écran d'ordinateur à une distance d'environ 50 cm. Les stimuli auditifs étaient délivrés au moyen d'écouteurs compatibles EEG à un niveau sonore confortable et égal pour tous les sujets. Le logiciel E-Prime contrôlait la présentation des stimuli auditifs et visuels ainsi que l'enregistrement des réponses manuelles.

Préalablement à l'expérience, les participantes ont été informées des contraintes liées à l'acquisition du signal EEG et des artefacts musculaires électriques. Il était ainsi demandé aux personnes de minimiser leurs mouvements durant l'expérience, plus spécifiquement au niveau de leur activité musculaire oro-faciale : soit limiter tant que possible les clignements et mouvements oculaires, de déglutition, etc. La personne devait fixer le centre de l'écran et ne pas détourner le regard, son attention étant vivement sollicitée.

Le protocole expérimental se décompose en 3 étapes : une phase d'entraînement de reconnaissance de visèmes, une tâche de décision phonémique (DP) et une tâche de décision lexicale (DL). Les personnes ayant participé à l'expérience ont été réparties alternativement dans 2 groupes pour varier l'ordre d'apparition des 2 dernières tâches. L'étape de reconnaissance de visèmes a été présentée systématiquement à toutes les participantes afin

d’effectuer un entraînement quant aux associations entre les visèmes et les phonèmes/graphèmes correspondants.

La **tâche d’entraînement à la reconnaissance de visèmes** dure 10 minutes. Elle se décompose en 2 parties : une phase d’exposition aux stimuli visuels labiaux, d’une durée de 2 minutes, suivie de 2 blocs de 4 minutes d’association active de ces mêmes stimuli au phonème associé. Lors de la première phase, le stimulus visuel labial est présenté à l’écran en même temps que le stimulus auditif congruent et le graphème correspondant. Il est indiqué aux participantes de bien faire attention au visuel présenté et d’essayer de reproduire les configurations articulatoires proposées pour mieux les intégrer. Ces consignes avaient pour but de favoriser un effet d’imprégnation avant une éventuelle mémorisation. Vient ensuite l’étape d’association active. L’association des stimuli se fait via l’utilisation de la souris de l'ordinateur. Toutes les 4 secondes, le stimulus visuel facial est affiché. La personne doit alors décider quel son de la parole est prononcée en choisissant la lettre qui lui correspond parmi 7 possibilités présentées à l’écran : a, i, o, p, t, g, f. Les stimuli visuels utilisés sont les images de visèmes correspondant aux phonèmes /a/, /i/, /o/, /p/, /t/, /ʒ/ et /f/ (voir la figure 3 pour une illustration). Une fois la réponse enregistrée ou 4 secondes passées, un message s'affiche à l’écran (“correct”, "incorrect", ou “répondez plus vite s’il vous plaît”) et le visème, la lettre correcte et le son associé correspondant sont présentés pendant 1.5 secondes. La tâche est divisée en 2 blocs de 35 items. Chaque visème est présenté 5 fois au sein d’un bloc. L’ordre de présentation des visèmes est aléatoire.

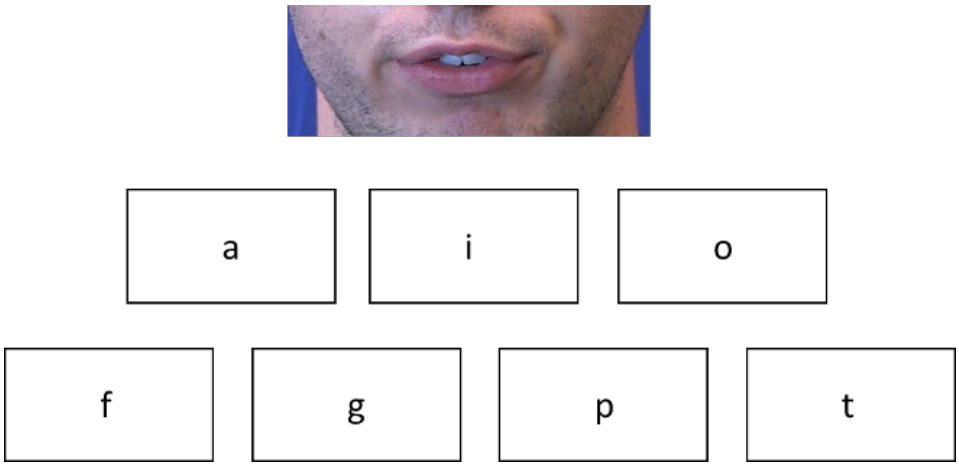


Figure 3. Tâche de reconnaissance de visème

Les stimuli décrits dans la partie 2 ont été utilisés pour générer les différents essais ("Go" et "NoGo"). L'assemblage des différents stimuli (visuels et auditifs) nous ont permis de générer **7 conditions expérimentales** :

- Auditive (A) : le mot est présenté de manière auditive seulement.
- Visuelle labiale (Vl) : une photo des lèvres du locuteur lors de la production d'un des 7 phonèmes cités plus haut est affichée sans aucun son (i.e. un des 7 visèmes mentionnés).
- Visuelle orthographique (Vo) : une des 7 lettres mentionnées plus haut est affichée sans aucun son.
- Audiovisuelle labiale (AVl) : une photo des lèvres du locuteur lors de la production du premier phonème d'un mot est affichée en même temps que le stimulus auditif correspondant à ce même mot prononcé.
- Audiovisuelle orthographique (AVo) : la première lettre d'un mot est affichée en même temps que le stimulus auditif correspondant à ce même mot prononcé.
- Audiovisuelle labiale incongruent (AVli) : une photo des lèvres du locuteur lors de la production du premier phonème d'un mot est affichée en même temps qu'un stimulus auditif correspond à un autre mot.
- Audiovisuelle orthographique incongruent (AVoi) : une lettre différente de la première lettre d'un mot est affichée en même temps que le stimulus auditif correspondant à ce même mot.

Les 7 modalités ont été utilisées pour créer les essais "NoGo" des deux épreuves. Pour les essais "Go", seules les 5 modalités présentant une entrée auditive ont été présentées : les modalités Vl et Vo n'ont donc pas été utilisées.

Les tâches de décision phonémique et de décision lexicale se différencient par leurs consignes. La **tâche de décision phonémique (DP)** consiste à cliquer sur une touche de boîtier de réponse lorsqu'un mot commençant par le son /ʒ/ est entendu (essais "Go"), quels que soient les stimuli visuels présentés. Pour la **tâche de décision lexicale (DL),** il s'agit désormais pour la personne de cliquer lorsqu'un non-mot (essais "Go") est entendu, quels que soient les stimuli visuels présentés. A ces essais s'ajoutent des essais "NoGo", c'est-à-dire des configurations où la participante n'aura pas à cliquer sur le clavier.

Pour chaque tâche, les participantes se sont vues présenter un total de 378 essais “NoGo” et 60 essais “Go”. Cela correspond à 54 essais “NoGo” pour chacune des 7 conditions expérimentales et à 12 essais “Go” pour chacune des 5 conditions expérimentales décrites plus haut. Dans les conditions Vl et Vo, la forme statique des lèvres et la lettre correspondante ont été présentées 9 fois pour chaque phonème. Ces derniers étaient présentés dans un ordre aléatoire, toutes les deux secondes, et ont été répartis en 5 blocs de 88 essais (le premier bloc a commencé avec 2 essais “tampons” ayant pour but de permettre à la participante de se préparer à la tâche). Les mêmes mots bisyllabiques parlés (voir annexe des stimuli cités plus haut) ont été utilisés dans toutes les conditions impliquant une entrée auditive ou audiovisuelle, ce qui nous a permis de contrôler les impacts possibles des caractéristiques acoustiques et psycholinguistiques des stimuli.

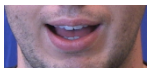
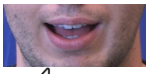
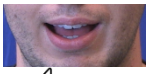
: /a/		a	 : /a/	a : /a/	 : /i/	i : /a/
A	Vl	Vo	AVl	AVo	AVli	AVoi

Figure 4. Exemple de déclinaison des 7 modalités pour le phonème /a/

La personne qui fait les passations peut interagir avec la participante entre les blocs, afin de vérifier que les consignes sont comprises et que l’expérience se déroule au mieux. Elle suit l’expérience et les réponses des électrodes via un ordinateur dans la salle concomitante et veille à la présence réduite d’artefacts. En effet, s’ils sont nombreux, ceux-ci peuvent rendre l’interprétation du signal difficile, voire impossible.

4. Enregistrement EEG

Le système d’acquisition EEG BIOSEMI a été utilisé pour enregistrer l’activité cérébrale. Celui-ci utilise des électrodes dites actives qui amplifient le signal au niveau du scalp et permettent une réduction importante des nuisances électriques et électromagnétiques externes. Le casque EEG, positionné sur le scalp du sujet selon le système international 10-20, se compose de 64 électrodes, auxquelles s’ajoutent 6 électrodes exogènes (EXG). Deux électrodes exogènes servant de référence étaient positionnées sur les os mastoïdes gauches et droits. Les mouvements oculaires horizontaux et verticaux étaient contrôlés par les quatre autres électrodes exogènes, positionnées sur les canthi extérieurs de chaque œil ainsi qu’au-

dessus et en dessous de l'œil gauche. Avant le début de l'enregistrement EEG, l'impédance de l'ensemble des électrodes était contrôlée comme stable et de faible tension de décalage. La pose des électrodes était réalisée via l'application d'un gel conducteur spécifique. La fréquence d'enregistrement était de 512 Hz, et l'enregistrement s'effectuait de manière continue et synchrone sur les 70 électrodes citées plus haut.

5. Situation sanitaire

Ces expérimentations ont eu lieu entre septembre et octobre 2020, lors de la pandémie du COVID-19. Dans ce contexte, différentes mesures avaient été mises en place par le Laboratoire Parole et Langage. Toutes les personnes souhaitant participer ont signé un document stipulant qu'elles n'étaient pas considérées comme "personnes à risque". De plus, des mesures strictes en lien avec la situation sanitaire ont été respectées : port du masque, désinfection de toutes les surfaces entre chaque participante, aération, désinfection régulière des mains, etc. Il est à noter qu'une fois dans la salle où se déroule l'expérience, les participantes quittent leurs masques.

