Apport des informations visuelles dans la perception et la production du phonème non natif /v/ chez les apprenants thaïlandais

Supansa Tusnyingyong

▶ **To cite this version:**

Supansa Tusnyingyong. Apport des informations visuelles dans la perception et la production du phonème non natif /v/ chez les apprenants thaïlandais. Sciences de l'Homme et Société. 2022. dumas-03794537

HAL Id: dumas-03794537

https://dumas.ccsd.cnrs.fr/dumas-03794537

Submitted on 3 Oct 2022

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

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

Apport des informations visuelles dans la perception et la production du phonème non natif /v/ chez les apprenants thaïlandais

Supansa
TUSNYINGYONG

Sous la direction de Thi Thuy Hien TRAN

Laboratoire : GIPSA-lab

UFR LLASIC
Département Sciences du langage et Didactique des Langues

Mémoire de master 2 mention sciences du langage - 30 crédits

Parcours : Linguistique

Année universitaire 2021-2022

Apport des informations visuelles dans la perception et la production du phonème non natif /v/ chez les apprenants thaïlandais

Supansa
TUSNYINGYONG

Sous la direction de Thi Thuy Hien TRAN

Laboratoire : GIPSA-lab

UFR LLASIC
Département Sciences du langage et Didactique des Langues

Mémoire de master 2 mention sciences du langage - 30 crédits

Parcours : Linguistique

Année universitaire 2021-2022

Remerciements

Qu'il me soit permis d'exprimer ici mes remerciements les plus sincères à ma directrice du mémoire, Madame Thi Thuy Hien TRAN, pour sa sympathie, ses conseils et son accompagnement tout au long de ce mémoire. Sans elle, ce mémoire n'aurait pas pu voir le jour.

Mes sincères remerciements vont également à toutes les personnes qui ont participé à ce projet. Un grand merci à Monsieur Coriandre VILAIN pour l'enregistrement des sons et des vidéos et pour toutes les étapes techniques depuis la constitution du corpus, la création de l'interface de test avec *Presentation* jusqu'à l'extraction des données. Sans lui, ce projet n'aurait pas été réussi. Je tiens également à le remercier pour avoir accepté d'évaluer ce travail.

Ma reconnaissance va également à Madame Anne VILAIN et Monsieur Sylvain GERBER pour leur aide au niveau d'analyses statistiques sous *R*. Merci à eux pour le temps qu'ils ont consacrés à la formation ainsi qu'à la constitution des scripts d'analyse. De plus, je remercie vivement Madame Anne VILAIN pour toutes les aides techniques et matérielles.

J'adresse aussi un grand merci à Madame Laura ABOU HAIDAR pour ses conseils et ses commentaires précieux. Merci également à elle d'avoir accepté de faire partie de mon jury de mémoire.

Je tiens à remercier les quatre participants (Chanyuth THAMWAN, Ku Ammar YAMIRUDENG, Smith PATARAPRASITPON et Wichayaporn FAIKHAW) pour leur disponibilité, leur participation et leur respect des consignes de l'expérience. Sans eux, je n'aurais pu établir ce mémoire avec succès.

Je remercie vivement Danielle HERITIER qui n'est jamais fatiguée de corriger ma rédaction. Merci à elle pour le temps qu'elle a consacré à la lecture minutieuse de ce mémoire. J'ai énormément apprécié son aide.

Je voudrais également exprimer mes remerciements à mes amies au laboratoire : Morgane ROSSIGNOL, Océane BELLENGER et Paula CÓRDOBA CANO pour leur aide au niveau d'analyse des données ainsi que leur amitié pendant mon stage au laboratoire.

Enfin, j'adresse ma gratitude à ma famille qui a toujours cru en moi. Merci de m'avoir soutenue pour tout ce qui me passionne.

DÉCLARATION ANTI-PLAGIAT

1. Ce travail est le fruit d'un travail personnel et constitue un document original.
2. Je sais que prétendre être l'auteur d'un travail écrit par une autre personne est une pratique sévèrement sanctionnée par la loi.
3. Personne d'autre que moi n'a le droit de faire valoir ce travail, en totalité ou en partie, comme le sien.
4. Les propos repris mot à mot à d'autres auteurs figurent entre guillemets (citations).
5. Les écrits sur lesquels je m'appuie dans ce mémoire sont systématiquement référencés selon un système de renvoi bibliographique clair et précis.

PRENOM : Supansa

NOM : TUSNYINGYONG

DATE : 14/07/2022

Sommaire

Introduction **9**

Chapitre 1 La perception auditive **12**

1.1 Perception auditive de la parole : de la langue native aux phonèmes non natifs 13

1.1.1 Perception de la parole chez l'enfant pendant la première année de vie 13

1.1.2 Perception de la parole en apprentissage d'une langue étrangère 15

1.2 Perception auditive en langue étrangère : modèles d'assimilation phonologique 17

1.2.1 Modèle d'assimilation perceptive (PAM) 17

1.2.2 Modèle d'apprentissage des langues (SLM) 19

Chapitre 2 La perception visuelle **22**

2.1 Perception visuelle : la reconnaissance du visage 23

2.2 Perception visuelle dans le traitement du langage : lecture dite labiale 24

2.3 Visèmes 25

Chapitre 3 La perception audiovisuelle **28**

3.1 Perception audiovisuelle dans le traitement de la langue native 29

3.2 Perception audiovisuelle dans le traitement d'une langue étrangère 31

3.2.1 Chez les bébés 31

3.2.2 Chez les adultes 32

3.2.3 Facteurs impactant l'utilisation des informations visuelles lors du traitement d'une langue étrangère 36

3.2.3.1 Saillance du visème *36*

3.2.3.2 Organisation de l'espace phonologique de la L1 ..... *37*

3.2.3.3 Relation entre les visèmes de la L1 et ceux de la LE *37*

3.2.3.4 Performance dans la LE *38*

Chapitre 4 Confusion entre /v/ et /w/ chez les apprenants thaïlandais **40**

4.1 Le thaï 41

4.1.1 Le système consonantique 41

4.1.2 Le système d'écriture du thaï 42

4.1.3 Les visèmes du thaï 43

4.2 La confusion /v/ ~ /w/ chez les apprenants thaïlandais 44

4.2.1 A l'origine de la confusion /v/ ~ /w/ chez les Thaïlandais 45

4.2.2 L'impact direct de la translittération franco-thaïlandaise 46

4.2.3 L'impact indirect de la translittération anglo-thaïlandaise 47

Chapitre 5 Étude expérimentale sur l'apport des informations visuelles dans la perception et la production de la consonne /v/ chez les apprenants thaïlandais **49**

5.1 Objectifs, questions de recherche et hypothèses 50

5.2 Méthodologie 51

5.2.1 Participants 51
5.2.2 Constitution du corpus 53
5.2.2.1 Préparation des stimuli de l'expérience *54*
5.2.2.2 Corpus pour la tâche de perception *55*
5.2.2.2 Corpus pour la tâche de production *56*
5.2.3 Déroulement de l'expérience 57
5.2.3.1 Test de perception *57*
5.2.3.2 Test de production *60*
5.2.4 Traitement des données 61
5.2.4.1 Perception *61*
5.2.4.2 Production *61*
5.2.4.3 Analyses statistiques *63*
5.3 Résultats 64
5.3.1 Tâche de perception 64
5.3.1.1 Taux de réussite *64*
5.3.1.2 Réponses incorrectes *67*
5.3.1.3 Temps de réponse *68*
5.3.2 Tâche de production 71
5.3.2.1 Taux de réussite *71*
5.3.2.2 Taux d'erreurs *74*
5.3.2.3 Stratégies de prononciation de /v/ *79*
5.3.3 Relation entre la perception et la production 82
5.4 Discussion 86
Conclusion et perspectives 92
Bibliographie 94
Abréviations utilisées 104
Liste des figures 105
Liste des tables 107
Tables des annexes 108
Tables des matières 122

Introduction

L'individu s'habitue dès l'enfance à ce qui lui est familier. Au cours de la première année de vie, l'enfant se spécialise aux sons dans sa langue première (L1) et devient progressivement « insensible » aux sons qui ne sont pas pertinents dans sa L1 (Trehub & Robinovitch, 1972 ; Trehub,1976 ; Werker et al., 1981 ; Werker & Tees, 1984 ; Best & McRoberts, 2003 ; Kuhl et al., 2006). Ce phénomène qui se produit dès l'âge de 10 mois environ est maintenu à l'âge adulte et impacte l'apprentissage d'une langue étrangère (LE). Lors de l'exposition aux sons non pertinents dans la L1, le système phonologique de la L1 fonctionne comme un « crible » qui laisse passer uniquement les éléments sonores d'une LE présents tels quels dans la L1 (Troubetzkoy, 1939). Les apprenants d'une LE rencontrent donc des difficultés à percevoir et à prononcer des sons non pertinents dans leur L1. Par exemple, les apprenants francophones de l'anglais rencontrent des difficultés à percevoir le phonème interdental de l'anglais /θ/ et ont tendance à l'assimiler à un des phonèmes plus proches au niveau articulatoire qui existe dans leur L1 (Brannen, 2002) (ex. /θ/ dans *thanks* /θæŋks/ prononcé [s]).

De même, il a été démontré que les apprenants thaïlandais du français rencontrent des difficultés à percevoir et à prononcer les fricatives sonores /z v ʒ/ qui n'existent pas en thaï (Ngammana, 2011 ; Le Corre, 2013 ; Promkesa, 2014). Le dévoisement des consonnes fricatives sonores est la stratégie largement utilisée par les apprenants thaïlandais ; pour prononcer [z] et [ʒ], les apprenants les remplacent par [s] et [ʃ] respectivement. Toutefois, pour prononcer [v] en position initiale de syllabe, cette consonne n'est pas remplacée par son homologue sourde [f] comme c'était le cas pour les fricatives [z] et [ʒ]. Les apprenants thaïlandais remplacent la labiodentale [v] par la labiovélaire [w] existant en thaï. Le remplacement de [v] par [w] semble provenir de l'association erronée entre le graphème latin < v > /v/ et le graphème < ว > /w/ du thaï à travers la romanisation du thaï et la translittération des mots d'emprunt en caractères thaïs (Tusnyingyong & Tran, 2022).

De nombreuse études montrent que la présentation audiovisuelle peut améliorer le traitement des sons non natifs (Hazan et al., 2006 ; Wang et al., 2008 ; Wang et al., 2009 ; Burfin, 2015). La performance dans le traitement de la parole en LE est plus élevée lorsque les apprenants sont exposés aux informations audiovisuelles de la parole que lorsqu'ils sont confrontés uniquement aux informations auditives. Cependant, les apprenants de différentes L1s ne

bénéficient pas des informations audiovisuelles de la parole de la même manière et les informations audiovisuelles de chaque phonème ne sont pas non plus exploitées de la même manière par les apprenants d'une LE.

Dans le cadre de cette présente étude, nous nous intéressons aux difficultés dans le traitement de la consonne /v/ en position initiale de syllabe chez les apprenants thaïlandais. Nous avons donc effectué une étude expérimentale visant à vérifier si la présentation audiovisuelle pourrait améliorer la perception et la production de la consonne /v/ en position initiale de syllabe chez les apprenants thaïlandais.

Ce mémoire se divise en cinq chapitres.

Le premier chapitre est consacré à *la perception auditive*. Dans ce chapitre, sont exposés la perception auditive du langage humain chez les enfants et les adultes ainsi que les modèles de la perception auditive en acquisition d'une LE. Deux modèles d'assimilation phonologique qui sont les plus influents dans le domaine de la perception auditive en LE sont présentés dans cette partie.

Le deuxième chapitre *la perception visuelle* vise à comparer la trajectoire développementale entre la perception auditive (cf. chapitre 1) et la perception visuelle de l'humain. Pour comparer les deux perceptions, nous présenterons dans ce chapitre la trajectoire développementale de la perception visuelle des visages ainsi que celle de la perception visuelle dans le traitement du langage humain. La notion « visème » est également abordé dans ce chapitre.

Après avoir abordé séparément la perception auditive (cf. chapitre 1) et la perception visuelle (cf. chapitre 2), nous présenterons *la perception audiovisuelle* dans le troisième chapitre. Ce chapitre vise à montrer l'usage différent des informations audiovisuelles dans le traitement de la parole en L1 et en LE.

Le quatrième chapitre aborde la problématique récurrente de cette présente étude : *la confusion entre /v/ et /w/ chez les apprenants thaïlandais*. Nous présenterons dans ce chapitre la littérature des études antérieures sur l'acquisition de la consonne /v/ chez les apprenants thaïlandais ainsi que la source et l'impact de la confusion entre /v/ et /w/ en position initiale de syllabe chez les apprenants thaïlandais en montrant au début de ce chapitre le système de la langue thaïe.

Le dernier chapitre est consacré à l'étude expérimentale visant à vérifier l'apport des informations audiovisuelles dans la perception et dans la production de /v/ chez les apprenants thaïlandais. Les résultats sont discutés à la fin de ce chapitre.

Chapitre 1

La perception auditive

1.1 Perception auditive de la parole : de la langue native aux phonèmes non natifs

1.1.1 Perception de la parole chez l'enfant pendant la première année de vie

Les premiers travaux sur les capacités de perception des nouveau-nés suggèrent que l'homme dispose de prédispositions pour percevoir les traits phonétiques (pour une revue, voir par exemple de De Boyssons-Bardies, 1996). Il a été démontré que les nourrissons peuvent discriminer les contrastes phonétiques de manière catégorielle dès l'âge d'un mois (Eimas, Siqueland, Jusczyk & Vogorito, 1971), y compris ceux qui ne sont pas utilisés dans leur environnement linguistique ambiant (Trehub & Robinovitch, 1972 ; Trehub, 1976 ; Werker et al., 1981 ; Werker & Tees, 1984 ; Best & McRoberts, 2003 ; Kuhl et al., 2006). Les nourrissons peuvent traiter de façon identique des segments phonétiques physiquement différents mais appartenant à la même catégorie. Ainsi, un nouveau-né distinguera la syllabe /ba/ de la syllabe /pa/ selon la durée du délai d'établissement du voisement (VOT) mais non deux exemplaires de /ba/ ou de /pa/ à écarts de VOT constants (Eimas et al., 1971).

Dans une autre étude, Trehub (1976) teste la capacité de discrimination des contrastes vocalique /pa/ ~ /pɑ̃/ du français et consonantique /za/ ~ /ʐa/ du tchèque chez les nouveau-nés anglophones canadiens âgés de 1 à 4 mois. Les phonèmes /p/, /a/ et /z/ existent en anglais, mais ce n'est pas le cas de /ɑ̃/ et de /ʐ/. La perception des deux contrastes des deux groupes de sujets (un expérimental et un de contrôle) est mesurée par le taux de succion. Pour le groupe expérimental, un stimulus de chaque contraste (par exemple, /pa/) est d'abord présenté jusqu'à ce que le taux de succion diminue et atteigne le seuil d'habituation. Ensuite, durant la phase de test, ils sont exposés à un autre stimulus (par exemple /pɑ̃/). Quant au groupe de contrôle, aucun changement des sons n'est effectué durant la phase de test. Les résultats montrent que le taux de succion du groupe expérimental augmente significativement entre les deux phases (habituation *vs* test) indiquant ainsi que les nouveau-nés réagissent à la nouveauté et sont donc capables de percevoir une différence entre les deux sons. Chez les enfants du groupe de contrôle, l'absence de reprise du taux de succion indique que la différence entre les deux phases (habituation *vs* test) n'a pas été perçue puisqu'un seul stimulus est présenté durant l'expérience. Ces nourrissons de 1-4 mois sont donc capables d'établir une relation entre l'apparition d'un nouveau son avec leur succion, et de discriminer les deux sons même si ceux-ci n'existent pas dans leur langue.

Les capacités initiales à percevoir l'ensemble des distinctions phonétiques déclinent cependant, au cours de la première année de vie, en l'absence d'expérience linguistique appropriée (Werker & Tees, 1984 ; Kuhl et al., 2006). Werker et Tees (1984) testent la perception des contrastes non natifs (en hindi et en salish) auprès de trois groupes d'enfants anglophones (6-8 mois, 8-10 mois et 10-12 mois). Les résultats montrent que les enfants de 6 à 8 mois sont capables de discriminer parfaitement ces contrastes inexistants dans leur langue. Cette capacité diminue progressivement à l'âge de 8-10 mois et disparait complètement à l'âge de 10-12 mois. Les enfants hindis et salishs de 11-12 mois (groupe contrôle) perçoivent parfaitement bien la différence entre les sons de leur langue. L'étude de Kuhl et collègues (2006) sur la perception du contraste consonantique de l'anglais /ra/ ~ /la/ chez les enfants anglophones et japonophones âgés de 6 à 12 mois confirme ces résultats en montrant que les enfants japonophones de 6-8 mois sont aussi performants que les anglophones dans la discrimination entre /ra/ et /la/ alors que ce contraste n'appartient pas au répertoire de leur langue puisque les deux sons ne sont que des variantes allophoniques en japonais. Cependant, cette capacité de discrimination diminue chez les enfants japonophones de 10 à 12 mois, alors qu'elle augmente chez les enfants anglophones de 10 à 12 mois.

Ainsi, la première année de vie d'un nourrisson est caractérisée par une adaptation rapide des mécanismes perceptifs aux inputs sensoriels spécifiques qu'il rencontre dans sa vie quotidienne. L'enfant se spécialise dans son environnement et sa capacité dans la perception des sons natifs augmente au cours de cette première année de vie. Mais il devient progressivement « sourd » ou « insensible » à ce qui ne lui est pas familier à partir de 10 mois. Ce phénomène peut survenir aussi bien pour les voyelles (Polka et Werkers, 1994) que pour les consonnes (Werkers & Tees, 1984) et pour les propriétés prosodiques comme les contrastes de tons lexicaux (par exemple, Mattock & Burnham, 2006 ; Yeung et al., 2013 ; Götz et al., 2018) ou encore de l'accent lexical (Höhle et al., 2009). Ce processus, connu sous le nom de « rétrécissement perceptif » (*perceptual narrowing* en anglais), entraine une diminution de la capacité à discriminer des sons qui ne sont pas présents ou pertinents dans l'environnement du nourrisson.

Il est à noter qu'il existe une dimension qui échappe au phénomène de « rétrécissement perceptif ». Best et al. (1988) montrent que les enfants de 12-14 mois et les adultes anglophones sont tout à fait capables de discriminer un contraste non natif en zoulou, une langue bantoue parlée en Afrique du Sud (clic dental [kǀ] *vs* clic latéral [kǁ]). Les auteurs suggèrent que les clics

constituent une dimension secondaire qui n'est en concurrence avec aucune autre et qu'il n'est donc pas nécessaire à l'enfant d'opposer cette dimension aux dimensions pertinentes pour percevoir les sons de sa langue.

1.1.2 Perception de la parole en apprentissage d'une langue étrangère

L'homme s'habitue dès l'enfance à son système de L1. Cette habitude impacte de manière conséquente la perception d'une LE. Cette idée a été évoquée pour la première fois par Polivanov en 1931 : « En entendant un mot inconnu étranger […] nous tâchons d'y retrouver un complexe de nos représentations phonologiques, de le décomposer en des phonèmes propres à notre langue maternelle, et même en conformité de nos lois de groupement en phonèmes. » (Polivanov, 1931, p. 79-80, cité par Borrell, 1996, p.106).

En 1939, Troubetzkoy a repris l'idée de Polivanov et a conceptualisé le rôle de la L1 dans la perception d'une LE comme le « crible phonologique ».

