

TABLE DES MATIERES

REMERCIEMENTS	1
TABLE DES MATIERES	2
I. INTRODUCTION THEORIQUE	3
1. Evolution de la capacité de perception de la parole	3
2. Cas spécifique de la perception du phonème /θ/ : comment les francophones perçoivent ce phonème ?	4
3. Apport de l'entraînement auditif et des informations visuelles sur le traitement de la parole	5
a. Intérêt de l'entraînement auditif dans la perception d'un phonème de L2	6
b. Intérêt de l'information articulatoire dans le traitement de la parole.....	6
c. Influence de l'orthographe dans le traitement de la parole	9
4. Impact du sommeil sur le traitement de la parole	11
II. PROBLEMATIQUE ET HYPOTHESES	12
1. Problématique	12
2. Description de l'étude	12
3. Hypothèses	13
III. EXPERIMENTATION	14
1. Population	14
a. Description	14
b. Critères d'inclusion ou d'exclusion	14
c. Considérations éthiques.....	14
2. Matériel	15
3. Procédure	17
a. Session 1	17
b. Session 2	20
IV.PRESENTATION DES RESULTATS	21
1. Résultats de la tâche de discrimination AX	21
2. Résultats de la tâche d'association mot-image	25
V. DISCUSSION DES RESULTATS	28
1. Objectifs de l'expérimentation	28
2. Commentaires des résultats	29
a. Tâche perceptive	29
b. Tâche de reconnaissance et de mémorisation.....	32
3. Réflexions sur l'expérimentation	34
a. Limitations et biais	34
b. Implications de l'étude et perspectives.....	35
VI. REFERENCES	36
RESUME	39

I. INTRODUCTION THEORIQUE

1. Evolution de la capacité de perception de la parole

Dès la naissance, les nouveau-nés présentent des capacités surprenantes dans le traitement des sons de la parole. Les nourrissons humains naissent avec la capacité de discriminer l'ensemble des contrastes phonémiques utilisés dans toutes les langues du monde (Dehaene-Lambertz & Dehane, 1994). Kuhl, Williams, Lacerda, Stevens et Lindblom (1992) expliquent que tous les nourrissons présentent à la naissance le même modèle de perception phonétique, et ce quel que soit l'environnement linguistique dans lequel ils sont nés. Ils possèdent des aptitudes précoces et innées qui leur permettent de discerner des différences entre les unités phonétiques des différentes langues, y compris des langues non natives, qu'ils n'ont jamais entendues auparavant. Néanmoins de nombreuses expériences ont démontré que cette capacité de distinction universelle ne perdurait pas dans le temps et qu'on assiste à un déclin progressif de cette capacité au cours de la première année de vie.

En conséquence de leur expérience linguistique spécifique à leur langue maternelle, les nourrissons perdent progressivement cette capacité de discrimination universelle initiale dans la première année de vie, pour ne discriminer plus que les unités sonores appartenant à leur langue maternelle. Les sons des langues non-maternelles, eux, ne sont alors plus discriminés (Gervain & Mehler, 2010). Voici un exemple extrait des travaux de Werker et Tees (1984) pour l'illustrer. L'anglais n'a qu'un son /d/ dentaire, alors que l'hindi fait la distinction entre un /d/ rétroflexe et un /d/ dentaire. Les nourrissons nés dans des environnements anglophones font facilement la différence entre ces deux sons hindis ; mais il a été constaté qu'après huit mois d'exposition à l'anglais, où les deux catégories de /d/ ne sont pas dissociées, les nourrissons apprenant l'anglais commençaient à perdre cette discrimination.

Kuhl et al. (1992) explique ce changement perceptif par un processus dans lequel l'expérience linguistique donnerait naissance à un inventaire de sons propres à la langue maternelle. Cet inventaire aiderait les nourrissons à organiser les sons de la parole en catégories : les sons de sa langue maternelle (prototypiques) et les sons des autres langues (non prototypiques). Les nourrissons sont automatiquement plus attirés par les sons de leur langue maternelle et vont progressivement réduire la distance perceptive entre les sons prototypiques et non-prototypiques faisant partie de la même catégorie phonémique de la langue maternelle ; ainsi un son n'appartenant pas à la langue maternelle sera assimilé au son de la langue maternelle le plus proche. Cet effet peut aider à expliquer pourquoi les adultes et les enfants dès l'âge d'un an, ne parviennent pas à distinguer deux sons d'une langue étrangère lorsque les deux sons ressemblent à un prototype de leur langue maternelle. Les prototypes phonétiques sont en place vers l'âge d'un an, lorsque les nourrissons commencent à acquérir le sens des mots. Les nourrissons passent ainsi d'un modèle de perception phonétique universel à un modèle de perception phonétique qui est spécifique à leur langue maternelle (Werker & Tees, 1984).

Kuhl et al. (1992) ont mené une étude auprès de nourrissons de six mois afin d'établir la période de développement au cours de laquelle l'expérience linguistique affecte la perception phonétique. Ils ont comparé les capacités de perception de nourrissons de six mois de deux pays, les États-Unis et la Suède, en utilisant à la fois des sons appartenant à leur langue maternelle et des sons de langue étrangère. Les résultats ont confirmé que l'exposition à une langue spécifique au cours des premiers mois de vie, modifie la perception des sons de la parole chez les nourrissons dès l'âge de six mois. En effet, les nourrissons des deux pays ont montré un effet d'aimant beaucoup plus fort pour leur prototype de langue maternelle. L'expérience linguistique entraîne donc une réorganisation perceptive à un âge précoce.

A l'âge adulte, l'expérience linguistique précoce a donc eu un effet profond sur la perception de la parole. La recherche auprès d'adultes a ainsi montré que, dans de nombreux cas et contrairement aux nourrissons, ils ne sont plus en mesure d'identifier ou de distinguer les contrastes sonores non natifs, alors qu'ils peuvent facilement percevoir les différences sonores pertinentes dans leur langue maternelle. La capacité perceptive de la parole des adultes devient donc plus restreinte, se limitant uniquement à la discrimination des contrastes phonémiques qui appartiennent à leur langue maternelle (Werker & Tees, 1984).

2. Cas spécifique de la perception du phonème /θ/ : comment les francophones perçoivent ce phonème ?

Comme nous l'avons vu précédemment, les nourrissons naissent avec une capacité de discrimination phonétique universelle, puis subissent un rétrécissement perceptif pour les sons de sa langue maternelle. Ainsi, nous pouvons nous demander comment les adultes perçoivent et produisent les sons d'une nouvelle langue (L2).

Dans le cadre de leur mémoire d'orthophonie, M-S. Villain Bailly et E. Clot (2019) ont mené une expérience sur des participants français adultes, afin de mesurer la manière dont ces participants catégorisent et discriminent le son /θ/, une consonne fricative dentale anglaise, en se référant au Perceptual Assimilation Model (PAM ; Best, 1995) et son extension au L2 (PAM-L2, Best & Tyler, 2007). Ce modèle est admis par le plus grand nombre pour prédire les capacités de perception des phonèmes de langue étrangère, en pronostiquant la capacité de discrimination de ces phonèmes par les participants.

Dans leur expérience, M-S. Villain Bailly et E. Clot ont utilisé une tâche de catégorisation afin d'identifier en quel(s) son(s) du français est catégorisé le son /θ/. Ensuite, elles se sont basées sur les résultats de cette tâche afin de prédire la capacité de perception de ce son /θ/ pour chaque participant. Ces deux tâches sont brièvement décrites dans les paragraphes ci-dessous.

Pour identifier la manière dont le son /θ/ est catégorisé en un son français, une tâche de catégorisation de sons est proposée aux participants. Dans la première partie de l'expérience, le participant entend une syllabe consonne-voyelle (/da/, /fa/, /ga/ ...) et il doit sélectionner la

consonne française la plus proche du son entendu dans la grille des consonnes françaises (Figure 1), avant d'estimer le niveau de confiance de sa réponse sur une échelle de 1 à 7. Cette tâche a ainsi montré que le son /θ/ est perçu 75,5% du temps en un /f/, et 19,7% en un /s/.

B	CH chat	D	DJ jazz	F
G gant	GN gnôle	HU huit	J	K
L	M	N	P	R
S	T	TCH	V	W wapiti
Y yaourt		Z		

Figure 1 - Grille des consonnes françaises (M.S. Villain-Bailly, E. Clot)

Les résultats de cette tâche de catégorisation ont permis de prédire la capacité de discrimination du phonème /θ/ par rapport aux phonèmes du français /f/, /s/, /t/ évaluée dans une tâche de discrimination perceptive de sons AXB. Dans cette tâche de discrimination, deux sons de référence sont proposés (sons A et B), et le participant doit dire si le son cible (X), présenté entre les sons de référence, est plus proche du son A ou du son B (voir exemple Figure 2).

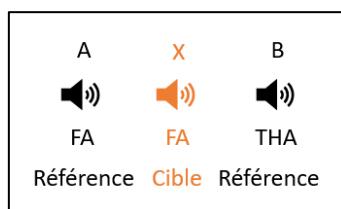


Figure 2 - Exemple de tâche AXB

Comme prédict par les résultats de la tâche de catégorisation, cette tâche de discrimination a montré que, chez les auditeurs français européens, la discrimination entre les sons /θ/ et /f/ est moins bonne que celle sons /θ/ et /s/, ou encore des sons /θ/ et /t/. Le phonème /θ/ est ainsi souvent perçu comme un /f/. Cette confusion perceptive entre les sons /θ/ et /f/ est à l'origine du choix d'utilisation du contraste phonémique /θ/-/f/ dans notre étude.

3. Apport de l'entraînement auditif et des informations visuelles sur le traitement de la parole

En nous basant sur les résultats du mémoire de M-S. Villain Bailly et E. Clot suggérant que le son /θ/ était la plupart du temps perçu comme un /f/ par les participants francophones européens, nous nous sommes demandées s'il existait des facteurs pouvant influencer positivement la capacité de distinguer ces deux phonèmes. De nombreuses études ont examiné l'apport d'un entraînement auditif, ainsi que l'influence de l'orthographe et des gestes articulatoires dans la compréhension de la parole.

a. Intérêt de l'entraînement auditif dans la perception d'un phonème de L2

Le Speech Learning Model (Santiago, 2012) est un modèle d'apprentissage des sons d'une L2 à destination des professeurs, ayant pour but d'améliorer la capacité des apprenants à percevoir et produire un nouveau son de la L2 étudiée. Les auteurs de ce modèle partent du postulat que les caractéristiques spécifiques des sons de la langue maternelle sont conservées dans la mémoire à long terme. Ainsi, une nouvelle catégorie peut être créée pour un nouveau son de la L2 qui diffère phonétiquement du son similaire dans la langue maternelle, à condition que les locuteurs arrivent à discerner les petits détails phonétiques entre les sons en question. La difficulté d'apprentissage d'un nouveau son réside alors dans la ressemblance entre le son de la langue maternelle et celui de la L2 : plus les deux sons sont proches et plus il est difficile pour l'apprenant de percevoir, de mettre en mémoire et donc de produire ce nouveau son.

L'auteur de cet article défend l'importance du travail de la perception auditive pour améliorer la prononciation d'un son en L2. L'entraînement auditif qui consiste à structurer le phonème étranger par le biais d'un travail perceptif, serait ainsi suffisant pour permettre la discrimination entre un son appartenant à la langue maternelle et un son de L2, puisqu'il permet un entraînement perceptif pour la création de nouvelles catégories dans la L2 en structurant autant les traits phonétiques que les traits phonologiques du son.

b. Intérêt de l'information articulatoire dans le traitement de la parole

La première rencontre du fœtus avec la parole au cours du premier trimestre de la vie fœtale est purement auditive. Vers l'âge de 2 mois, les nourrissons deviennent capables de faire correspondre un visage articulant silencieusement la voyelle avec la voyelle auditive appropriée (Bristow et al., 2009). Vers l'âge de 3 à 5 mois, ils produisent des vocalisations s'approchant de la voyelle cible lorsqu'ils sont en présence d'une information audio-visuelle (Kuhl & Meltzoff, 1982). Toutes ces performances auditives surviennent alors que les nourrissons ont encore des connaissances linguistiques rudimentaires, ce qui va dans le sens d'un langage inné. Le développement du langage parlé ne serait pas issu d'un processus d'apprentissage mais d'un héritage génétique, qui se révèle avec l'exposition aux stimuli langagiers. (Dorn, Weinert & Falck-Ytter, 2018).

Les nourrissons perçoivent les sons de la parole ainsi que les traits phonologiques et articulatoires de chaque son. Ils en extraient les propriétés, et créent ensuite une cartographie des repères des indices audio-visuels associés à la parole. Ainsi, la parole devient rapidement un stimulus multimodal à mesure que les capacités motrices et visuelles se développent (Patterson & Werker, 2003). D'une manière générale, pour maximiser l'efficacité du traitement de la parole, les auditeurs nourrissons puis adultes se fient non seulement au signal acoustique, mais aussi aux informations linguistiques (y compris le contexte lexical, syntaxique et sémantique) et aux autres signaux sensoriels. Parmi les indices non acoustiques, l'information visuelle disponible pendant la

conversation en face à face est d'une importance primordiale (Peelle & Sommers, 2015). De nombreuses études ont ainsi montré que lorsque les participants se concentrent sur le contenu de la parole, ils regardent davantage la bouche que les yeux, et ce d'autant plus lorsque la perception vocale est rendue difficile, à cause de l'écoute dans le bruit ou de l'écoute d'une langue non maternelle par exemple (Buchan, Pare, & Munhall, 2007 ; Lansing & McConkie, 2003 ; Yehia, Rubin, & Vatikiotis-Bateson, 1998).

La parole visuelle fait référence à l'information disponible en voyant les gestes articulatoires d'un locuteur. Voir ces mouvements faciaux améliore notre capacité de détection et de compréhension de la parole dans les conditions d'écoute optimales, mais surtout dans des conditions d'écoute dégradée. Cette information visuelle vient parfois modifier la perception des phonèmes, comme le montre l'effet McGurk (McGurk & McDonald, 1976). Par exemple, un auditeur entend une syllabe /ba/ et voit un visage réalisant le mouvement articulatoire /ga/. L'auditeur percevra alors le son /da/. Ce phénomène confirme ainsi l'intégration multimodale automatique de la perception de la parole chez l'adulte. (Grant & Seitz, 2000 ; Erber, 1975).