6. Traitement des données EEG

Les données EEG ont été traitées à l'aide du logiciel EEGLAB (Delorme et Makeig, 2004 ; version 2020.0) fonctionnant sur Matlab (Mathworks, Natick, USA ; version R2019a). Pour chaque participante et chaque tâche, les données EEG ont d'abord été re-référencées à la moyenne des mastoïdes gauche et droite et ont été filtrées en utilisant un filtrage passe-bande 1-30 Hz (méthode FIR). Le bruit sinusoïdal résiduel des canaux du cuir chevelu a été estimé et supprimé à l'aide du plug-in EEGLAB CleanLine (version 2012). Les canaux ont ensuite été automatiquement inspectés et les mauvais canaux interpolés à l'aide du plug-in EEGLAB Clean_rawdata (version 0.34). Sur tous les canaux, les mouvements articulatoires, les clignements des yeux, les mouvements des yeux et autres artefacts possibles de mouvement ont été détectés et supprimés à l'aide du plug-in EEGLAB Artifact Subspace Reconstruction (version 0.13). Basé sur une analyse en composantes principales à fenêtre glissante, cet algorithme a rejeté les périodes de données erronées à variance élevée en déterminant des seuils basés sur des segments propres des données EEG.

Pour chaque participante, chaque tâche (DP, DL) et chaque condition expérimentale (A, Vl, Vo, AVl, AVo, AVli, AVoi), les données EEG de tous les essais "NoGo" ont été

moyennées ensemble et segmentées en périodes de 700 ms (à partir de -100 ms à 600 ms par rapport au début du stimulus), corrigée d'une ligne de base de -100 ms à 0 ms. Les périodes avec un changement d'amplitude supérieur à ± 100 μV sur n'importe quel canal (y compris les canaux EXG) ont été rejetées. En moyenne, l'ensemble du pipeline de prétraitement a rejeté 17% (± 4 ET) et 16% (± 5 ET) dans les tâches de décision phonémique et lexicale. Comme les composantes N1/P2 ont une réponse maximale sur les sites fronto-centraux (Scherg et Von Cramon, 1986 ; Näätänen et Picton, 1987), les données EEG ont ensuite été moyennées sur F1, Fz, F2, FC1, FCz, FC2, C1, Cz, C2 électrodes fronto-centrales du cuir chevelu.

Nous avons utilisé un modèle additif pour tester l'intégration AV, dans lequel le signal EEG auditif était comparé à la différence entre les signaux EEG audiovisuels et visuels. À cette fin, les signaux EEG obtenus dans les conditions visuelle labiale (Vl) et visuelle orthographique (Vo) ont été soustraits de ceux obtenus dans les conditions AV correspondantes de la manière suivante : AVo - Vo ; AVoi - Vo ; AVl - Vl ; AVli - Vl. Chacune des ondes de différence des PEs résultantes (nommée à partir de maintenant difPE) a été comparée au signal obtenu dans la condition auditive seule (A), sur la base de l'hypothèse du modèle additif (Baart, 2016) selon laquelle l'intégration AV se produit chaque fois que les signaux difPEs étaient différents du signal obtenu dans la condition A (AV-V ≠ A) dans l'une ou l'autre direction (supra-additive ou sous-additive).

De plus, l'impact du type d'entrée visuelle (orthographe vs. gestes articulatoires) sur l'intégration AV a été examiné en comparant le signal difPE AVo - Vo au signal difPE AVl - Vl. Enfin, l'impact de la congruence AV pour chaque type d'entrée visuelle a été obtenu en comparant le signal difPE AVo - Vo au signal difPE AVoi - Vo, et le signal difPE AVl - Vl au signal difPE AVli - Vl, respectivement.

Sur la base de la littérature et de l'inspection visuelle du signal moyen, deux fenêtres temporelles séparées correspondant aux composantes N1 et P2 ont été sélectionnées : N1 (70-150 ms), P2 (150-250 ms). Dans chaque fenêtre temporelle, les pics d'amplitude et de latence des signaux obtenus dans la condition A et les signaux difPEs décrits ci-dessus ont été extraits de 9 électrodes fronto-centrales (F1, Fz, F2, FC1, FCz, FC2, C1, Cz, C2).

Pour chaque composante des PEs, une ANOVA (analyse de variance) considérant la tâche (DP ; DL) et la condition (un PE et quatre difPEs) comme facteurs intra-sujet a été menée sur les pics maximaux de l'amplitude et la latence. La correction de Greenhouse-Geisser a été appliquée (Greenhouse & Geisser, 1959), et les degrés de liberté et les valeurs p corrigés sont rapportés. L'effet significatif de la condition et son interaction avec la tâche ont été analysés plus en détail en utilisant des comparaisons par paires planifiées pour examiner les effets d'intérêt décrits ci-dessus. Au besoin, des analyses post-hoc non planifiées ont été effectuées avec des corrections de Bonferroni. La figure 7 montre les formes d'onde des PEs obtenues dans la condition auditive et les quatre difPEs.

PRÉSENTATION DES RÉSULTATS

1. Analyse des données comportementales

a. Reconnaissance des visèmes

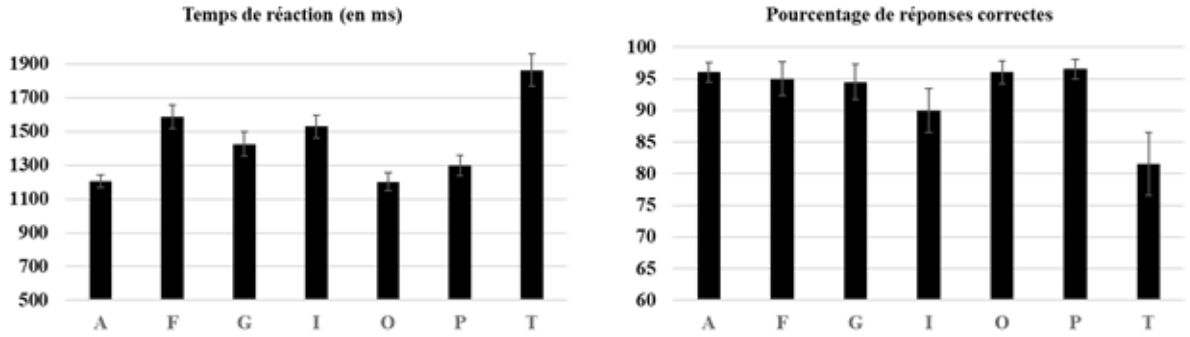


Figure 5 : Pourcentage moyen de réponses correctes et temps de réaction moyens (en ms) obtenus pour les visèmes présentés dans la tâche de reconnaissance des visèmes. Les barres d'erreur représentent l'erreur standard des moyennes.

Une ANOVA traitant le visème comme un facteur intra-sujet a été menée sur le pourcentage de réponses correctes des participantes. L'analyse a montré une différence significative [$F(6, 114) = 4.65$, $p = .0003$, $p\eta2 = .20$] qui était due à un score de précision plus faible obtenu pour le visème /t/ par rapport aux autres visèmes ($ps < 0.01$, corrigés pour les comparaisons multiples en utilisant la correction de Bonferroni) sauf /i/ (90%, $p = 0.40$). Bien que les instructions n'aient pas mis l'accent sur la vitesse de réponse, la tendance générale des données relatives au temps de réaction (TR) a confirmé que certains visèmes étaient plus difficiles à traiter que les autres [$F(6, 114) = 21,17$, $p < .00001$, $p\eta2 = .53$]. Le TR moyen obtenu pour le visème de /t/ était significativement plus long que ceux obtenus sur les autres visèmes (tous $ps \leq .005$). Les TR moyens pour les visèmes /f/ et /i/ étaient significativement plus longs que ceux obtenus pour les visèmes /a/, /o/ et /p/ ($ps. < 05$, les valeurs p ont été corrigées pour les comparaisons multiples en utilisant la correction de Bonferroni).

b. Tâches principales

Des analyses statistiques ont été réalisées sur les performances obtenues lors des essais “Go”. Des ANOVAs à mesures répétées séparées ont été réalisées sur le pourcentage de réponses correctes et les TR aux essais corrects. Pour chaque tâche, les TR inférieurs ou supérieurs au TR moyen de tous les participants ±2,5 ET ont été exclus de l'analyse. La tâche

(DP, DL) et la condition (A, AVl, AVli, AVo, AVoi) ont été traitées comme des facteurs intra-sujet.

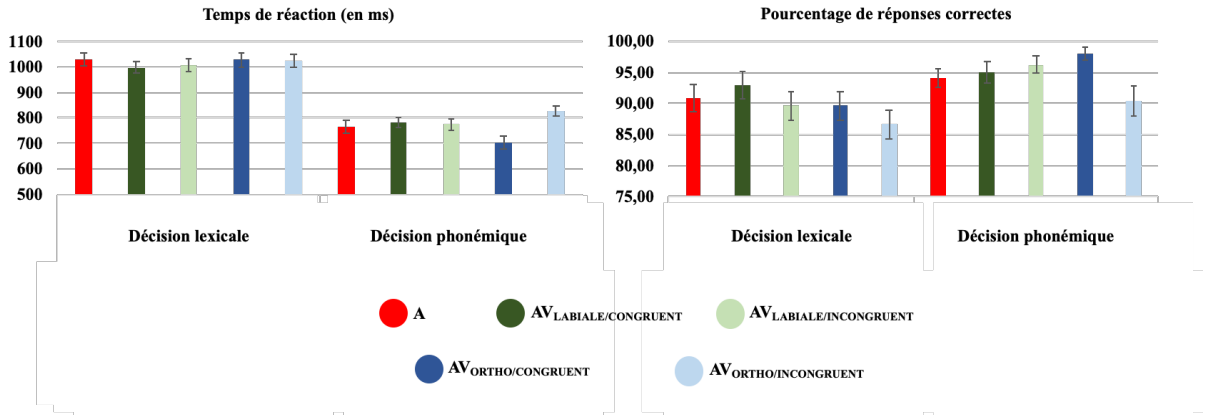


Figure 6 : Pourcentage moyen de réponses correctes et temps de réaction moyens (en ms) sur les essais corrects obtenus dans les tâches de décision phonémique et de décision lexicale. Les barres d'erreur représentent l'erreur standard des moyennes.