> Chaque homme s'habitue dès l'enfance à analyser ainsi ce qui est dit et cette analyse se fait d'une façon tout à fait automatique et inconsciente. Mais en outre le système des cribles, qui rend cette analyse possible, est construit différemment dans chaque langue. L'homme s'approprie le système de sa langue maternelle. Mais s'il entend parler une autre langue, il emploie involontairement pour l'analyse de ce qu'il entend le « crible phonologique » de sa langue maternelle qui lui est familier. Et comme ce crible ne convient pas pour la langue étrangère entendue, il se produit de nombreuses erreurs et incompréhensions. (Troubetzkoy, 1939, traduit par Cantineau, 1949 p. 54).

Ce concept considère le système phonologique de la L1 comme un « crible » qui laisse passer uniquement les éléments sonores d'une LE présents tels quels dans la L1. Lors de la perception d'un son inconnu de la langue cible, ce crible est activé et les apprenants cherchent le son qui semble le plus proche dans le système de leur L1.

Ce concept a été largement étudié dès la deuxième moitié du XX^{e} siècle. De nombreuses études expérimentales montrent des difficultés des apprenants dans la perception des sons inexistants en L1 (Trehub, 1976 ; Werker et al, 1981 ; Werker, Frost & McGurk, 1992 ; Best, McRoberts & Goodell, 2001 ; Le Corre, 2013 ; Promkesa, 2014 ; Tyler et al., 2019 entre autres). Werker et collègues (1992) ont effectué une étude perceptive de la consonne interdentale /ð/ chez les locuteurs anglophones et francophones. Cette consonne existe dans le système phonologique de l'anglais, mais pas du français. Les locuteurs écoutent les stimuli de test /ða/ ainsi que ceux de contrôle /ba va da ʒa ga/, dont les consonnes initiales existent dans les deux langues, avant de les

répéter et ensuite d'écrire sur une feuille la consonne initiale des syllabes entendues. Les résultats montrent que, pour les stimuli de contrôle, le pourcentage des réponses correctes est élevé chez les deux groupes (91 % chez les anglophones, 100 % chez les francophones). Cependant, une différence de réponses correctes est attestée pour le stimulus de test /ða/ entre les anglophones (93 %) et les francophones (70 %). Les auteurs interprètent que les unités non natives sont plus difficiles à percevoir par les francophones que celles qui existent dans le système de départ.

Les résultats similaires ont été montrés dans l'étude de Promkesa (2014) auprès des apprenants thaïlandais. Il a été montré que les Thaïlandais rencontrent des difficultés dans la perception et la production des fricatives non natives du français. Promkesa a effectué une étude expérimentale sur la perception et la production du contraste de voisement des fricatives du français /f/ ~ /v/, /s/ ~ /z/ et /ʃ/ ~ /ʒ/ en position finale de syllabe chez les francophones (groupe de contrôle) et chez les apprenants thaïlandais (groupe d'étude). En perception, les sujets écoutent les stimuli et décident si deux séquences entendues sont identiques ou différentes. En production, les sujets doivent répéter les stimuli après l'écoute. Les résultats montrent que les erreurs en perception et en production sont significativement plus élevées chez les apprenants thaïlandais que chez les francophones. L'étude montre également que le contraste labiodental /f/ ~ /v/ pose plus de problème chez les apprenants que les contrastes alvéolaire /s/ ~ /z/ et post-alvéolaire /ʃ/ ~ /ʒ/.

Des difficultés résident non seulement dans la perception des sons non natifs, mais aussi dans celle des sons allophoniques en L1. En 1971, Goto a effectué une étude avec la tâche d'identification d'un contraste /r/ ~ /l/ chez les locuteurs anglophones et japonophones. Ce contraste est attesté en anglais, mais les deux sons sont des variantes libres en japonais (Labrune, 2006). Les stimuli sont les mots en anglais, par exemple « collect » *vs* « correct », « glass » *vs* « grass », etc. Les résultats montrent que les sujets japonophones sont moins performants que les anglophones en identification du contraste /r/ ~ /l/. Miyawaki et collègues (1975) ont également étudié ce contraste à travers la tâche de discrimination des syllabes /ra/ ~ /la/ chez les anglophones et les japonophones. Les résultats vont dans le même sens que l'étude de Goto (1971) ; le taux de réussite chez les japonophones est moins élevé (58 %) que celui des anglophones (presque 100 %). Ces deux études (Goto, 1971 ; Miyawaki et al., 1975) montrent que les japonophones rencontrent des difficultés dans la discrimination des sons allophoniques /r/ ~ /l/ dans leur L1.

1.2 Perception auditive en langue étrangère : modèles d'assimilation phonologique

Plusieurs modèles ont été établis pour modéliser la perception auditive de la parole. Dans le cadre de l'apprentissage d'une LE, les modèles d'assimilation phonologique bien connus sont le modèle d'Assimilation Perceptive (*Perceptual Assimilation Model*) (PAM) (Best et al., 1988 ; Best, 1995a ; Best, 1995b) et le modèle d'Apprentissage des Langues (*Speech Learning Model*) (SLM) de Flege (1995). Le PAM se focalise sur la perception d'un contraste des phonèmes non natifs par des individus inexpérimentés, alors que le SLM s'intéresse aux difficultés de production et de perception de phonèmes isolés par les apprenants avancés (Burfin, 2015).

1.2.1 Modèle d'assimilation perceptive (PAM)

Le modèle PAM, visant à expliquer des performances perceptives différentes des contrastes non natifs, prédit que l'assimilation perceptive des sons non natifs à des catégories natives est basée sur des gestes articulatoires (articulateurs, lieu d'articulation, degré de constriction). L'apprenant perçoit dans l'input non natif des similarités ainsi que des différences articulatoires entre des sons non natifs et des sons natifs. Ce modèle propose donc trois cas d'assimilation perceptive. Premièrement, si l'individu détecte des similarités, des sons non natifs seront assimilés à des catégories natives ; les sons sont « catégorisables ». Deuxièmement, lorsqu'aucune similarité n'est détectée, mais les sons non natifs sont encore perçus comme langage humain. Ces sons ne seront pas assimilés à des catégories spécifiques de la L1 (« non-catégorisable »), mais resteront dans l'espace phonétique du langage humain. Troisièmement, les sons non natifs perçus ne sont pas considérés comme langage humain. Par conséquent, ils seront placés en dehors de l'espace phonétique du langage humain.

Rappelons que le modèle PAM se focalise sur la perception d'un contraste des phonèmes non natifs. Six scénarios d'assimilation ont été établis quand un individu perçoit un contraste des phonèmes non natifs (cf. figure 1).

1. *Deux Catégories* : lorsque l'individu perçoit la différence articulatoire des phonèmes non natifs, chacun des deux phonèmes sera assimilé dans des catégories phonologiques natives différentes.

2. *Qualité Catégorie* : deux phonèmes non natifs sont assimilés à une même catégorie native, mais ils diffèrent en distance prototypique ; l'un est proche du prototype, l'autre est jugé déviant.
3. *Catégorie Unique* : c'est le cas dans lequel deux phonèmes non natifs perçus sont assimilés à une même catégorie native, mais ils sont considérés comme aussi acceptables ou déviants du prototype natif l'un que l'autre.
4. *Catégorisé/non-catégorisé* : lorsque l'un des deux phonèmes est assimilé à une catégorie native, l'autre ne l'est pas et reste en dehors de l'espace phonologique natif.
5. *Non-catégorisable* : un contraste non natif n'est pas assimilé à des catégories natives, mais il reste toujours dans l'espace phonétique du langage humain, mais en dehors de l'espace phonologique natif.
6. *Non-assimilable* : un contraste des phonèmes non natifs n'est pas perçu comme étant le langage humain. Il s'installe donc en dehors du domaine phonétique.

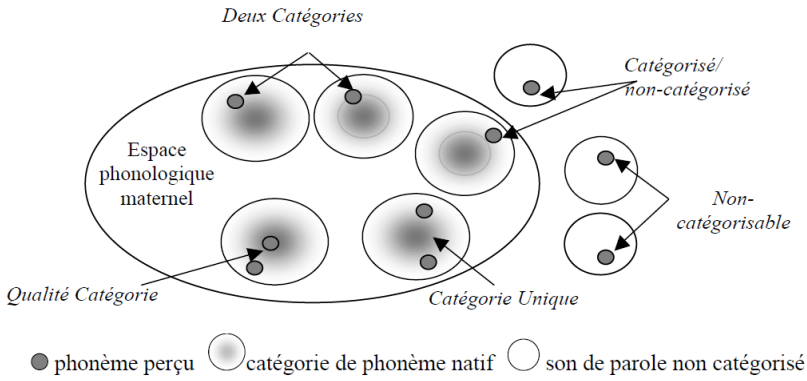

Figure 1. Représentation des scénarios d'assimilation du modèle PAM de Best (1995a) (figure empruntée à Burfin, 2015, p. 20)

Best & McRoberts (2003) ont complété le modèle PAM en rajoutant l'hypothèse des organes articulatoires (AO). L'hypothèse des AO postule que ce que les enfants détectent sur le langage humain est des organes articulatoires. Par conséquent, les enfants rencontreront plus de difficultés à discriminer un contraste impliquant les mêmes articulateurs (*within-organ*) qu'un contraste articulé par des organes articulatoires différents (*between-organ*). Avec l'hypothèse des AO, a été établi le modèle PAM/AO qui postule que la capacité de discrimination d'un contraste non natif qui partage les mêmes organes articulatoires décline plus tôt et plus rapidement que celle d'un contraste produit par les organes articulatoires différents.

Cette hypothèse a été vérifiée chez les enfants anglophones de 6-8 et de 10-12 mois à travers des contrastes qui partagent les mêmes articulateurs (par exemple, les fricatives latérales voisée et non voisée de la langue zoulou [ɬ] ~ [ɮ] et les fricatives alvéolaires de l'anglais [s] ~ [z]) et un contraste articulé par des organes différents (c'est-à-dire, les plosives éjectives bilabiale et alvéolaire du tigrigna, langue parlée en Éthiopie [p'] ~ [t']). Le temps de fixation du regard a été mesuré et comparé entre les phases d'habituation et de test. Les résultats montrent que le contraste du tigrigna est discriminé aussi bien pour les deux groupes d'enfants. En revanche, concernant les contrastes [ɬ] ~ [ɮ] et [s] ~ [z], les enfants de 10-12 mois les discriminent significativement moins bien que ceux de 6-8 mois. Les auteurs suggèrent en utilisant l'hypothèse de PAM/AO que les articulateurs auraient un effet sur la capacité de discrimination des enfants. Aucune différence significative n'est trouvée chez les deux groupes d'enfants pour le contraste *between-organ* ([p'] ~ [t']), alors que pour des contrastes *within-organ,* la capacité des enfants plus âgés (10-12 mois) diminue par rapport aux plus jeunes (6-8 mois). Ce déclin apparaît non seulement avec des contrastes non natifs ([ɬ] ~ [ɮ]), mais aussi avec des contrastes natifs [s] ~ [z].

Dans ce modèle PAM/AO, les organes articulatoires pourraient donc jouer un rôle important dans la discrimination des phones : les sons qui partagent les mêmes articulateurs seraient plus difficiles à discriminer que ceux produits avec les organes articulatoires différents.

1.2.2 Modèle d'apprentissage des langues (SLM)

Contrairement aux modèles PAM et PAM/AO qui se focalisent sur la perception d'un contraste non natif, le modèle SLM de Flege (1995) s'intéresse à l'acquisition d'une LE. Le principe de base de ce modèle est que de nombreuses erreurs de production d'une LE proviennent de la perception qui est affinée par l'expérience de la L1. Cependant, ces erreurs peuvent évoluer au cours de l'apprentissage d'une LE. Le modèle SLM propose les quatre postulats suivants (Flege, 1995, traduit par Tran, 2011) :

- P1 : Les processus et les mécanismes utilisés lors de l'apprentissage du système des sons de la L1, y compris la formation des catégories, restent intacts tout au long de la vie, et peuvent être appliqués pour apprendre une LE.
- P2 : Les aspects, spécifiques à chaque langue, des sons de parole sont spécifiés dans les représentations de la mémoire à long terme appelées catégories phonétiques.

- P3 : Les catégories phonétiques établies pour les sons de la L1 dans l'enfance évoluent toute la vie afin de refléter les propriétés de tous les phones de la L1 ou de la LE identifiés comme une réalisation de chaque catégorie.
- P4 : Les apprenants bilingues s'efforcent de maintenir le contraste entre les catégories phonétiques de la L1 et de la LE qui existent dans un espace phonologique commun.

A partir des quatre postulats, ont été établies sept hypothèses de l'acquisition d'une LE (H1-H7).

- H1 : La première hypothèse suppose que les phones de la L1 et de la LE seraient perceptivement liés au niveau allophonique. Mais cette relation allophonique varie selon leurs positions syllabiques. Autrement dit, les phones de la L1 et de la LE pourraient être placés dans une même catégorie perceptive et seraient considérés comme allophones. Un même phone avec une position syllabique différente pourrait être classé dans des catégories différentes.
- H2 : Ce modèle formule la deuxième hypothèse sur la capacité d'identification des différences phonétiques chez les apprenants d'une LE ; les apprenants pourraient établir une nouvelle catégorie pour un phone de la LE.
- H3 : La troisième hypothèse est basée sur le principe de similarité et de différence ; plus un phone de la LE est différent de celui de la L1, plus des différences phonétiques entre les phones sont perçues.
- H4 : Cette hypothèse s'intéresse à l'âge de l'apprentissage d'une LE ; la capacité d'identification des différences phonétiques, soit entre les phones de la L1 et ceux de la LE, soit entre les phones de la LE qui ne sont pas contrastés en L1, varierait selon l'âge de l'apprentissage. Il est probable que les apprenants plus âgés soient moins performants que les apprenants plus jeunes.
- H5 : La cinquième hypothèse met l'accent sur le *mécanisme de classification par équivalence*. Ce mécanisme pourrait bloquer la formation d'une nouvelle catégorie pour des phones de la LE. Dans ce cas, les phones de la L1 et de la LE qui sont perceptivement liés seraient classés dans une même catégorie. La production des phones serait donc similaire.
- H6 : Cette hypothèse suppose que la catégorie phonétique établie pour les phones d'une LE par un bilingue serait différente de celle d'un monolingue si : 1) la catégorie chez

le bilingue est « déviée » par rapport à une catégorie de la L1 afin de maintenir un contraste phonétique entre des catégories dans un espace phonologique commun à L1 et LE ; ou 2) la représentation chez le bilingue est fondée sur des traits différents ou sur des pondérations de traits différentes par rapport à celle du monolingue (Flege, 1995, traduit par Tran, 2011). C'est le cas par exemple d'un bilingue français-anglais qui prononce les plosives /p t k/ différemment dans chaque langue, le VOT de ces trois plosives étant différent en français (66 ms) et en anglais (108 ms) (Mack, 1990, cité par Flege, 1995). Dans ce cas, la catégorie phonétique établie pour les phones d'une LE par un bilingue parait différente de celle d'un monolingue pour maintenir le contraste phonétique entre deux langues.

- H7 : La dernière hypothèse parle de relation entre la perception et la production ; la production d'un phone correspond à la représentation de la catégorie phonétique. Par conséquent, la majorité des erreurs de production résulte, selon SLM, de la mauvaise perception.

Les deux modèles (PAM et SLM) sont les modèles les plus influents dans le domaine de la perception auditive du langage humain, notamment la perception des phones non natifs (Burfin, 2015). Le modèle SLM converge vers le modèle PAM dans le sens où les deux modèles mettent l'accent sur le rôle de la L1 à travers laquelle est affinée la perception du langage humain. La L1 fonctionne comme *un mécanisme de classification par équivalence* qui pourrait perturber la perception des phones non natifs. Cependant, alors que le modèle PAM est centré uniquement sur la perception d'un contraste non natif, le modèle SLM s'étend sur la relation entre la perception et la production dans la perspective de l'acquisition d'une LE.

La perception auditive humaine est généralement affinée par l'expérience linguistique. Chaque individu devient « insensible » à ce qui ne lui est pas familier. Rappelons que les enfants sont capables de différencier des contrastes phonétiques inexistants en L1, mais que cette capacité diminue en fonction de leur propre expérience linguistique vers l'âge de 8 mois. Quant à la perception visuelle, suivrait-elle la même trajectoire développementale que celle de la perception auditive ? Nous la présenterons dans le chapitre suivant.

Chapitre 2

La perception visuelle

2.1 Perception visuelle : la reconnaissance du visage

Pascalis, de Hann et Nelson (2002) ont effectué une étude visant à vérifier si la trajectoire développementale de la perception visuelle correspond à celle de la perception auditive. Ils testent la capacité des adultes et des enfants de 6 mois et de 9 mois à discriminer les différents visages de singes et aussi ceux des humains. Les sujets ont été testés dans la tâche de comparaison visuelle par paire (*visual paired-comparison procedure*). Un visage (de singe ou d'humain) avec lequel les sujets avaient été familiarisés, a été présenté à côté d'un nouveau visage jamais vu avant. Le temps de fixation sur les deux visages (familiarisé et nouveau) a été mesuré. Les résultats montrent que, les enfants de 6 mois fixent le regard sur le nouveau visage, que ce soit du singe ou d'humain, significativement plus longuement que celui familiarisé. Les 6 mois sont donc capables de discriminer les différents visages de chaque espèce. Cependant, les 9 mois et les adultes semblent perturbés en regardant les deux visages de singe (familiarisé et nouveau) avec la même durée, alors que pour les visages d'humain, ils fixent le nouveau significativement plus longuement que celui familiarisé. Ces résultats montrent que, de la même manière que la perception auditive, la perception visuelle (capacité de discrimination des visages) est également affinée au cours de la première année de vie.

Cette étude de Pascalis, de Hann et Nelson (2002) a été confirmée par l'étude de Kelly et collègues (2007). Ces derniers étudient la façon dont les visages observés dans l'environnement visuel des enfants affectent le développement du système de traitement du visage au cours de la première année de vie. Ils testent la capacité de différencier des visages humains d'une même ethnie et de ceux d'autres ethnies chez les enfants caucasiens âgés de 3, de 6 et de 9 mois. Les résultats montrent que les enfants de 3 mois sont capables de discriminer tous les différents visages quel que soit le type d'ethnie (caucasienne, chinoise, africaine ou moyen-orientale). Ceux de 6 mois sont moins performants. Ils sont capables de discriminer uniquement les visages des deux ethnies : les Caucasiens et les Chinois. En ce qui concerne les enfants de 9 mois, seuls les visages des Caucasiens (leur propre ethnie) peuvent être discriminés. Le schéma de préférences indique donc que l'effet de « l'autre groupe ethnique » est émergé à l'âge de 6 mois et bien présent à 9 mois chez ces enfants. L'étude de Kelly et al. (2007) montre bien que l'expérience joue un rôle crucial dans le développement de la reconnaissance du visage.

La ressemblance entre la capacité perceptive de la parole et celle des visages chez les enfants et les adultes est pointée par Nelson (2001). Les enfants sont capables de discriminer un large éventail de sons jusqu'à l'âge de 10 mois avant de rejoindre les adultes à ne plus pouvoir discriminer les sons inexistants ou non pertinents dans leur L1. Ce phénomène de « rétrécissement perceptif » attesté en parole est observé aussi dans la perception des visages où les enfants sont plus performants que les adultes dans la discrimination des visages d'une autre espèce (des singes) et d'une autre ethnie. Mais une dégradation de cette capacité perceptive visuelle est constatée également avant la fin de leur première année. Nelson (2001) suggère que le système de perception des visages est « large » et non-spécifique et que ce système se développe rapidement au cours de la première année de vie en fonction de l'expérience individuelle.