Dès le plus jeune âge, nous intégrons rapidement et automatiquement les gestes articulatoires lors que nous percevons la parole. L'intégration précoce et continue des repères visuels dans les représentations phonétiques a été confirmée dans l'étude réalisée par Bristow et al. (2009). Les auteurs ont examiné les oscillations corticales chez des nourrissons de 10,5 semaines en moyenne lorsqu'ils étaient confrontés à une parole audio-visuelle. Ils ont présenté aux nourrissons des visages articulant silencieusement une voyelle /a/ ou /i/ deux fois de suite, puis une voyelle /a/ ou /i/ est présentée uniquement auditivement. Lorsque la voyelle auditive n'était pas conforme aux mouvements articulatoires observés précédemment, ils ont observé un phénomène de "négativité d'inadéquation", c'est à dire un changement d'oscillation corticale dû à un changement brusque d'ordre phonologique. ("mismatch response" - MMR). Cette réponse au changement phonologique présente des propriétés fonctionnelles similaires à celles décrites chez l'adulte, au niveau de la perception catégorielle des sons de la parole, de la normalisation et de l'intégration des signaux faciaux.

Cependant, l'évolution de l'utilisation de la parole visuelle est différente selon la langue maternelle du locuteur. Dans leur étude, Sekiyama et Burnham (2008) ont présenté des syllabes produites auditivement, visuellement, et auditivement et visuellement à des locuteurs anglais et japonais âgés de 6, 8, 11 ans et adultes. Les résultats ont montré que les enfants anglais et japonais utilisent de la même façon l'information visuelle jusqu'à l'âge de 6 ans. Après l'âge de 6 ans, l'utilisation de l'information visuelle n'augmente plus chez les locuteurs japonais, alors que l'utilisation de l'information visuelle continue d'augmenter chez les locuteurs anglophones, avant de stabiliser à l'âge de 8 ans. Toutefois, cette observation semble contredire l'étude de Erdener et Burnham (2013) qui suggère que les capacités de lecture labiale et d'intégration de la parole auditive et visuelle n'évoluerait pas entre 5 et 8 ans, mais augmenteraient linéairement avec l'âge à partir de 8 ans. Ils expliquent également que l'utilisation de l'information visuelle dépendrait de la langue utilisée, et non de la maîtrise de cette langue par le locuteur.

Les gestes articulatoires peuvent servir de complément essentiel à l'information auditive en fournissant deux types d'indices.

Premièrement, les gestes articulatoires fournissent des indices sur son contenu. L'aide visuelle fournit une information sur le lieu et le mode d'articulation, ce qui aide à résoudre l'identité des phonèmes individuels et permet de mieux comprendre ce qui est dit en décodant plus finement les catégories syllabiques et lexicales (Carlyon, Cusack, Foxton & Robertson, 2001). Cette aide influe donc sur la compétition lexicale. Lorsque nous écoutons la parole, notre but est d'identifier correctement les mots entendus afin de déduire le sens du mot. Parfois, des mots sont très proches phonologiquement (voisins phonologiques), une compétition lexicale se met donc en place pour savoir quel est le mot réellement entendu parmi tous ses voisins phonologiques (Luce & Pisoni, 1998 ; Marslen-Wilson & Tyler, 1980). Ainsi, un stimulus verbal entendu active un ensemble de candidats lexicaux phonologiquement similaires et la cible correcte doit être choisie parmi cet ensemble de concurrents ou de voisins. Dans le cas d'une reconnaissance uniquement auditive, l'élimination des concurrents phonologiques se ferait seulement grâce au stimulus auditif, laissant ainsi des concurrents proches au niveau des mouvements articulatoires (candidats visuellement proches). A l'inverse, une reconnaissance uniquement visuelle éliminerait uniquement les concurrents visuellement, laissant ainsi des concurrents proches auditivement. La reconnaissance audio-visuelle est donc généralement meilleure que la reconnaissance uniquement auditive ou visuelle, puisqu'elle ne se limite pas à l'élimination des candidats potentiels dans une seule modalité. Selon les informations visuelles disponibles dans la parole, des informations supplémentaires par rapport à la parole purement auditive sont données. Ces informations visuelles fournissent donc une aide supplémentaire pour résoudre ce problème de compétition lexicale en améliorant la précision de la reconnaissance lexicale, notamment lorsque le signal auditif est dégradé (bruit de fond, malentendant, etc.) (Peelle & Sommers, 2015). Les mots compatibles avec les informations auditives et visuelles déterminent fortement le succès de la perception de la parole pendant l'écoute quotidienne.

Deuxièmement, les gestes articulatoires fournissent des indices temporels, sur la synchronisation du signal acoustique entrant. L'aide visuelle donne des informations sur l'amplitude du signal qui influe sur l'attention et la sensibilité perceptive. Les mouvements articulatoires fournissent aux auditeurs une information temporelle qui correspond aux propriétés acoustiques du signal vocal cible, et qui permet notamment de savoir quand leur locuteur va commencer à parler. Par exemple, dans un restaurant bruyant, il est utile de savoir quand notre partenaire conversationnel parle, car cela augmente notre attention au signal auditif, ce qui facilite l'intégration de ce dernier (Carlyon, Cusack, Foxton, & Robertson, 2001).

Nous avons vu que la parole audio-visuelle jouait un rôle primordial lors de la perception d'un signal auditif dégradé. Voici un exemple de situation où la présence des gestes articulatoires contribue au traitement de la parole. Une personne souffrant d'une perte auditive légère à modérée possède des difficultés de compréhension de la parole en particulier dans des conditions d'écoute défavorables. L'expérience de Rosemann et Thiel (2018) sur deux populations normo- et mal-

entendant comportait deux étapes. Dans une première étape, après avoir prononcé une phrase dans différentes modalités (auditive, visuelle, audiovisuelle incongrue et congruente), le participant devait, selon un choix multiple, retrouver le nom prononcé dans la phrase. Lors de la deuxième étape, le test de McGurk a été réalisé pour évaluer l'intégration audiovisuelle. Les résultats indiquent que les sujets mal-entendants légers à modérés sont plus sensibles au discours prononcé selon une modalité audiovisuelle ainsi qu'à l'illusion de McGurk (discours audiovisuel incongru) que les sujets normo-entendants, ce qui indique une plus grande sensibilité à l'intégration audiovisuelle.

c. Influence de l'orthographe dans le traitement de la parole

Contrairement au caractère inné de l'association entre la parole et les gestes articulatoires qui est une faculté produit de l'évolution biologique, la lecture et l'écriture, c'est-à-dire, les activités s'appuyant sur l'association entre la parole et le code écrit, sont issues d'inventions culturelles assez récentes dans l'histoire de l'humanité (Liberman, 1992). L'apprentissage des correspondances artificiellement créées entre les unités sonores de la parole (phonèmes) et leur représentation par des symboles visuels (graphèmes) constituent la base de la lecture. Cette étape cruciale dans l'acquisition de la lecture est réalisée avec une étonnante efficacité pour un phénomène purement culturel (Van Atteveldt, Formisano, Goebel & Blomert, 2004).

Grainger et Ziegler (2007) proposent un « modèle d'activation interactive bimodale » pour illustrer les mécanismes cognitifs soutenant la communication entre la phonologie et l'orthographe, éléments essentiels dans l'acquisition de la lecture (Figure 3). L'architecture de ce modèle est divisée en 3 codes : l'orthographe (connaissance de l'identité et de la position des lettres), la phonologie (code articulatoire nécessaire pour générer une prononciation), la sémantique (connaissance de la signification des mots). Ces différents codes qui interviennent à la fois dans la lecture silencieuse et la lecture à voix haute fonctionnent en interaction. Ils s'influencent mutuellement lors de la perception des mots imprimés ainsi que lors du traitement de la parole, et ont conjointement des répercussions sur les performances observables. Selon ce modèle, les connections entre les codes orthographique et phonologique interviennent aux niveaux de traitement lexical et sous-lexical. Ces deux codes du langage communiquent via une interface centrale permettant aux informations phonologiques d'influencer le processus de reconnaissance visuelle des mots et aux informations orthographiques d'influencer le traitement de la parole. Dans ce modèle d'activation interactive, la compatibilité entre le code orthographique et phonologique affecte le traitement des informations (Grainger & Ziegler, 2007).

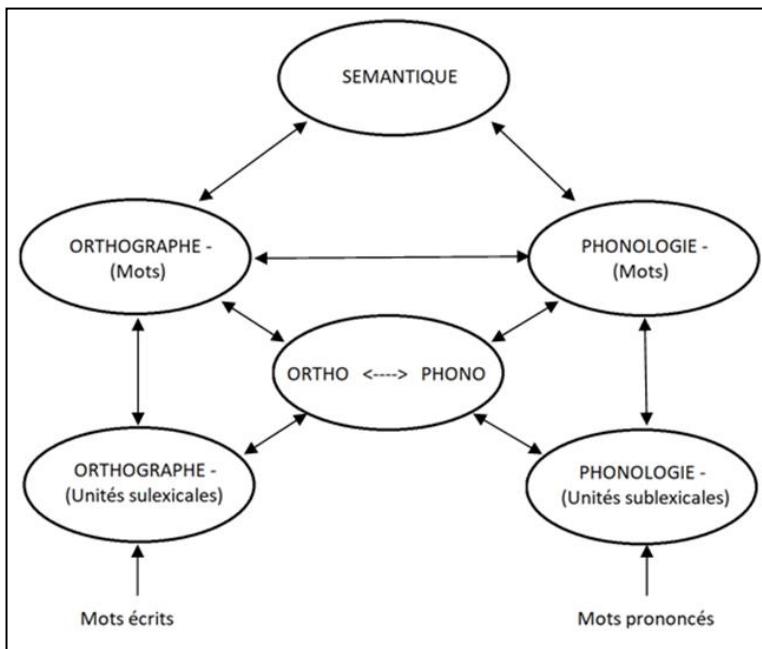


Figure 3 - Modèle bimodal d'activation interactive de la reconnaissance de mots (Grainger & Ziegler, 2007)

Plusieurs études ont démontré l'influence de l'information orthographique sur les différents processus de traitement de la parole.

Dans leur étude, Van Atteveldt, Formisano, Goebel et Blomert (2004) ont réalisé une tâche de perception passive où les participants ont écouté et/ou regardé passivement des lettres et des sons de la parole présentés de manière unimodale ou bimodale (les stimuli bimodaux étaient soit congruents soit incongrus). La comparaison entre les différentes combinaisons a mis en évidence une influence du code écrit sur l'activité du cortex auditif primaire, dépendante de la congruence entre les lettres et les sons de parole présentés simultanément. Cette influence s'exprimait par une augmentation (en cas de congruence) ou une diminution (en cas d'incongruence) de la réponse cérébrale aux sons de la parole. Ces analyses des résultats d'imagerie indiquent que l'information visuelle des lettres peut influencer le traitement auditif des sons de la parole à un stade très précoce. Selon Frost, Repp et Katz (1988), la présentation simultanée de textes imprimés permettrait d'améliorer la perception de la parole.

L'étude réalisée par Keetels, Bonte, et Vroomen (2018) a montré que lorsqu'un auditeur entend un son ambigu, il peut ajuster son interprétation perceptive de ce son en fonction de l'information contextuelle, comme la présence d'un texte d'accompagnement. Dans cette étude, les participants ont été exposés à des consonnes ambiguës (à mi-chemin entre /aba/ et /ada/) associées à la présentation écrite du non-mot "aba" ou "ada". Après cette phase d'exposition audiovisuelle, les participants devaient classer les sons ambigus, présentés cette fois uniquement de façon auditive, en indiquant si les sons entendus étaient plus proches de /aba/ ou /ada/. Les résultats obtenus ont montré que l'information écrite présentée lors de la phase d'exposition a eu une influence sur la perception des sons ambigus de la parole lors de la deuxième phase de l'expérience. Cette étude a ainsi permis de démontrer que l'information orthographique pouvait modifier l'interprétation de sons de la parole ambigus.

Une étude réalisée par Ricketts, Bishop et Nation (2009) illustre l'idée que les représentations orthographiques favoriseraient l'apprentissage, mais aussi la mémorisation des formes phonologiques. Les auteurs ont réalisé une expérimentation afin de déterminer si l'exposition à l'orthographe faciliterait l'apprentissage du vocabulaire oral chez les enfants. Pour cela, cinquante-huit enfants âgés de 8 à 9 ans, devaient réaliser l'apprentissage de 12 non-mots. Pour la moitié des non-mots, les enfants ont été exposés à l'orthographe des non-mots. Les résultats ont montré que les enfants ont fait preuve d'un meilleur apprentissage des non-mots après avoir été exposés à l'orthographe. Les chercheurs ont constaté que la présence de l'orthographe a facilité l'apprentissage des non-mots, tant pendant l'expérimentation qu'après celle-ci.

4. Impact du sommeil sur le traitement de la parole

Après avoir étudié les facteurs qui pourraient influencer l'apprentissage d'un nouveau son, nous nous sommes demandées si le sommeil pouvait faciliter l'apprentissage en permettant de consolider les connaissances apprises.

Selon McQueen, Norris et Cutler (1994), la présence d'une compétition lexicale indique qu'une séquence phonologique est intégrée dans le lexique mental (le processus de lexicalisation). Ainsi, dans leur première expérience, Gaskell et Dumay (2003) ont cherché à déterminer si le sommeil nocturne joue un rôle dans le processus de lexicalisation. Lors de la présentation de nouveaux mots à des sujets, les nouveaux-mots entendus sont comparés aux voisins phonologiques appartenant au lexique mental. Les auteurs ont constaté que juste après l'exposition aux nouveaux mots, aucune compétition lexicale n'avait démarré, mais que celle-ci a bien été constatée après la nuit de sommeil. Dans l'expérience, les mots appris à 20h ont montré un effet de compétition lexicale dans les 12h comprenant un temps de sommeil. En ce qui concerne les mots appris à 8h, ils ont montré cet effet dans les 24h, après la nuit de sommeil suivante. La compétition lexicale ne dépend donc pas de l'heure d'apprentissage, mais bien de la présence d'une nuit de sommeil. La performance des sujets lors de l'épreuve de rappel libre des mots après le sommeil s'est également améliorée. Suite à ces résultats, les auteurs ont suggéré que l'apprentissage de nouveaux mots passe par deux phases : (1) une phase d'apprentissage rapide durant laquelle les nouvelles connaissances sont stockées dans la mémoire épisodique. (2) Ensuite, ces connaissances sont intégrées dans le lexique mental après une période de consolidation. Toutefois, la durée exacte de cette période reste controversée (Bakker, Takashima, Van Hell, Janzen & Mc Queen, 2014).

II. PROBLEMATIQUE ET HYPOTHESES

1. Problématique

Les nourrissons, dans leur première année de vie, passent d'un modèle de perception phonétique universelle à un modèle spécifique à leur langue maternelle (Werker & Tees, 1984). A l'âge adulte, la capacité de percevoir un nouveau phonème n'appartenant pas à leur langue maternelle est donc réduite. Le mémoire d'orthophonique de M-S. Villain-Bailly et E. Clot a montré que les adultes francophones européens perçoivent majoritairement le phonème /θ/ comme un /f/. Nous pouvons ainsi nous demander si la capacité de perception du nouveau son /θ/ ainsi que la capacité à reconnaître et à mémoriser des mots contenant ce phonème peuvent être améliorées grâce à un entraînement.