L'analyse effectuée sur le pourcentage de réponses correctes a montré des effets principaux significatifs de la tâche [$F(1, 19) = 7.33$, $p = .013$, $p\eta2 = .28$] et de la condition [$F(4, 76) = 4.33$, $p = .003$, $p\eta2 = .19$]. L'interaction entre les deux facteurs n'était pas significative [$F(4, 76) = 1,72$, $p = 0,155$, $p\eta2 = 0,08$]. Comme l'illustre la figure 6 (panneau de droite), la performance obtenue lors de la décision phonémique (94,75 %) était supérieure à celle obtenue lors de la tâche de décision lexicale (89,82 %). L'effet de condition a reflété une performance plus faible obtenue dans la condition AVoi par rapport aux autres conditions ($ps < .01$).

L'analyse des TR a montré des effets significatifs de la tâche [$F(1, 19) = 119.73$, $p < .0001$ $p\eta2 = .86$], de la condition [$F(4, 76) = 7.85$, $p < .0001$, $p\eta2 = .29$] et de leur interaction [$F(4, 76) = 9.06$, $p < .0001$, $p\eta2 = .32$]. Comme l'illustre la figure 6 (panneau de gauche), les participantes étaient plus rapides à identifier le phonème initial qu'à identifier le statut lexical des entrées parlées, ce qui est probablement dû au fait que les décisions phonémiques pouvaient être prises sans attendre la fin du signal vocal. Les analyses de l'interaction entre la tâche et la condition ont indiqué que l'effet de la condition était significatif uniquement dans la tâche de décision phonémique [$F(4, 76) = 21.90$, $p < .0001$, $p\eta2 = .53$] où le TR moyen obtenu dans la condition AVoi était plus long que ceux observés dans les autres conditions, et

où le TR moyen obtenu dans la condition AVo était plus court que ceux observés dans les autres conditions (ps< .001).

Dans l'ensemble, les mesures comportementales ont montré que les performances obtenues dans les deux tâches étaient sensibles à la relation entre les sons de la parole et les indices orthographiques. En revanche, aucune preuve de l'impact du visème n'a été révélée dans les mesures comportementales.

2. Analyse des données EEG

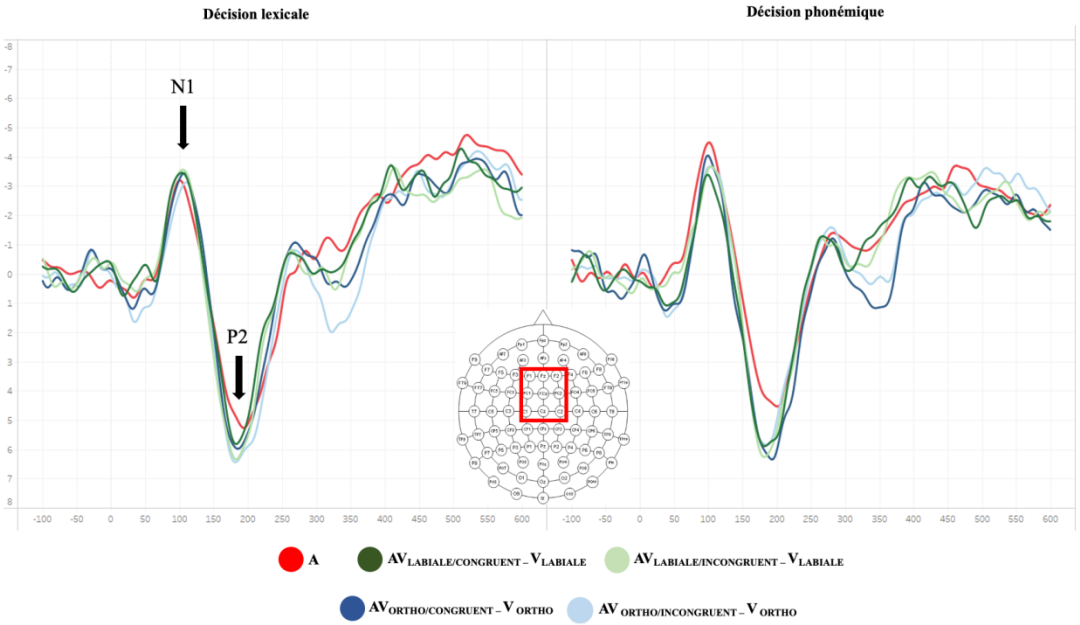


Figure 7 : Les PEs obtenus dans les tâches de DP et de DL dans la condition auditive et les différents PEs suivants (AV-V) : AVl - Vl ; AVli - Vl ; AVo - Vo ; AVoi - Vo. Les électrodes d'intérêt sont indiquées dans le carré rouge.

a. N1 : 70-150 ms

L'ANOVA effectuée sur l'amplitude de la composante N1 n'a montré aucun effet significatif (ps > 0,16). Un résultat similaire a été obtenu dans l'analyse menée sur la latence de la composante (ps > 0,37). Les résultats suggèrent clairement une absence d'intégration AV sur cette composante précoce.

b. P2 : 150-250 ms

L'ANOVA effectuée sur l'amplitude de la composante P2 a montré un effet principal significatif de la condition [$F(2.641, 50.170) = 4.29$, $p = .012$, $p\eta2 = .18$]. Son interaction avec la tâche n'était pas significative ($F < 1$). Étant donné que ces effets n'ont pas interagi

avec la tâche, nous avons combiné les données obtenues dans les tâches de décision phonémique et lexicale ensemble et examiné l'effet de condition. Comme l'illustre la figure 8, les effets significatifs de la condition étaient dus à une augmentation de l'amplitude de la composante P2 déclenchée par les stimuli AV (difPE) par rapport au PE déclenché par les stimuli A seuls, reflétant ainsi une intégration AV supra-additive (ps ≤ .025 pour toutes les comparaisons). L'analyse comparant le degré d'intégration AV induit par l'orthographe et par les gestes articulatoires n'a pas montré de différence significative (p = .38). Enfin, le degré d'intégration n'était pas sensible à la congruence entre les entrées A et V pour les deux types de repères visuels (ps > .25).

L'ANOVA effectuée sur la latence de la composante P2 n'a pas montré d'effet significatif (ps > 0,22).

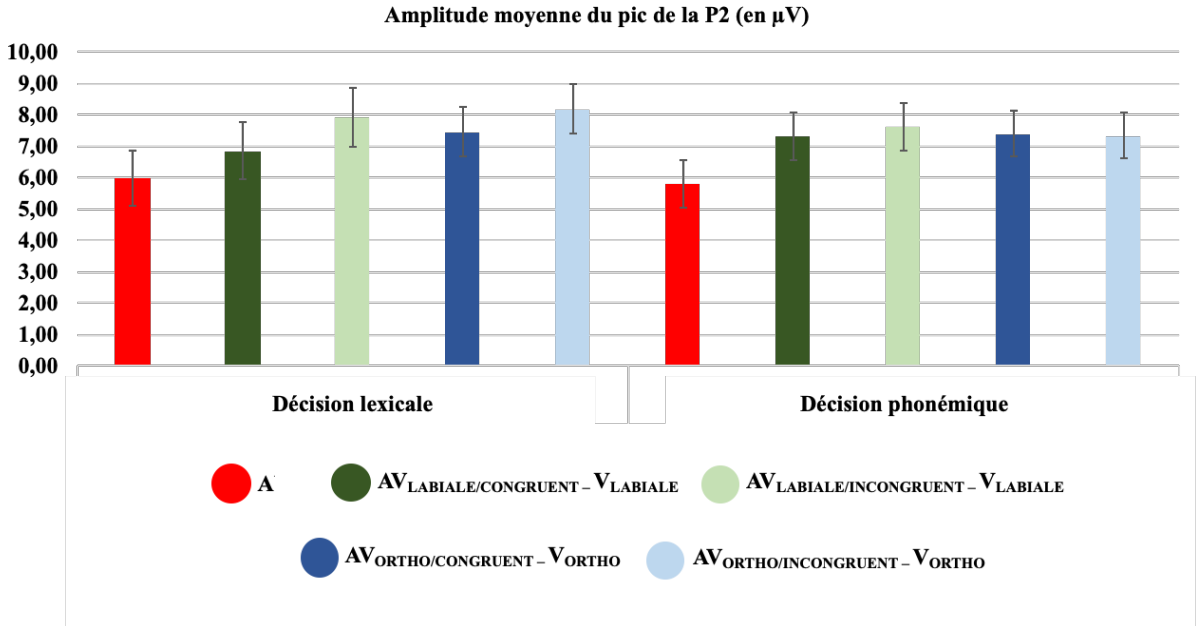


Figure 8 : Amplitudes moyennes des composantes P2 (en µv) correspondant aux PEs obtenus dans la condition auditive et aux différents PEs (AV-V) suivants : AVl - Vl ; AVli - Vl ; AVo - Vo ; AVoi - Vo. Les résultats obtenus dans la DP et la DL sont moyennés. Les barres d'erreur représentent l'erreur standard des moyennes.

DISCUSSION DES RÉSULTATS

Suite à ce protocole expérimental, nous avons pu extraire deux types de résultats. Nous avons d'une part les résultats en EEG, qui correspondent à l'enregistrement de l'activité électrique neuronale enregistrée sur les électrodes fronto-centrales (selon la littérature, l'activité dans cette région trouve son origine dans le cortex auditif), et d'autre part les résultats comportementaux, correspondant aux temps de réaction des sujets et à leurs pourcentages de réponses correctes. Il convient alors de discuter de ces deux éléments et de les comparer. Les résultats EEG correspondent à l'enregistrement des PEAs lors de la présentation des stimuli "NoGo", tandis que les résultats comportementaux s'appuient sur les réponses des sujets, c'est-à-dire leurs réactions face aux stimuli "Go", pour lesquels elles devaient cliquer sur le boîtier.