2.2 Perception visuelle dans le traitement du langage : lecture dite labiale

La lecture labiale ne constitue pas uniquement la lecture sur les lèvres. En réalité, il s'agit d'un « procédé permettant de comprendre ce que dit une autre personne en tenant compte d'indications visuelles telles que les mouvements des muscles faciaux, des lèvres, des mains et du corps en général » (l'UNESCO, 1983 citée par Dumont & Calbour, 2002, p. 116). Par conséquent, le terme « lecture labiale » paraît restrictif et il a tendance à être remplacé par un autre terme plus général : « la lecture de la parole » (Dumont & Calbour, 2002). Cependant, même si la lecture de la parole implique plusieurs mouvements faciaux et corporels, « l'identification des mots est essentiellement réalisée sur la partie basse du visage » (Dumont & Calbour, 2002, p. 8).

Il a été démontré dans l'étude de Field et collègues (1982) que les nouveau-nés peuvent discriminer en perception des expressions faciales (sourire-joie, tristesse, surprise) et les reproduire dès l'âge de 36 heures après la naissance. Melzoff et Moore (1977) attestent aussi que les enfants âgés de 12 jours reproduisent plus souvent un geste facial (protrusion de la langue, protrusion des lèvres, ouverture de la bouche) lorsque les mêmes gestes adultes sont présentés devant eux que si l'adulte lui présente une expression faciale neutre. Un enfant qui n'a jamais vu sa propre bouche peut donc discriminer et imiter les mouvements de celle d'autrui. Il est capable de mettre en correspondance le mouvent de l'adulte avec son propre appareil musculaire. L'hypothèse avancée par ces auteurs est que l'imitation dépendrait d'un processus de transfert intermodal actif (*Active intermodal mapping*) (Melzoff et Moore, 1997). Les enfants auraient donc la capacité de se

représenter visuellement et proprioceptivement l'information perçue sous une forme commune aux deux modalités.

Weikum et collègues (2007) montrent que les enfants de 4 et de 6 mois sont capables de discriminer les informations visuelles de la parole native de celles de la parole non native. Les enfants anglophones ont été testés dans deux conditions : condition de test et condition de contrôle. Dans la condition de test, ils sont familiarisés avec une vidéo silencieuse dans laquelle un locuteur articule des phrases d'une langue (par exemple, l'anglais). Lorsque la durée de regard diminue, les enfants sont exposés par la suite à une autre vidéo silencieuse dans laquelle un même locuteur articule des phrases d'une autre langue (par exemple, le français). Dans la condition de contrôle, lorsque la durée de regard diminue, les enfants sont exposés à une autre vidéo silencieuse dans laquelle un locuteur articule des phrases d'une même langue que dans la phase de familiarisation. Les résultats montrent que les enfants de 4 et de 6 mois sont capables de discriminer visuellement la différence d'articulation entre deux langues (native et non native). Cependant, les enfants de 8 mois qui sont testés dans la même étude n'en sont pas capables.

Il a été démontré que les informations visuelles sont exploitées lors du traitement de la parole. Comment les informations visuelles sont-elles catégorisées lors du traitement du langage ? Si la plus petite unité distinctive des sons d'une langue est appelée *phonème*, comment appelle-t-on la plus petite unité visuelle lors du traitement de la parole ?

2.3 Visèmes

En lecture de la parole, notamment la lecture des mouvements articulatoires, la plus petite unité distinctive est appelée « visème ». Le visème fait référence à un « phonème visuel » (*visual phoneme*) (Fisher, 1968 cité par Borel et al., 2016). Ce terme désigne « les phonèmes dont l'articulation visible procède du même geste facial » (Dumont & Calbour, 2002, p.11). Par exemple, les consonnes /p b m/ sont toutes articulées avec le même organe, les lèvres. L'articulation visuelle de ces trois consonnes constitue donc un visème. Les trois consonnes /p b m/ qui ont la même image labiale sont appelées « sosies labiaux », c'est à dire « un ensemble de phonèmes que l'on ne peut différencier avec les seules informations de la lecture labiale » (Istria et al., 1982 ; Borel et al., 2016, p.28). Cependant, l'articulation de certains phonèmes est

« invisible », par exemple les consonnes /k ɡ/ qui sont articulées dans la partie postérieure de la cavité buccale.

Classifications de visèmes consonantiques du français

Istria et collègues (1982) proposent une classification de visèmes consonantiques du français en trois catégories selon leur visibilité : 1) les consonnes invisibles, 2) les consonnes stables et 3) les consonnes variables.

Les consonnes invisibles /k ɡ ʁ/ sont les consonnes dont l'articulation se situe au fond de la gorge, tandis que les consonnes stables sont les consonnes les plus visibles ; le mouvement des lèvres pour les prononcer est assez apparent. Les consonnes stables se divisent en trois groupes selon le lieu d'articulation (cf. table 1). Concernant les consonnes variables, ce sont les consonnes dont la visibilité est intermittente : c'est-à-dire qu'elles seront visibles lorsqu'elles se combinent avec une voyelle ouverte, mais elles ne le seront pas lors de l'articulation avec une voyelle fermée. Les consonnes variables sont les alvéolaires et elles sont classées en trois groupes selon le mode d'articulation (cf. table 1).

Table 1. Classification de visèmes consonantiques du français proposée par Istria et al. (1982)[1]

Consonnes invisibles		/k ɡ ʁ/
Consonnes stables ou visibles	Groupe P	/p b m/
	Groupe F	/f v/
	Groupe CH	/ʃ ʒ/
Consonnes variables	Groupe T	/t d n/
	Groupe S	/s z/
	-	/l/

Dumont et Calbour (2002) proposent également la classification des visèmes consonantiques du français en fonction de leur visibilité et de leur position dans la syllabe (cf. table 2).

[1] La consonne /ɲ/ n'est pas indiquée dans cette classification.

Table 2. Classification de visèmes consonantiques du français proposée par Dumont et Calbour (2002)

Visèmes consonantiques du français en position initiale de syllabe	
Visème des occlusives bilabiales	[p b m]
Visème des constrictives labiodentales	[f v]
Visème des constrictives postalvéolaires	[ʃ ʒ]
Visème des consonnes intrabuccales	[t d n k g ɲ s z l r]
Visèmes consonantiques du français en position finale de syllabe	
Visème des occlusives bilabiales	[p b m]
Visème des constrictives labiodentales	[f v]
Visème des constrictives postalvéolaires	[ʃ ʒ]
Visème des consonnes intrabuccales	[k g ɲ r]
Visème des occlusives	[t d n]
Visème des constrictives dentales	[s z]
Visème de [l]	[l]

Dans les deux premiers chapitres, la perception auditive et la perception visuelle sont abordées séparément. Cependant, lors de la communication, l'individu est généralement exposé à l'input auditif et à l'input visuel de manière simultanée. Que se passe-t-il quand la perception auditive et la perception visuelle se combinent ? Quels sont les avantages de la présentation audiovisuelle ?

Chapitre 3

La perception audiovisuelle

> La perception de la parole n'est donc pas qu'auditive. Elle concerne aussi la vision, qui dans ce processus ne réalise pas la simple addition d'indices visuels aux percepts acoustiques, mais qui en synchrone avec l'audition s'intègre dans l'acte de perception unitaire de la parole qu'est « l'audiovision »
>
> (Dumont & Calbour, 2002, p. 8)

3.1 Perception audiovisuelle dans le traitement de la langue native

La langue orale est généralement perçue en modalité audiovisuelle ; on entend la voix et on voit la personne qui parle. Comment les enfants et les adultes utilisent-ils les informations audiovisuelles dans le traitement de la parole native ?

Plusieurs études montrent que les enfants dès 4-5 mois sont capables de reconnaître les mouvements articulatoires correspondant aux sons qu'ils entendent (Kuhl & Meltzoff, 1982 ; Legerstee, 1990 ; Patterson & Werker, 1999 ; Kubicek et al., 2014). A l'aide d'un protocole de préférence visuelle, Kuhl et Meltzoff (1982) montrent que les enfants âgés de 18 à 20 semaines, devant un écran présentant deux visages articulant [a] et [i], regardent plus longtemps celui dont le mouvement de la bouche correspond au son [a] ou [i] présenté simultanément. Ces enfants peuvent donc établir une correspondance entre les gestes articulatoires vus sur les lèvres et les sons entendus. Ils savent que le son [a] correspond à une bouche ouverte et le son [i] à une position rétractée des lèvres. Patterson et Werker (1999) confirment les résultats de Kuhl et Meltzoff (1982) et attestent que les enfants de 4,5 mois sont capables de reconnaître les mouvements articulatoires des lèvres correspondant au son vocalique entendu ([i] ou [a]), que les stimuli soient le visage d'un homme ou d'une femme. Lorsque les informations vues sur les lèvres et entendues sont cohérentes, la concordance des informations auditive et visuelle favorise l'imitation chez les enfants de 3 mois (Legerstee, 1990). Ces derniers vocalisent et font des mouvements de bouche appropriés beaucoup plus dans une condition appariée (par exemple visage 'a' + son [a]) que dans une condition non appariée (visage 'a' + son [u]).

Chez l'adulte, il a été démontré que la présentation audiovisuelle facilite l'intelligibilité de la parole native (L1) en situation bruyante (Sumby & Pollack, 1954 ; Benoît, Mohamadi & Kandel, 1994 ; Wang, Behne & Jiang, 2008). Sumby et Pollack (1954) testent l'aptitude de compréhension des anglophones, en deux modalités auditive seule et audiovisuelle. Les stimuli auditifs (les mots en anglais), utilisés dans les deux modalités, sont noyés dans le bruit avec de différents niveaux d'intensité. En modalité auditive seule, les locuteurs doivent identifier ce qu'ils entendent. En modalité audiovisuelle, les locuteurs effectuent la même tâche d'identification mais avec les stimuli auditifs présentés avec les images (visage) de la personne qui produit ces sons. Les résultats

montrent que, au même niveau d'intensité de la parole (rapport signal sur bruit[2] = -30 dB), le pourcentage des réponses correctes en modalité audiovisuelle est plus élevé qu'en modalité auditive. Autrement dit, la présentation audiovisuelle facilite l'intelligibilité de la parole native, notamment en situation bruyante.

L'étude de Benoît, Mohamadi et Kandel en 1994 confirme également l'impact positif de la présentation audiovisuelle sur la perception de la parole lorsque les informations acoustiques sont réduites. Cette étude a testé l'aptitude à comprendre des francophones en deux modalités (auditive et audiovisuelle) avec différents niveaux de bruit : le rapport signal sur bruit varie entre -24 et 0 dB. Les stimuli sont les pseudo-mots qui comportent six consonnes [b v z ʒ ʁ l] et trois voyelles [i a y] du français sous la structure [VCVCVz] (par exemple, [iviviz]). Les sujets ont dû choisir une voyelle et une consonne qui correspondent à chaque séquence. Les résultats montrent que le pourcentage de réponses correctes en modalité audiovisuelle est significativement plus élevé qu'en modalité auditive. La visibilité du visage des locuteurs augmente la capacité de perception de la parole, notamment dans la situation bruyante. La parole qui est perturbée par le bruit peut être récupérée partiellement par l'information visuelle.

Toutefois, l'intégration incohérente des informations auditives et des informations visuelles peut distordre la perception de la parole. McGurk et MacDonald en 1976 ont effectué une étude pour tester la perception des enfants (3-4 ans et 7-8 ans) et des adultes anglophones en deux modalités : auditive et audiovisuelle. Il s'agit d'une tâche de répétition. En modalité auditive, les sujets répètent ce qu'ils entendent. En modalité audiovisuelle, les sujets sont exposés aux stimuli auditifs et visuels qui ne sont pas congruents. Par exemple, le stimulus auditif [ba] est présenté simultanément avec la vidéo sur laquelle un locuteur articule la syllabe [ga]. La prononciation des sujets qui correspond aux stimuli auditifs est considérée comme réponse correcte. Les résultats montrent que, en modalité auditive, le pourcentage des réponses correctes est élevé : plus de 90 % des réponses correctes pour les deux groupes de sujets (enfants et adultes). En modalité audiovisuelle, les sujets produisent beaucoup plus d'erreurs ; 59 % d'erreurs pour les enfants de 3-4 ans, 52 % pour les enfants de 7-8 ans et 92 % pour les adultes. Une des erreurs

[2] Le rapport signal sur bruit désigne le rapport entre la puissance du signal auditif et la puissance du signal parasite. Cela permet de quantifier le niveau de bruit inséré dans le signal acoustique d'intérêt. Plus le rapport signal sur bruit est grand, plus le bruit est faible par rapport au signal et plus le signal est perceptible. Ainsi un rapport signal sur bruit de -18dB masquera bien plus le signal acoustique qu'un signal rapport sur bruit de -6dB. (Burfin, 2015)

remarquables est celle des réponses dites « fusionnées » : par exemple, quand les sujets sont exposés au stimulus auditif [ba] et au stimulus visuel [ga], ils perçoivent [da]. Cette fusion entre les stimuli auditif et visuel est appelée « l'effet McGurk ».

Ainsi, l'information visuelle peut influencer fortement la perception de la parole native. Du côté positif, la présentation visuelle facilite l'intelligibilité de la parole lorsque l'audition est perturbée (en milieu bruyant). Cependant, lorsque la vision n'est pas cohérente avec l'audition, la perception de la parole est donc déviée comme l'effet McGurk mentionné ci-dessus. Quant à la perception de la parole d'une LE, comment et dans quelle mesure est-elle influencée par les informations visuelles ?

3.2 Perception audiovisuelle dans le traitement d'une langue étrangère

3.2.1 Chez les bébés

Rappelons que les enfants plus jeunes possèdent une capacité remarquable de discrimination des sons natifs et non natifs (cf. chapitre 1) et des gestes articulatoires visuels entre L1 et LE (cf. chapitre 2). Que se passe-t-il lors de la présentation audiovisuelle de la parole en LE ?

Kubicek et al. (2014) étudient la capacité d'identification de parole native et non native chez les enfants germanophones de 4,5 mois et de 6 mois dans la condition où les informations auditives et visuelles ne sont pas présentées simultanément (tâche asynchrone). Dans la phase de test, les enfants sont exposés premièrement aux stimuli auditifs : des phrases prononcées soit en allemand, soit en français. Deuxièmement, ils sont exposés aux stimuli visuels. Il s'agit de deux vidéos silencieuses placées côte à côte : l'une présente une locutrice qui articule les mêmes phrases que celles présentées dans la phase d'audition, l'autre présente une même locutrice qui articule les phrases de l'autre langue. La durée de regard a été mesurée. Les résultats globaux montrent que les enfants sont capables de reconnaître les informations visuelles liées aux sons entendus malgré l'absence de synchronie mais les enfants de 4,5 mois sont plus performants que ceux de 6 mois. Les plus jeunes à 4,5 mois sont capables d'identifier la cohérence des mouvements articulatoires correspondant aux sons entendus, que ce soient les sons natifs ou non natifs, tandis que ceux de 6 mois sont capables d'identifier uniquement les mouvements articulatoires qui correspondent aux sons existants dans leur langue.

Dans la même étude, Kubicek et al. (2014) étudient également la capacité perceptive des enfants germanophones de 6 mois et de 12 mois lorsqu'ils sont exposés aux stimuli auditifs et visuels simultanément (tâche synchrone). Dans la phase de test, les enfants sont exposés à un stimulus auditif (les phrases lues en allemand ou en français) avec deux stimuli visuels placés côte à côte (les vidéos silencieuses dans lesquelles un locuteur bilingue prononce les phrases en allemand et en français). La durée de regard a été mesurée. Les résultats montrent que la capacité d'identification des enfants de 6 mois est reliée à la synchronie des informations. Ces enfants sont capables de reconnaître les mouvements articulatoires correspondant aux sons entendus, que ce soient les sons natifs ou non natifs. Autrement dit, l'exposition simultanée aux indices auditifs et visuels facilite la capacité d'identification des stimuli non natifs chez les enfants de 6 mois. Chez les enfants de 12 mois, les résultats sont intrigants, ceux-ci étant seulement capables de reconnaître les mouvements articulatoires de la LE. Les auteurs, Kubicek et collègues (2014), expliquent que les résultats chez les enfants de 12 mois vont dans le même sens que l'étude de Lewkowicz et Hansen-Tift (2012) qui montre que les enfants de 12 mois regardent plus longtemps la bouche lorsqu'un locuteur parle une LE. Par ailleurs, Kubicek et collègues (2013) attestent également dans une étude antérieure que lorsque les enfants de 12 mois sont exposés aux sons non natifs, ils regardent la bouche du locuteur plus que les yeux. Autrement dit, lors de l'exposition à la parole non native, l'attention des enfants de 12 mois est portée sur la bouche et ils peuvent bénéficier de la synchronie des informations audiovisuelles pour identifier la parole non native.

En récapitulatif, les enfants dès 4-5 mois sont capables d'identifier les mouvements articulatoires de la parole native et non native même si l'input auditif et l'input visuel ne sont pas présentés simultanément. Les enfants de 6 mois, en absence de synchronie audiovisuelle, peuvent identifier uniquement des mouvements articulatoires de la parole native. Leur capacité peut cependant être remédiée par la présence de l'information audiovisuelle synchronisée de la parole. Quant aux enfants de 12 mois, les informations visuelles sont plus utilisées lors du traitement de la parole non native. Ces enfants sont attirés plutôt par la bouche et sont capables de reconnaître les gestes articulatoires de la LE.

3.2.2 Chez les adultes

La capacité d'identification de la parole attestée chez les enfants semble maintenue à l'âge adulte. Dans une tâche asynchrone dont l'input auditif et l'input visuel sont présentés l'un après

l'autre, Pons et collègues (2009) montrent que les adultes peuvent identifier les mouvements articulatoires correspondant aux sons natifs entendus. Mais ils n'en sont pas capables pour les sons non natifs. Dans cette étude, Pons et collègues étudient la capacité des adultes anglophones et hispanophones à identifier les phonèmes /v/ et /b/. L'opposition /v/ ~ /b/ est attestée en anglais, mais pas en espagnol. Les participants doivent choisir la vidéo (silencieuse) correspondant au son précédemment entendu. Les résultats montrent que, de manière significative, le score des réponses correctes chez les anglophones est largement plus élevé (91,9 %) que celui des hispanophones (51,56 %). Autrement dit, comme les enfants à partir de 10 mois, les adultes ne sont plus capables de reconnaître les mouvements articulatoires correspondant aux sons non natifs entendus si les stimuli auditif et visuel sont présentés de manière asynchrone.

Hazan et collègues (2006) étudient l'usage des indices visuels dans la perception du lieu d'articulation (labial /p b/ *vs* labiodental /v/) chez les apprenants hispanophones et japonophones de l'anglais. Les stimuli sont les pseudo-mots de structures CV, VCV ou VC dont C sont /p b v/ et V /i ɑ u/. Les bilabiales /p b/ existent en espagnol et en japonais, mais ce n'est pas le cas de /v/. Cette dernière est cependant l'allophone de /f/ en coda en espagnol. Les apprenants sont exposés aux stimuli en trois modalités : auditive, visuelle et audiovisuelle. Ils ont pour consigne de choisir une des trois consonnes (*p*, *b* et *v*) qui correspond à chaque stimulus présenté (tâche d'identification). Les résultats montrent que l'effet de modalité est statistiquement significatif. La capacité à détecter les 3 lieux d'articulation chez les deux groupes d'apprenants en modalité audiovisuelle est significativement plus élevée qu'en modalité auditive et qu'en modalité visuelle. Cependant, les résultats des deux groupes d'apprenants ne sont pas au même niveau ; le score des réponses correctes des hispanophones est plus élevé que celui des japonophones (cf. figure 2). L'effet de l'expérience linguistique est statistiquement significatif. Selon Hazan et collègues (2006), ces résultats pourraient résulter du fait que les hispanophones sont exposés au son [v] qui est allophone de /f/ en coda en espagnol.