Dans la littérature, de nombreuses études ont suggéré que le soutien visuel labial et orthographique pouvaient accroître l'efficacité de l'apprentissage de nouveaux mots parlés. Les gestes articulatoires fournissent des indices sur le contenu de la parole (Carlyon, Cusack, Foxton & Robertson, 2001) ainsi que des indices temporels soutenant la capacité de perception (Peelle & Sommers, 2015). L'aide orthographique, elle, a des influences sur le traitement du langage parlé (Grainger & Ziegler, 2007). Elle permettrait notamment d'améliorer la perception de la parole (Atteveldt & al., 2004) et de favoriser l'apprentissage et le maintien en mémoire de séquences verbales (Ricketts, Bishop & Nation, 2009).

Ainsi, pour cette étude, nous avons comparé l'efficacité d'un entraînement purement auditif à l'efficacité de deux types d'entraînement audio-visuel afin d'étudier la contribution de l'indice orthographique et des gestes articulatoires à la capacité de perception du phonème /θ/, ainsi qu'à la capacité de reconnaissance et de mémorisation des nouveaux mots contenant ce phonème chez des participants francophones européens.

2. Description de l'étude

Notre protocole expérimental réalisé sur deux jours comprenait sept épreuves réparties en plusieurs phases. Le premier jour, regroupait une phase de pré-apprentissage, une phase d'apprentissage puis une phase post-apprentissage, et, le lendemain, à nouveau une phase de post-apprentissage. La phase de pré-apprentissage comprenait une tâche de discrimination « AX » permettant de mesurer les performances initiales des participants à discriminer les sons /θ/ et /f/. Puis, la phase d'apprentissage avait pour but d'apprendre aux participants de nouveaux mots qui sont des paires minimales contenant les phonèmes /θ/ et /f/. Lors de cette phase, trois types d'entraînement ont été utilisées :

- Un entraînement purement auditif pendant lequel les participants étaient exposés à des pseudo-mots présentés auditivement et associés à leur image correspondante.

- Un entraînement audio-orthographique pendant lequel les participants étaient exposés à des pseudo-mots présentés auditivement, accompagnés de leur transcription écrite et de l'image correspondante.
- Un entraînement audio-articulatoire pendant lequel les participants étaient exposés à des pseudo-mots présentés auditivement, accompagnés de leurs gestes articulatoires et de l'image correspondante.

Enfin, la phase de post-apprentissage comprenait une tâche de discrimination « AX » et une tâche d'association mot-image, afin d'évaluer l'impact de l'apprentissage d'une part sur la capacité de discrimination des phonèmes /θ/ et /f/, et d'autre part sur la capacité de reconnaissance et de mémorisation des paires minimales contenant ces deux phonèmes. Cette dernière phase était à nouveau réalisée le lendemain, afin d'évaluer l'impact du sommeil sur la consolidation des connaissances apprises.

3. Hypothèses

Le Speech Learning Model (Santiago, 2012), un modèle d'apprentissage auditif des sons d'une nouvelle langue défend l'importance de l'entraînement auditif pour améliorer la capacité de discrimination de deux sons et créer de nouvelles catégories des sons de la L2. En conséquence, nous nous attendons à une amélioration de la perception du son /θ/ pour chacun des trois types d'entraînement grâce à la présence d'un entraînement auditif.

Selon Luce et Pisoni (1998) et Peelle et Sommers (2015), l'information articulatoire associée à l'entraînement auditif va fournir des indices temporels sur le signal vocal et influer sur la compétition lexicale. Dans notre expérience, nous nous attendons donc à ce que le niveau de perception (processus de bas niveau) évalué dans la tâche de discrimination AX ainsi que le niveau de reconnaissance et de mémorisation (processus de haut-niveau) évalué dans la tâche d'association mot-image soient significativement meilleurs après un entraînement auditif avec information articulatoire qu'après un entraînement purement auditif.

Également, d'après Frost, Repp et Katz (1988), l'information orthographique permettrait d'améliorer la perception de la parole. Ricketts, Bishop et Nation (2009) ajoutent que les représentations orthographiques favoriseraient l'apprentissage mais aussi la mémorisation des formes phonologiques. Ainsi, nous nous attendons à ce que le niveau de perception évalué dans la tâche de discrimination AX, ainsi que le niveau de reconnaissance et de mémorisation évalués dans la tâche d'association mot-image soient significativement meilleurs après un entraînement auditif avec information orthographique qu'après un entraînement purement auditif.

En référence aux études de Gaskell et Dumay (2003) ainsi que Bakker, Takashima, Van Hell, Janzen et Mc Queen (2014), une nuit de sommeil permet la consolidation de nouvelles connaissances en mémoire à long terme ainsi que l'intégration d'un nouveau mot dans le lexique mental. Ainsi, nous supposons que, quel que soit le type d'entraînement, les résultats des participants à la tâche d'association mot-image sera significativement meilleure après une nuit de sommeil.

III. EXPERIMENTATION

1. Population

a. Description

101 volontaires sains (70 femmes et 31 hommes) ont participé à notre expérience. Ils avaient entre 19 et 47 ans, avec une moyenne de 22,95 ans et un écart-type de 3,39 ans. 92% des participants étaient droitiers et 8% des participants étaient gauchers.

Les participants ont étudié l'anglais à partir de l'âge de 9 ans en moyenne (moyenne : 8,97 ; écart-type : 2,17) et pendant environ 10 ans (moyenne : 10,26 ; écart-type : 2,69). Tous les participants avaient un niveau d'études allant de bac à bac +5. Ces volontaires ont participé à l'expérience de M-S Villain-Bailly et E. Clot dans le cadre de leur mémoire d'orthophonie. Les participants ont été partagés en trois groupes pour observer l'impact de différents types d'entraînement, en les répartissant équitablement selon leur capacité à percevoir le phonème /θ/ rapportée dans le mémoire de M-S Villain-Bailly et E. Clot. En moyenne, les trois groupes avaient donc la même capacité de percevoir ce phonème avant l'entraînement :

- Groupe « auditif » : soumis à un entraînement purement auditif, 34 participants (8 hommes et 26 femmes)
- Groupe « audio-orthographique » : soumis à un entraînement auditif avec information orthographique, 33 participants (11 hommes et 22 femmes)
- Groupe « audio-articulatoire » : soumis à un entraînement couplant information auditive et la visualisation des mouvements labiaux correspondants sous forme de vidéo, 34 participants (12 hommes et 22 femmes)

b. Critères d'inclusion ou d'exclusion

Les participants ont le français comme langue maternelle, ils ont grandi dans un milieu francophone européen. Aucun n'avait de parent anglophone, aucun n'avait vécu dans un pays anglophone pendant plus de six mois, aucun n'avait de trouble auditif, visuel non-corrigé ou bien de troubles du langage oral ou écrit.

c. Considérations éthiques

Notre étude s'inscrit dans le projet *Perceptual assimilation in second-language learning* mené par Chotiga Pattamadilok, chercheuse au Laboratoire Parole et Langage (Aix-en-Provence), Pauline Welby chercheuse au Laboratoire Parole et Langage (Aix-en-Provence) et Michael Tyler, chercheur à la Western Sydney University.

Tous les volontaires ayant participé à notre expérimentation ont lu et signé librement un formulaire de consentement sur l'étude et étaient libres de cesser l'expérience à tout moment. Ils ont été rémunérés à hauteur de 20€ pour une passation de 2 heures sur deux jours. Enfin, toutes les données recueillies, à l'exception du formulaire de consentement, ont été anonymisées.

2. Matériel

Pour la tâche de discrimination AX, nous avons générée deux productions sonores ayant la structure consonne-voyelle : /fa/ et /θa/. L'enregistrement audio des productions sonores a été réalisé dans une chambre sourde au *Centre d'Expérimentation sur la Parole, Laboratoire Parole et Langage* par une locutrice australienne. Chacune de ces deux syllabes a été enregistrée en plusieurs exemplaires afin de pouvoir sélectionner ensuite les quatre meilleurs pour chaque syllabe.

Nous avons utilisé, pour les enregistrements, une carte son externe USB EDIROL UA-25EX afin d'isoler la prise de son et d'éviter les interférences avec les composantes de l'ordinateur, et un microphone BEYERDYNAMIC TG H55c placé sur le côté de la bouche de la locutrice pour ne pas capter des bruits de souffle nasal et buccal. L'enregistrement a été réalisé grâce au logiciel Audacity à un taux d'échantillonnage de 44,1 kHz et une résolution de 16-bit.

Un auditeur était présent pour contrôler la qualité de l'enregistrement de manière auditive (clarté sonore et vocale, prosodie descendante, pas de bruits parasites) et visuelle (allure de la courbe spectrale). La sélection s'est faite de cette façon afin de privilégier les variations acoustiques naturelles entre les enregistrements.

Nous avons ensuite utilisé le logiciel Praat pour appliquer un filtre passe-haut afin d'éliminer les fréquences basses parasites de l'enregistrement et conserver uniquement les fréquences à partir de 70 Hz, avec 10 Hz de lissage. Puis, nous avons découpé l'enregistrement complet dans le but d'isoler chaque production sonore, et avons procédé à un tri à l'oreille et en tenant compte de la qualité acoustique des signaux par l'intermédiaire du logiciel Praat. Les productions trop courtes, mal articulées, avec une mauvaise intonation ou masquées par des bruits parasites ont été éliminées afin de ne garder que les quatre meilleurs enregistrements pour chaque item. Les enregistrements sélectionnés ont ensuite été découpés à nouveau afin d'égaliser leur longueur : les voyelles ont été tronquées à 80 ms, et une rampe de 5 ms a été appliquée pour minimiser l'influence de la catégorie de la voyelle.

Les fichiers sonores ont été sous-échantillonnés à 22,05 kHz et soumis à un filtre passe-haut à partir de 300 Hz pour analyser les moments spectraux, qui ont été calculés sur la base d'une tranche de Hamming de 25,6 ms centrée au point de frication d'une durée de 50%. La puissance était réglée à 1.

Pour la phase d'apprentissage et la tâche d'association mot-image : nous avons créé 8 pseudo-mots bisyllabiques associés à 8 images d'objets qui n'existent pas (Tableau 1). Ces 8 pseudo-mots constituent les 4 paires minimales suivantes : /fɪnt/-/θɪnt/, /fedʒ/-/θedʒ/, /fald/-/θald/, /felk/-/θelk/. Le choix des pseudo-mots a été fait de façon à ce que ces derniers ne se rapprochent pas d'un mot existant anglais ou français. Chaque pseudo-mot est associé à une image différente.

Pseudo-mot (forme orale)	Pseudo-mot (forme écrite)	Image correspondante
/fɪnt/	fint	
/θɪnt/	thint	
/fedʒ/	fedge	
/θedʒ/	thedge	
/fɔld/	fald	
/θɔld/	thald	
/felk/	felk	
/θelk/	thelk	

Tableau 1- Tableau mettant en correspondance les pseudo-mots utilisés avec leur image associée

Les stimuli audio-visuels (vidéo) utilisés dans la phase d'apprentissage ont été enregistrés par la même locutrice australienne dans une cabine insonorisée à l'Institut MARCS for Brain, Behavior and Development à l'Université de Western Sydney. Elle était assise devant un fond blanc, éclairé par deux lumières de studio recouvertes de papier de diffusion (Studio-Lite Photon Beard Highlight 110). La tête et les épaules de la locutrice ont été enregistrées à l'aide d'une caméra vidéo SONY HXR-NX30P (résolution 1080p, 25 images par seconde, codec H.264). Simultanément, l'audio a été enregistré à l'aide d'un microphone fusil SONY ECM-XM1 (monté sur la caméra) à une fréquence d'échantillonnage de 44,1 kHz via une carte son MOTU ultra-légère MK3 connectée à Adobe Audition sur un PC Windows. Au début de la vidéo, un claquement de main a été utilisé pour synchroniser a posteriori les pistes audio et vidéo.

Un ordinateur portable a été utilisé pour présenter les stimuli (fint, thint, fedge, thedge, felk, thelk, fald, thald), avec une présentation contrôlée par le Psyscope X B77 (<http://psy.ck.sissa.it/>). Avant de commencer, la locutrice lit à haute voix chaque élément, écrit sur une feuille de papier, pour s'assurer qu'elle savait comment chaque élément devait être prononcé (/fɪnt/, /θɪnt/, /fedʒ/, /θedʒ/, /felk/, /θelk/, /fɔld/, /θɔld/) respectivement (voir Cox & Palethorpe, 2007, pour une description des voyelles anglaises australiennes). Les éléments ont été présentés en cinq blocs aléatoires. L'expérimentateur passait à l'essai suivant en appuyant sur "y" sur le clavier. Si un élément devait être présenté à nouveau, l'expérimentateur appuyait sur la touche "n" et l'élément était réintroduit dans la séquence aléatoire.

Après l'enregistrement, le fichier audio était filtré en passe-haut à 100 Hz à l'aide du logiciel Praat (Boersma, 2001) afin d'éliminer tout grondement de basse fréquence et de corriger la composante continue. Les quatre meilleurs enregistrements de chaque élément ont été sélectionnés sur la base d'une inspection auditive et d'une analyse visuelle de la forme d'onde et du spectrogramme. Un fondu de 5 ms a été appliqué au début et à la fin de chaque énoncé.

Le stimulus vidéo devait commencer avant le début de l'énoncé pour s'assurer que la bouche de la locutrice soit en position neutre. Pour éviter toute information auditive non vocale qui pourrait influencer la reconnaissance avant le début du discours, une seconde avant et après chaque énoncé a été convertie en silence.

L'enregistrement audio filtré passe-haut a été importé dans Adobe Premiere Pro CC 2018 sur MacOS 10.14 et aligné sur la piste audio de la vidéo. Une fois aligné, l'audio de la caméra vidéo a été retiré. Les échelles temporelles de chaque énoncé, obtenues à partir de Praat, ont été converties en minutes, secondes et en échantillons audio de 1 à 44100 cycles par seconde. Ces valeurs ont été utilisées pour localiser avec précision le début et la fin de chaque vidéo dans Adobe Premiere Pro CC 2018. Chaque vidéo a été recadrée manuellement pour que la bouche soit au centre de l'image, puis enregistrée au format Quicktime (résolution 720 x 576, 25 images par seconde), ce qui a permis de préserver le format audio brut (PCM wave).