1. Résultats EEG

Les résultats obtenus en EEG sont contraires aux études EEG portant sur l'intégration audio-visuelle-labiale dans son aspect dynamique. Celles-ci montrent une diminution de l'amplitude et de la latence du N1 pour la modalité AV et ce par rapport à la somme des modalités A et V seules (Klucharev, Möttönen et Sams, 2003 ; Stekelenburg et Vroomen, 2007 ; Treille, 2017). En effet, lors de notre expérience, aucun effet n'a pu être observé au niveau de la composante N1, attestant ainsi d'une absence d'intégration audiovisuelle. De plus, il n'y a pas eu d'effet de la tâche, ni de la congruence des stimuli. Les précédents résultats vus dans la littérature suggèrent que ce résultat pouvait être attendu (Pinto et al., 2019). En effet, une étude portant sur l'impact des prédictions phonétiques, temporelles et articulatoire ("quoi, quand, comment") sur la perception auditive de la parole (Pinto et al., 2019) a consisté en la présentation de stimuli auditifs, visuels et audiovisuels selon différentes conditions. Plus précisément, les situations d'intégration audio-visuo-labiales étaient composées d'une vidéo de parole dynamique et congruente. Le "quand" (prédiction temporelle) était défini par la présentation d'un trait vertical se réduisant jusqu'à disparaître à l'arrivée du stimulus auditif. Le "quoi" (prédiction orthographique du contenu phonétique) consistait en la présentation de la transcription orthographique avant l'arrivée du son. Enfin, ces deux modalités pouvaient être présentées ensemble (i.e. modalité "quand"-"quoi").

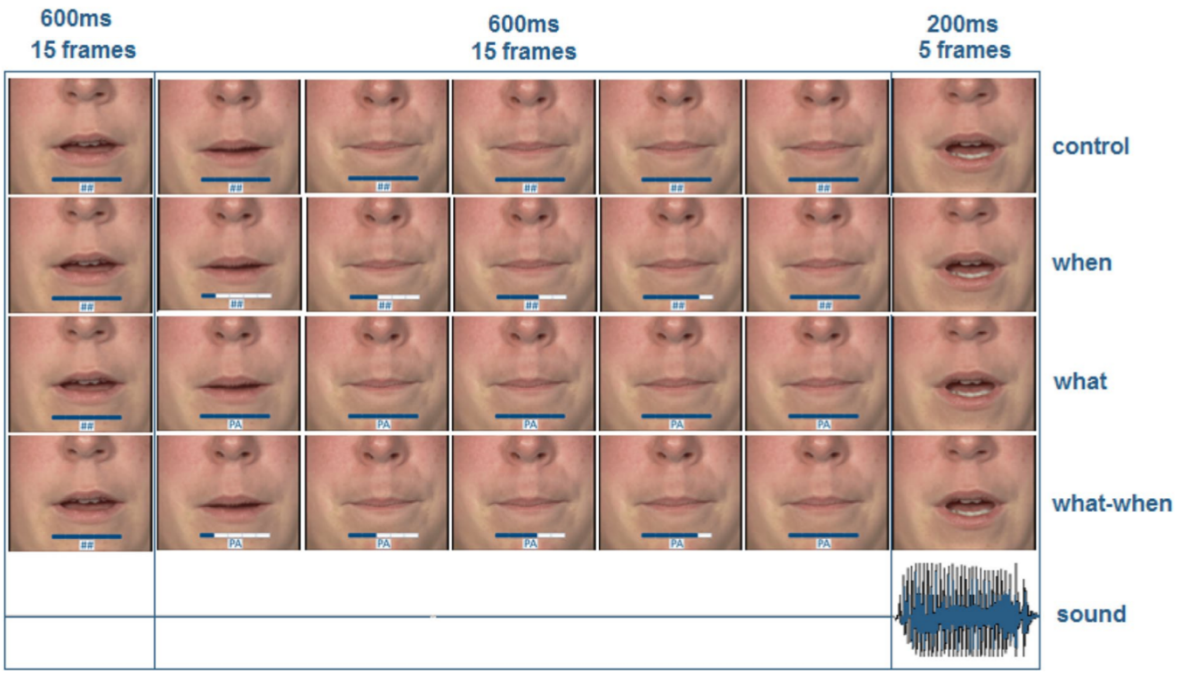


Figure 9 : Exemple des 4 conditions de prédictions pour la syllabe /pa/ présentées dans l'étude de Pinto et al., 2019

Lorsque des stimuli auditifs sont combinés à des indices visuels temporels ou orthographiques, nous pouvons observer une diminution de l'amplitude de la composante N1. Néanmoins, lors de la présentation dynamique des articulateurs, le fait d'ajouter ces indices n'induit pas de modulation de l'amplitude du N1. Nous pouvons alors supposer la nécessité d'indices prédictifs dynamiques pour qu'émerge une modification d'amplitude de la composante N1 dans les mécanismes d'intégration audio-visuelle. Ainsi, n'ayant pas de prédiction temporelle dans notre étude mais bien un stimulus orthographique et visémique figé, nous pouvions nous attendre à ce qu'il n'y ait pas d'influence sur le N1. De plus, dans cette étude, le contenu phonétique est présenté en amont du stimulus auditif. Au sein de notre expérience, on cherche à rapprocher la parole du code orthographique, le stimulus orthographique est ainsi présenté simultanément au stimulus auditif. Dans l'étude de Pinto et collègues (2009), c'est le code orthographique qui semble rapproché de la parole, en l'insérant dans une dynamique prédictive selon les aspects décrits plus haut dans ce manuscrit.

Au sujet du potentiel évoqué P2, une supra-additivité a pu être décrite (i.e. une augmentation d'amplitude de la modalité AV en comparaison des modalités A et V seules). Cet effet a été décrit sans que l'on puisse identifier de différences entre les tâches de DP et de

DL entre conditions audio-visuo-orthographique et audio-visuo-labiale, ou encore entre modalités congruentes et incongruentes. Nous pouvons donc conclure que cette augmentation de l'amplitude n'est ni dépendante de la tâche, ni de la condition, ni de la congruence. Cet effet d'intégration peut notamment être mis en lien avec une étude ayant décrit les modifications d'amplitudes lors de présentations audiovisuelles de stimuli non vocaux (Molholm et al., 2002). L'étude de Molholm et collègues a consisté en la présentation de disques rouges sur fonds noirs, ces derniers étaient associés, ou non, à un stimulus auditif correspondant à un son de 1000 Hz (Molholm et al., 2002). Deux ensembles étaient alors comparés. D'un côté, nous avions la somme des PEs correspondant à la présentation des stimuli auditifs et visuels seuls ; de l'autre, les PEs obtenus suite à la présentation simultanée des deux types de stimuli. Les résultats de cette étude permettent d'observer un phénomène de supra-additivité lors de la présentation simultanée des stimuli auditifs et visuels non vocaux (AV > A +V). Ainsi, nous pouvons supposer que le paradigme utilisé lors de notre expérience a pu rapprocher la parole des caractéristiques de stimuli non vocaux. Cette hypothèse peut notamment s'expliquer par la perte des aspects prédictifs de la parole : il est ainsi possible qu'en présentant une image figée et donc non dynamique, les mécanismes d'intégration se soient naturellement rapprochés de ceux observés en situation non vocale. Cela pourrait notamment expliquer l'absence de différences entre tâches, conditions et modalités (i.e. congruente ou non). Ces aspects n'auraient alors potentiellement pas d'influence sur l'émergence d'un phénomène d'intégration, nous permettant ainsi d'argumenter dans le sens d'un traitement des informations multisensorielles proche de celui de stimuli non vocaux.

2. Résultats comportementaux

Au sujet des résultats comportementaux, nous décrirons dans un premier temps les effets observés lors de l'entraînement à la reconnaissance de visèmes, puis lors des deux tâches principales.

A l'exception du phonème /t/, nous remarquons une très bonne identification des autres visèmes (>90%), attestant ainsi d'un effet d'entraînement efficace. Le but était de nous assurer de la bonne identification des visèmes qui seraient proposés ultérieurement, pour pallier d'éventuelles difficultés qui auraient pu biaiser la manière d'intégrer le message AV. Au sujet des temps de réponse (TR), le traitement du phonème /t/ semble plus long. Ce

résultat semble cohérent, au vu du pourcentage de réponses correctes de ce même-phonème : le visème proposé pour le phonème /t/ pourrait être moins saillant que les autres. Cette hypothèse peut également être mise en lien avec la proximité visémique entre les phonèmes /t/ et /i/ (non significativement différents). Il est à noter que les TR sont influencés par le fait qu'il n'était pas demandé aux sujets de répondre rapidement. De plus, cette épreuve présentait une modalité à choix multiple, induisant ainsi une modification du temps de réponse due au trajet de la souris. Enfin, on peut aussi se demander si la proportion élevée d'étudiantes en orthophonie lors des passations a eu ou non une influence sur le taux de réponses correctes important.

Durant les deux tâches principales, plusieurs effets émanant de composantes comportementales sont apparus. Un **effet de tâche** a pu être observé lors de la comparaison des pourcentages de réponses correctes. En effet, les performances étaient en moyenne plus faibles lors de la tâche de DL. Cet effet nous permet de supposer que la tâche de DP était plus accessible pour les sujets. Le traitement phonétique et l'identification du premier phonème semblent plus aisés en comparaison de processus de plus haut niveau comme la DL. Cet effet de tâche a également été retrouvé lors de la comparaison des TR, puisque les TR en DP étaient inférieurs à ceux en DL. Cette différence nous permet d'argumenter dans le sens d'un traitement simplifié pour la tâche phonémique. Cette identification plus rapide peut également s'expliquer par le fait que, lors de la tâche de DP, le sujet n'a pas besoin d'attendre d'avoir entendu le mot entier pour répondre à la consigne qui lui est donnée : l'identification du premier phonème suffit.

De plus, en termes de réponses correctes, il a été retrouvé un **effet de condition** : le pourcentage de réponses correctes est en effet moindre pour la condition audio-visuo-orthographique incongruente. Cette observation nous permet d'émettre l'hypothèse d'une importance du stimulus visuel orthographique lors de la perception audio-visuo-orthographique. En effet, que ce soit pour la tâche de DL ou de DP, la présentation d'un stimulus orthographique incongruent entraîne plus de mauvaises réponses. Pour rappel, le seul effet obtenu en EEG lors de la présentation de situations audio-visuo-orthographiques était une augmentation de l'amplitude de la composante P2 lorsque nous comparions la modalité AV, à laquelle nous avons donc soustrait la modalité V, à la condition A seule (i.e. AV-V > A). Cet effet avait été observé sans effet de tâche, de modalité ou de congruence. En EEG, la modalité orthographique n'avait donc pas eu d'influence spécifique sur l'intégration

observée. Néanmoins, ces données comportementales nous permettent tout de même d'envisager que les stimuli orthographiques sont bien pris en compte : ils induisent un changement sur le traitement du signal acoustique. De plus, comme seule la condition incongruente a affecté la performance des participantes, nous pouvons supposer que ces dernières étaient sensibles à la congruence des stimuli (A et Vo). Un effet de condition est également retrouvé lors de l'analyse des TR, pour la tâche de DP uniquement. Lorsqu'un stimulus orthographique congruent est présenté, les TR se trouvent réduits. A l'inverse, lorsqu'un stimulus orthographique incongruent est associé au signal acoustique, les TR observés sont plus élevés. Ainsi, ces données nous permettent de supposer que, lors de la tâche de DP, la présentation d'une lettre influence la vitesse de traitement de l'information et l'identification du premier son. En effet, l'usage d'un stimulus orthographique congruent représenterait ainsi un indiçage permettant d'accélérer le traitement de l'information, à l'inverse d'un graphème incongruent. Pour rappel, cette épreuve demande notamment au sujet de porter son attention sur le premier phonème d'un mot. Nous pourrions notamment expliquer ce phénomène par l'existence de certains processus associant phonèmes et graphèmes.