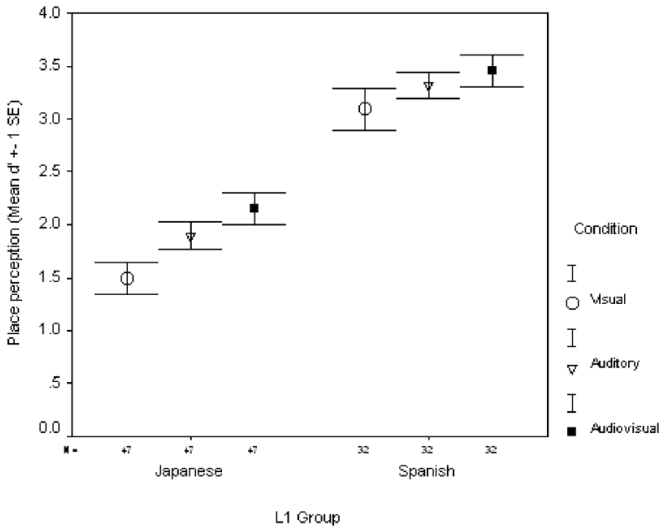

Figure 2. Détectabilité (*d'*)[3] du lieu d'articulation en trois modalités (visuelle, auditive et audiovisuelle) chez les apprenants japonophones et hispanophones de l'anglais (selon Hazan et al., 2006)

L'étude de Wang et collègues (2009) montrent qu'en modalité audiovisuelle, les apprenants d'une LE sont plus performants en identification du lieu d'articulation inexistant en L1 (visème non natif) que d'un autre lieu articulatoire présent en L1 (visème natif). Dans cette étude, les apprenants sinophones (mandarins) et coréens ainsi que les anglophones (groupe de contrôle) sont testés en trois modalités : auditive, visuelle et audiovisuelle. Les stimuli sont divisés en trois groupes en fonction du lieu d'articulation : labiodental /f v/, interdental /θ ð/ et alvéolaire /s z/. Chaque consonne est accompagnée d'une des trois voyelles /i a u/ en structure CV. Il est à noter que le visème interdental n'existe pas en mandarin ni en coréen et que le mandarin possède le visème labiodental /f/ et le visème alvéolaire /s/, alors que le coréen ne possède que le visème alvéolaire /s/. Les sujets doivent identifier quelle consonne correspond aux stimuli auxquels ils sont exposés. Les résultats (cf. table 3) montrent que, pour le lieu interdental (visème non natif), le pourcentage des réponses correctes en modalité audiovisuelle chez les apprenants sinophones et coréens est beaucoup plus élevé qu'en modalité auditive. Cependant, l'augmentation des réponses correctes est moins évidente pour le lieu labiodental (visème natif). De plus, aucune augmentation n'est trouvée pour le lieu alvéolaire : le visème natif moins visible. Ces résultats impliquent que les apprenants sinophones et coréens utilisent davantage les informations visuelles lors du traitement du visème non natif que du visème natif.

[3] *d'* : une mesure de sensibilité ou de discriminabilité qui n'est pas affectée par les biais de réponse

Table 3. Pourcentage des réponses correctes (identification du lieu d'articulation) en modalités auditive et audiovisuelle chez les anglophones, les apprenants coréens et les apprenants sinophones de l'anglais (selon Wang et al., 2009) (Table adaptée de l'original)

		Stimuli présentés selon le lieu d'articulation								
		Labiodental			Interdental			Alvéolaire		
Modalité	Perçus comme	Anglo.	Cor.	Sino.	Anglo.	Cor.	Sino.	Anglo.	Cor.	Sino.
Auditive	Labiodentale	**87**	**93**	**85**	11	29	31	0	0	0
	Interdentale	13	6	9	**89**	**67**	**50**	0	1	13
	Alvéolaire	0	1	2	0	3	19	**100**	**98**	**85**
Audiovisuelle	Labiodentale	**93**	**96**	**90**	6	14	13	0	0	0
	Interdentale	7	4	7	**94**	**84**	**67**	0	1	16
	Alvéolaire	0	0	1	0	2	18	**99**	**98**	**81**

L'étude de Burfin (2015) confirme également l'apport des informations audiovisuelles dans le traitement de la parole non native, notamment le visème inexistant en L1. Burfin (2015) étudie la perception du contraste /f/ ~ /θ/ chez les francophones. Il est à remarquer que /f/ existe en français, mais ce n'est pas le cas de /θ/ ainsi que son visème (interdental). Les sujets ont été testés dans les deux modalités : auditive et audiovisuelle. En modalité auditive, les sujets sont exposés dans un premier temps à un son considéré comme mot de référence (par exemple, /paθo/) et ensuite aux stimuli de test (par exemple, /pafo/, /paso/ et /paθo/). Les sujets doivent mémoriser le mot de référence entendu et juger si les stimuli de test sont identiques ou différents par rapport au mot de référence. En modalité audiovisuelle, les sujets suivent les mêmes consignes, mais ils sont exposés à la fois à l'audition et à la vision. Les sons sont accompagnés des vidéos dans lesquelles un locuteur articule les mots en question. Les résultats montrent que le pourcentage des réponses correctes pour la consonne /θ/ chez les francophones est significativement plus élevé en modalité audiovisuelle qu'en modalité auditive (cf. figure 3).

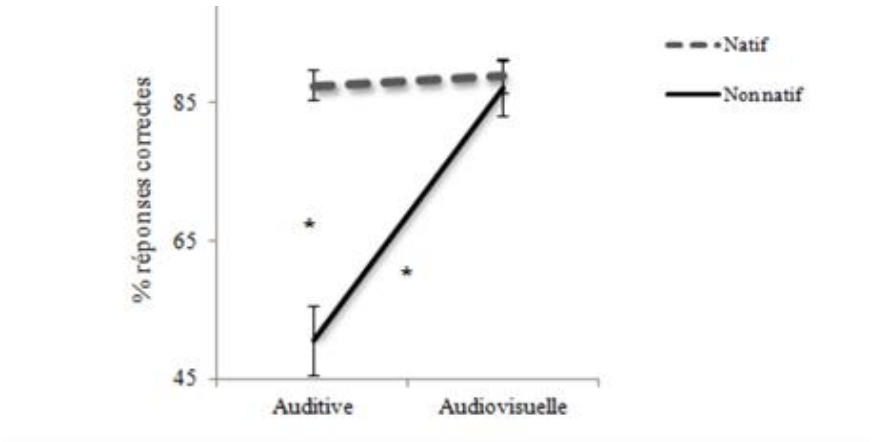

Figure 3. Pourcentage des réponses correctes chez les francophones dans la tâche de discrimination des sons /f/ ~ /θ/ en modalités auditive et audiovisuelle (selon Burfin, 2015) (figure adaptée de l'original)

En résumé, les adultes maintiennent la capacité perceptive audiovisuelle attestée chez les enfants de 6-12 mois. Lorsqu'ils sont exposés à l'input auditif et à l'input visuel de manière asynchrone, ils ne sont pas capables d'identifier les mouvements articulatoires correspondant aux sons non natifs précédemment entendus. Au contraire, lorsque les stimuli auditif et visuel sont présentés simultanément, la capacité d'identification/discrimination de la parole non native est remédiée. Autrement dit, la présentation audiovisuelle facilite la perception de la parole non native. Cependant, il a été démontré que les apprenants d'une LE n'utilisent pas les informations visuelles de la même manière. Quels seraient les facteurs qui impactent l'utilisation des informations visuelles lors de l'apprentissage d'une LE ?

3.2.3 Facteurs impactant l'utilisation des informations visuelles lors du traitement d'une langue étrangère

Dans la section précédente (cf. 3.2.2), il a été démontré que, lors du traitement de la parole non native, les apprenants d'une LE n'utilisent pas les informations visuelles de la même manière. Quatre facteurs suivants pourraient impacter l'utilisation des informations visuelles de la parole lors du traitement de la LE (Burfin, 2015).

3.2.3.1 Saillance du visème

Tous les visèmes n'ont pas le même niveau de saillance, certains visèmes étant plus ou moins visibles que d'autres. La saillance ou la visibilité des visèmes pourrait être un des facteurs qui impacte l'utilisation des informations visuelles lors du traitement de la LE : plus le visème est saillant, plus les apprenants sont capables d'utiliser les informations visuelles. Wang et al. (2009)

montrent que, en modalité audiovisuelle, la capacité perceptive des interdentales (visème bien visible ou saillant) chez les apprenants de l'anglais augmente de manière conséquente par rapport à la modalité auditive. Aucune amélioration n'est cependant attestée dans l'identification des alvéolaires (visème moins saillant) entre les deux modalités audiovisuelle et auditive seule.

3.2.3.2 Organisation de l'espace phonologique de la L1

Les apprenants de différentes L1s n'utilisent pas les informations visuelles de la même manière. Wang et collègues (2009) montrent que les apprenants coréens et mandarins de l'anglais bénéficient des informations visuelles lors du traitement des fricatives interdentales non natives /θ ð/, mais les Coréens sont plus performants que les mandarins. La différence de la capacité perceptive entre les apprenants coréens et mandarins pourrait résulter de l'organisation de l'espace phonologique de la L1. Le mandarin possède les fricatives labiodentale /f/ et alvéolaire /s/ considérées comme les « voisins » proches des interdentales de l'anglais. Ce voisinage pourrait engendrer de l'ambiguïté chez les apprenants mandarins lors de la perception des fricatives interdentales. Cependant, le coréen possède uniquement l'alvéolaire /s/ qui n'a pas de lien direct avec l'articulateur dental. Par conséquent, les apprenants coréens rencontrent moins de difficultés que les apprenants mandarins lors de la perception audiovisuelle des interdentales.

3.2.3.3 Relation entre les visèmes de la L1 et ceux de la LE

En se basant sur le modèle PAM (cf. section 1.2.1), Hazan et al. (2006) établissent trois catégories de relation entre les visèmes de la L1 et ceux de la LE :

1. Les visèmes qui sont identiques en L1 et en LE et qui sont utilisés pour marquer la même distinction phonémique en deux langues seront assimilés à des catégories visémiques natives. Par exemple, le visème labiodental de l'anglais (LE) qui est utilisé pour deux phonèmes /f v/ sera assimilé à la même catégorie visémique de la L1 chez les apprenants français de l'anglais. Dans ce cas, les apprenants ne rencontrent pas de difficulté à utiliser des indices visuels.
2. Les visèmes d'une LE qui n'existent pas en L1 ne seront pas assimilés à des catégories natives. Par exemple, le visème interdental de l'anglais qui n'existe pas en français ne sera pas assimilé à une catégorie visémique native des apprenants français.
3. Les visèmes existent en L1 et en LE, mais ils ne sont pas utilisés pour la même distinction phonémique entre L1 et LE. Par exemple, le visème labiodental existe en thaï (L1) et en

français (LE), mais il n'est pas utilisé pour la même distinction phonémique dans les deux langues. En thaï, un seul phonème /f/ appartient au visème labiodental, alors que le français possède le visème labiodental pour deux phonèmes /f v/.

Dans ces deux derniers cas, les apprenants d'une LE seraient moins sensibles aux visèmes inexistants en L1 et aux visèmes qui ne sont pas utilisés pour faire la même distinction phonémique en L1. Cependant, certaines études s'opposent à cette considération.

3.2.3.4 Performance dans la LE

Le poids donné aux informations visuelles de la parole non native pourrait varier en fonction du niveau des apprenants. Les apprenants plus expérimentés d'une LE ont tendance à utiliser plus efficacement les informations visuelles. Wang et collègues (2008) analysent la capacité perceptive des trois groupes de sujets : a) les apprenants sinophones qui habitent dans une zone anglophone du Canada depuis environ 10 ans (longue durée de séjour LDS) ; b) les apprenants sinophones avec une durée moyenne de séjour au Canada de 2 ans (courte durée de séjour CDS) et c) les anglophones canadiens (group de contrôle). Les trois groupes sont exposés aux stimuli /θi θa θu ði ða ðu/ en trois modalités auditive, visuelle et audiovisuelle. Il s'agit d'une tâche d'identification dans laquelle les sujets doivent choisir une des six fricatives de l'anglais /s z f v θ ð/ correspondant aux stimuli auxquels ils sont exposés. Les résultats montrent que, en modalité audiovisuelle, les apprenants LDS sont plus performants que les apprenants CDS avec un pourcentage des réponses correctes significativement plus élevé. Autrement dit, l'utilisation des informations visuelles dans le traitement de la parole non native pourrait être modulée en fonction du niveau des apprenants.

En dehors des quatre facteurs mentionnés ci-dessus, Hazan et al. (2006) propose un facteur lié à la culture. Le fait qu'un individu utilise moins d'indices visuels que d'autres pourrait résulter du fait que, dans sa culture, il est irrespectueux de regarder directement le visage des locuteurs. L'utilisation des indices visuels est donc moins performante lors du traitement de la parole non native.

Même s'il y a plusieurs facteurs qui peuvent impacter négativement l'utilisation des informations visuelles dans la parole non native, plusieurs études démontrent que la présentation audiovisuelle facilite le traitement de la LE. Dans le cadre de l'apprentissage d'une LE, les

informations audiovisuelles pourraient jouer un rôle important dans la perception ainsi que dans la production : elles pourraient amener les apprenants à bien percevoir et à bien prononcer les sons non natifs. Par conséquent, dans cette présente étude, nous visons à vérifier si la présentation audiovisuelle pourrait faciliter la perception et la production de la consonne /v/ en position initiale de syllabe chez les apprenants thaïlandais du français. La consonne /v/ n'existe pas en thaï. Les apprenants thaïlandais rencontrent beaucoup de difficultés dans la perception et dans la production de cette consonne /v/ en la confondant très souvent avec /w/ en initiale de syllabe (Tusnyingyong & Tran, 2022). Nous présenterons dans le chapitre suivant des difficultés que rencontrent les apprenants ainsi que les facteurs qui entraînent la confusion entre /v/ et /w/ chez les apprenants thaïlandais.

Chapitre 4

Confusion entre /v/ et /w/ chez les apprenants thaïlandais

4.1 Le thaï

Le thaï appartient au groupe taï, un sous-groupe de la famille tai-kadaï (Maspero, 1911 ; Rousset, 2004). Il s'agit d'une langue tonale qui comporte cinq tons : moyen, bas, descendant, haut et montant (Abramson, 1963).

Il existe quatre langues régionales en Thaïlande en fonction des zones géographiques et administratives : le Centre, le Nord, le Nord-Est et le Sud (Premsrirat et al., 2004). La variété du Centre appelée le « thaï standard » est la langue officielle de la Thaïlande. Le thaï standard « a un statut formel utilisé dans l'administration, dans le domaine de l'éducation et dans les médias » (Promkesa, 2014, p. 40).

4.1.1 Le système consonantique

Le thaï possède 21 phonèmes consonantiques dont deux glides (cf. table 4). Ces 21 phonèmes peuvent apparaître tous en attaque, tandis que seuls 9 d'entre eux peuvent se positionner en coda /p t k m n ŋ ʔ w j/ (Pacquement, 2007). Les consonnes sourdes qui existent majoritairement en thaï sont contrastées par l'aspiration (/p/ ~ /pʰ/, /t/ ~ /tʰ/, /k/ ~ /kʰ/, /tɕ/ ~ /tɕʰ/).

Table 4. Phonèmes consonantiques du thaï
(pour chaque lieu d'articulation, les non-voisés en colonne de gauche et les voisés de droite)

	Bilabiales		**Labio-dentale**		**Alvéolaires**		**Alvéo-palatales**		**Palatale**		**Vélaires**		**Glottales**		**Labio-vélaire**	
Plosives	p pʰ	b			t tʰ	d					k kʰ		ʔ			
Nasales		m				n						ŋ				
Fricatives			f		s								h			
Affriquées							tɕ tɕʰ									
Vibrante						r										
Approximantes										j						w
Latéral-approximante						l										

Comparaison avec le système consonantique du français

Le système consonantique du français est présenté dans la table 5. Les consonnes en fond gris sont les consonnes du français qui n'existent pas en thaï : /v/, /z/, /ʃ/, /ʒ/, /ɲ/, /g/, /ʁ/ et /ɥ/. En

comparaison entre les deux systèmes, nous constatons que les fricatives du français (/f/, /v/, /s/, /z/, /ʃ/, /ʒ/ et /ʁ/) sont plus nombreuses que celles du thaï (/f/, /s/ et /h/). La sonorité est le trait distinctif dominant pour les fricatives du français, alors que les fricatives du thaï s'opposent entre elles par le lieu d'articulation.

Table 5. Phonèmes consonantiques du français (à gauche, les non-voisés et à droite, les voisés)

	Bilabiales		**Labio-dentale**		**Alvéolaires**		**Post-alvéolaires**		**Palatale**		**Vélaires**		**Uvulaire**		**Labio-palatale**		**Labio-vélaire**	
Plosives	p	b			t	d					k	ɡ						
Nasales		m				n				ɲ		ŋ						
Fricatives			f	v	s	z	ʃ	ʒ						ʁ				
Approximantes										j						ɥ		w
Latéral-approximante						l												

4.1.2 Le système d'écriture du thaï

Le thaï possède sa propre écriture provenant d'une forme ancienne d'écriture khmère. Le système d'écriture du thaï est un système alphabétique ; chaque graphème représente la plus petite unité de la syllabe.

Le thaï possède 44 graphèmes consonantiques dont 2 sont obsolètes : < ฃ > et < ฅ > (Kosawat, 2003). Rappelons que le thaï possède 21 phonèmes consonantiques. Un phonème peut donc être représenté par plus qu'un graphème. De plus, un même graphème si présenté dans des positions syllabiques différentes ne représente pas nécessairement un même phonème (Danvivathana, 1981). Par exemple, le graphème < บ > en initiale de syllabe représente le phonème /b/, mais ce même graphème trouvé en finale représente le phonème /p/.

Table 6. Graphèmes consonantiques du thaï (adaptés de Kosawat, 2003)

Graphèmes	**Phonèmes**		**Graphèmes**	**Phonèmes**	
	En initiale	**En finale**		**En initiale**	**En finale**
ก	/k/	/k/	ท	/tʰ/	/t/
ข	/kʰ/	/k/	ธ	/tʰ/	/t/
ฃ	/kʰ/	-	น	/n/	/n/
ค	/kʰ/	/k/	บ	/b/	/p/

ค	/k^{h}/	-	ป	/p/	/p/
ฆ	/k^{h}/	/k/	ผ	/p^{h}/	-
ง	/ŋ/	/ŋ/	ฝ	/f/	-
จ	/tɕ/	/t/	พ	/p^{h}/	/p/
ฉ	/tɕ^{h}/	-	ฟ	/f/	/p/
ช	/tɕ^{h}/	/t/	ภ	/p^{h}/	/p/
ซ	/s/	/t/	ม	/m/	/m/
ฌ	/tɕ^{h}/	-	ย	/j/	/j/
ญ	/j/	/n/	ร	/r/	/n/
ฎ	/d/	/t/	ล	/l/	/n/
ฏ	/t/	/t/	ว	/w/	/w/
ฐ	/t^{h}/	/t/	ศ	/s/	/t/
ฑ	/t^{h}/ ou /d/	/t/	ษ	/s/	/t/
ฒ	/t^{h}/	/t/	ส	/s/	/t/
ณ	/n/	/n/	ห	/h/	-
ด	/d/	/t/	ฬ	/l/	/n/
ต	/t/	/t/	อ	/ʔ/	-
ถ	/t^{h}/	/t/	ฮ	/h/	-

4.1.3 Les visèmes du thaï

En se basant sur la classification des visèmes consonantiques du français proposée par d'Istria et collègues (1982), nous proposons une classification des visèmes du thaï en les divisant en trois catégories en fonction de leur visibilité (cf. table 7).