3. Procédure

La passation de l'expérience a été réalisée sur un ordinateur portable, utilisant le logiciel E-Prime version 3.0 qui contrôlait la présentation des stimuli et la collecte des réponses durant les épreuves proposées. Pour l'ensemble des épreuves, les stimuli auditifs étaient présentés via un casque audio. Les consignes étaient présentées à l'écrit sur l'ordinateur, et ont été récapitulées oralement par les expérimentatrices.

Les passations de l'expérience ont été réalisées dans un environnement calme : 64,4% au domicile des expérimentatrices et 35,6% dans une salle de travail fermée au sein de la bibliothèque universitaire de la faculté des sciences médicales et paramédicales de la Timone à Marseille. Les participants étaient testés individuellement.

L'expérience, qui se déroulait sur 2 jours consécutifs, durait environ 1h30 le premier jour et 30 minutes le deuxième jour. Notre protocole expérimental comprenait sept épreuves regroupées en quatre phases : une phase de pré-apprentissage, une phase d'apprentissage, une phase de post-apprentissage qui suivait immédiatement la phase d'apprentissage et une phase de post-apprentissage le lendemain afin d'observer l'impact du sommeil sur la consolidation des connaissances apprises le jour précédent.

Les participants passaient les épreuves dans l'ordre présenté ci-dessous.

a. Session 1

Le premier jour, l'expérimentation était composée de trois parties : une phase de pré-apprentissage, une phase d'apprentissage et une phase de post-apprentissage.

❖ **Phase de pré-apprentissage** : les participants devaient réaliser une tâche de discrimination « AX ». Il s'agit d'une tâche purement perceptive, permettant d'évaluer la capacité initiale de chaque participant à discriminer les sons /θ/ et /f/.

Dans cette tâche, quatre paires de syllabes suivantes ont été utilisées : /θa/-/fa/, /fa/-/θa/, /θa/-/θa/, /fa/-/fa/. Chaque paire était répétée 12 fois (au sein de chaque paire, il y avait quatre versions d'enregistrement, chacun des enregistrements a été répété trois fois). Dans les paires où la même syllabe était répétée (/θa/-/θa/ et /fa/-/fa/), deux enregistrements différents ont été utilisés. Les 48 paires étaient présentées dans un ordre aléatoire.

A chaque essai, le participant entendait deux syllabes de manière consécutive avec une seconde d'intervalle. Il devait les écouter attentivement et décider si la consonne initiale de la deuxième syllabe était de la même catégorie que celle de la première. Un exemple concret était donné (/ta/-/ta/ = même catégorie ; /ta/-/da/ = différente catégorie). Le participant était prévenu que les consonnes qu'il allait entendre venaient de l'anglais et que certaines pouvaient être difficiles à distinguer. S'il n'était pas certain de sa réponse, il devait répondre au hasard. Le participant devait cliquer sur la touche 1 s'il les jugeait similaires, et sur la touche 5 s'il les jugeait différentes. Aucun feed-back correctif n'apparaissait ensuite. Le participant avait deux secondes pour répondre. Une fois qu'une réponse avait été enregistrée, l'essai suivant commençait après une seconde. Dans le cas où le participant ne répondait pas dans les deux secondes, le message d'avertissement "veuillez répondre plus rapidement" s'affichait à l'écran et l'item était par la suite réintroduit de manière aléatoire. Une croix de fixation était présentée au centre de l'écran pendant l'intervalle entre deux essais.

Pour familiariser les participants à la tâche, nous les avions d'abord entraînés en les exposants à dix essais utilisant des syllabes issues du français.

❖ **Phase d'apprentissage** soumettait le participant à un entraînement spécifique selon son groupe d'appartenance (« auditif », « audio-orthographique », « audio-articulatoire »). Nous avons cherché à comparer l'efficacité des différents types d'entraînement. La phase d'apprentissage était composée de deux parties. La première partie consistait en une exposition passive où le participant était exposé de façon répétitive aux huit pseudo-mots durant environ 15 minutes. La deuxième partie, consistait en un entraînement actif avec un feed-back correctif où le participant devait associer le pseudo-mot entendu avec l'image correspondante.

- *Partie 1 : Exposition passive*

Lors de cette phase d'exposition passive, le participant devait apprendre huit pseudo-mots associé chacun à une image d'objet inexistant (Tableau 1). Les huit pseudo-mots formaient quatre paires minimales décrites dans la section *Matériel*. Chaque pseudo-mot avait quatre versions d'enregistrement. Chaque enregistrement était présenté six fois. Au total, cette tâche comportait 192 essais présentés aux participants de manière aléatoire.

A chaque essai, un pseudo-mot était présenté accompagné ou non de son information visuelle selon le groupe. Dans le groupe « auditif », les participants étaient soumis à une présentation purement auditive des pseudo-mots. Dans le groupe « audio-orthographique », les participants étaient soumis à une présentation auditive des pseudo-mots associée à leur

représentation orthographique. Dans le groupe « audio-articulatoire », les participants étaient soumis à une présentation auditive des pseudo-mots associée à une vidéo représentant les gestes articulatoires correspondants (Figure 4). Les participants devaient apprendre les associations entre les 8 pseudo-mots et les 8 images d'objets. Les participants étaient prévenus qu'ils seraient testés sur leurs connaissances de ces associations mot-image dans la suite de l'expérience. Cette phase d'exposition durait environ 15 minutes.

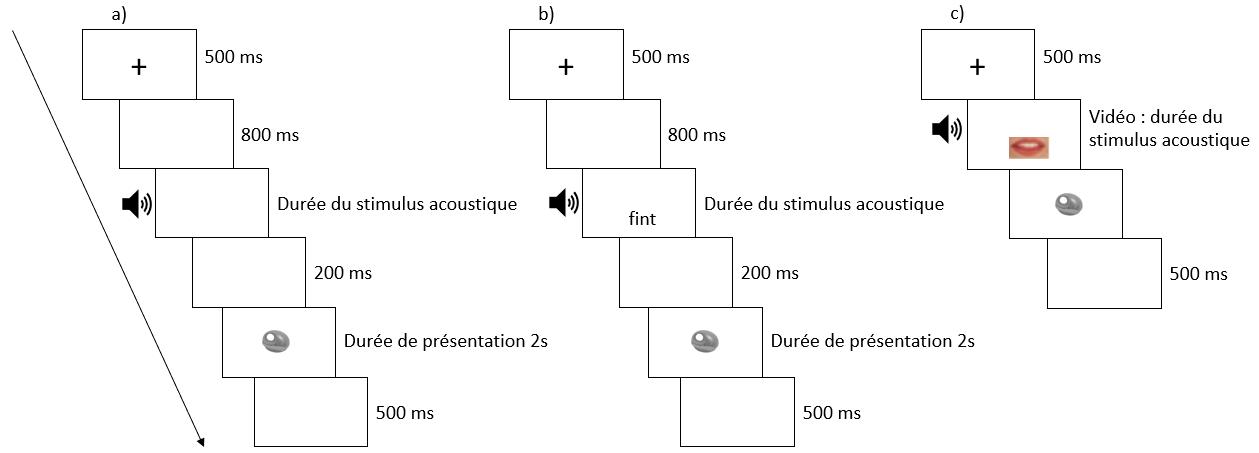


Figure 4 - Illustration du déroulement de la partie exposition passive de la phase d'apprentissage pour les groupes « auditif » (a), « audio-orthographique » (b) et « audio-articulatoire » (c).

- *Partie 2 : Entraînement avec feed-back*

Lors de cette phase d'apprentissage actif, le participant devait réaliser une tâche d'appariement mot-image. Dans chaque essai, un pseudo-mot était présenté au participant ; uniquement oralement pour le groupe « auditif », associé aux gestes articulatoires pour le groupe audio-articulatoire et associé la représentation orthographique pour le groupe « audio-orthographique ». Ensuite, deux images apparaissaient à gauche et à droite de l'écran : une image correspondant au pseudo-mot prononcé, l'autre correspondant à sa paire minimale. Le participant devait cliquer sur l'image correspondant au pseudo-mot entendu dans les 3000 ms imparties. Puis un feed-back de correction : « correct », « incorrect » ou « répondez plus vite » s'affichait à l'écran. Ce feed-back était également accompagné de l'image de l'objet qu'il fallait sélectionner et du pseudo-mot présenté uniquement oralement ou avec l'information visuelle selon le groupe d'appartenance. Le message feed-back et l'image de l'objet restaient sur l'écran pendant 2 secondes (Figure 5). Chaque pseudo-mot était répété 24 fois (six fois par enregistrement de chaque pseudo-mot) soit 192 essais au total, présentés dans un ordre aléatoire. Le participant était prévenu qu'il devait cliquer le plus vite possible sur l'image correspondant au pseudo-mot entendu, et que la réponse correcte lui serait présentée après avoir répondu. Cette tâche durait environ 20 minutes.

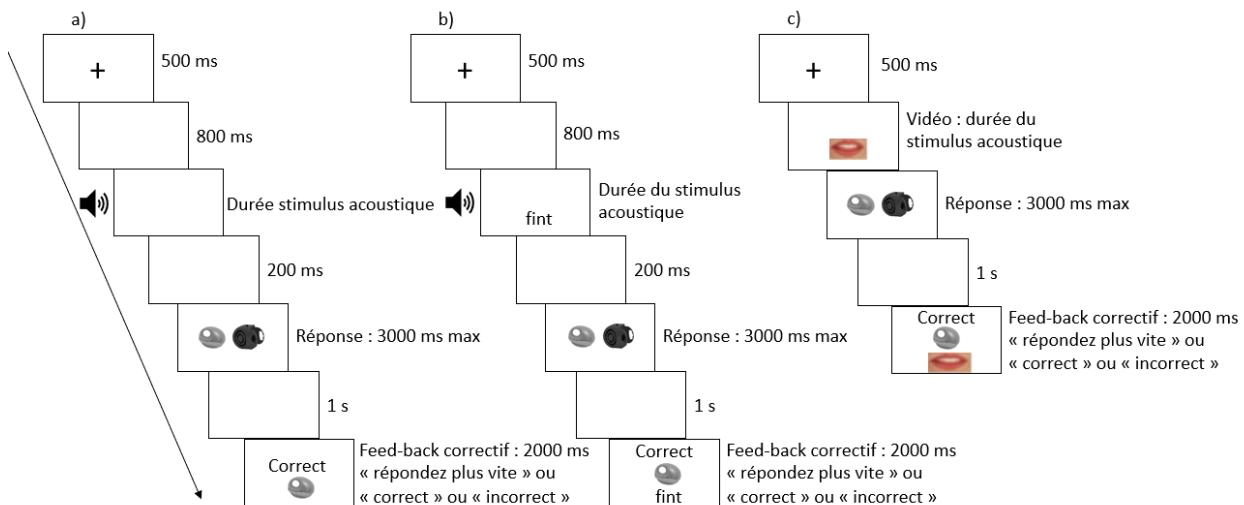


Figure 5 - Illustration du déroulement de la partie entraînement avec feed-back de la phase d'apprentissage pour les groupes « auditif » (a), « audio-orthographique » (b) et « audio-articulatoire » (c).

- ❖ **Phase de post-apprentissage** était composée par la tâche de discrimination AX décrite ci-dessus et une tâche d'association mot-image.

Dans la tâche d'association mot-image, nous cherchions à évaluer la capacité des participants à reconnaître correctement des pseudo-mots qui différaient par leurs consonnes initiales /θ/ et /f/, et leur capacité à associer les huit pseudo-mots aux huit objets. Lors de cette tâche, un pseudo-mot était présenté oralement, suivi des huit images présentées pendant dix secondes. Les images étaient disposées sur deux lignes et quatre colonnes. La position des images changeait entre chaque présentation. Le participant devait cliquer sur l'image correspondant au pseudo-mot entendu (Figure 6). Le participant était prévenu qu'il devait cliquer le plus vite possible sur l'image associée au pseudo-mot entendu, et qu'aucun feedback ne lui serait donné à la fin de l'essai. Chaque pseudo-mot avait 4 enregistrements, chaque enregistrement était présenté deux fois, menant à 64 essais au total. Les essais étaient présentés dans un ordre aléatoire.

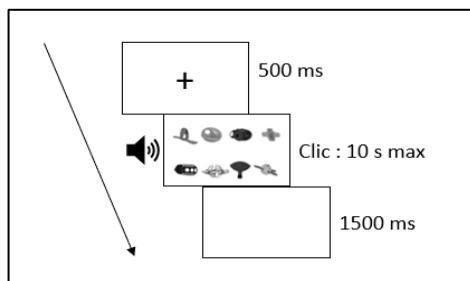


Figure 6 - Illustration du déroulement de la tâche d'association mot-image

b. Session 2

- ❖ **Phase de post-apprentissage** : le deuxième jour, le participant accomplissait à nouveau les différentes tâches réalisées la veille lors de la phase post-apprentissage, c'est-à-dire la tâche de discrimination AX et la tâche d'association mot-image.

IV. PRESENTATION DES RESULTATS

1. Résultats de la tâche de discrimination AX

Dans un premier temps, nous avons récolté les scores bruts de chaque candidat dans la tâche de discrimination AX. Ensuite, nous nous sommes appuyées sur le calcul de la valeur d'A-prime proposé par Snodgrass, Levy-Berger & Haydon (1985). Cette valeur prenant en considération le taux de détection correcte et le taux de fausse-alarme permet d'estimer le niveau de sensibilité de chaque participant à la similitude/différence entre deux phonèmes.

Si le taux de détection correcte était égal au taux de fausse-alarme, alors le participant répondait au hasard et A-prime valait 0,5. Plus le taux de détection correcte dépassait le taux de fausse-alarme, plus la sensibilité de discrimination du participant était bonne. Un score de 1 indiquait donc une sensibilité parfaite, tandis qu'un score de 0,5 ou moins indiquait un manque de sensibilité.

Dans notre expérience, la création des trois groupes s'est basée sur les performances de chaque participant lors de la tâche de discrimination réalisée dans le cadre du mémoire d'orthophonie de M.S Villain-Bailly et E.Clot (2019). Les trois groupes ont été créés de manière à ce qu'ils aient globalement un niveau de discrimination similaire avant le début de l'expérimentation. A l'issue de notre expérience, nous avons cherché à vérifier ce paramètre dans les trois groupes.

En pré-apprentissage, la valeur A-prime moyenne des participants dans le groupe auditif était de 0,69, celle des groupes audio-orthographique et audio-articulatoire de 0,68. En comparant les performances entre les participants des groupes auditif et audio-orthographique, nous n'avons pas observé de différences significatives [$t(64) = 0,45, p=0,65$]. Il en a été de même lors de la comparaison des groupes auditif et audio-articulatoire [$t(66) = 0,43, p=0,66$], ainsi qu'entre les groupes audio-orthographiques et audio-articulatoires [$t(64) = 0,04, p=0,97$]. Ces résultats nous ont permis de confirmer que les trois groupes étaient appariés au niveau de leur capacité perceptive évaluée par la tâche AX avant la phase d'apprentissage.