Il convient alors de mentionner le fait qu'un **effet d'interaction** a également pu être décrit. En effet, l'effet de condition observé sur les TR ne s'observe qu'en tâche de DP, puisque la présentation d'un graphème n'induit aucune différence significative en tâche de DL. Le fait que ces résultats ne soient pas retrouvés lors de la tâche de DL nous permet également d'appuyer l'hypothèse d'une comparaison entre phonème entendu et graphème observé lorsque l'attention est portée sur des critères phonologiques. Néanmoins, en suivant cette logique, nous aurions également pu nous attendre à un effet semblable pour les conditions audio-visuo-labiales. Or, aucune observation similaire n'a pu être effectuée, puisque cet effet n'a pas été retrouvé lors de la comparaison des modalités audio-visuo-labiales congruente et incongruente. La présentation d'un stimulus visuel orthographique influerait donc sur la perception du premier phonème, contrairement à la présentation d'un stimulus visémique. Le fait de demander aux participantes de porter leur attention sur des composantes phonologiques semblerait alors nécessaire pour qu'émerge l'influence de la modalité audio-visuo-orthographique.

Ces résultats ne peuvent pas s'expliquer par un problème d'identification des visèmes labiaux présentés lors de l'expérience, puisque les participantes obtiennent, en moyenne, plus

de 90% de bonnes réponses lors de la tâche d'identification de visèmes. Il convient alors d'envisager l'influence de certains processus attentionnels pour chercher à expliquer ces résultats. En effet, ceux-ci pourraient être liés au fait que les participantes auraient porté leur attention sur la modalité A lors de la présentation de conditions audio-visuo-labiales. Lors de l'épreuve de DP, l'analyse des stimuli visuels n'est pas essentielle à la bonne perception des mots, puisque le signal acoustique est clair. Ainsi, il suffisait d'entendre le premier phonème pour fournir une réponse correcte. Cette hypothèse nous permet d'appuyer l'éventuelle intégration audio-visuo-orthographique observée plus haut : si les stimuli visuels labiaux ne sont pas pris en compte, le fait que les stimuli orthographiques aient pu influencer le pourcentage de bonnes réponses et les TR représente une donnée particulièrement intéressante. De plus, le code orthographique étant figé et non ambigu, il a pu être utilisé de manière privilégiée par rapport à l'information articulatoire, qui, comme on a pu le voir dans les résultats ci-dessus, était moins saillante perceptivement par les participantes. Une autre hypothèse serait donc que l'extraction du contenu phonémique à partir des visèmes soit plus élaborée et prenne plus de temps que l'extraction du contenu phonémique à partir des lettres. En effet, dans notre étude, le signal de parole est suffisamment clair, ce qui nous amène à penser que la qualité des stimuli n'influe pas sur cet effet. Il semblerait que, malgré cela, les participantes n'aient donc pas eu besoin de faire appel à d'autres informations que les stimuli A pour percevoir le signal qui leur était présenté. A l'inverse, le fait que l'orthographe montre une influence dans ces situations nous permet d'appuyer l'hypothèse que la modalité orthographique est moins ambiguë et plus automatique qu'une photo de configuration labiale figée, notamment chez des normo-lecteurs (Kolinsky et al., 2012).

Cette expérience nous apporte donc des résultats discordants entre les données en EEG, où l'on observe une intégration AV sur la composante P2 quelle que soit la nature de la tâche et de l'indice visuel, et les données comportementales, où seule l'orthographe affecte la performance (avec un effet de tâche et un impact de la congruence entre les infos A et V). Ceux-ci pourraient être dus aux composantes choisies pour étudier l'intégration AV. Il se pourrait que d'autres éléments apparaissent plus tard dans le traitement, notamment la dissociation entre les deux types d'informations visuelles ou l'effet de congruence/l'effet de tâche. Afin d'être étudiés, il faudrait alors étudier les PEAs sur une durée plus importante du signal.

Pour conclure, cette étude nous permet donc d'agrémenter les connaissances déjà établies au sujet de l'intégration AV. Nous observons une absence d'intégration AV au niveau de la composante N1 et un phénomène de supra-additivité pour la composante P2, et ce en situation articulatoire non dynamique. Lorsque l'on situe les stimuli visuels labiaux dans un contexte figé, nous retrouvons donc des mécanismes similaires à la présentation de situations AV d'éléments non vocaux, ce qui nous suggère que les aspects prédictifs et dynamiques de la parole sont essentiels à l'émergence d'une intégration audio-visuo-labiale. Ainsi, les modalités audio-visuo-orthographique et audio-visuo-labiale présentent les mêmes résultats lorsque les stimuli visuels labiaux présentés sont figés. Les données comportementales nous permettent néanmoins d'observer que la présentation de stimuli orthographiques et auditifs non congruents induit une baisse des pourcentages de bonnes réponses, que ce soit en tâche de DP ou de DL. De plus, nous remarquons également une influence de la congruence des stimuli orthographiques, et ce sur les temps de réponse inhérents à la perception des situations d'intégration AV en tâche de DP. Cette influence étant absente lors de la présentation de stimuli visuels labiaux, ces résultats nous suggèrent alors que le code orthographique est plus facilement pris en compte, notamment lors de tâches nécessitant de porter l'attention sur des critères phonologiques.

3. Limites et extensions de l'étude

Plusieurs éléments nous permettent d'explorer les limites et les biais de notre expérience. En effet, nous avons utilisé un paradigme spécifique qui nous permet d'obtenir les résultats présentés plus haut. La modification de certains aspects de notre protocole pourrait donc induire des modifications quant à ces résultats.

Premièrement, à la suite de cette expérience et en l'absence de résultats significatifs en neuroimagerie (EEG), nous pouvons nous questionner sur le fait d'avoir ciblé uniquement le complexe N1/P2, celui-ci correspondant aux composantes précoces de l'intégration du langage. Si d'autres composantes du signal EEG avaient été examinées, un effet plus tardif aurait éventuellement pu être mis en évidence. La littérature étant plus orientée vers le complexe N1/P2, nous avons ciblé ce complexe dans notre étude. Nous disposions donc de plusieurs hypothèses quant aux résultats attendus, contrairement aux autres composantes. Néanmoins, il pourrait être intéressant d'étendre notre analyse des PEAs à ces composantes plus tardives.

De plus, dans cette situation expérimentale, les stimuli auditifs étaient saillants. Pour autant, lors d'une situation bruitée, la modalité visuelle faciale entre davantage en jeu pour la compréhension du message, ces résultats pourraient donc être différents si les stimuli auditifs étaient bruités ou difficilement compréhensibles.

Enfin, l'usage de l'électroencéphalographie a été justifié par le recueil de données typiques de l'intégration AV, comme l'analyse du complexe N1/P2. Néanmoins, d'autres paradigmes expérimentaux peuvent nous permettre d'appréhender les mécanismes mis en jeu lors de la présentation simultanée de stimuli V et A. Par exemple, l'usage de la MMN (i.e. Mismatch Negativity) aurait permis l'exploration des résultats liés à un phénomène d'habituation, visant à décrire certaines influences entre ces deux types de stimuli.

4. Apports pour la pratique de l'orthophonie

Bien qu'éloignés de la pratique clinique de l'orthophonie, ces résultats permettent tout de même de participer à l'approfondissement de nos connaissances liées à l'intégration du langage. Le métier d'orthophoniste est récent et les apports de la recherche sont multiples pour développer les bases sur lesquelles s'appuie notre exercice clinique. Ainsi, nous pouvons établir des liens entre l'expérience effectuée dans le cadre de ce mémoire et certaines pratiques orthophoniques, notamment via la perception visuelle de la parole, qu'elle soit orthographique ou articulatoire.

Tout d'abord, la phase d'entraînement à la reconnaissance de visèmes nous donne des indices intéressants sur les perceptions des schèmes articulatoires que nous utilisons pour parler. En effet, nous avons pu observer que les réponses étaient meilleures lorsque les visèmes sont particulièrement saillants, comme nous le montrent les difficultés observées pour distinguer le visème correspondant au phonème /t/. Au cours de la prise en soin orthophonique de personnes bilingues, nous pouvons être confrontés à des situations de confusions phonémiques, qui peuvent alors être travaillées via des supports écrits ou encore par un entraînement à une auto-perception articulatoire, reproduite par mimétisme.

La situation sanitaire actuelle (i.e. pandémie due au COVID-19) implique aussi le port d'un masque couvrant le nez et la bouche lors des séances d'orthophonie, ce qui peut induire

des difficultés de compréhension. Dans le cadre des troubles du langage oral, la perception du visage de l'interlocuteur ou de l'interlocutrice peut être essentielle pour pallier certaines difficultés. En revanche, si le visage n'est pas visible en entier, ou si le code écrit pose problème, des difficultés accrues peuvent exister.