La première catégorie se compose des consonnes invisibles /tɕ tɕ^{h} j k k^{h} ŋ ʔ h/ dont l'articulation n'est pas apparente. La deuxième catégorie constitue les consonnes stables dont l'articulation est bien visible. Cette catégorie possède trois visèmes : bilabiales /p p^{h} b m/, labiodentale /f/ et labiovélaire /w/. La dernière catégorie constitue les consonnes variables dont la visibilité est intermittente : c'est-à-dire qu'elles seront visibles lorsqu'elles se combinent avec la voyelle ouverte, mais elles ne le seront pas lors de l'articulation avec la voyelle fermée.

Table 7. Classification de visèmes consonantiques du thaï

Consonnes invisibles		/tɕ tɕʰ j k kʰ ŋ ʔ h/
Consonnes stables ou visibles	Bilabiale	/p pʰ b m/
	Labiodentale	/f/
	Labiovélaire	/w/
Consonnes variables		/t tʰ d n s r l/

4.2 La confusion /v/ ~ /w/ chez les apprenants thaïlandais

Il a été montré que chez les apprenants thaïlandais, la prononciation erronée dans la LE provient, dans la plupart des cas, des interférences avec le système phonologique de la L1 (Le Corre, 2013 ; Sridhanyarat, 2017 ; Purinthrapibal, 2018). C'est le cas par exemple du dévoisement de certaines consonnes sonores du français, une stratégie largement utilisée par les apprenants thaïlandais (ex. *choisir* prononcé [ʃwaz̥iʁ]). En effet, les consonnes du français s'opposent principalement par le voisement (/p/ ~ /b/, /t/ ~ /d/, /k/ ~ /g/, /f/ ~ /v/, /s/ ~ /z/, /ʃ/ ~ /ʒ/), tandis que dans la langue thaïe, les obstruantes sonores sont beaucoup moins présentes et les consonnes sourdes qui existent majoritairement dans la langue peuvent être aussi contrastées par l'aspiration (/p/ ~ /pʰ /, /t/ ~ /tʰ /, /k/ ~ /kʰ /, /tɕ/ ~ /tɕʰ/). Il a été montré que les apprenants thaïlandais, ne maîtrisant pas la prononciation des consonnes voisées, ont tendance à les dévoiser : la plosive vélaire sonore /g/ prononcée comme son homologue sourd [k] (Ngammana, 2011) ou les fricatives /z/ et /ʒ/ en initiale de la syllabe réalisées comme [s] et [ʃ] respectivement dans la plupart des cas (Le Corre, 2013).

Cependant, pour la consonne /v/, les apprenants n'appliquent pas la même stratégie de dévoisement comme c'était le cas pour les fricatives sonores /z/ et /ʒ/. Cette consonne, n'existant pas en thaï, est confondue avec [w] en position initiale de syllabe et très souvent réalisée déformée par rapport à la cible (par exemple *visa* [wiːsâː][4]) (Nacaskul, 1979). La labiodentale voisée n'est donc pas prononcée comme son homologue non voisé [f], alors que ce dernier existe bien en initiale en thaï. En comparaison entre plusieurs études antérieures sur l'acquisition de /v/ de

[4] L'accent représentent la transcription des tons.

l'anglais et du français chez les apprenants thaïlandais (Chunsuvimol & Ronnakiat, 2001 ; Ngammana, 2011 ; Le Corre, 2013 ; Sridhanyarat, 2017), nous constatons que la confusion entre /v/ et /w/ en initiale de syllabe se produit quand la tâche demandée se présente avec l'orthographe directement (lecture) ou indirectement (entretien autour d'une biographie ou récitation préparée) ; cette confusion est moins présente si la tâche est celle d'imitation où les apprenants avaient pour consigne « écouter et répéter » (sans accès à un support écrit). Par conséquent, le remplacement de /v/ en initiale de syllabe par [w] pourrait résulter de la confrontation à l'orthographe.

Nous avons donc effectué une étude pilote expérimentale visant à tester l'effet de l'orthographe sur la prononciation de /v/ chez les apprenants thaïlandais (Tusnyingyong & Tran, 2022). Deux types de tâche ont alors été mis en œuvre : un sans input écrit (imitation) et un avec (lecture oralisée). En contexte de la pandémie, cette étude pilote n'a pu être menée qu'auprès de quelques apprenants lycéens de niveaux A1-A2 à Bangkok. Les résultats basés sur un apprenant de niveau A2 montrent un taux de réussite très faible dans la prononciation de /v/ (0,54 %), quel que soit le type de stimuli utilisé (mot isolé ou phrase, mot courant ou pseudo-mot, mot emprunté ou pas), quel que soit le contexte vocalique et quelle que soit la position de /v/ dans le mot observé (initiale absolue ou médiane). La présence de l'orthographe augmente considérablement le taux de confusion entre /v/ et /w/ (95,24 % en lecture *vs* 65,48 % en imitation).

4.2.1 A l'origine de la confusion /v/ ~ /w/ chez les Thaïlandais

Ferlus (1997) montre un lien historique entre le graphème < ว > du thaï qui se prononce [w] et un graphème de l'ancien khmer ᩅ᩠ᩅ qui se prononce [v]. Ce lien qui date du XIII^{e} siècle influence considérablement les processus de romanisation du thaï ainsi que de translittération des mots d'emprunt en caractères thaïs. Plusieurs systèmes de romanisation du thaï (De La Loubère, 1691 ; Pallegoix, 1854 ; Vajiravudh, 1913) proposent de latiniser le graphème < ว > [w] en initiale de syllabe par le graphème < v > [v]. À l'inverse, lors de la translittération des mots d'emprunt en caractères thaïs depuis la Seconde Guerre mondiale, tous les emprunts de l'anglais à /v/ initiale s'écrivent avec < ว > et se prononcent par conséquent avec la labiovélaire [w] (ex. *vaccine* > วัคซีน [wáksi:n], *vitamin* > วิตามิน [wíta:min], TV > ทีวี [t^{h}i:wi:], etc.) (Nacaskul, 1979 ; Tri Endarto, 2015). Cette manière de romanisation (< ว > [w] → < v > [v]) et de translittération (< v > [v] → < ว > [w])

est actuellement en vigueur en Thaïlande et impacte négativement l'apprentissage du français chez les apprenants thaïlandais.

4.2.2 L'impact direct de la translittération franco-thaïlandaise

La translittération des mots du français en caractères thaïs n'est pas courante dans les établissements scolaires, mais elle se manifeste plus fréquemment dans les manuels que nous trouvons à la librairie. Plusieurs manuels d'auto-apprentissage proposent de manière automatique la forme de translittération en thaï pour que les apprenants ou les lecteurs puissent communiquer en français facilement sans prendre des cours de langue (cf. figure 4). L'apprentissage des langues se fait donc dans ce cas en toute autonomie. Les apprenants ne sont pas accompagnés par un enseignant qui pourrait surveiller la prononciation. Ils sont confrontés directement à la translittération des mots du français en caractères thaïs et les prononcent par conséquent de manière erronée.

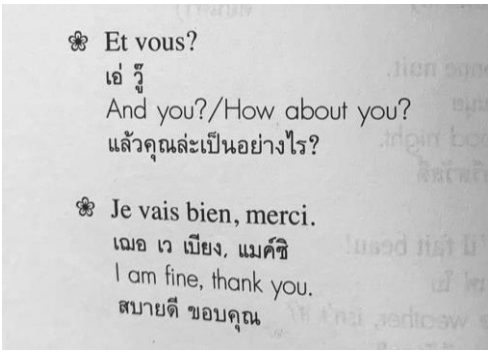

Figure 4. Phrases du français translitérées en caractères thaïs avec la traduction en thaï et en anglais (Chitkla, 2017, p. 16)

La figure ci-dessus est un exemple de la translittération des phrases du français en caractères thaïs d'un manuel d'auto-apprentissage acheté en librairie. La consonne < v > est translittérée avec le graphème < ว > qui représente le son /w/. Les mots *vous* [vu] et *vais* [vɛ] sont translittérés comme « วู๊ » [wú][5] et « เว » [we] respectivement. L'association erronée entre < v > et [w] impacte clairement ici de manière négative la prononciation chez les apprenants thaïlandais.

[5] L'accent ‹ ˊ › représente le ton haut en thaï.

4.2.3 L'impact indirect de la translittération anglo-thaïlandaise

La difficulté dans la prononciation de /v/ en position initiale de syllabe est constatée non seulement chez les apprenants thaïlandais du français, mais aussi de l'anglais (Chunsuvimol & Ronnakiat, 2001 ; Kanokpermpoon, 2007 ; Sridhanyarat, 2017). Il faut noter que l'anglais est la première langue étrangère apprise en Thaïlande dès l'école maternelle. L'alphabet de l'anglais constitue très souvent l'objet de la première leçon dans l'apprentissage de cette langue. Pour permettre aux enfants de prononcer et de mémoriser l'alphabet anglais, le nom de chaque lettre est translitteré en caractères thaïs, comme par exemple la lettre < A > translittérée comme « เอ » [ʔeː], la lettre < B > comme « บี » [biː]. Pour l'épellation de chaque lettre, les Thaïlandais ajoutent aussi le nom d'un animal ou d'un objet concret en anglais qui commence par la lettre concernée ainsi que la translittération de ce nom en caractères thaïs (cf. figure 5). Par exemple, la lettre < A > est accompagnée du mot *ant* (fourmi) qui est translittéré comme « แอ็นท » [ʔɛ́n], la lettre < B > accompagnée du mot *bird* (oiseau) et translittéré comme « เบิร์ด » [bɤ́ːt] (Le Corre, 2013). Les enfants sont amenés à mémoriser en répétant, pour chaque lettre de l'alphabet, une séquence qui contient la lettre, un animal/objet en anglais et sa traduction en thaï, par exemple « A ant มด » [ʔeːʔɛ́nmót], « B bird นก » [biːbɤ́ːtnók]. L'association erronée entre le graphème latin < v > et le graphème thaï < ว > /w/ est introduite dès le début de l'apprentissage de l'anglais à l'école maternelle. En effet, la lettre < V > est traitée de la même manière : translittérée comme « วี » [wiː] et accompagnée du mot *van* แวน [wɛːn] (minibus). La séquence à mémoriser pour la lettre < V > est donc « V van รถตู้ » [wiːwɛːnróttûː].

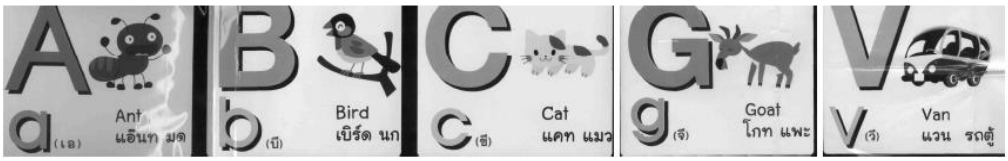

Figure 5. Introduction de l'alphabet de l'anglais dans l'école maternelle (Le Corre, 2013)

La confusion entre /v/ et /w/ ainsi imprégnée pourrait tout à fait être transférée de manière inconsciente dans une deuxième langue étrangère apprise ultérieurement chez les Thaïlandais, comme le cas du français étudié dans le cadre de cette étude. La façon de translittération de

l'anglais en thaï entraînerait de mauvaises habitudes de prononciation et impacterait par conséquent indirectement l'apprentissage du français chez les apprenants thaïlandais.

Pour récapituler, les apprenants thaïlandais rencontrent des difficultés dans la prononciation de /v/ du français (Debyser, 1969 ; Ngammana, 2011 ; Le Corre, 2013). Ils la remplacent par [w], même si la consonne /f/ (homologue sourd) existe en thaï. Nous avons montré dans notre recherche antérieure (Tusnyingyong & Tran, 2022) que (1) cette confusion entre /v/ et /w/ est liée aux processus de la romanisation du thaï et de la translittération des mots d'emprunt en caractères thaïs, et que (2) la présence de l'orthographe (tâche de lecture) augmente considérablement le taux de confusion entre /v/ et /w/.

Il a été montré que la présentation audiovisuelle facilite le traitement de la LE (cf. section 3.2.2). Dans le chapitre suivant, nous proposons d'étudier l'apport des informations visuelles dans la perception et dans la production de la consonne /v/ chez les apprenants thaïlandais.

Chapitre 5

Étude expérimentale sur l'apport des informations visuelles dans la perception et la production de la consonne /v/ chez les apprenants thaïlandais

5.1 Objectifs, questions de recherche et hypothèses

Plusieurs études démontrent l'apport des informations audiovisuelles dans le traitement de la parole en LE (Hazan et al., 2006 ; Wang et al., 2008 ; Wang et al., 2009 ; Burfin, 2015). En nous basant sur ces études antérieures, nous avons mené une étude expérimentale visant à tester si la présentation audiovisuelle pourrait amener les apprenants thaïlandais à mieux percevoir et à mieux prononcer la consonne /v/ en position initiale de syllabe.

Pour vérifier l'apport des informations audiovisuelles dans la perception et la production de /v/ chez les apprenants thaïlandais, deux modalités de test sont mises en œuvre dans notre étude : modalité auditive (A) et modalité audiovisuelle (AV). Nous avons choisi de tester les 3 consonnes /f v w/ dans cette étude parce que (1) /f/ et /v/ partagent le même lieu articulatoire labiodental à la seule différence du voisement et (2) la confusion problématique entre /v/ et /w/ est largement attestée chez les apprenants thaïlandais.

En perception, une tâche de discrimination est choisie dans laquelle les sujets sont exposés d'abord à un stimulus de référence (par exemple, [va]) et ensuite aux stimuli de test (par exemple, [fa], [va] et [wa]). Ils doivent juger si les stimuli de test sont identiques ou différents au stimulus de référence. En modalité AV, les stimuli sonores et les stimuli visuels (vidéos dans lesquelles un locuteur natif du français articule les stimuli en question) sont présentés simultanément, tandis qu'en modalité A, les stimuli sonores sont diffusés avec l'image statique d'un visage neutre du locuteur. Concernant la tâche de production, les sujets répètent les stimuli après les avoir entendus (modalité A) ou entendus et vus (modalité AV).

Trois hypothèses sont formulées pour cette expérience :

Hypothèse 1 *(en fonction de la **consonne** et de la **modalité** de test)* :

Consonne /f/ : Les études antérieures montrent que les locuteurs anglophones sont performants dans l'identification des consonnes natives /f v θ ð s z/ quelle que soit la modalité testée (A ou AV) (cf. table 3) (Wang et al., 2008, 2009 ; Burfin, 2015). Puisque la labiodentale sourde /f/ existe en thaï mais elle n'est jamais confondue avec son homologue sonore /v/ en initiale de syllabe (Chunsuvimol & Ronnakiat, 2001 ; Ngammana, 2011 ; Le Corre, 2013 ; Sridhanyarat, 2017 ; Tusnyingyong & Tran, 2022), le taux de réussite de /f/ en perception et en production devrait être élevé, et en égalité entre la modalité A et la modalité AV.

Consonne /v/ : Plusieurs études démontrent l'apport des informations audiovisuelles dans la perception des phonèmes non natifs (Hazan et al., 2006 ; Wang et al., 2008 ; Wang et al., 2009 ; Burfin, 2015). Puisque /v/ fait partie des consonnes stables ou visibles (cf. section 2.3), à la lecture labiale du geste articulatoire, cette consonne devrait être mieux perçue, mieux produite et moins confondue avec /w/ en modalité AV qu'en modalité A chez les apprenants.

Consonne /w/ : La labiovélaire /w/ existe en thaï et très souvent confondue avec /v/ en initiale de syllabe chez les apprenants thaïlandais (Chunsuvimol & Ronnakiat, 2001 ; Ngammana, 2011 ; Le Corre, 2013 ; Sridhanyarat, 2017 ; Tusnyingyong & Tran, 2022). En étant exposés aux informations audiovisuelles de [w], le visème faisant partie de la catégorie visible (Istria et al., 1982), les participants devraient bien percevoir et produire [w]. Le taux de réussite en modalité AV serait plus élevé que celui en modalité A à la fois en perception et en production.

Hypothèse 2 (*en fonction du **temps** de **réponse***) : Burfin (2015) montre qu'en perception, le temps de réponse en modalité AV est significativement plus court qu'en modalité A, que les stimuli soient natifs ou non natifs. La parole est multimodale et la présence des deux inputs auditif et visuel devrait donc accélérer le processus de traitement d'informations en perception, quelle que soit la consonne de test (native ou non native). Le temps de réponse en modalité AV serait moins long qu'en modalité A.

Hypothèse 3 (*le lien entre perception et production*) : Flege (1995) postule avec la septième hypothèse de son modèle SLM que la production d'un phone correspond à la représentation de la catégorie phonétique et que par conséquent, la majorité des erreurs de production résulte de la mauvaise perception (cf. section 1.2.2). Si les participants rencontrent des difficultés dans la perception de /v/ puisque cette consonne non native /v/ est assimilée à la catégorie de la consonne native /w/, une corrélation entre perception et production devrait être trouvée dans les scores de bonne réponse chez les mêmes participants, quelle que soit la modalité de test.

5.2 Méthodologie

5.2.1 Participants

Avant de sélectionner les participants pour l'expérience, nous avons effectué un pré-test avec 7 étudiants thaïlandais (5 hommes, 2 femmes) à Grenoble. Ce pré-test s'est déroulé dans

une pièce calme. Les étudiants doivent lire les mots et les phrases en français (cf. table 8) qui sont affichés sur le diaporama. Les mots de test (contenant /v/ en initiale absolue et en médiane) et ceux de contrôle servant de distracteur sont présentés en ordre aléatoire. Les phrases (sauf la première utilisée pour l'entraînement) se composent d'un ou plusieurs mots contenant /v/ en initiale de syllabe. Les productions (mots et phrases) de ces 7 étudiants ont été analysées sous *Praat*. Quatre étudiants (3 hommes et 1 femme), âgés de 23-36 ans, qui rencontrent le plus de difficulté dans la prononciation de /v/ ont été sélectionnés pour participer à l'expérience.

Table 8. Stimuli utilisés dans le pré-test

Mots		**Phrases**[6]
Mots de test	**Mots de contrôle**	
vase	plat	J'aime les pommes.
tra**v**ail	cou	**Voici** la photo des **viandes**
voiture	moi	Je **vais** à la mer.
dé**v**elopper	poulet	**Vous** allez à l'école.
venir	choisir	Ils **vont** à Paris.
ri**v**ière	sportive	**Écrivez**-les sur le tableau **vide**.
télé**v**ision	rêve	Posez des questions à **votre** professeur sur les plats **suivants**.
visa		**Voudriez-vous** les goûter ?
		Amandine et André **vont** manger dans un restaurant. Lisez le menu, écoutez leur **conversation avec** le **serveur** et prenez la commande.