Dans un second temps, nous avons transformé la valeur d'A-prime en pourcentage de changement de performance par rapport au niveau initial des participants évalué lors de la tâche AX effectuée en pré-apprentissage, en utilisant la formule :

$$\left(\frac{A' \text{ post-apprentissage} - A' \text{ pré-apprentissage}}{A' \text{ pré-apprentissage}} \right) \times 100 = \text{pourcentage de changement de performance}$$

Le but était de connaître le taux d'amélioration du participant par rapport à son niveau de discrimination initial. Ce calcul nous a permis d'observer le pourcentage de changement de performance entre la tâche AX effectuée en pré-apprentissage et celle effectuée en post-apprentissage dans les 2 sessions (Jour 1 et Jour 2).

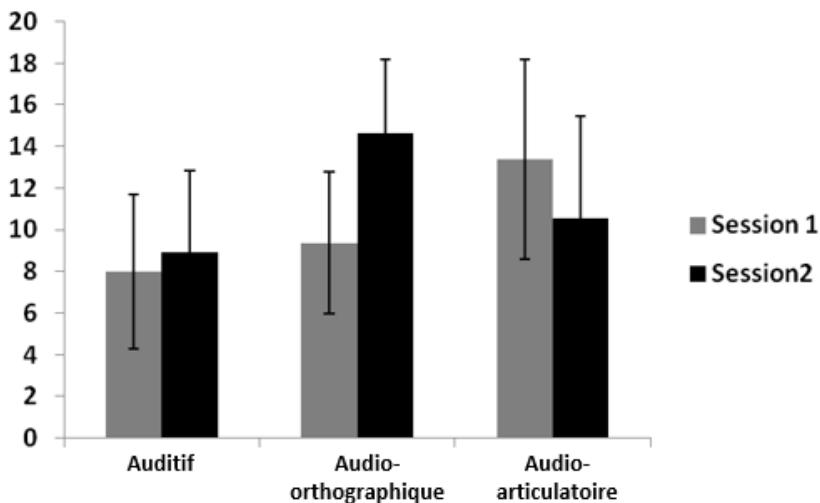


Figure 7 - Pourcentage de changement de performance dans la tâche de discrimination pour les trois types d'apprentissage et par session. Les barres d'erreur représentent les erreurs standards

Cette figure 7 illustre le pourcentage de changement de performance des participants dans la tâche de discrimination, en fonction de leur groupe et de la session d'expérimentation. Le niveau 0 signifiait que le résultat A-prime obtenu en pré-apprentissage était égal au résultat A-prime obtenu en post-apprentissage, et donc qu'il n'y avait aucune amélioration par rapport au niveau de départ mesuré avant la phase d'apprentissage. En revanche, un taux d'amélioration positif indiquait que le participant avait progressé entre les épreuves de pré- et de post-apprentissage. Dans cette figure, nous avons observé un taux d'amélioration positif dans chaque groupe et pour chaque session. Cela signifiait que pour chaque groupe et pour chaque session le participant possédait une meilleure capacité de discrimination après la phase d'apprentissage qu'avant.

Nous avons ensuite regardé si les taux d'amélioration illustrés dans la figure 7 étaient influencés par le type d'entraînement, la session ou encore l'interaction entre ces deux paramètres. Nous avons ainsi effectué une analyse de variance (ANOVA) mixte sur les taux d'amélioration, en considérant le type d'entraînement (auditif, audio-orthographique et audio-articulatoire) comme variable inter-participant, et la session (session 1 et session 2) comme variable intra-participant. L'effet principal de session [$F(1, 195) = 0,1, p=0,75$] n'était pas significatif, il n'y avait également pas d'effet significatif du type d'entraînement utilisé [$F(2, 195) = 0,52, p=0,96$]. De plus, l'interaction entre ces deux facteurs n'était pas significative [$F(2, 195) = 0,48, p = 0,62$]. Contrairement à ce que nous pouvions constater en observant la Figure 7, l'analyse statistique a suggéré que les taux d'amélioration ne variaient pas selon le type d'entraînement, la session ou l'interaction entre ces deux facteurs.

Toutefois, lorsque nous avons calculé les valeurs A-prime de chaque participant, nous nous sommes rendues compte qu'il y avait deux profils de participants. Certains participants n'avaient pas de sensibilité de discrimination entre le /θ/ et le /f/ avec A-prime équivalent à 0,5, alors que d'autres semblaient avoir une bonne sensibilité de discrimination /θ/-/f/ avec A-prime supérieur à > 0,7. Nous nous sommes demandées si cette variabilité inter-participants était l'origine de l'absence de résultat significatif dans l'ANOVA ci-dessus.

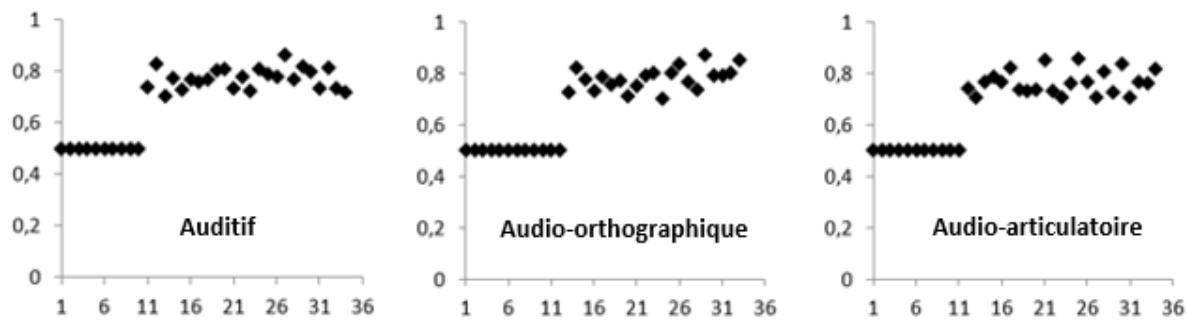


Figure 8 - Résultat de chaque participant en phase de pré-apprentissage en fonction du type d'entraînement.

Afin d'explorer si la capacité initiale de nos participants à discriminer les phonèmes /θ/ et /f/ avant la phase d'apprentissage jouait un rôle important, nous avons séparé chaque groupe de participants en deux sous-groupes selon leur score A-prime obtenu lors de la phrase de pré-apprentissage : ceux ayant les valeurs d'A-prime = 0,5 (groupe peu sensible) et ceux ayant les valeurs d'A-prime > 0,5 (groupe sensible).

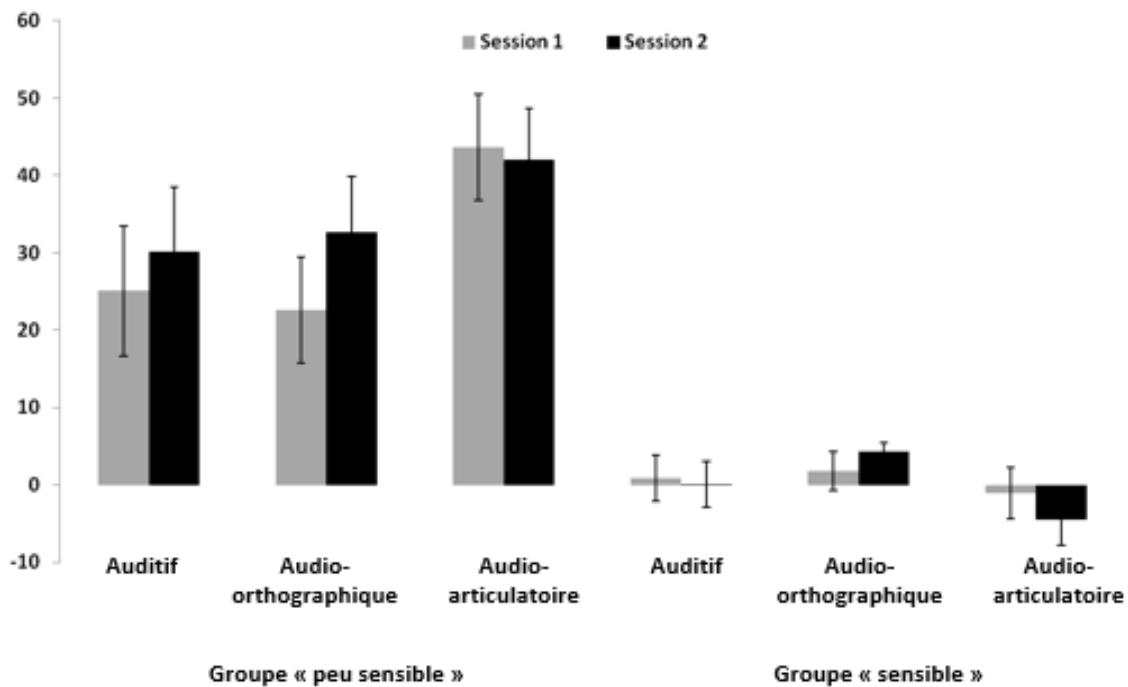


Figure 9 - Pourcentage de changement de performance dans la tâche de discrimination par niveau de sensibilité, pour les trois types d'entraînement et par session. Les barres d'erreur représentent les erreurs standards

Cette figure 9 illustre le pourcentage de changement de performance des participants dans la tâche de discrimination, en fonction de leur sensibilité à la discrimination /θ/- /f/, du type d'entraînement et de la session d'expérimentation. Dans cette figure, nous observons que, tout de suite après l'entraînement, il y une amélioration de la capacité de discrimination /θ/-/f/ chez les participants à la base peu sensibles ($p < 0,02$ pour les trois groupes de participants) mais pas chez les participants sensibles ($p > 0,05$ pour les trois groupes de participants).

Nous avons ensuite effectué une ANOVA à trois facteurs sur le pourcentage de changement de performance, avec le groupe de niveau de sensibilité (peu sensible et sensible) et le type d'entraînement (auditif, audio-orthographique et audio-articulatoire) comme variables inter-participant, et la session (session 1 et session 2) comme variable intra-participant. L'effet principal du niveau de sensibilité [$F(1, 189)= 146,04, p=0,0001$] était significatif, mais l'effet d'entraînement [$F(2, 189)=0,35, p=0,70$] n'était pas significatif, de même que l'effet de session [$F(1, 189)=0,18, p=0,67$]. En ce qui concerne les résultats des interactions entre ces facteurs, nous avons observé que l'interaction entre le niveau de sensibilité et l'entraînement [$F(2, 189)=6,18, p=0,003$] était significative. En revanche, l'interaction entre le niveau de sensibilité et la session [$F(1, 189)=0,99, p=0,32$], l'interaction entre l'entraînement et la session [$F(2, 189)=0,81, p=0,44$] et l'interaction entre le niveau de sensibilité, l'entraînement et la session [$F(2, 189)=0,11, p=0,9$] n'étaient pas significatives. L'interaction significative entre le niveau de sensibilité et l'entraînement suggérait que l'influence du type d'entraînement variait en fonction du niveau de sensibilité à la discrimination /θ/- /f/. observée avant la phase d'apprentissage. Pour comprendre cette interaction, nous avons par la suite comparé l'efficacité des différents types d'entraînement au sein des groupes « peu sensible » et « sensible ».

Nous avons commencé par analyser les résultats des participants faisant partie du groupe "peu sensibles" à la discrimination /θ/- /f/. Nous avons comparé les résultats des participants peu sensibles ayant reçu un entraînement auditif avec ceux ayant reçu un entraînement audio-orthographique. Les deux moyennes étaient équivalentes : ils avaient tous augmenté de 27%, et il n'y avait pas de différence significative entre les deux groupes [$t(39)= -0,01, p=0,99$]. Ainsi, dans le groupe peu sensible, le fait d'être exposé à l'orthographe en plus de l'information auditive n'a pas induit une plus grande amélioration que le fait d'être exposé à un entraînement auditif simple.

Une comparaison entre les participants « peu sensibles » ayant reçu un entraînement auditif et ceux ayant reçu un entraînement audio-articulatoire a montré une plus forte amélioration chez ceux qui ont reçu un entraînement audio-articulatoire (27,58% vs. 42,84% d'amélioration pour les groupes auditif et audio-articulatoire respectivement : $t(37)= -2,05, p=0,05$). Ce résultat signifie que dans le groupe peu sensible, le fait d'être exposé à un entraînement audio-articulatoire induisait une amélioration significativement plus grande que le fait d'être exposé à un entraînement purement auditif. Enfin, une comparaison entre les participants peu sensibles qui ayant reçu un entraînement audio-orthographique et ceux ayant reçu un entraînement audio-articulatoire (27,64% vs. 42,84% d'amélioration pour les groupes audio-orthographique et audio-articulatoire respectivement) a également montré une plus forte amélioration dans le groupe audio-articulatoire [$t(44)= -2,23, p=0,03$].

En résumé, les participants peu sensibles à la discrimination /θ/-/f/ avait un taux d'amélioration significativement meilleur lorsqu'ils étaient soumis à un entraînement audio-articulatoire plutôt qu'à un entraînement audio-orthographique ou purement auditif. Les entraînements auditifs et audio-orthographiques ont tout de même permis une amélioration, même si celle-ci restait moins importante que dans le cadre de l'entraînement audio-articulatoire. Nous

avons également constaté que l'entraînement audio-orthographique ne permettait pas d'améliorer la capacité de discrimination des participants par rapport à un entraînement purement auditif. Dans cette tâche de discrimination AX, l'information orthographique ne constituait donc pas une aide supplémentaire à l'augmentation de la performance de discrimination. Ces résultats coïncidaient avec les résultats visibles dans la figure 9.

Ensuite, nous avons analysé les résultats des participants présentant un bon niveau de sensibilité à la discrimination /θ/-/f/ avant la phase d'apprentissage. Nous avons comparé les résultats des participants ayant reçu un entraînement auditif et ceux ayant reçu un entraînement audio-orthographique. Les deux moyennes étaient légèrement différentes, avec 0,4% d'augmentation pour le groupe auditif et 3% pour le groupe audio-orthographique. La différence entre les deux groupes n'était pas significative [$t(80) = -1,03, p=0,31$]. Nous avons ensuite fait de même entre les participants sensibles ayant reçu un entraînement auditif et ceux ayant reçu un entraînement audio-articulatoire. Les deux moyennes étaient peu différentes, avec 0,4% d'amélioration pour les participants du groupe auditif contre une baisse de performance de 2,7% pour ceux du groupe audio-articulatoire. La différence n'était pas significative [$t(90) = 1,03, p=0,30$]. En ce qui concerne la comparaison entre les résultats des participants sensibles ayant reçu un entraînement audio-orthographique et ceux ayant reçu un entraînement audio-articulatoire, les deux moyennes étaient également peu différentes. Nous avons observé 3% d'amélioration pour le groupe audio-orthographique contre une baisse de performance de 2,7% pour le groupe audio-articulatoire. Nous avons observé une différence significative [$t(72) = -2,12, p=0,04$]. Cependant, cette différence significative était probablement due au fait que les deux types d'entraînement montrent deux tendances inverses.