De plus, pour des troubles inhérents au langage écrit ou au langage oral, nous pouvons nous questionner sur la modalité à privilégier, entre supports visuels faciaux et supports orthographiques, pour travailler certains domaines comme la phonologie par exemple. Dans ce cas, le support visuel articulatoire ne serait pas figé comme dans notre expérience ; nous pourrions utiliser des images de configurations labiales pour soutenir un apprentissage. Plusieurs techniques de soutien à la communication utilisent d'ailleurs des supports visuels. On retrouve parmi elles le MAKATON©. Le MAKATON© permet d'associer un concept à un pictogramme et un signe en LSF (Langue des Signes Française). C'est une approche pédagogique multimodale et structurée, qui progresse de la manière suivante : usage des gestes, des signes, de l'expression faciale, de la posture, de la vocalisation, de la parole, puis de l'écrit. Une autre méthode de soutien à la communication s'appuyant sur des supports visuels et gestuels est la DNP (Dynamique Naturelle de la Parole). Elle se définit comme une approche ludique et sensorielle de la parole, dans laquelle les ressentis corporels et la perception sensorielle ont une grande importance. La parole devient corporellement ancrée et la sensorialité globale est sollicitée pour retrouver ses mouvements par imprégnation. Divers éléments de la parole sont mis en évidence : les voyelles, les consonnes, les mots, les structures grammaticales, le rythme, ces derniers étant associés à des sens, couleurs, dessins, mouvements etc. Pour définir les phonèmes, elle présente notamment un usage de gestes, un code couleur spécifique, etc. Il pourrait d'ailleurs être intéressant d'approfondir l'intégration de ce type d'indices. En effet, pour des patients ayant bénéficié d'une approche visuelle gestuelle comme la DNP, nous pouvons nous questionner sur l'intégration audio-visuo-gestuelle lorsque sont présentés des stimuli auditifs et gestuels congruents et incongruents.

La compréhension de l'intégration AV sera également essentielle dans le cadre de troubles neurologiques comme l'apraxie de la parole, c'est-à-dire "un trouble acquis de la capacité à programmer le positionnement de l'appareil bucco-phonatoire et la séquence des mouvements musculaires nécessaires à la production volontaire des phonèmes, non relié à une paralysie, une akinésie ou à une ataxie de l'appareil articulatoire" (Deal, Darley, 1972). Les exercices pouvant être proposés utilisent diverses stratégies et notamment l'utilisation de

facilitations visuelles, comme le miroir, ou d'indices verbaux (Sabadell et al., 2018). Ainsi, la perception des mouvements articulatoires ou du code orthographique pourra servir d'indice intéressant.

Enfin, les orthophonistes exercent également dans le domaine de la surdité, dans laquelle la lecture labiale a une grande importance pour compenser les déficits auditifs. En effet, l'usage de l'information visuelle permet aux personnes présentant une perte d'audition de pallier le manque du contenu présent au sein de l'information auditive. C'est d'ailleurs pour cette raison que, dans le cadre d'une prise en soin orthophonique, l'apprentissage de la lecture labiale pourra représenter un axe essentiel de la rééducation. Notre étude se place ainsi dans une démarche proche de ce domaine théorique.

La pratique de l'orthophonie s'appuie sur de nombreuses disciplines scientifiques liées aux sciences cognitives, du langage et aux neurosciences, ce qui lui permet d'évoluer grâce à l'acquisition de connaissances nouvelles à ces sujets. Cette expérience a donc participé à approfondir nos connaissances liées à l'intégration du langage et pourrait renforcer les bases théoriques qui soutiennent la pratique orthophonique.

RÉFÉRENCES

Baart, M. (2016). Quantifying lip-read-induced suppression and facilitation of the auditory N1 and P2 reveals peak enhancements and delays. *Psychophysiology*, *53*(9), 1295-1306. https://doi.org/10.1111/psyp.12683

Bernstein, L. E., Tucker, P. E., & Demorest, M. E. (2000). Speech perception without hearing. *Perception & Psychophysics*, *62*(2), 233-252. https://doi.org/10.3758/BF03205546

Boersma, P., & Weenink, D. (2019). *Praat : Doing phonetics by computer (Version 6.0.33).*

Brin, F., Courrier, C., Lederlé, E., & Masy, V. (2011). *Dictionnaire d'orthophonie*. Ortho Edition.

Bunch, M. (1982). *Dynamics of the singing voice. Vienna : Springer-Verlag*.

Calvert, G. A., Campbell, R., & Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Current Biology*, *10*(11), 649-657. https://doi.org/10.1016/S0960-9822(00)00513-3

Deal, J. L., & Darley, F. L. (1972). The influence of linguistic and situational variables on phonemic accuracy in apraxia of speech. *Journal of Speech and Hearing Research*, *15*(3), 639-653. https://doi.org/10.1044/jshr.1503.639

Dehaene, S., Pegado, F., Braga, L. W., Ventura, P., Filho, G. N., Jobert, A., Dehaene-Lambertz, G., Kolinsky, R., Morais, J., & Cohen, L. (2010). How Learning to Read Changes the Cortical Networks for Vision and Language. *Science*, *330*(6009), 1359-1364. https://doi.org/10.1126/science.1194140

Delorme, A., & Makeig, S. (2004). EEGLAB : An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9-21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Fisher Cletus G. (1968). Confusions Among Visually Perceived Consonants. *Journal of Speech and Hearing Research*, *11*(4), 796-804. https://doi.org/10.1044/jshr.1104.796

Fort, M. (2011). *L'accès au lexique dans la perception audiovisuelle et visuelle de la parole* [Phdthesis, Université Grenoble Alpes]. https://tel.archives-ouvertes.fr/tel-00716384

Giovanni, A. (2014). *Physiologie de la Phonation*. 63.

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, *24*(2), 95-112. https://doi.org/10.1007/BF02289823

Jordan, T. R., & Thomas, S. M. (2011). When half a face is as good as a whole : Effects of simple substantial occlusion on visual and audiovisual speech perception. *Attention, Perception, & Psychophysics*, *73*(7), 2270-2285. https://doi.org/10.3758/s13414-011-0152-4

Klucharev, V., Möttönen, R., & Sams, M. (2003). Electrophysiological indicators of phonetic and non-phonetic multisensory interactions during audiovisual speech perception. *Brain Research. Cognitive Brain Research*, *18*(1), 65-75. https://doi.org/10.1016/j.cogbrainres.2003.09.004

Kolinsky, R., Morais, J., Cohen, L., Dehaene-Lambertz, G., & Dehaene, S. (2014). L'influence de l'apprentissage du langage écrit sur les aires du langage. *Revue de neuropsychologie*, *6*(3), 173. https://doi.org/10.3917/rne.063.0173

Kolinsky, R., Pattamadilok, C., & Morais, J. (2012). The impact of orthographic knowledge on speech processing. *Ilha Do Desterro A Journal of English Language, Literatures in English and Cultural Studies*, *0*(63), 161-186. https://doi.org/10.5007/2175-8026.2012n63p161

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

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

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*(1), 115-128. https://doi.org/10.1016/S0926-6410(02)00066-6

Morais, J., Cary, L., Alegria, J., & Bertelson, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously? *Cognition*, *7*(4), 323-331. https://doi.org/10.1016/0010-0277(79)90020-9

Näätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound : A review and an analysis of the component structure. *Psychophysiology*, *24*(4), 375-425. https://doi.org/10.1111/j.1469-8986.1987.tb00311.x

New, B., Pallier, C., Brysbaert, M., & Ferrand, L. (2004). Lexique 2 : A new French lexical database. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 516-524. https://doi.org/10.3758/BF03195598

Perfetti, C. A. (2000). Reading Optimally Builds on Spoken Language : Implications for Deaf Readers. *Journal of Deaf Studies and Deaf Education*, *5*(1), 32-50. https://doi.org/10.1093/deafed/5.1.32

Pinto, S., & Sato, M. (2016). *Traité de neurolinguistique : Du cerveau au langage*. De Boeck Superieur.

Pinto, S., Tremblay, P., Basirat, A., & Sato, M. (2019). The impact of when, what and how predictions on auditory speech perception. *Experimental Brain Research*, *237*(12), 3143-3153. https://doi.org/10.1007/s00221-019-05661-5

Rosenblum, L. D. (2019). Audiovisual Speech Perception and the McGurk Effect. In L. D. Rosenblum, *Oxford Research Encyclopedia of Linguistics*. Oxford University Press.

https://doi.org/10.1093/acrefore/9780199384655.013.420

Sabadell, V., Tcherniack, V., Michalon, S., Kristensen, N., & Renard, A. (2018). *Pathologies neurologiques : Bilans et interventions orthophoniques*. De Boeck Superieur.

Scherg, M., & Von Cramon, D. (1986). Evoked dipole source potentials of the human auditory cortex. *Electroencephalography and Clinical Neurophysiology*, *65*(5), 344-360. https://doi.org/10.1016/0168-5597(86)90014-6

Schwartz, J.-L. (2011). *Analyse audiovisuelle des scènes de parole. Rencontre Jeunes Chercheurs en Parole, RJCP2011*.

Stekelenburg, J. J., Keetels, M., & Vroomen, J. (2018). Multisensory integration of speech sounds with letters vs. visual speech : Only visual speech induces the mismatch negativity. *European Journal of Neuroscience*, *47*(9), 1135-1145. https://doi.org/10.1111/ejn.13908

Stekelenburg, J. J., & Vroomen, J. (2007). *Neural Correlates of Multisensory Integration of Ecologically Valid Audiovisual Events*. *19*(12), 11.

Strelnikov, K., Rouger, J., Barone, P., & Deguine, O. (2009). Role of speechreading in audiovisual interactions during the recovery of speech comprehension in deaf adults with cochlear implants. *Scandinavian Journal of Psychology*, *50*(5), 437-444. https://doi.org/10.1111/j.1467-9450.2009.00741.x

Treille, A. (2017). *Percevoir et agir : La nature sensorimotrice, multisensorielle et prédictive de la perception de la parole* [Phdthesis, Université Grenoble Alpes]. https://tel.archives-ouvertes.fr/tel-01693084

van Atteveldt, N., Formisano, E., Goebel, R., & Blomert, L. (2004). Integration of Letters and Speech Sounds in the Human Brain. *Neuron*, *43*(2), 271-282. https://doi.org/10.1016/j.neuron.2004.06.025

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural

processing of auditory speech. *Proceedings of the National Academy of Sciences*, *102*(4), 1181-1186. https://doi.org/10.1073/pnas.0408949102

van Wassenhove, Virginie, Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia*, *45*(3), 598-607. https://doi.org/10.1016/j.neuropsychologia.2006.01.001

Vroomen, J., & Stekelenburg, J. J. (2010). Visual Anticipatory Information Modulates Multisensory Interactions of Artificial Audiovisual Stimuli. *Journal of Cognitive Neuroscience*, *22*(7), 1583-1596. https://doi.org/10.1162/jocn.2009.21308

ANNEXES

Annexe 1 - Fiche information COVID-19 (D04) avant participation à une expérimentation, Laboratoire Parole et Langage