Tous les quatre suivent leurs études (cours de français, master Génie industriel et master Informatique) à Grenoble. Ils sont en France depuis la rentrée universitaire en septembre 2021, donc moins d'un an au moment du test. Leur niveau du français est débutant (A1-A2). Ils parlent tous le thaï « standard » du Centre comme L1. Trois participants parlent aussi d'autres langues comme L1 (laotien, malais et une langue régionale du Sud de la Thaïlande), mais la consonne /v/ n'existe pas dans le répertoire phonologique de leurs L1s. L'anglais est la première LE des quatre participants. Trois participants sont aussi apprenants au niveau débutant du coréen, du mandarin et de l'arabe.

[6] Certaines phrases sont empruntées au manuel « Phonétique : 350 exercices avec 6 cassettes » (Abry & Chalaron, 1994)

Deux participants sont myopes et portent des lunettes ou des lentilles de contact durant l'expérience. Un des participants est atteint de daltonisme. Celui-ci passe l'expérience en modalité auditive seule. Aucun participant n'a de problème d'audition.

Les 4 sujets sont divisés en deux groupes : deux en modalité A (sujets nommés S2, S3) et deux en modalité AV (sujets nommés S1, S4).

5.2.2 Constitution du corpus

Nous avons choisi de travailler avec les pseudo-mots dans cette expérience. Notre étude antérieure (Tusnyingyong & Tran, 2022) montre que les corrélations phonosymboliques entre <v> et [w] établies avec les vrais mots sont bien présentes dans le lexique mental de l'apprenant. Puisque les pseudo-mots sont dépourvus de sens, nous voulons être sûrs que les apprenants n'utilisent que le son des pseudo-mots et ne laissent pas influencer par la sémantique ni par la familiarité lexicale comme pour le cas d'un vrai mot.

Puisque la confusion entre /v/ et /w/ est attestée dans deux positions (initiale absolue et médiane) dans le mot (Le Corre, 2013 ; Tusnyingyong & Tran, 2022), deux structures syllabiques CV et VCV ont été choisies. Les pseudo-mots sont donc des monosyllabes CV (ex. [fi]) et des dissyllabes VCV (ex. [ifi]).

Les voyelles utilisées pour cette étude sont celles qui existent à la fois en français et en thaï. Il s'agit des voyelles non arrondies antérieures /i e ɛ/ et centrale /a/. Les voyelles arrondies communes aux deux langues /u o ɔ/ dont l'articulation labiale (la protrusion des lèvres) pouvant donc impacter la visibilité des consonnes ne sont pas retenues pour le test.

Vingt-quatre stimuli de test ont été sélectionnés (12 monosyllabes, 12 dissyllabes) (cf. table 9). Vingt-quatre autres stimuli de contrôle (contenant les consonnes communes aux deux langues /p b t/) ont été insérées dans l'expérience pour éviter que les participants puissent deviner l'objet de la recherche. Nous avons aussi prévu douze stimuli avec d'autres consonnes /m s k/ pour la phase d'entrainement avant le test.

Table 9. Trois groupes de stimuli : 24 stimuli de test, 24 de contrôle et 12 d'entraînement

Stimuli de test		**Stimuli de contrôle**		**Stimuli d'entraînement**	
monosyllabes	**dissyllabes**	**monosyllabes**	**dissyllabes**	**monosyllabes**	**dissyllabes**
/fi/	/ifi/	/pi/	/ipi/	/ma/	/eme/
/fe/	/efe/	/pe/	/epe/	/mi/	/ɛmɛ/
/fɛ/	/ɛfɛ/	/pɛ/	/ɛpɛ/	/sa/	/ese/
/fa/	/afa/	/pa/	/apa/	/si/	/ɛsɛ/
/vi/	/ivi/	/bi/	/ibi/	/ka/	/eke/
/ve/	/eve/	/be/	/ebe/	/ki/	/ɛkɛ/
/vɛ/	/ɛvɛ/	/bɛ/	/ɛbɛ/		
/va/	/ava/	/ba/	/aba/		
/wi/	/iwi/	/ti/	/iti/		
/we/	/ewe/	/te/	/ete/		
/wɛ/	/ɛwɛ/	/tɛ/	/ɛtɛ/		
/wa/	/awa/	/ta/	/ata/		

Chaque item a été lu 5 fois par un locuteur français, 49 ans, originaire du Sud de la France. L'enregistrement sonore ainsi qu'audiovisuel s'est déroulé dans la chambre anéchoïque du Pôle Parole et Cognition du laboratoire GIPSA-lab de l'Université Grenoble Alpes. Les vidéos ont été enregistrées avec la caméra Sony HDR-XR500E. L'enregistrement sonore s'est fait simultanément avec l'enregistreur Marantz PMD 670, le micro AKG C1000S. Les stimuli ont été numérisés à 44,1 kHz, 16 bit sous format .wav.

5.2.2.1 Préparation des stimuli de l'expérience

Après l'enregistrement, nous avons choisi, pour chaque item, un ou deux meilleurs exemplaires où la vidéo commence et se termine avec la bouche fermée et la direction du regard centrée. Par exemple, pour l'item /awa/, nous avons choisi la première et la troisième prononciation (awa_1 et awa_3). Les deux entrées audio (enregistré avec le PMD 670) et vidéo (enregistré avec la caméra Sony) ont été synchronisées, ce qui a permis de remplacer le son de la vidéo par celui qui était enregistré séparément afin d'avoir une meilleure qualité des stimuli audiovisuels utilisés pour le test.

Au final, nous avons 78 stimuli audiovisuels fabriqués pour la modalité AV. Le même nombre de stimuli ont été créés pour la modalité A où chaque stimulus sonore a été associé à une image statique (visage neutre) du locuteur.

Les interfaces de test (perception et production) ont été fabriqués avec Presentation® (version 23.0) (NeuroBehavioral Systems, Berkeley, California). Ce logiciel de conception d'expérimentations, très couramment utilisé en neurosciences, fonctionne sur les PC Windows et administre des stimuli auditifs, visuels et multimodaux avec une précision temporelle de quelques millisecondes.

5.2.2.2 Corpus pour la tâche de perception

Dans la tâche de perception, les sujets sont exposés aux 4 séquences d'entraînement, 72 séquences de test (24 séquences × 3 répétions) et 24 séquences de contrôle (cf. annexe 1). Chaque séquence contient : (1) un item de référence (par exemple, /vi/) qui est présenté 6 fois et (2) trois items de test dont un est identique à l'item de référence (/vi/) et deux sont différents (/fi/ et /wi/) (cf. tables 10 et 11). Il est à noter qu'à l'intérieur de chaque séquence de test, les voyelles de chaque item sont inchangées. Pour que la position des items de test soit diversifiée, nous avons établi trois séquences pour chaque item de référence où l'ordre de présentation des items de test est varié (cf. tables 10 et 11). Ainsi, dans un bloc de trois séquences pour un item de référence (par ex. [vi]), chaque item de test (par ex. [vi], [wi], [fi]) se trouve une fois tout de suite après la présentation de l'item de référence [vi]. Par conséquent, trois séquences ont été établies pour chacun des 24 stimuli de test (cf. table 9). Nous avons au total 72 séquences de test.

Table 10. Exemples de séquences de test pour la tâche de perception.

Séquences de test			
Item de référence	**Items de test**		
vi	vi	fi	wi
vi	wi	vi	fi
vi	fi	wi	vi
awa	awa	ava	afa
awa	afa	awa	ava
awa	ava	afa	awa

Table 11. Exemples de séquences de contrôle pour la tâche de perception.

Séquences de contrôle			
Items de référence	**Items de test**		
pe	pe	te	be
pe	be	pe	te
pe	te	be	pe
ipi	ipi	ibi	iti
ipi	iti	ipi	ibi
ipi	ibi	iti	ipi

L'ensemble des séquences de test et de contrôle a été mis en ordre aléatoire en respectant le modèle 3-1 : toutes les 3 séquences de test sont suivies d'une séquence de contrôle. Les exemples sont présentés dans la table 12.

Table 12. Exemples de séquences de test et de contrôle mises en ordre aléatoire pour la tâche de perception.

Types de séquence	**Item de référence**	**Items de test**		
Séquences de test	/ava/	/afa/	/ava/	/awa/
	/fi/	/fi/	/vi/	/wi/
	/vɛ/	/fɛ/	/wɛ/	/vɛ/
Séquence de contrôle	/pa/	/ta/	/ba/	/pa/
Séquences de test	/fi/	/vi/	/wi/	/fi/
	/ewe/	/eve/	/efe/	/ewe/
	/va/	/wa/	/fa/	/wa/
Séquence de contrôle	/ɛpɛ/	/ɛpɛ/	/ɛbɛ/	/ɛtɛ/

En raison de la variabilité des items de référence qui sont présentés 6 fois, deux exemplaires de stimuli ont été mis l'un après l'autre. Par exemple, pour la présentation de l'item de référence /fi/ 6 fois, deux exemplaires fi_2 et fi_4 (les meilleurs exemplaires choisis après l'enregistrement du corpus) sont mis l'un après l'autre : fi_2, fi_4, fi_2, fi_4, fi_2, fi_4.

5.2.2.2 Corpus pour la tâche de production

Chaque 24 stimulus de test (cf. table 9) est présenté 5 fois en ordre aléatoire. Un stimulus de contrôle est inséré après tous les quatre stimuli de test. Un exemple de l'ordre de présentation

des stimuli est présenté à la table 13. Au total, nous avons 4 stimuli d'entraînement (/ma/, /eke/, /si/ et /ɛmɛ/), 120 stimuli de test et 30 stimuli de contrôle (cf. annexe 2).

Table 13. Exemples de stimuli de test et de contrôle mis en ordre aléatoire pour la tâche de production.

Stimuli de test	/awa/
	/ve/
	/ɛfɛ/
	/vi/
Stimuli de contrôle	/te/
Stimuli de test	/afa/
	/wɛ/
	/vi/
	/fe/
Stimuli de contrôle	/apa/

5.2.3 Déroulement de l'expérience

Rappelons que pour chaque modalité de test (A et AV), nous avons deux apprenants qui participent à deux sessions de test (une de perception et une de production). Chacun des participants est testé sur une seule modalité (soit auditive seule, soit audiovisuelle). L'ordre des sessions de test est contrebalancé entre participants au sein d'une même modalité (perception puis production et vice versa, production puis perception).

L'expérience s'est déroulée dans la chambre anéchoïque du Pôle Parole et Cognition de GIPSA-lab, Université Grenoble Alpes. Les stimuli sont présentés à l'aide du logiciel *Présentation*. Les sujets sont assis à 0,5 mètre de l'écran. Les sons sont diffusés par le haut-parleur placé derrière l'écran. Le niveau de son est adapté selon le besoin de chaque sujet. Avant de passer les tests, tous les participants doivent remplir un questionnaire sociolinguistique (cf. annexe 3) à partir duquel les informations récoltées sont présentées en section 5.2.1.

5.2.3.1 Test de perception

Les sujets sont exposés dans un premier temps à un stimulus de référence pour 6 répétitions. Ensuite, le symbole + apparaît sur l'écran pour informer les sujets que la présentation d'un stimulus de référence est terminée. Par la suite, trois stimuli de test sont présentés. Les sujets doivent juger si chaque stimulus de test est pareil ou différent du stimulus de référence en appuyant sur le

clavier le bouton ← pour la réponse « pareil » et le bouton → pour la réponse « différent ». Pour passer à la séquence suivante, les sujets appuient sur le bouton « espace ». Une phase d'entrainement à l'interface et au type de stimuli leur avait préalablement proposée, avec des stimuli différents du corpus de test.

Rappelons que, en modalité AV, les sujets ont accès à la vidéo à l'écran où le locuteur natif prononce les stimuli. En modalité A, les stimuli sonores sont présentés mais seule une image statique du visage du locuteur apparait à l'écran. Au total, il y a 4 séquences d'entraînement et 96 séquences dans la phase de test. Cinq pauses sont insérées toutes les 16 séquences. La durée de la tâche de perception est environ 30 minutes. La représentation schématique de la tâche de perception est présentée à la figure 6.

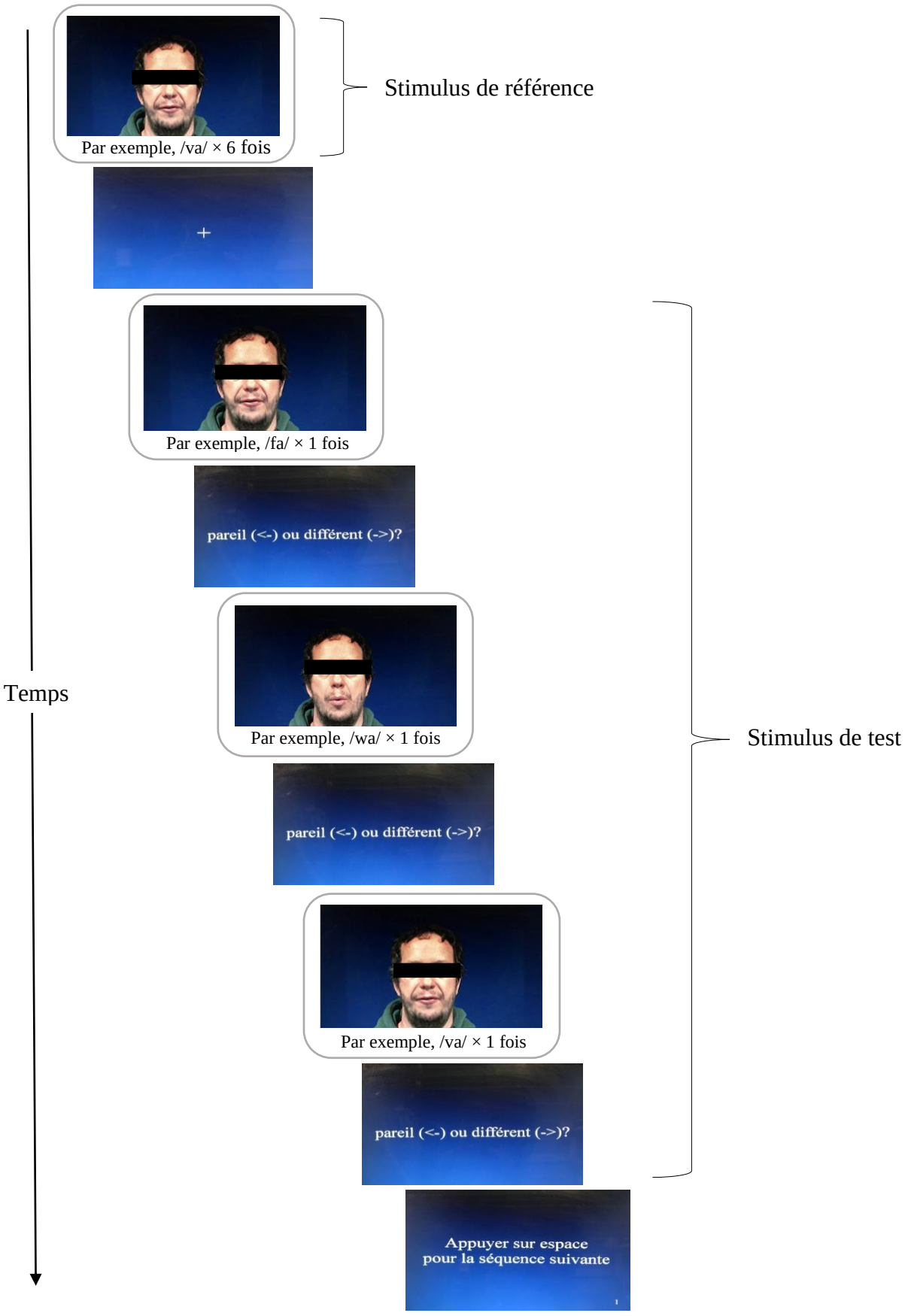

Figure 6. Représentation schématique d'une séquence de test dans la tâche de perception en modalité AV

5.2.3.2 Test de production

Pour la tâche de production, les sujets ont pour consigne de répéter les sons perçus. Ils sont exposés aux vidéos en modalité AV ou seulement aux sons couplés à une image statique du visage neutre du locuteur (bouche fermée) en modalité A.

Chaque stimulus est présenté une fois. Après la présentation de chaque stimulus, le mot « Répétez » apparaît sur l'écran pour que les sujets répètent les sons perçus. L'expérimentateur contrôle l'expérience en cliquant sur le bouton pour passer au stimulus suivant. L'enregistrement a été réalisé avec l'enregistreur Marantz PMD 670, le micro AKG C1000S à directivité cardioïde. Rappelons qu'il y a 4 stimuli d'entraînement et 150 stimuli dans la phase de test. Les pauses sont insérées tous les 25 stimuli. La durée de la tâche de production est environ 10 minutes. La représentation schématique de la tâche de production est présentée à la figure 7.

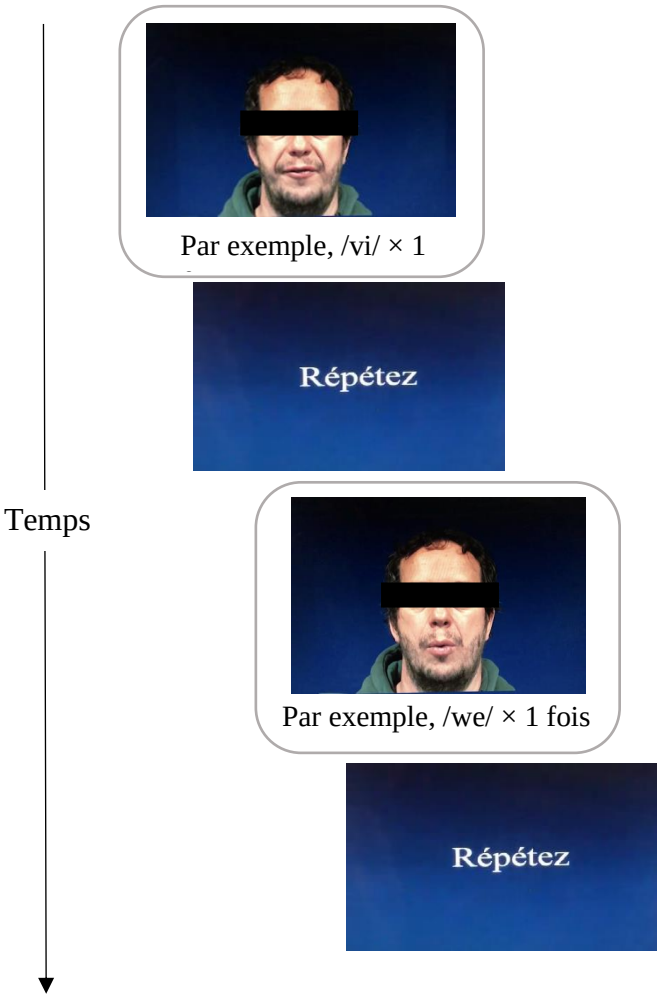

Figure 7. Représentation schématique de la tâche de production en modalité AV

5.2.4 Traitement des données

5.2.4.1 Perception

À partir des fichiers de résultats obtenus par le logiciel *Presentation*®, l'extraction des données a été faite à l'aide du logiciel *Matlab* (cf. figure 8). Des analyses quantitatives et statistiques ont été menées par la suite avec *Microsoft Excel* et *R*.