En résumé, contrairement aux résultats observés dans le groupe « peu sensible », le taux d'amélioration des participants sensibles à la discrimination /θ/-/f/ avant la phase d'apprentissage dépendait peu du type d'entraînement reçu.

2. Résultats de la tâche d'association mot-image

D'après la Figure 10, nous observons que tous les groupes de participants ont des résultats supérieurs au seuil du hasard (12,5% correspondant au pourcentage de chance de sélectionner la bonne image parmi les 8 images proposées), et ce pour les trois groupes de participants ($p < 0,03$). Ceci semble être le cas également lorsque nous avons considéré séparément les participants « sensibles » et « peu sensibles » au contraste /θ/-/f/ (Tableau 2). Ainsi, les différents entraînements ont permis aux participants d'apprendre les nouveaux mots.

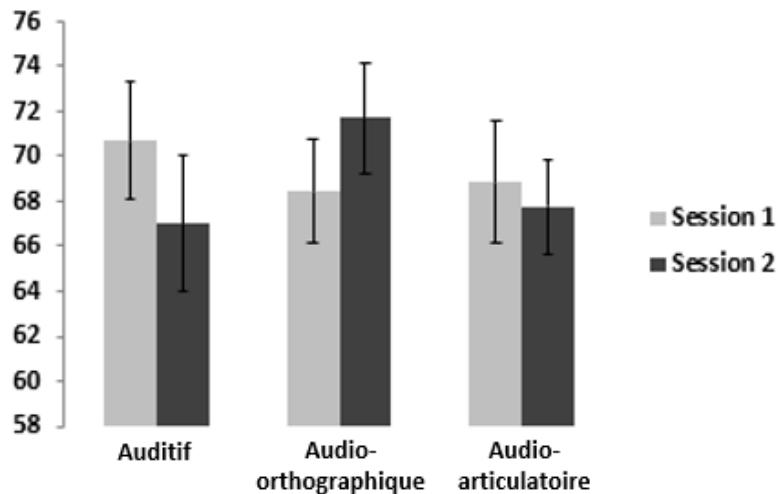


Figure 10 - Pourcentage de bonnes réponses des participants dans la tâche d'association mot-image en fonction du type d'entraînement et de la session.

Sensibilité au contraste /θ/-/f/	Entrainement	Session	Moyenne	Ecart-type
Peu sensible	Auditif	1	67,66%	17,65
		2	62,34%	17,59
	Audio-orthographique	1	64,84%	11,06
		2	65,76%	15,32
	Audio-articulatoire	1	61,36%	11,76
		2	63,07%	9,41
	Auditif	1	71,94%	14,44
		2	68,95%	17,54
	Audio-orthographique	1	70,46%	14,10
		2	75,07%	12,67
	Audio-articulatoire	1	72,42%	16,26
		2	70,40%	12,98

Tableau 2 - Moyenne et écart-type du pourcentage de bonnes réponses dans la tâche d'association mot-image en fonction du niveau de sensibilité du participant et de la session pour chaque type d'entraînement.

Nous avons effectué une ANOVA à trois facteurs sur le pourcentage de bonnes réponses, avec le niveau de sensibilité (peu sensible et sensible) et le type d'entraînement (auditif, audio-orthographique et audio-articulatoire) comme variables inter-participantes, et la session (session 1 et session 2) comme variable intra-participante.

Le principal effet du niveau de sensibilité [$F(1, 95)= 6,05, p=0,02$] était significatif. L'effet d'entraînement [$F(2, 95)=0,21, p=0,81$] n'était pas significatif, de même que l'effet de session [$F(1, 95)=0,43, p=0,52$]. En ce qui concerne les résultats des interactions entre ces facteurs, nous avons observé que l'interaction entre l'entraînement et la session [$F(2, 95)=5,10, p=0,01$] était significatif. En revanche, l'interaction entre le niveau de sensibilité et la session [$F(1, 95)=0,13, p=0,72$], l'interaction entre le niveau de sensibilité et l'entraînement [$F(2, 95)=0,12, p=0,89$] et l'interaction entre le niveau de sensibilité, l'entraînement et la session [$F(2, 95)=1,87, p=0,16$] n'étaient également pas significatives.

Afin d'examiner l'interaction significative entre la session et l'entraînement, nous avons par la suite comparé l'effet de session sur chaque type d'entraînement.

Nous avons analysé les résultats des participants ayant bénéficié d'un entraînement auditif en comparant le pourcentage de bonnes réponses obtenu à la session 1 avec celui obtenu à la session 2. La différence entre les performances des deux jours était significative [$F(1, 33)= 7,61$, $p=0,01$]. La mise en relation de ce résultat avec ceux de la figure 10 a permis de déduire que, suite à un entraînement auditif, le pourcentage de bonnes réponses de la première session était significativement meilleur que celui de la deuxième session.

Nous avons ensuite fait de même pour les participants ayant bénéficié d'un entraînement audio-orthographique, en comparant le pourcentage de bonnes réponses obtenu à la session 1 avec celui obtenu à la session 2. La différence entre les performances des deux jours était significative [$F(1, 32)= 5,05$, $p=0,03$]. La mise en relation de ce résultat avec ceux de la figure 10 a permis de déduire que suite à un entraînement audio-orthographique le pourcentage de bonnes réponses de la deuxième session est significativement meilleur que celui de la première session.

Ensuite, nous avons analysé les résultats des participants ayant bénéficié d'un entraînement audio-articulatoire en comparant le pourcentage de bonnes réponses obtenu à la session 1 avec celui obtenu à la session 2. La différence entre les performances des deux jours n'était pas significative [$F(1, 33)= 0,48$, $p=0,49$]. Ainsi, pour les participants ayant reçu un entraînement audio-articulatoire, le pourcentage de bonnes réponses de la première session est approximativement similaire à celui de la deuxième session.

En résumé, l'évolution du pourcentage de bonnes réponses entre les 2 jours évolue différemment selon le type d'entraînement reçu par les participants. Lorsque le participant a reçu un entraînement auditif, nous constatons une chute des performances entre les deux jours. Inversement, pour les participants ayant reçu un entraînement audio-orthographique, nous constatons une amélioration des performances. En ce qui concerne les participants ayant reçu un entraînement audio-articulatoire, nous constatons aucune évolution.

Par la suite, nous avons également cherché à comparer les pourcentages de bonne réponse entre les trois types d'entraînement en fonction de la session. Pour la session 1, le pourcentage de bonnes réponses des participants ne diffère pas significativement entre les trois types d'entraînement [$F(2, 98)=0,22$, $p=0,80$]. Pour la session 2, de la même manière, le pourcentage de bonnes réponses des participants ne diffère pas significativement entre les trois types d'entraînement [$F(2, 98)=0,95$, $p=0,39$]. Ainsi, pour les deux sessions, le pourcentage de bonnes réponses à la tâche d'association mot-image est approximativement similaire quel que soit l'entraînement dont a bénéficié le participant.

V. DISCUSSION DES RESULTATS

1. Objectifs de l'expérimentation

De nombreuses recherches ont examiné la capacité des adultes à percevoir des sons d'une langue étrangère. Dans de nombreux cas, les apprenants d'une deuxième langue perçoivent les sons de cette langue différemment des locuteurs natifs. Des études comme celle de Guion et al. (2000) ont néanmoins montré que l'entraînement peut faire évoluer la capacité de perception des apprenants. Le Speech Learning Model (Santiago, 2012) défend l'importance d'un entraînement purement auditif pour améliorer la perception d'un son de L2. A cela s'ajoutent des études montrant l'utilité de l'information articulatoire et orthographique dans le traitement de la parole (Luce & Pisoni, 1998 ; Peelle & Sommers, 2015 ; Frost, Repp & Katz, 1988 ; Ricketts, Bishop & Nation, 2009). L'objectif de cette présente étude était de comparer l'efficacité d'un entraînement purement auditif à l'efficacité de deux types d'entraînement audiovisuel, afin d'étudier la contribution de l'information orthographique et des gestes articulatoires à la capacité de perception du phonème /θ/, ainsi que la contribution de ces informations visuelles à la capacité de reconnaissance et de mémorisation des nouveaux-mots contenant ce phonème chez des participants francophones européens.

Pour cela, nous avons élaboré une expérimentation permettant d'évaluer l'effet des trois types d'entraînements (auditif, audio-articulatoire et audio-orthographique) dans deux types de tâches. Dans une tâche de discrimination (tâche de bas-niveau), nous avons évalué la capacité de perception du phonème anglais /θ/ par rapport à son voisin articulatoire, le phonème français /f/. Puis, dans une tâche d'association mot-image, nous avons évalué la capacité de reconnaissance et de mémorisation de nouveaux-mots, des paires minimales contenant les phonèmes /θ/ et /f/, appris lors de la phase d'apprentissage.

En nous basant sur les résultats des études citées précédemment, nous avons émis l'hypothèse d'une amélioration de la perception du son /θ/ pour chacun des trois types d'entraînement grâce à l'exposition aux stimuli auditifs. Également, nous nous attendions à ce que la capacité de perception évaluée dans la tâche de discrimination AX, ainsi que l'apprentissage des nouveaux mots évalué dans la tâche d'association mot-image soient significativement meilleurs après un entraînement audio-articulatoire qu'après un entraînement purement auditif, ou encore après un entraînement audio-orthographique qu'après un entraînement purement auditif. Pour finir, nous avons supposé que, quel que soit le type d'entraînement, les résultats des participants à la tâche d'association mot-image soient significativement meilleurs après une nuit de sommeil puisque cette dernière favorise la reconnaissance et la mémorisation.

2. Commentaires des résultats

a. Tâche perceptive

L'analyse des résultats de la tâche de discrimination AX, une tâche de bas niveau, réalisée lors de la phase de pré-apprentissage a montré que les participants des trois groupes d'entraînement présentaient une capacité initiale de discrimination des phonèmes /θ/ et /f/ statistiquement similaire. Les résultats de la tâche de discrimination AX réalisée lors de la phase de post-apprentissage ont montré que les participants ont amélioré leur capacité de perception du phonème /θ/ quel que soit le type d'entraînement dont ils ont bénéficié (auditif seul, audio-orthographique ou audio-articulatoire). En revanche, le taux d'amélioration des participants à la tâche de discrimination /f/-/θ/ était dépendant de leur sensibilité initiale. Ainsi, pour l'analyse de cette tâche, les participants ont été séparés en 2 sous-groupes en fonction de leur niveau de sensibilité initial à la discrimination /f/-/θ/ : les "peu sensibles" et les "sensibles" à cette discrimination. L'analyse des résultats a montré que dans cette tâche de bas niveau, les participants ayant un taux de sensibilité élevé au début de notre expérience ne montraient aucune augmentation significative de la capacité de discrimination des phonèmes /f/ et /θ/ après avoir reçu un entraînement. En revanche, les participants initialement peu sensibles à cette discrimination bénéficiaient des trois types d'entraînement et progressaient dans leur capacité de discrimination.

Nous avons observé dans notre expérimentation que, chez les participants initialement peu-sensibles, l'exposition aux stimuli purement auditifs permet l'augmentation des capacités de discrimination du participant par rapport à son niveau de base mesuré lors de la tâche de discrimination AX réalisée avant la phase d'apprentissage. Ce résultat est conforme à la prédiction du Speech Learning Model, un modèle d'apprentissage auditif des sons d'une nouvelle langue défendant l'importance de l'entraînement de la perception auditive pour améliorer la capacité de discrimination de deux sons et pour créer de nouvelles catégories de sons de la L2 (Santiago, 2012).

La comparaison des résultats de la tâche de discrimination AX pour l'entraînement purement auditif avec les entraînements audio-orthographiques et audio-articulatoires nous a permis d'examiner l'influence des informations orthographiques et articulatoires sur la capacité de perception des participants.

L'analyse des résultats pour les participants ayant reçu un entraînement audio-orthographique a montré un taux d'amélioration similaire à ceux ayant reçu un entraînement purement auditif, quel que soit leur niveau de sensibilité. Ainsi, la présence de l'information orthographique n'entraîne pas de bénéfice supplémentaire dans cette tâche de discrimination par rapport à un entraînement purement auditif. Ces résultats ne correspondent pas notre prédiction qui supposait que la capacité perceptive évaluée dans la tâche de discrimination AX serait meilleure après un entraînement audio-orthographique qu'après un entraînement purement auditif. Cela pourrait être expliqué par le statut lexical de stimuli utilisés dans notre expérimentation par rapport à ceux utilisés dans l'étude de Frost, Repp et Katz (1988). En effet, Frost a démontré l'utilité

du code orthographique dans la capacité de détection de la parole uniquement si les mots écrits étaient des mots existants. Dans notre expérience, l'entraînement s'est basé sur l'apprentissage de non-mots n'existant dans aucune des deux langues. L'influence de l'orthographe dans la perception de la parole pourrait donc être dépendante du statut lexical des stimuli utilisés.

La similitude des résultats dans la tâche de discrimination pour les participants ayant reçu un entraînement auditif et audio-orthographique pourrait également s'expliquer par le fait que pendant l'entraînement les participants du groupe audio-orthographique se seraient reposés sur la fiabilité et la clarté de l'information orthographique au détriment d'une information auditive plus ambiguës. En effet, de nombreux participants du groupe audio-orthographique estimaient que l'orthographe constituait une aide précieuse à l'apprentissage, et ont rapporté s'être appuyé davantage sur l'information écrite que sur l'information auditive pour apprendre les nouveaux-mots contenant les phonèmes /f/-/θ/ lors de la phase d'apprentissage. Les participants se seraient énormément appuyés sur l'orthographe lors de la phase d'apprentissage, sans pour autant que leur capacité de perception dans la tâche de discrimination AX ne progresse plus que celle des participants ne disposant pas de cette information orthographique. Cela montre, qu'au moins dans le protocole utilisé dans notre étude, l'information orthographique n'apporte pas de bénéfice supplémentaire à la capacité perceptive par rapport à l'entraînement auditif seul.