Laboratoire Parole et Langage
UMR 7309
www.lpl-aix.fr
5 avenue Pasteur
13100 Aix-en-Provence
France
T. +33 (0)4 13 55 36 34
F. +33 (0)4 13 55 37 44
secretariat@lpl-aix.fr

FICHE INFORMATION COVID-19 (D04) AVANT PARTICIPATION A UNE EXPERIMENTATION LABORATOIRE PAROLE ET LANGAGE

Je soussigné.e

NOM : ……

PRENOM : : ……

Date de Naissance : ……

atteste ne pas être considéré.e comme personne « à risque » telle que définie par le décret n° 2020-521 du5 mai 2020, c'est-à-dire ne pas :

1° Etre âgé de 65 ans et plus ;

2° Avoir des antécédents (ATCD) cardiovasculaires : hypertension artérielle compliquée (avec complications cardiaques, rénales et vasculo-cérébrales), ATCD d'accident vasculaire cérébral ou de coronaropathie, de chirurgie cardiaque, insuffisance cardiaque stade NYHA III ou IV ;

3° Avoir un diabète non équilibré ou présentant des complications ;

4° Présenter une pathologie chronique respiratoire susceptible de décompenser lors d'une infection virale: (broncho-pneumopathie obstructive, asthme sévère, fibrose pulmonaire, syndrome d'apnées du sommeil, mucoviscidose notamment) ;

5° Présenter une insuffisance rénale chronique dialysée ;

6° Etre atteint de cancer évolutif sous traitement (hors hormonothérapie) ;

7° Présenter une obésité (indice de masse corporelle (IMC) > 30 kgm2) ;

8° Etre atteint d'une immunodépression congénitale ou acquise :

- médicamenteuse : chimiothérapie anti-cancéreuse, traitement immunosuppresseur, biothérapie et/ou corticothérapie à dose immunosuppressive ;

- infection à VIH non contrôlée ou avec des CD4 < 200/mm3 ;

- consécutive à une greffe d'organe solide ou de cellules souches

hématopoïétiques ;

- liée à une hémopathie maligne en cours de traitement ;

9° Etre atteint de cirrhose au stade B du score de Child Pugh au moins ;

10° Présenter un syndrome drépanocytaire majeur ou ayant un antécédent de splénectomie ;

11° Etre au troisième trimestre de la grossesse.

J'atteste, de plus, ne présenter aucun symptôme lié à la Covid-19.

Dans le cas contraire, je m'engage à en informer l'investigateur de l'étude pour laquelle j'ai été sollicité.e et à ne pas me rendre sur le lieu de la recherche.

Fait à ……., le ……..

Signature

Annexe 2 - Renseignements des informations des volontaires à l'expérience

Renseignements Volontaire

Date :

Nom :

Prénom :

Date de naissance :

Sexe :

Dernier diplôme obtenu :

Langue maternelle :

Problèmes antécédents (vision, audition, perception et production de la parole et du langage, neuropsychologique, neurologique) :

Test de latéralité :

Prière d'indiquer votre préférence manuelle pour chacune des activités ci-dessous en inscrivant un signe + dans la colonne appropriée. Si la préférence est si forte que vous n'utilisez l'autre main que si vous y êtes absolument forcé(e), inscrivez ++. Si vous utilisez l'une ou l'autre main indifféremment, inscrivez un + dans chaque colonne. Répondez à chaque question.

		Gauche	Droite
1	Ecrire		
2	Dessiner		
3	Coudre (main tenant l'aiguille)		
4	Tenir une paire de ciseaux		
5	Se brosser les dents		
6	Tenir un couteau		
7	Tenir une cuillère		
8	Tenir un balai (main supérieure)		
9	Allumer une allumette (main tenant l'allumette)		
10	Ouvrir une boîte (main tenant le couvercle)		

Annexe 3 – Synthèse des données des sujets

	Sujet	Groupe	Ordre	Sexe	Age	Education (bac=0)	Latéralité
1	S9	1	DP-DL	F	28	3	0,90
2	S10	1	DL-DP	F	24	2	0,78
3	S12	2	DL-DP	F	20	1	0,76
4	S13	1	DP-DL	H	28	5	1,00
5	S14	1	DL-DP	F	24	5	1,00
6	S15	2	DP-DL	F	21	4	0,80
7	S16	2	DL-DP	F	23	4	0,70
8	S17	1	DP-DL	F	21	3	0,56
9	S18	1	DL-DP	F	22	0	1,00
10	S19	2	DP-DL	H	18	1	1,00
11	S20	2	DL-DP	F	19	1	1,00
12	S21	1	DP-DL	F	20	0	0,80
13	S22	1	DL-DP	F	19	0	1,00
14	S23	2	DP-DL	F	25	5	0,80
15	S24	2	DL-DP	F	19	1	0,47
16	S25	1	DP-DL	F	21	2	1,00
17	S26	1	DL-DP	F	21	1	1,00
18	S27	2	DP-DL	H	25	5	0,90
19	S28	2	DL-DP	F	18	0	0,80
20	S29	1	DP-DL	F	20	3	1,00

F	17	moyenne	22	2	0,86
H	3	ecart-type	3	2	0,16

min	18
max	28

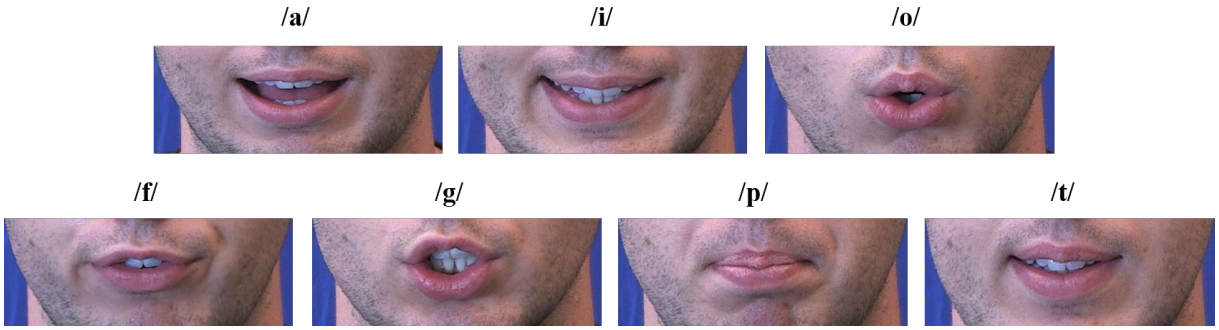
Annexe 4 - Liste des stimuli utilisés

Stimuli NoGo

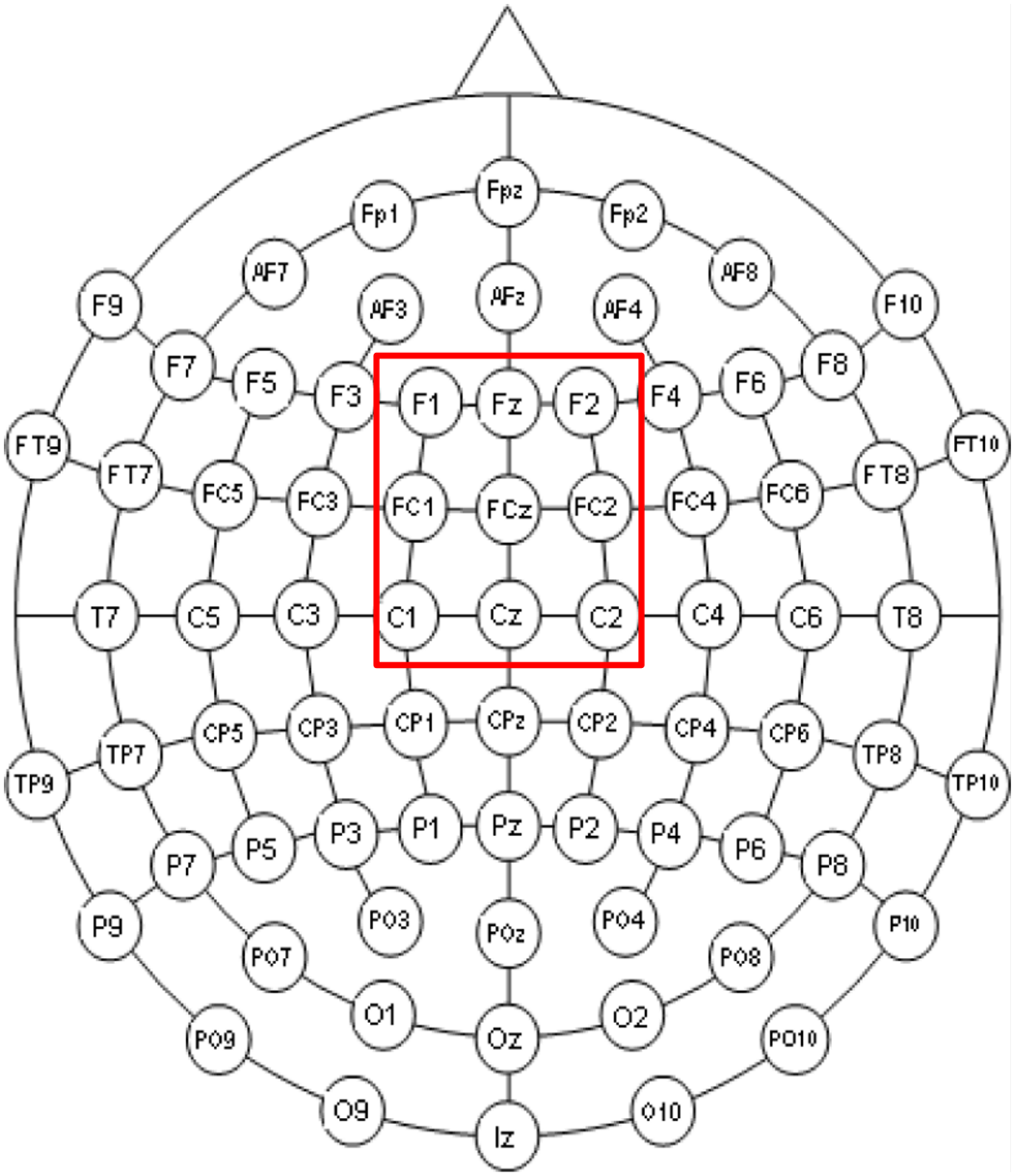
abat	ibis	oblique	palette	tabou	fenouil
acquis	idiome	obus	péage	talus	fabrique
agile	idole	office	perplexe	tarif	falaise
alcôve	inerte	olive	pétale	teinture	félin
agneau	inox	opium	pilon	terroir	ferveur
ardu	iris	orange	pochette	têtard	fiction
aride	issue	osseux	posture	tirage	finance
athlète	ivrogne	otage	pourri	traîneau	flambeau
aveu	icône	ozone	primate	triangle	frayeur
abeille	igloo	obèse	parade	tactile	faisable
accueil	illustre	ogive	pécule	taquin	faïence
acide	ignoble	omis	persil	tassé	farine
admis	inné	opaque	pétrin	terreau	férié
aneth	inouï	orage	pirate	têtu	fermier
arène	islam	orient	poli	timbale	filière
asile	item	osier	poteau	tisane	fissure
atout	ivoire	otite	poussin	traiteur	flanelle
azur	ivresse	ovale	préface	tribune	freinage