Sujet	trial	syllabe_reference			syllabe_test	reponse_attendue	réponse_obtenue	is_correct
1	1	ma_4	ka_3	64	64	1	3.858	
1	1	ma_4	sa_2	64	64	1	2.454	
1	1	ma_4	ma_4	32	32	1	2.690	
1	2	si_3	mi_4	64	64	1	2.680	
1	2	si_3	si_3	32	32	1	2.723	
1	2	si_3	ki_4	64	64	1	2.540	
1	3	éké_3	éké_3	32	32	1	2.960	
1	3	éké_3	émé_3	64	64	1	2.816	
1	3	éké_3	ésé_1	64	64	1	2.253	
1	4	èsè_1	èkè_3	64	64	1	2.845	
1	4	èsè_1	èmè_2	64	64	1	2.233	
1	4	èsè_1	èsè_1	32	32	1	2.734	

Figure 8. Exemple de l'extraction des données.

5.2.4.2 Production

Les données acoustiques ont été traitées à l'aide des deux logiciels Phon (Rose et al., 2006) et Praat (Boersma & Weenick, 2020). Sous Phon, les données ont été segmentées, et chaque segment a été annoté manuellement sur la tire « Orthography » et la tire « IPA Target » qui correspondent respectivement à l'orthographe et à la transcription phonétique du mot cible (cf. figure 9).

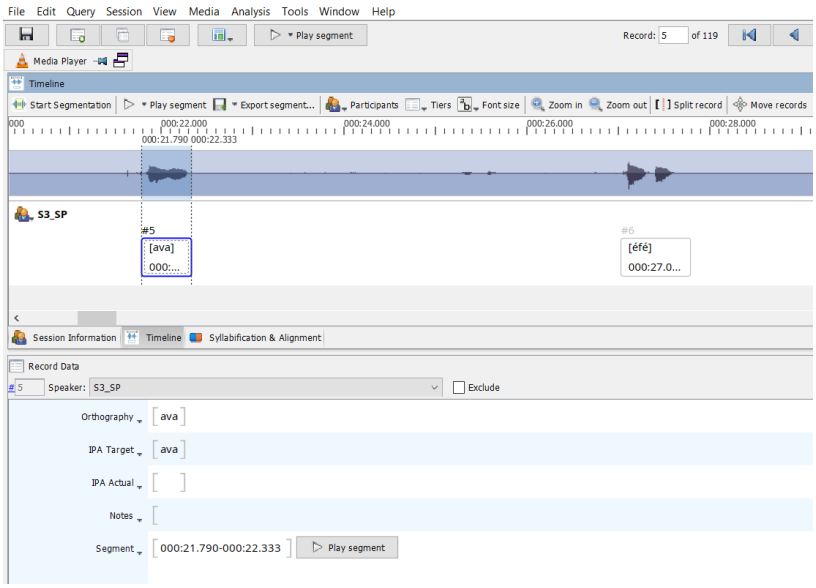

Figure 9. Affichage de Phon avec les données segmentées, annotées et transcrites en API.

Avant de compléter la tire « IPA Actual » de *Phon* qui correspond à la transcription phonétique de la prononciation réelle, les analyses acoustiques des productions des apprenants ont été faites au préalable sous *Praat* (cf. figure 10), en nous basant sur les descriptions acoustiques des consonnes de Calliope et al. (1989) et de Vaissière (2006).

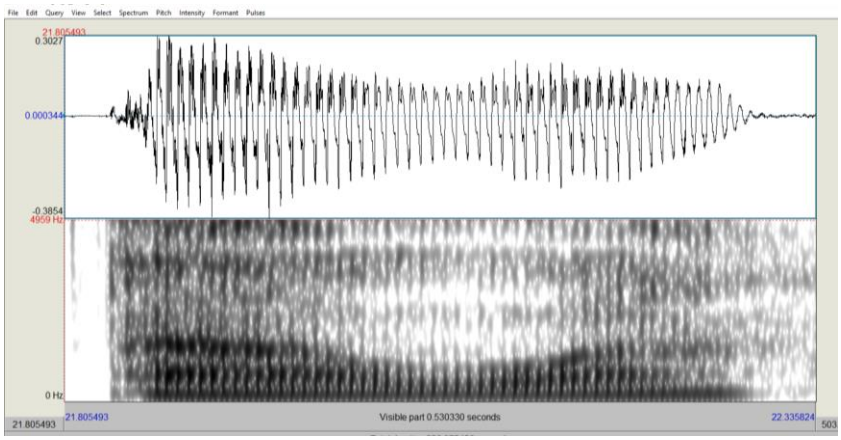

Figure 10. Affichage de Praat : l'analyse acoustique de la prononciation réelle du stimulus [ava] en modalité auditive (prononcé comme [awa]).

Les segments de la fricative labiodentale [v] sont identifiés d'après la présence du bruit de friction durant la tenue sur le signal acoustique. Des pics diffus de faible intensité sont visibles vers 1,5 kHz ainsi qu'une barre de voisement dans les très basses fréquences doit également être déterminés sur le spectrogramme. La prononciation correcte de [v] est par conséquent attestée par la présence du bruit de friction et la barre de voisement. En cas d'une réalisation de [w] à la place de [v], les segments de cette approximante labiovélaire sont identifiés par l'absence du bruit de friction et par la présence d'une structure formantique de faible amplitude en variation constante en fonction du contexte vocalique avoisinant. En cas de doute, nous avons demandé aux deux locutrices francophones natives d'écouter les sons et de juger quelle consonne elles entendent.

Une fois que la transcription de la prononciation réelle (IPA Actual) est terminée, l'étape de syllabification est exigée pour l'alignement de la structure syllabique entre la prononciation cible (Target Syllable) et la prononciation réelle (Actual Syllable) (cf. figure 10). Cette étape s'est faite de manière semi-automatique avec notre vérification.

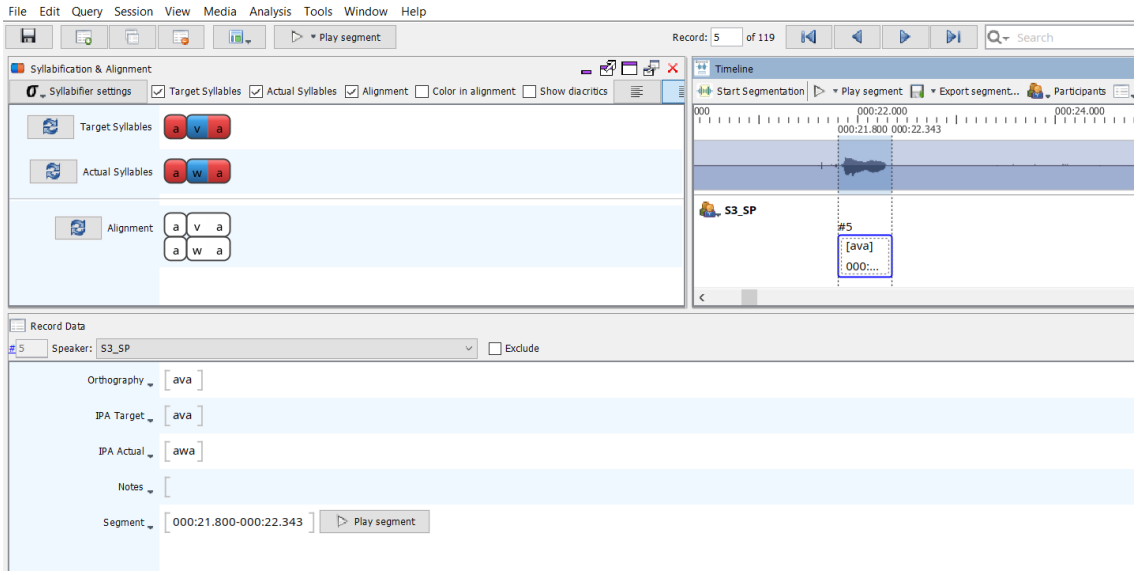

Figure 11. Affichage de la fenêtre session Phon avec, comme exemple ici, le traitement du stimulus [ava] : les tiers « Orthography » (entrée orthographique) et « IPA Target » (transcription cible [ava]), « IPA Actual » (transcription de la prononciation réelle de l'apprenant [awa]) et en haut à gauche, la syllabification et l'alignement entre la syllabe cible (Target Syllable) et la syllabe réalisée (Actual Syllable).

Lorsque l'étape de syllabification a été terminée, nous avons extrait les résultats sous le format « .xls » qui sont ensuite analysés quantitativement sous *Excel*. Les scores de bonne réponses (en perception et en prononciation) ont été codés 1 si réponse correcte et 0 si ce n'est pas le cas. L'ensemble des résultats a été par la suite traité statistiquement avec *R*.

5.2.4.3 Analyses statistiques

Nous souhaitons étudier l'influence de 4 variables explicatives et de leurs interactions sur 2 variables de réponse « scores » et « temps de réponse ». Les 4 variables explicatives sont les modalités de test (2 modalités : A et AV), les consonnes de référence (3 modalités : f, v, w), les contextes vocaliques (4 modalités : a, e, ɛ, i) et les structures syllabiques (2 modalités : CV et VCV).

La variable réponse « scores » dans les tâches de perception et de production suit une distribution binomiale (soit 1, soit 0). Compte tenu qu'un participant est sollicité à plusieurs reprises, nous introduisons la variable « sujets » comme effet aléatoire dans le modèle. Nous avons donc choisi d'utiliser une régression linéaire mixte avec distribution binomiale. La méthode a été réalisée à l'aide de la fonction *glmer* du package *lme4* du logiciel *R*.

Quant à la variable « temps de réponse » dans la tâche de perception, notre protocole ayant permis de recueillir plusieurs valeurs de cette variable pour un même sujet, il ne nous garantit donc pas l'indépendance des observations. Notre choix s'est porté sur le modèle linéaire à effets mixtes (avec la fonction *lme* du package *nlme*) pour permettre de respecter l'hypothèse selon laquelle les résidus suivent une loi normale (condition d'application des modèles mixtes), nous avons choisi de transformer la variable réponse en son logarithme.

Pour analyser la différence entre les deux modalités (A et AV) à l'intérieur de chaque modalité de consonnes, voyelles, structures syllabiques, nous appliquons la méthode de Hothorn, Bertz et Westfall (2008) qui permet de réaliser les comparaisons multiples de moyennes avec le modèle mixte. La méthode a été appliquée aux données avec la fonction *emmeans* du package *emmeans* du logiciel R.

5.3 Résultats

5.3.1 Tâche de perception

5.3.1.1 Taux de réussite

Rappelons qu'il y a 72 séquences de test pour la tâche de perception et que chaque séquence comporte 3 stimuli à distinguer du stimulus de référence. Par conséquent, chaque participant a dû donner au total 216 réponses.

Sur l'ensemble des 864 réponses des quatre participants, le score de réponses correctes en modalité AV (98,84 %) est significativement plus élevé qu'en modalité A (92,82 %) (z = -2.388, $p < 0.0169$). Le taux de non-réussite est de 4,17 % avec 36 réponses incorrectes parmi lesquelles 31 réponses sont en modalité A et seules 5 réponses en modalité AV.

a. En fonction de la consonne de référence

Le score de réponses correctes en fonction de la modalité de test (A ou AV) et de la consonne de référence est présenté à la figure 12. Les résultats montrent que les participants réussissent bien lorsqu'il s'agit de la consonne /f/ comme référence, quelle que soit la modalité de test (taux de réussite de 98,61% en moyenne).

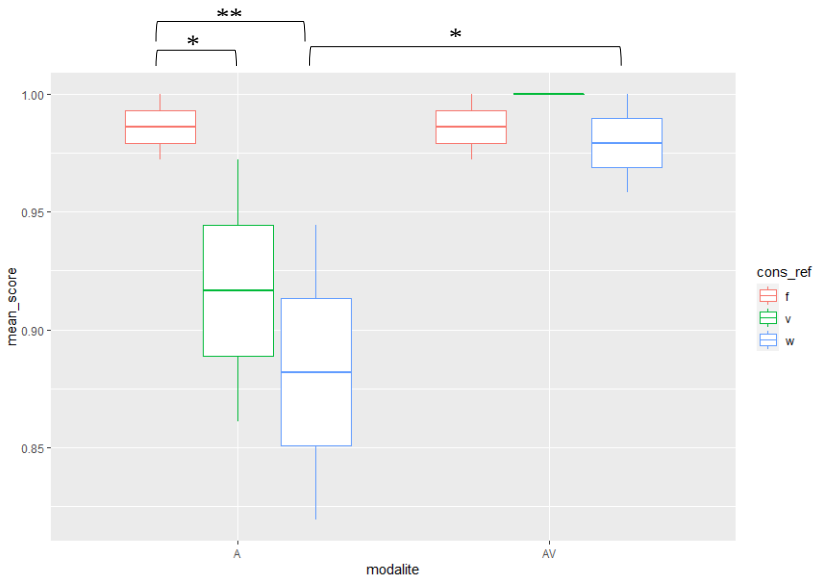

Figure 12. Pourcentage des réponses correctes dans la tâche de perception en fonction de la modalité (A *vs* AV) et de la consonne de référence (/f v w/)

Quant aux consonnes /v/ et /w/, les réponses correctes sont plus élevées en modalité AV qu'en modalité A. Les participants perçoivent correctement /v/ beaucoup plus en modalité AV (100 %) qu'en modalité A (91,67 %). Cette différence, qui aurait pu être significative, mais ne l'a pas été attestée statistiquement par le modèle, uniquement en raison du fait qu'il n'y a pas de variabilité dans les réponses de /v/ en modalité AV (100 %). Lorsque /w/ est de référence, la différence de score de bonnes réponses en fonction de la modalité de test est bien significative ($z = -2.073$, $p < 0.0382$). La labiovélaire /w/ est moins bien perçue en modalité A (88,19 %) qu'en modalité AV (97,92 %).

En comparaison des réponses correctes en fonction de la consonne à l'intérieur de la modalité, les résultats montrent qu'il n'y a pas de différence significative dans la perception des 3 consonnes /f v w/ en modalité AV. Quant à la modalité A, /f/ est significativement mieux perçu (98,61 %) que /v/ (91,67 %) ($z = 2.451$, $p < 0.0379$) et que /w/ (88,19 %) ($z = 3.027$, $p < 0.0070$). Il n'y a pas de différence significative entre les scores de bonnes réponses de /v/ et de /w/.

b. En fonction du contexte vocalique

Globalement, le taux de réponses correctes est observé plus élevé en modalité AV qu'en modalité A pour chacune des 4 voyelles (cf. figure 13) : 98,15 % *vs* 95,37 % en contextes /a/ ; 99,07 % *vs* 96,30 % en contexte /e/ ; 98,15 % % *vs* 89,81 % en contextes /ɛ/ ; 100 % *vs* 89,81 % en contexte /i/.

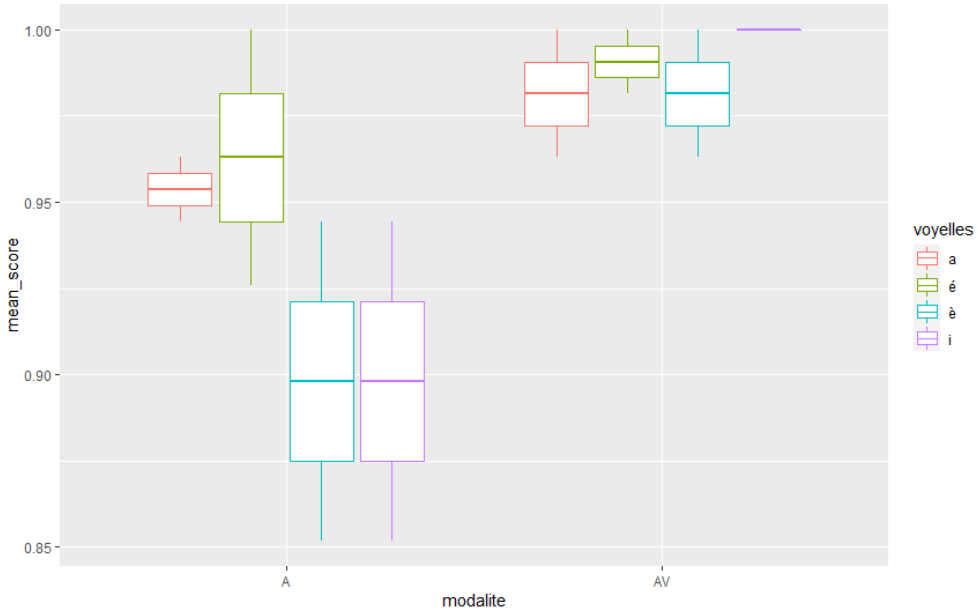

Figure 13. Pourcentage des réponses correctes dans la tâche de perception en fonction de la modalité (*A* vs AV) et de la voyelle (/i e ɛ a/). « é » et « è » dans la légende correspondent aux voyelles /e/ et /ɛ/ respectivement.

Des comparaisons entre les deux modalités pour chacun de contexte vocalique n'ont pas révélé des différences significatives. Concernant la voyelle /i/ où le graphique montre une différence entre deux modalités A vs AV, les résultats ne sont pas statistiquement significatifs en raison de l'absence de variabilité des réponses (100 %) dans ce contexte vocalique en modalité AV. Quand nous regardons à l'intérieur de chaque modalité A et AV, aucune différence significative de scores n'est observée entre les 4 voyelles.

c. En fonction de la structure syllabique

Quel que soit la structure syllabique (CV ou VCV), le taux de réussite en modalité AV est plus élevé qu'en modalité A (cf. figure 14). En modalité AV, les participants répondent correctement 100 % sous la structure CV. Le score est légèrement moins élevé (97,69 %) en structure VCV. Il n'y pas de différence significative de scores entre les deux structures.

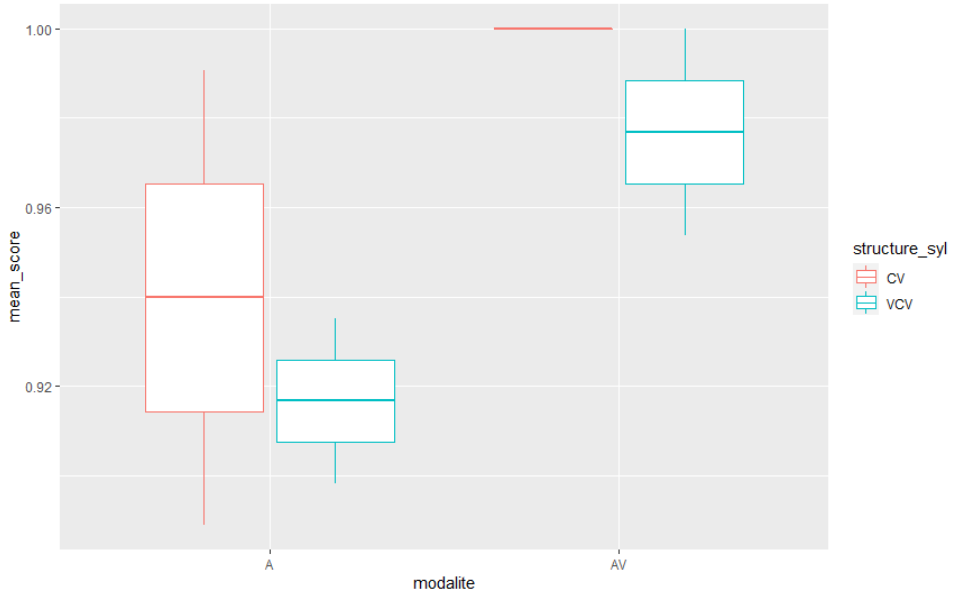

Figure 14. Pourcentage des réponses correctes dans la tâche de perception en fonction de la modalité (A *vs* AV) et de la structure syllabique (CV et VCV)

Comme dans la modalité AV, le score de réponses correctes en CV (93,98 %) est légèrement plus élevé que celui en VCV (91,67 %) en modalité A mais la différence n'est pas significative ($z = 0.944$, $p < 0.3450$).