Cette idée peut être illustrée par le modèle d'activation interactive de Grainger et Ziegler (2007), qui affirme que la compatibilité entre le code orthographique et phonologique affecte le traitement de la parole. Lors du traitement de la parole, la répartition des ressources attentionnelles donnée aux différentes informations sensorielles est ajustée en fonction des informations présentes dans le stimulus. Ainsi, les participants ayant reçu un entraînement purement auditif bénéficieraient uniquement de l'information auditive et se concentreraient entièrement sur le signal acoustique, et mettraient donc toute leur attention à l'écoute des informations phonologiques données par le signal acoustique. En revanche, lors de la présentation d'un stimulus acoustique accompagné de l'information orthographique correspondante, les participants bénéficieraient de 2 informations sensorielles : l'information acoustique et l'information orthographique. Ils partageraient donc leur attention, en dirigeant moins d'attention sur l'information auditive que sur l'information orthographique.

Dans notre expérience, les participants ayant reçu un entraînement audio-orthographique ont ainsi partagé leur attention entre les deux informations orthographique et acoustique. Ainsi, la présence de l'orthographe pendant l'apprentissage pourrait avoir une influence résiduelle sur la tâche AX, en apportant une réorganisation des entrées sensorielles : plus d'attention allouée à l'information orthographique pour moins d'attention allouée au signal acoustique. Cela pourrait expliquer les résultats équivalents entre les participants ayant bénéficié d'un entraînement audio-orthographique et ceux ayant reçu un entraînement auditif.

Pour les participants peu sensibles ayant reçu un entraînement audio-articulatoire, l'analyse des résultats a montré que leur taux d'amélioration est significativement meilleur que pour ceux

ayant reçu un entraînement purement auditif ou accompagné d'une information orthographique, et ce même avec un apprentissage implicite de la discrimination. Ainsi, comme prédit par notre hypothèse, la présence de l'information articulatoire a permis l'augmentation des performances des capacités de discrimination des sons d'une L2, au moins chez les apprenants peu sensibles.

Ces résultats peuvent être expliqués par l'étude de Peelle et Sommers (2015) qui affirme que pour optimiser le traitement de la parole, nous avons besoin non seulement du signal acoustique, mais aussi des autres signaux sensoriels et des informations linguistiques (contexte lexical, syntaxique et sémantique). Dans l'entraînement audio-articulatoire reçu par les participants, l'ensemble de ces éléments était donné. La présentation des traits articulatoires ajoutés aux traits phonologiques de chaque son constitue un signal sensoriel visuel qui permet la création d'une cartographie des repères des indices audio-visuels associés à la parole en extrayant les traits de chaque son (Patterson & Werker, 2003). Ainsi, chez les participants initialement peu sensibles, l'information articulatoire couplée à l'information auditive permettrait de maximiser le traitement des traits phonétiques associés aux phonèmes ambigus et conduit ainsi à l'amélioration de la perception de ces phonèmes. Cette notion est rapportée dans l'étude de Fenwick, Best, Davis, Tyler (2017), affirmant qu'un grand nombre d'informations articulatoires saillantes peut mettre en évidence les différences phonétiques.

Pour les participants sensibles ayant reçu un entraînement audio-articulatoire, l'analyse des résultats n'a pas montré d'amélioration significative du taux de discrimination suite à cet entraînement. Cela nous permet de mitiger notre hypothèse, en affirmant que la présence de l'information articulatoire permet l'augmentation des capacités de discrimination des apprenants des sons d'une L2, mais ceci uniquement pour des personnes ayant un niveau de sensibilité initialement faible. Pour expliquer cette distinction d'amélioration de la discrimination entre les participants sensibles et peu sensibles, nous supposons que les participants possédant initialement un haut niveau de sensibilité auraient besoin d'un entraînement plus spécifique et individualisé pour leur permettre une progression dans leurs capacités de discrimination déjà élevées. Cette supposition serait renforcée par le fait que chez les participants initialement sensibles aucun des trois types d'entraînements ne permettait l'augmentation significative de cette capacité de discrimination.

En résumé, dans cette tâche de bas niveau, les participants ayant un taux de sensibilité élevé au début de notre expérience ne bénéficiaient d'aucune augmentation de la capacité de discrimination des phonèmes /f/ et /θ/ grâce à l'entraînement. En revanche, les participants initialement peu sensibles amélioraient significativement leur capacité de discrimination des phonèmes /f/ et /θ/ grâce aux trois types d'entraînement, même si l'entraînement audio-visuel paraissait manifestement le plus bénéfique des trois. L'amélioration de la perception d'un son de L2 par un apprenant est donc dépendante de la capacité perceptive initiale du participant, ainsi que de l'entraînement reçu par ce dernier. Par ailleurs, la comparaison des résultats entre les deux sessions montre que la capacité de discrimination des participants est maintenue le lendemain.

b. Tâche de reconnaissance et de mémorisation

L'analyse des résultats de la tâche d'association mot-image, une tâche de haut niveau, a montré que les participants ont réussi à apprendre les nouveaux mots contenant les phonèmes /f/ et /θ/ étant donné que leur performance était au-dessus du seuil du hasard.

Lors de la première session, les trois groupes de participants avaient un taux de bonnes réponses supérieur au seuil du hasard. Les apprenants de L2 peuvent ainsi apprendre un vocabulaire de 8 mots de L2 contenant un contraste phonémique qui n'existe pas dans leur langue maternelle en seulement 30 minutes. Ils sont donc capables de distinguer deux paires minimales, ce qui implique qu'ils étaient capables de percevoir la différence entre /θ/ et /f/ après une très courte période d'apprentissage. Ce niveau de performance significativement supérieur au niveau du hasard était également observé dans la deuxième session.

Ces résultats indiquent que les apprenants avaient préservé les connaissances lexicales des mots appris la veille. Cependant, l'absence de différence significative entre le niveau de performance observé dans les trois groupes de participants ne nous a pas permis de valider l'hypothèse selon laquelle le niveau de reconnaissance et de mémorisation serait meilleur après un entraînement audio-articulatoire ou audio-orthographique qu'après un entraînement purement auditif.

Nous avons toutefois observé que les différents types d'entraînement n'avaient pas le même impact sur la manière dont les connaissances apprises étaient consolidées. L'analyse de la différence de taux de bonnes réponses obtenus dans les deux sessions pour chaque type d'entraînement nous a montré que le taux de bonnes réponses obtenu par les participants ayant reçu un entraînement purement auditif avait significativement diminué lors de la deuxième session. Le fait d'être exposé uniquement aux informations auditives n'a donc pas permis aux participants de maintenir ni de consolider les connaissances apprises. En ce qui concerne les participants ayant reçu un entraînement auditif avec information articulatoire avaient un taux de bonnes réponses équivalent entre les deux sessions. Bien que ce type d'entraînement ait permis le maintien des connaissances entre les deux sessions expérimentales, son apport à la consolidation des connaissances apprises semble être plus limité que celui de l'entraînement audio-orthographique. En effet, les participants ayant reçu un entraînement auditif avec information orthographique, le taux de bonnes réponses données était significativement meilleur lors de la deuxième session que lors de la première. La présence du code orthographique a donc permis la consolidation des acquis.

L'idée que l'information orthographique soutiendrait la consolidation est relevée dans de nombreuses études. L'expérimentation de Ricketts, Bishop et Nation (2009) a permis de mettre en avant l'idée que l'information orthographique favorise l'apprentissage ainsi que la mémorisation des formes phonologiques. L'écrit met en relation l'orthographe qui est une représentation abstraite, stable et non-ambiguë, issue de l'apprentissage de la lecture, avec les phonèmes contenant des informations acoustiques plus variables, moins délimitées et dépendantes du

locuteur. Il semble ainsi plus aisé de s'appuyer sur l'orthographe, pour apprendre des nouveaux mots, les intégrer dans le lexique mental et les stocker dans la mémoire à long terme. Dans leur étude, Escudero et al (2008) ont montré que les formes orthographiques facilitaient la mémorisation et la lexicalisation des mots de L2 comprenant des paires minimales difficilement distinguables en aidant les apprenants à établir des représentations lexicales séparées pour chaque mot de la paire minimale. Également, de nombreuses recherches (Darcy et al. 2012 ; Escudero et al. 2008 ; Weber et Cutler 2004) ont montré que les apprenants de L2 pouvaient établir des contrastes lexicaux s'ils disposaient suffisamment d'informations sensorielles de cette langue cible, même s'ils montraient encore des difficultés à discriminer les contrastes sonores. De plus, Perfetti et Hart (2002) ont montré que pour apprendre un nouveau mot, le participant devait acquérir et intégrer des informations sur l'orthographe, la phonologie et la sémantique des mots. Ils estimaient que la représentation d'un nouveau-mot était de bonne qualité et opérationnelle lorsque l'on pouvait récupérer les trois informations (orthographique, sémantique et phonologique). Ces éléments reflètent bien l'importance de l'information orthographique dans l'acquisition efficace de nouveaux-mots. En l'absence de cette information orthographique, l'apprenant construirait une représentation lexicale incomplète basée uniquement sur les informations phonologiques et sémantiques du nouveau-mot. Lors de la récupération de ce nouveau-mot, les apprenants ne pourraient donc que se baser sur ces deux informations, rendant la récupération moins optimale.

Dans notre étude, les participants ayant reçu un entraînement purement auditif ou accompagné d'une information articulatoire créeraient lors de la phase d'apprentissage une représentation lexicale incomplète qui ne leur permettrait pas de consolider leurs connaissances. L'entraînement purement auditif, se basant sur des représentations phonologiques potentiellement ambiguës pour nos participants, ne permettrait ni un bon maintien en mémoire, ni la consolidation des mots appris. L'exposition aux gestes articulatoires associés à des entrées auditives, lui, permettrait aux participants de construire des représentations phonologiques plus précises, comme l'ont montré les résultats de la tâche de discrimination AX. Ces représentations phonologiques de meilleure qualité ont également permis aux participants de maintenir les mots appris, même si l'absence d'informations orthographiques empêcherait une consolidation efficace de ces mots.

Notre étude avait pour objectif de comparer l'efficacité d'un entraînement purement auditif à l'efficacité de deux types d'entraînement audiovisuel afin d'étudier la contribution de l'indice orthographique et des gestes articulatoires à la capacité de perception du phonème /θ/, ainsi qu'à la capacité de reconnaissance et de mémorisation des nouveaux-mots contenant ce phonème chez des participants francophones européens. Nous avons ainsi pu constater que les deux types d'information visuelle (orthographique et articulatoire) apportent leur contribution à des niveaux différents du traitement de la parole. Ce qui est d'autant plus impressionnant, c'est que ces contributions persistent même lorsque ces informations visuelles ne sont plus physiquement présentes, comme lors de la phase de post-apprentissage. L'information articulatoire faciliterait la

perception d'un son de L2, ce qui montre son influence sur le traitement de bas niveau de la parole. De la même façon que ce qui a été démontré pour les sons de la langue maternelle (Patterson & Werker, 2003; Carlyon, Cusack, Foxon & Robertson, 2001), l'articulation permettrait la création d'une cartographie des repères articulatoires pour chaque son, et apporterait des indices sur le contenu de la parole, ainsi que des indices temporels, soutenant la capacité de perception d'un son n'appartenant pas à la langue maternelle. Ainsi, la présence de l'articulation peut faciliter la perception d'un nouveau son de L2, et donc l'apprentissage de la langue cible. De son côté, l'information orthographique faciliterait la reconnaissance et la consolidation des mots de L2 nouvellement acquis, et influence donc le traitement de haut niveau de la parole. De la même façon que ce qui a été démontré pour l'apprentissage des mots de la langue maternelle (Ricketts, Bishop & Nation, 2009), l'orthographe faciliterait l'apprentissage ainsi que la mémorisation des formes phonologiques de mots ou de non-mots, et donc la reconnaissance de mots parlés n'appartenant pas à la langue maternelle de l'apprenant. Nous pourrions donc suggérer qu'un entraînement ciblé sur le phonème cible de L2 accompagné de l'information orthographique pourrait permettre d'intégrer à long terme ce nouveau phonème dans l'inventaire de sons de l'apprenant.

3. Réflexions sur l'expérimentation

a. Limitations et biais

Les résultats obtenus dans ce mémoire sont néanmoins à nuancer, car ils comportent des biais méthodologiques notamment dus à la population étudiée.

Il aurait été préférable de recueillir davantage de données de participants initialement peu sensibles à la discrimination /f/-/θ/. Nos analyses sur la population peu sensible se basent sur un faible échantillon : 33 participants peu sensibles répartis entre les trois types d'entraînement (10 en auditif seul, 12 en audio-orthographique et 11 en audio-articulatoire). Nous supposons que ces participants ont été mis en difficulté lors de l'expérimentation précédente (M-S. Villain Bailly et E. Clot, 2019), et qu'ils ont été réticents à l'idée de participer à nouveau à une expérimentation sur la même thématique.

Notre population de participants n'était pas très variée, on retrouve notamment 26 étudiants en médecine et 40 étudiants en orthophonie. Ces derniers peuvent présenter une meilleure sensibilité à la perception phonémique du fait d'une exposition à la linguistique durant leur cursus universitaire. En outre, le niveau d'études de tous les participants était globalement élevé (données relevées dans le mémoire de M.S. Villain-Bailly & E.Clot, 2019) : 7 niveau bac, 15 bac +1, 26 bac+2, 21 bac+3, 24 bac+4, 8 bac+5, ce qui participe à la faible diversité de notre population.

Nous constatons également une grande variabilité inter-participant en ce qui concerne les horaires des passations de l'expérience. Les participants ont tous réalisé les passations sur deux jours, mais avec un délai entre les deux sessions compris entre 12 et 30 heures, rendant le temps d'intégration des apprentissages très variable. De plus, nous avons observé un degré de fatigabilité du participant différent en fonction de la plage horaire choisie par ce dernier.

Lors des passations, nous avons remarqué que le niveau d'attention et de fatigue semblait également varier selon le type d'apprentissage. Les participants ayant reçu un entraînement purement auditif ont fait preuve d'une attention soutenue et constante durant la phase d'apprentissage très coûteuse cognitivement. Leur positionnement maintenu près de l'écran durant toute la durée de l'épreuve et leur fatigue visible après l'expérience étaient assez révélateurs de l'investissement du participant à cette tâche. En revanche, les participants ayant reçu un entraînement auditif avec information articulatoire semblaient moins fatigués à la fin de l'expérience, même s'ils maintenaient une attention constante tout au long de la phase d'apprentissage. L'entraînement semblait donc moins coûteux cognitivement que l'entraînement purement auditif. De leur côté, les participants ayant reçu un entraînement auditif avec information orthographique trouvaient la tâche simple et semblaient avoir beaucoup plus de facilité à apprendre les nouveaux-mots que ceux ayant reçu un entraînement auditif ou audio-articulatoire. Ces participants ne jugeaient pas utile de disposer de 15 minutes pour réaliser cet entraînement. Un entraînement comprenant des informations visuelles complémentaires au signal acoustique permettraient donc de diminuer la charge mentale. En résumé, l'observation des participants pendant l'expérience a montré une mise en jeu de ressources cognitives et attentionnelles différente en fonction du type d'entraînement reçu par le participant. Cela renforce l'idée que la présence des informations visuelles facilite l'apprentissage.

b. Implications de l'étude et perspectives

Les résultats rapportés ne sont qu'une partie de ceux récoltés au sein d'une étude plus large. Des épreuves de production ont également été réalisées par les participants, afin d'évaluer l'évolution de leur capacité de production des fricatives dentales anglaises (/f/ et /θ/) après différents types d'entraînement. Il serait intéressant d'analyser ces données collectées.