Stimuli Go - décision phonémique	**Stimuli Go – décision lexicale**
Gendarme	Fékase
gencive	pifare
gelée	panlu
genèse	imur
géant	aphir
gélule	ulon
gibier	sotin
germain	talou
gestion	todylle
gésier	iceau
gigot	armate
geôlier	ulist

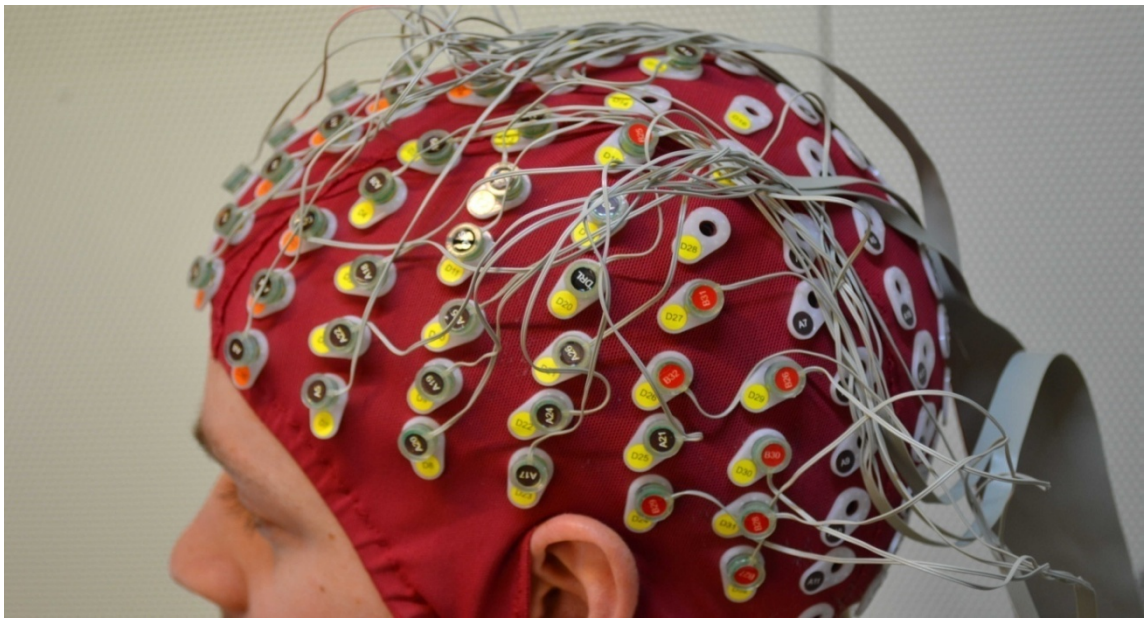
Annexe 5 - Photos des configurations articulatoires utilisées



Annexe 6 - Plan des électrodes

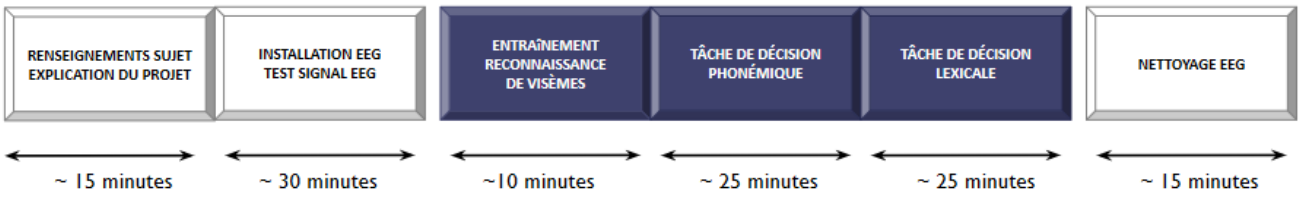


Annexe 7 - Photo du casque d'électrodes



Annexe 8 - Photo d'une participante lors de l'expérience, avec les électrodes

Annexe 9 – Explication du déroulé de l'expérience et consignes présentées oralement aux participantes



Déroulement de l'étude –

Renseignements sujet : Nom, prénom, date de naissance, sexe, langue maternelle, dernier diplôme obtenu, problèmes actuels et/ou passés de vision, audition, de perception et/ou de production de la parole et du langage, neuropsychologiques et/ou neurologiques, test de latéralité.

Explication du protocole : description de l'EEG (méthode non invasive, indolore, principe d'acquisition, contraintes liées aux mouvements oculaires et oro-faciaux), explication du protocole expérimental (tâche de décision phonémique et lexicale).

Installation EEG et test du signal : sélection du bonnet, insertion du gel conducteur, pose des électrodes de surface, pose des électrodes du bonnet, vérification de l'impédance et stabilité du signal de chaque électrode.

Entraînement à la reconnaissance de visèmes : rappel des consignes, court entraînement et vérification de la compréhension des consignes par le sujet, entraînement de 10 min

Tâche de décision phonémique : rappel de la tâche, court entraînement et vérification de la compréhension des consignes par le sujet, tâche de 25 min.

Tâche de décision lexicale : rappel de la tâche, court entraînement et vérification de la compréhension des consignes par le sujet, tâche de 25 min.

Nettoyage EEG : explication simple du but de l'étude, nettoyage du bonnet et des électrodes

Entraînement à la reconnaissance de visèmes -

Consigne : lors de chaque essai, toutes les 4 secondes, le visage du locuteur produisant une lettre sera affiché. Vous devrez décider quelle lettre est prononcée parmi 7 possibilités a, f, g, i, o, p, t. Pour ce faire, vous devrez cliquer sur le bouton correspondant à la lettre de votre

choix. En cas d'erreur, il vous sera indiqué quelle était la bonne réponse. La durée de la tâche est de 10 minutes.

Tâches de décision phonémiques et lexicales –

Consignes pour la tâche de décision phonémique :

Lors de chaque essai, toutes les 2 secondes, suite à l'affichage d'une croix de fixation, un mot vous sera présenté selon cinq modalités :

- Auditive: le mot est présenté de manière auditive.
- Visuelle labiale : le visage du locuteur lors de la production de la première lettre du mot est affiché sans aucun son.
- Visuelle orthographique : la première lettre du mot est présentée de manière orthographique sans aucun son.
- Audiovisuelle labiale : le visage du locuteur lors de la production de la première lettre du mot est affiché avec les sons.
- Audiovisuelle orthographique : la première lettre du mot est présentée de manière orthographique avec le son.

Durant toute la tâche, il vous est demandé de garder les yeux ouverts et de prêter attention aux visages et aux lettres affichées à l'écran. Pour chaque essai, vous devez appuyer sur la touche espace du clavier, **si et seulement vous entendez un mot qui débute par la lettre « g »** (comme dans « genou », « girafe », etc.).
La durée de la tâche est de 25 minutes. Durant toute la tâche :

- Gardez le regard fixé au milieu de l'écran et éviter au maximum de cligner ou de bouger les yeux
- Évitez au maximum les mouvements oraux (bâillements, déglutition, etc.)
- Si vous devez cligner des yeux ou produire un mouvement oral, vous pouvez néanmoins le faire entre deux essais (avant l'affichage de la croix de fixation).

Consignes pour la tâche de décision lexicale :

La consigne est identique à l'exception de la présentation de mots mais aussi de non-mots (mots sans sens). Vous devrez appuyer sur la touche espace du clavier **si et seulement vous entendez un non-mot.**

RÉSUMÉ

La perception de la parole est multisensorielle, dynamique et prédictive. Lorsque disponibles, les informations visuelles articulatoires permettent à l'auditeur ou l'auditrice de désambiguïser le signal acoustique de parole et, de là, facilitent la compréhension du message linguistique perçu. Le code orthographique, quant à lui, est un deuxième code visuel artificiel qui se traduit par ses aspects figés non prédictifs et une relation arbitraire avec le signal acoustique de parole.

Cette étude en électroencéphalographie a pour objectif d'explorer l'influence d'indices visuels, articulatoires et orthographiques, sur les mécanismes d'intégration audiovisuelle. En inscrivant les mouvements articulatoires dans un contexte figé et non dynamique proche du code orthographique, nous avons cherché à comparer l'influence possible de ces deux types d'indices visuels lors de la perception de mots parlés lors de deux tâches de décision phonémique (DP) et de décision lexicale (DL).

Nous nous sommes intéressés à l'étude des potentiels évoqués auditifs, et notamment du complexe N1/P2. Aucune intégration audiovisuelle n'a été observée pour la composante N1. Concernant la composante P2, un phénomène de supra-additivité a pu être observé. Ces résultats se rapprochent de ceux obtenus lors d'études de situations d'intégration audiovisuelle d'évènements non vocaux et non prédictibles. D'un point de vue comportemental, pour les deux tâches, une baisse des pourcentages de réponses correctes a été observée dans le cas d'une incongruence entre signal acoustique et signal visuel orthographique. De plus, lors de la seule tâche de DP, une telle influence de la congruence audiovisuelle sur les temps de réponse a également été observée lors de la présentation de stimuli orthographiques. Cette étude souligne indirectement l'importance des aspects prédictifs et dynamiques de la parole dans les mécanismes d'intégration audio-visuo-labiale. Les données comportementales permettent néanmoins d'observer une influence des seuls stimuli orthographiques sur la perception auditive et suggèrent une possible intégration audio-visuo-orthographique plus tardive.

Mots-clefs : *Perception de la parole - Intégration audiovisuelle - Intégration audio-visuo-labiale - Intégration audio-visuo-orthographique - Décision phonémique - Décision lexicale - EEG - N1/P2 - Orthophonie.*