Notre étude nous a permis d'apporter un autre regard sur nos pratiques orthophoniques, notamment sur l'utilisation de la multi-sensorialité dans nos rééducations. Pour la rééducation des troubles phonologiques, il paraît important de coupler autant que possible les différents canaux sensoriels, en s'appuyant sur le canal visuel comprenant le langage écrit et les gestes articulatoires en plus du canal oral. Afin de favoriser la perception, la reconnaissance et la mémorisation d'un phonème, il semble facilitateur de produire le phonème cible en mettant en valeur les gestes articulatoires et en y associant l'orthographe du phonème cible. Pour rendre la rééducation plus spécifique, il paraît important de cibler l'aide apportée au patient en fonction du plan de soin. Pour une rééducation de la perception, l'appui sur les gestes articulatoires semble recommandé, alors que pour une activité de plus haut niveau comme l'identification et la mémorisation de phonèmes, l'aide orthographique semble à privilégier. Ainsi, les conclusions de notre mémoire peuvent servir de base dans la mise en place de séances de rééducation des troubles phonologiques.

Plus personnellement, ce mémoire nous a permis de mieux comprendre le fonctionnement de la recherche expérimentale. Nous avons développé notre esprit critique grâce à la lecture des articles et à la rédaction de ce mémoire, qui nous permettra de pratiquer l'Evidence Based Practice avec plus d'efficacité et d'améliorer notre future pratique.

VI. REFERENCES

- Atteveldt, N. van, Formisano, E., Goebel, R., et Blomert, L. « Integration of Letters and Speech Sounds in the Human Brain ». *Neuron* 43, n° 2 (juillet 2004): 271-82. <https://doi.org/10.1016/j.neuron.2004.06.025>.
- Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. (pp. 171–206.). York Press, Baltimore.
- Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In M. J. Munro & O.-S. Bohn (Eds.), *Second language speech learning: The role of language experience in speech perception and production* (pp. 13–34). Amsterdam: John Benjamins.
- Bristow, D., Dehaene-Lambertz, G., Mattout, J., Soares, C., Gliga, T., Baillet, S., et Mangin, J.F. « Hearing Faces : How the Infant Brain Matches the Face It Sees with the Speech It Hears ». *Journal of Cognitive Neuroscience* 21, n° 5 (mai 2009) : 905-21. <https://doi.org/10.1162/jocn.2009.21076>.
- Buchan, J. N., Paré, M., & Munhall, K. G. « Spatial statistics of gaze fixations during dynamic face processing ». *Social Neuroscience* 2, no 1 (1 mars 2007): 1-13. <https://doi.org/10.1080/17470910601043644>.
- Carlyon R.P., Cusack R., Foxton J.M., Robertson IH. « Effects of Attention on Neuroelectric Correlates of Auditory Stream Segregation ». *Journal of Cognitive Neuroscience* 18, n° 1 (1 janvier 2006): 1-13. <https://doi.org/10.1162/089892906775250021>.
- Clot, E., Villain-Bailly, M., Pattamadilok, C., Tyler, M., Aix-Marseille Université Faculté de médecine, & Aix-Marseille Université. *Perception de phonèmes étrangers par des auditeurs francophones adultes : Le cas des fricatives dentales anglaises*
- Darcy, I., Dekydspotter, L., Sprouse, R. A., Glover, K. C., Kaden, C., McGuire, M., et al. (2012). Direct mapping of acoustics to phonology: On the lexical encoding of front rounded vowels in L1 English–L2 French acquisition. *Second Language Research*, 28, 5–40
- Dehane-Lamberts G, Dehaene S. 1994. Speed and cerebral correlates of syllable discrimination in infants. *Nature* 30(6487) : 292-95
- Dorn, K., Weinert, S., et Falck-Ytter, T. « Watch and Listen - A Cross-Cultural Study of Audio-Visual-Matching Behavior in 4.5-Month-Old Infants in German and Swedish Talking Faces ». *Infant Behavior & Development* 52 (2018): 121-29. <https://doi.org/10.1016/j.infbeh.2018.05.003>.
- Erber, N. P. « Auditory-Visual Perception of Speech ». *The Journal of Speech and Hearing Disorders* 40, n° 4 (novembre 1975) : 481-92. <https://doi.org/10.1044/jshd.4004.481>.
- Erdener, Doğu, et Denis Burnham. « The Relationship between Auditory–Visual Speech Perception and Language-Specific Speech Perception at the Onset of Reading Instruction in English-Speaking Children ». *Journal of Experimental Child Psychology* 116, n° 2 (octobre 2013): 120-38. <https://doi.org/10.1016/j.jecp.2013.03.003>.
- Escudero, Paola, Rachel Hayes-Harb, et Holger Mitterer. « Novel Second-Language Words and Asymmetric Lexical Access ». *Journal of Phonetics* 36, n° 2 (avril 2008): 345-60. <https://doi.org/10.1016/j.wocn.2007.11.002>.
- Fenwick, S.E., Best, C.T., Davis, C., Tyler, M.D. « The Influence of Auditory-Visual Speech and Clear Speech on Cross-Language Perceptual Assimilation ». *Speech Communication* 92 (septembre 2017): 114-24. <https://doi.org/10.1016/j.specom.2017.06.001>.

- Frost, R., Repp, B.H, Katz, L. « Can Speech Perception Be Influenced by Simultaneous Presentation of Print? ». *Journal of Memory and Language* 27, n° 6 (décembre 1988) : 741-55. [https://doi.org/10.1016/0749-596X\(88\)90018-6](https://doi.org/10.1016/0749-596X(88)90018-6).
- Gaskell, M. G., Dumay, N. « Lexical Competition and the Acquisition of Novel Words ». *Cognition* 89, n° 2 (1 septembre 2003): 105-32. [https://doi.org/10.1016/S0010-0277\(03\)00070-2](https://doi.org/10.1016/S0010-0277(03)00070-2).
- Gervain, J., Mehler, J. « Speech Perception and Language Acquisition in the First Year of Life ». *Annual Review of Psychology* 61, n° 1 (janvier 2010): 191-218. <https://doi.org/10.1146/annurev.psych.093008.100408>.
- Grainger, J., Ziegler, J. (2007). Cross-code consistency in a functional architecture for word recognition. Single word reading: Behavioral and biological perspectives. 129-157.
- Grant, K., Seitz, P. Grant, Ken, W., Seitz, P.F.. « The Recognition of Isolated Words and Words in Sentences: Individual Variability in the Use of Sentence Context », 2000, 12.
- Guion, S.G., Fledge, J.E., Akahane-Yamada, R., & Pruitt, J.C. (2000). An investigation of current models of second language speech perception : The case of Japanese adults' perception of English consonants. *The Journal of the Acoustical Society of America*, 107(5), 2711-2724. <https://doi.org/10.1121/1.428657>
- Han, Jeong-Im, et Sujin Oh. « The Role of Phonetic Similarity and Orthographic Information in Asymmetrical Lexical Encoding in Second Language ». *Journal of Psycholinguistic Research* 47, n° 5 (octobre 2018): 1015-33. <https://doi.org/10.1007/s10936-018-9574-7>.
- Keetels, M., Bonte, M., Vroomen, J. « A Selective Deficit in Phonetic Recalibration by Text in Developmental Dyslexia ». *Frontiers in Psychology* 9 (15 mai 2018): 710. <https://doi.org/10.3389/fpsyg.2018.00710>.
- Kuhl, P. K., et A. N. Meltzoff. « The Bimodal Perception of Speech in Infancy ». *Science* (New York, N.Y.) 218, no 4577 (10 décembre 1982): 1138-41. <https://doi.org/10.1126/science.7146899>.
- Kuhl, P.K., Williams, K.A., Lacerda, F., KN Stevens, et B Lindblom. « Linguistic experience alters phonetic perception in infants by 6 months of age ». *Science* 255, no 5044 (31 janvier 1992) : 606. <https://doi.org/10.1126/science.1736364>.
- Lansing, C.R., & McConkie, G.W. (2003). Word identification and eye fixation locations in visual and visual-plus-auditory presentations of spoken sentences. *Perception & Psychophysics*, 65, 536–552. <https://link.springer.com.lama.univ-amu.fr/article/10.3758/BF03194581>.
- Liberman, Alvin M. « Chapter 9 The Relation of Speech to Reading and Writing ». In *Advances in Psychology*, édité par Ram Frost et Leonard Katz, 94:167-78. Orthography, Phonology, Morphology, and Meaning. North-Holland, 1992. [https://doi.org/10.1016/S0166-4115\(08\)62794-6](https://doi.org/10.1016/S0166-4115(08)62794-6).
- Luce P.A., Pisoni D.B. « Recognizing spoken words: The neighborhood activation model ». *Ear and Hearing* 19, n°1 (février 1998); 1–36.
- Marslen-Wilson, W., Tyler, L.K.. « The Temporal Structure of Spoken Language Understanding ». *Cognition* 8, n° 1 (janvier 1980): 1-71. [https://doi.org/10.1016/0010-0277\(80\)90015-3](https://doi.org/10.1016/0010-0277(80)90015-3).
- McGurk, H., MacDonald, J. Hearing lips and seeing voices. *Nature* 264, 746–748 (1976). <https://doi.org/10.1038/264746a0>
- McQueen, J. M., Norris, D., & Cutler, A. (1994). « Competition in Spoken Word Recognition: Spotting Words in Other Words ». *Journal of Experimental Psychology: Learning, Memory, and Cognition* 20 (1994): 621-38.
- Patterson, M. L., & Werker, J. F. « Two-Month-Old Infants Match Phonetic Information in Lips and Voice ». *Developmental Science* 6, n° 2 (2003): 191-96. <https://doi.org/10.1111/1467-7687.00271>.

- Peelle JE, Sommers M (2015) Prediction and constraint in audiovisual speech perception. *Cortex* 68:169–181. doi:10.1016/j.cortex.2015.03.006
- Ying, L., Perfetti, C.A., Hart, L. « ERP Evidence for the Time Course of Graphic, Phonological, and Semantic Information in Chinese Meaning and Pronunciation Decisions ». *Journal of Experimental Psychology: Learning, Memory, and Cognition* 29, no 6 (novembre 2003): 1231-47. <https://doi.org/10.1037/0278-7393.29.6.1231>.
- Ricketts, J., Bishop D.V.M, Nation, K. « Orthographic Facilitation in Oral Vocabulary Acquisition ». *Quarterly Journal of Experimental Psychology* 62, no 10 (octobre 2009): 1948-66. <https://doi.org/10.1080/17470210802696104>.
- Rosemann S, Thiel CM (2018) Audio-visual speech processing in age-related hearing loss: stronger integration and increased frontal lobe recruitment. *NeuroImage* 175:425–437. S1053-8119 (18)30319-7. <https://www.ncbi.nlm.nih.gov/pubmed/29655940>.
- Santiago, F. La didactique de la prononciation de L2 et la perception auditive : vers une nouvelle approche. Synergies Mexique, GERFLINT, 2012, pp.57 - 70. hal-01737832
- Sekiyama, K., Burnham, D.. « Impact of Language on Development of Auditory-Visual Speech Perception ». *Developmental Science* 11, no 2 (2008): 306-20. <https://doi.org/10.1111/j.1467-7687.2008.00677.x>.
- Snodgrass J.G., Levy-Berger, G., & Haydon,M.(1985). Human experimental psychology. NewYork : Oxford University Press
- Takashima, A., Bakker I, van Hell, J.G., Janzen G, McQueen, J.M. « Richness of Information about Novel Words Influences How Episodic and Semantic Memory Networks Interact during Lexicalization ». *NeuroImage* 84 (2014): 265-78. <https://doi.org/10.1016/j.neuroimage.2013.08.023>.
- Weber, A., Cutler, A. « Lexical competition in non-native spoken-word recognition ». *Journal of Memory and Language* 50 (31 janvier 2004). [https://doi.org/10.1016/S0749-596X\(03\)00105-0](https://doi.org/10.1016/S0749-596X(03)00105-0).
- Werker, Janet F., et Richard C. Tees. « Cross-language speech perception: Evidence for perceptual reorganization during the first year of life ». *Infant Behavior and Development* 7, n° 1 (1 janvier 1984): 49-63. [https://doi.org/10.1016/S0163-6383\(84\)80022-3](https://doi.org/10.1016/S0163-6383(84)80022-3).
- Yehia, Hani, Philip Rubin, et Eric Vatikiotis-Bateson. « Quantitative Association of Vocal-Tract and Facial Behavior ». *Speech Communication* 26, n° 1 (1 octobre 1998): 23-43. [https://doi.org/10.1016/S0167-6393\(98\)00048-X](https://doi.org/10.1016/S0167-6393(98)00048-X).

RESUME

Les Hommes ont des difficultés à percevoir les phonèmes n'appartenant pas à leur langue maternelle. 75% des francophones européens perçoivent la consonne fricative dentale anglaise /θ/ comme la fricative dentale française /f/. Notre étude évalue l'efficacité de trois types d'entraînements (auditif, audio-orthographique, audio-articulatoire) dans l'amélioration de la capacité de perception du phonème /θ/ ainsi que dans la capacité de reconnaissance et de mémorisation des nouveaux mots contenant ce phonème, chez des participants francophones européens. Une tâche de discrimination et une tâche d'association mot-image ont été proposées à nos participants afin d'évaluer leur capacité de perception, de reconnaissance et de mémorisation du phonème cible /θ/. La capacité de perception des participants a été améliorée grâce aux trois types d'entraînements, même si l'entraînement audio-visuel paraît le plus bénéfique des trois. De son côté, la capacité de reconnaissance et de mémorisation des mots contenant le phonème /θ/ a augmenté grâce aux trois types d'entraînement, en sachant que seul l'entraînement orthographique a permis la consolidation de ces nouveaux-mots.

Mots clés : Perception de la parole - Perception phonémique - Discrimination phonémique - Reconnaissance phonémique - Mémorisation - Langue étrangère - Fricatives dentales - Entraînement auditif - Entraînement visuel