Concernant la comparaison entre les deux modalités, aucune différence significative de réponses correctes n'est attestée en fonction de structures syllabiques entre les deux modalités.

5.3.1.2 Réponses incorrectes

Rappelons que les réponses incorrectes sont beaucoup plus en modalité A (31) qu'en modalité AV (5) (cf. tables 14 et 15). En modalité A, les réponses incorrectes sont trouvées majoritairement lorsque les participants doivent distinguer /v/ de /w/ et vice versa (cf. table 14) : 10 réponses incorrectes lorsque la consonne de référence et la consonne de test sont /v/ et /w/ respectivement et 14 réponses incorrectes lorsque la consonne de référence et la consonne de test sont /w/ et /v/ respectivement.

Table 14. Réponses incorrectes en modalité A en fonction de la consonne de référence.

Modalité auditive			
Consonne de test / **Consonne de référence**	**/f/**	**/v/**	**/w/**
/f/	1	1	-
/v/	1	1	**10**
/w/	-	**14**	3

Table 15. Réponses incorrectes en modalité AV en fonction de la consonne de référence et de la consonne de test

Modalité audiovisuelle			
Consonne de test / **Consonne de référence**	**/f/**	**/v/**	**/w/**
/f/	1	1	-
/v/	-	-	-
/w/	-	3	-

En modalité AV, les 5 réponses incorrectes apparaissent uniquement en contexte VCV. Les 31 réponses incorrectes en modalité A sont trouvées dans les deux contextes : un peu plus lorsque les consonnes sont en intervocalique VCV (18) qu'en initiale CV (13).

En fonction de la voyelle, les 5 réponses incorrectes en modalité AV sont trouvées dans trois contextes vocaliques : 2 avec /a/, 1 avec /e/ et 2 avec /ɛ/. En modalité A, les 31 réponses incorrectes apparaissent dans les quatre contextes vocaliques : 11 en contextes /i/ et /ɛ/, 4 en contexte /e/ et 5 en contexte /a/.

5.3.1.3 Temps de réponse

Les résultats montrent que le temps moyen de réponse en modalité A (986 ms) est significativement plus important que celui en modalité AV (482 ms) ($t = 4.448$, $p < 0.0470$). Les participants testés en modalité A prennent plus de temps pour répondre que ceux qui sont testés en modalité AV.

a. En fonction de la consonne de référence

La figure 15 présente le temps moyen de réponse en fonction de la modalité et de la consonne de référence. Le temps moyen de réponse pour chaque consonne de référence (/f v w/) en modalité AV est significativement moins important que celui en modalité A (t = 5.187, p < 0.0352 pour /f/, t = 5.931, p < 0.0273 pour /v/ et t = 5.582, p < 0.0306 pour /w/).

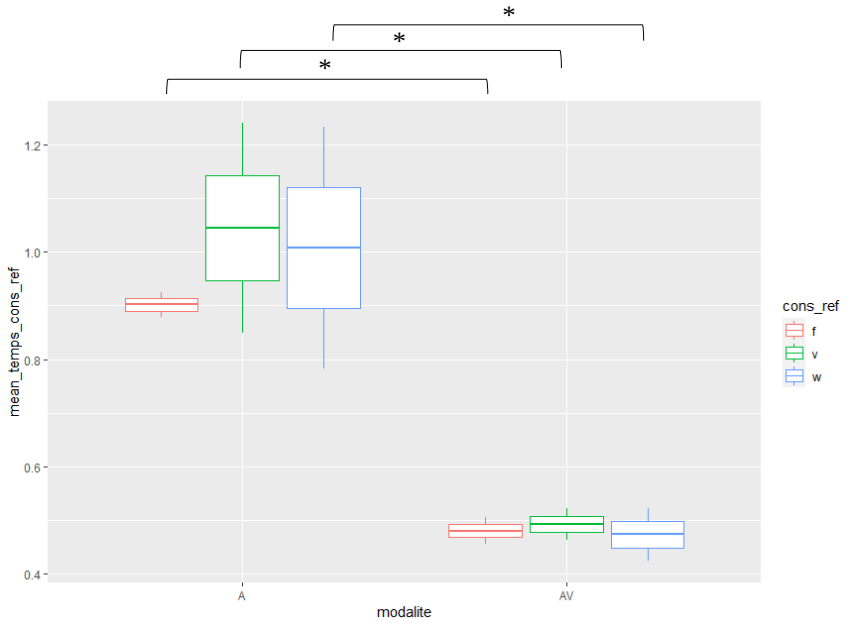

Figure 15. Temps moyen de réponse dans la tâche de perception en fonction de la modalité (A *vs* AV) et de la consonne de référence (/f v w/)

En modalité A, le temps de réponse en moyenne est de 0,92 seconde pour la consonne /f/, de 1,05 seconde pour la consonne /v/ et de 1,01 seconde pour la consonne /w/. Il n'y a pas de différence significative de temps de réponse entre ces trois consonnes à l'intérieur de la modalité A. Par ailleurs, nous pouvons remarquer que la distribution des données des consonnes /v/ et /w/ est bien plus grande que celle de la consonne /f/ en modalité A. Il y a donc beaucoup plus de variabilité dans le temps de réponse pour les consonnes /v/ et /w/.

En modalité AV, le temps moyen de réponse est de 0,480 seconde pour la consonne /f/, de 0,493 seconde pour la consonne /v/ et de 0,473 seconde pour la consonne /w/. Aucune différence significative n'est observée dans le temps de réponse entre les trois consonnes en modalité AV.

b. En fonction du contexte vocalique

Les résultats montrent que le temps de réponse est significativement plus court en modalité AV qu'en modalité A pour chaque contexte vocalique : pour /i/ (t = 6.215, p < 0.0249) ; pour /e/ (t = 6.101, p < 0.0258) ; pour /ɛ/ (t = 7.067, p < 0.0194) et pour /a/ (t = 5.706 p < 0.0294) (cf. figure 16). Par ailleurs, la variabilité dans le temps de réponse est bien plus marquée en modalité A qu'en AV, et ce quel que soit le contexte vocalique.

En comparaison des résultats à l'intérieur de chaque modalité, la différence du temps moyen de réponse entre les quatre voyelles en modalité AV n'est pas significative. En modalité A, seuls les temps de réponse entre /e/-/ɛ/ et entre /ɛ/-/a/ sont significativement différents (t = -2.658, p < 0.0399 pour /e/-/ɛ/, t = -2.697, p < 0.0358 pour /ɛ/-/a/).

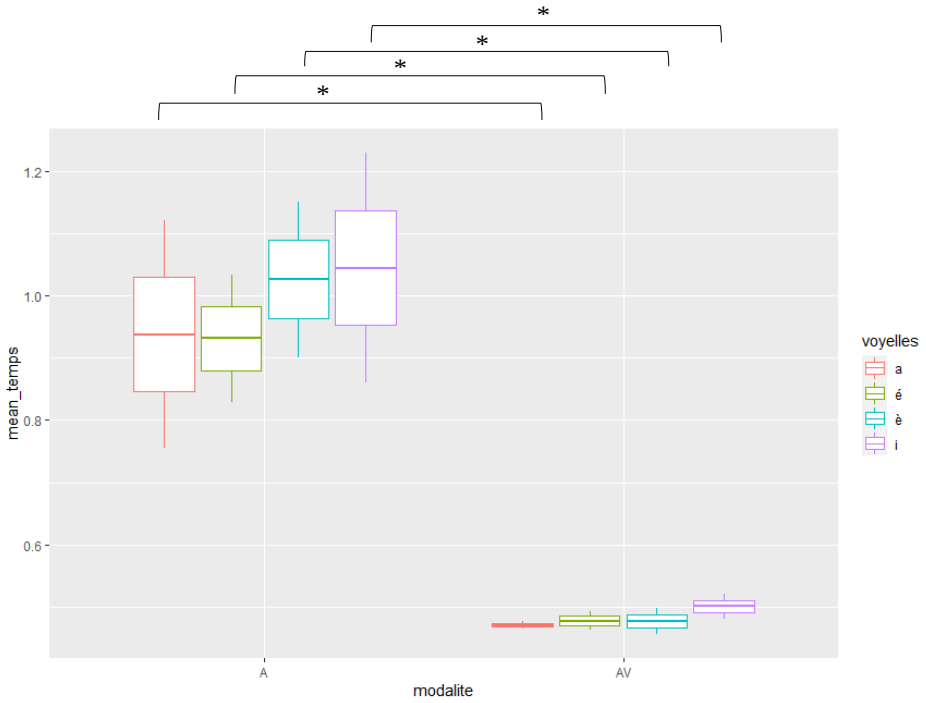

Figure 16. Temps moyen de réponse dans la tâche de perception en fonction de la modalité (A *vs* AV) et de la voyelle (/i e ɛ a/)

c. En fonction de la structure syllabique

Quelle que soit la structure syllabique, le temps de réponse en modalité AV est largement plus court que celui en modalité A (t = 7.199, p < 0.0188 pour CV et t = 6.188, p < 0.0251 pour VCV) (cf. figure 17). Par ailleurs, la variabilité dans le temps de réponse pour chaque structure syllabique en modalité AV est très faible par rapport à celle en modalité A.

En modalité A, le temps de réponse en moyenne pour la structure CV est significativement plus long (1,02 s) que celui en VCV (0,958 s) (t = 2.348, p < 0.0191). Concernant la modalité AV, aucune différence significative est observée pour le temps de réponse entre la structure CV (475 ms) et la structure VCV (489 ms).

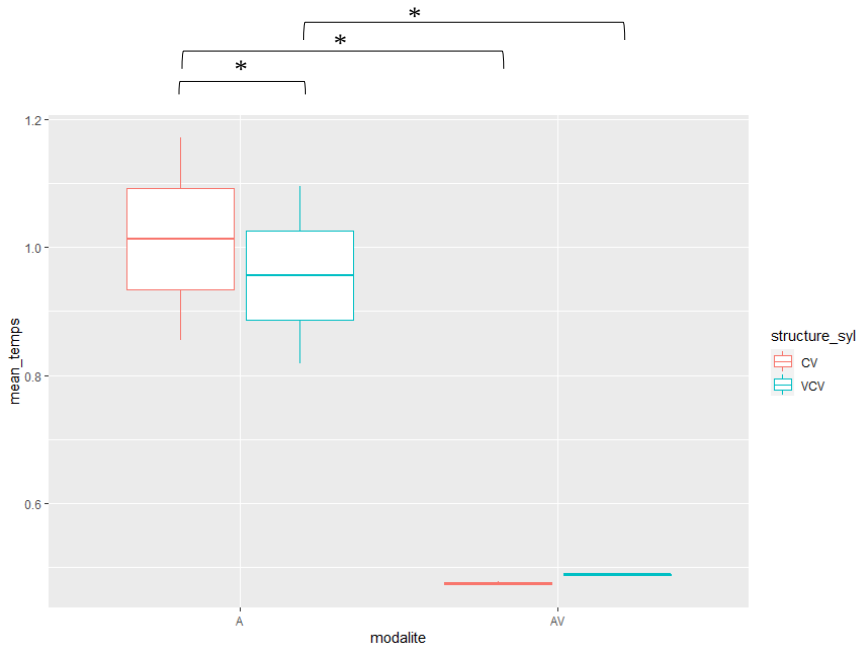

Figure 17. Temps de réponse en moyenne en perception en fonction de la modalité (A *vs* AV) et de la structure syllabique (CV ou VCV)

5.3.2 Tâche de production

Rappelons qu'il y a 120 items de test pour chaque participant en production, ce qui devrait faire 480 items à analyser pour l'ensemble des participants. Cependant, un problème technique s'est produit durant l'expérience. Il n'y avait pas de son pour un item /iwi/ lors de l'expérience d'un participant en modalité A. Cet item ne fait donc partie de l'analyse. Par conséquent, il y a au total 479 items à analyser.

5.3.2.1 Taux de réussite

Sur l'ensemble des 479 items analysés, les apprenants prononcent correctement 86,01 % des cas (soit 412 items bien produits) avec des taux de réussite significativement plus élevés en modalité AV (94,58 %) qu'en modalité A (77,41 %) (z = -2.595, p < 0.0095). Ils rencontrent bien plus de difficultés en modalité A (54 items mal prononcés) qu'en modalité AV (13 items).

a. En fonction de la consonne

Le score de prononciation correcte en fonction de la modalité et de la consonne est présenté dans la figure 18. Les résultats montrent que de manière générale, les apprenants réussissent mieux en modalité AV qu'en modalité A. Pour la consonne /v/, les apprenants testés en modalité AV ont mieux réussi à la prononcer (88,75 %) que ceux testés en modalité A (57,50 %). Cette différence dans le score de prononciation correcte de /v/ est significative (z = -2.454, p < 0.0141). Aucune différence significative n'est détectée entre les deux modalités A *vs* AV pour /f/ (91,25 % *vs* 100 % ; z = -92.587, p < 131.82) ni pour /w/ (83,54 % *vs* 95 % ; z = 1.716, p < 4.28).

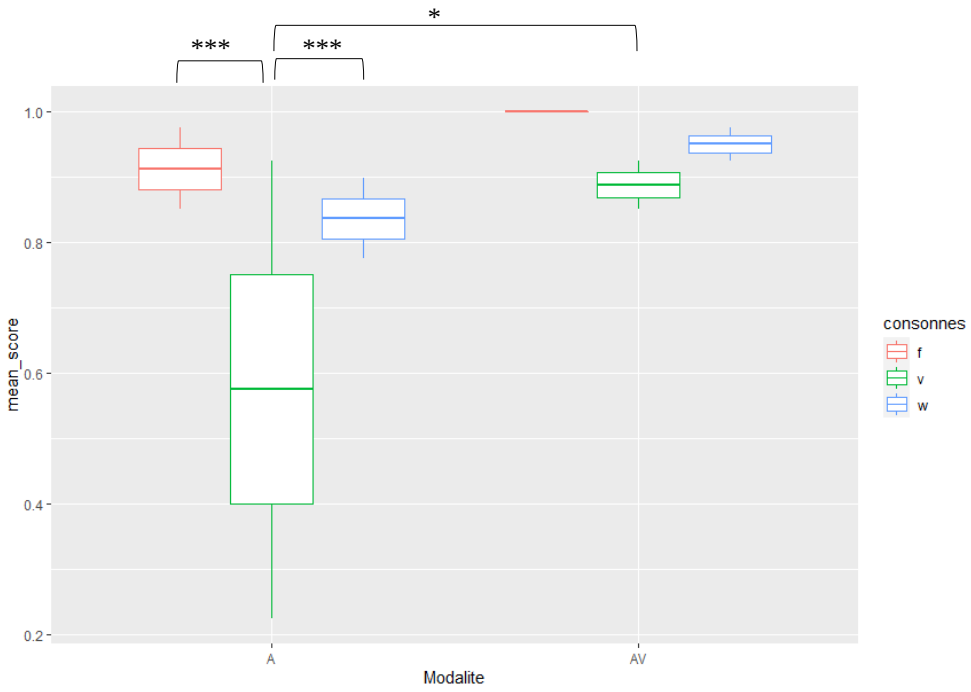

Figure 18. Pourcentage de la prononciation correcte en fonction de la modalité (A *vs* AV) et en fonction de la consonne de référence (/f v w/)

En comparaison de la prononciation correcte entre les trois consonnes à l'intérieur de chaque modalité, les résultats ne montrent pas de différence significative entre les scores de bonne production de /f/, /v/ et /w/ en modalité AV. En modalité A, la labiodentale /v/ est significativement la consonne la moins bien produite (57,5 %) par rapport à /f/ (91,25 %) (z = 4.654, p < 0.0001) et à /w/ (83,54 %) (z = -3.624, p < 0.0008). Il n'y a pas de différence significative de la prononciation correcte entre /f/ et /w/ en modalité A (z = 1.501, p < 0.2904).

b. En fonction du contexte vocalique

Le taux de réussite en fonction de la voyelle est présenté dans la figure 19. Globalement, le taux de la prononciation correcte dans chaque contexte vocalique est plus élevé en modalité AV qu'en modalité A. Mais les différences significatives ne sont attestées entre A et AV qu'en contexte vocalique /e/ (71,67 % *vs* 95 % ; z = -2.360, p < 0.0183) et /ɛ/ (75 % *vs* 96,67 % ; z = -2.375, p < 0.0175), tandis que celles en contextes /i/ (86,44 % *vs* 95 % ; *z* = 1.6368, *p* < 4.34) et /a/ (76,67 % *vs* 91,67 % ; *z* = 1.2851, *p* < 3.60) ne le sont pas.

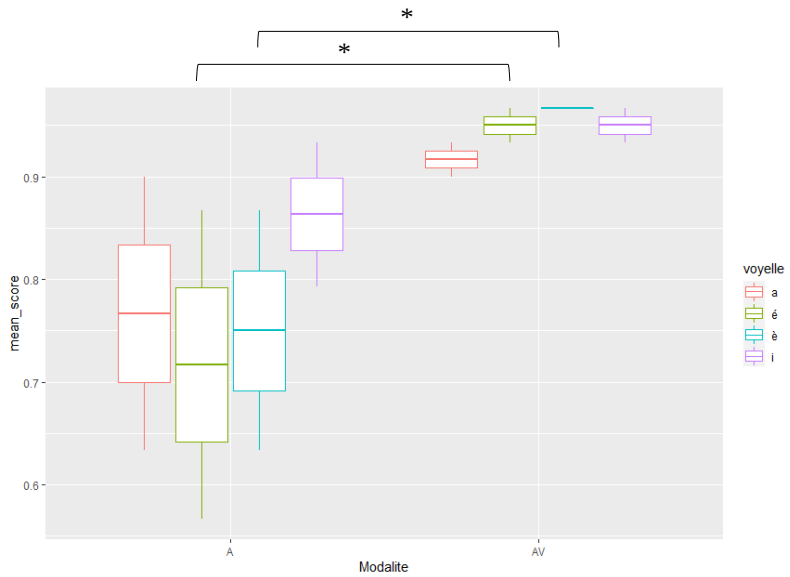

Figure 19. Pourcentage de la prononciation correcte en fonction de la modalité (A *vs* AV) et en fonction de la voyelle (/i e ɛ a/)

Concernant la comparaison à l'intérieur de chaque modalité, les résultats ne montrent pas de différence significative dans les scores de prononciation correcte entre les quatre contextes vocaliques, que ce soit en modalité A ou modalité AV.

c. En fonction de la structure syllabique

Le taux de la prononciation correcte en fonction de la structure syllabique est présenté à la figure 20. Nous pouvons remarquer que, quelle que soit la structure syllabique (CV ou VCV), les taux de réussite sont bien plus élevés en modalité AV. Il y a en plus beaucoup moins de variabilité dans les scores de production correcte lors que les apprenants ont accès à des stimuli audiovisuels. La différence de la prononciation correcte en CV entre les deux modalités est significative (79,17 % en modalité A *vs* 98,33 % en modalité AV ; *z* = -2.967, *p* < 0.0030), mais ce n'est pas le cas pour la structure VCV (75,63 % en modalité A *vs* 90,83 % en modalité AV ; *z* = -1.727, *p* < 0.0841).

En deux modalités, le taux de la prononciation correcte en CV est plus élevé qu'en VCV. En modalité A, le pourcentage de la prononciation correcte en CV et en VCV est de 79,17 % et de 75,63 % respectivement ; la différence n'est pas significative. Concernant la modalité AV, les participants prononcent correctement les consonnes de test en CV dans 98,33 % des cas et en VCV dans 90,83 % des cas ; la différence est significative ($z = 2.321$, $p < 0.0203$).

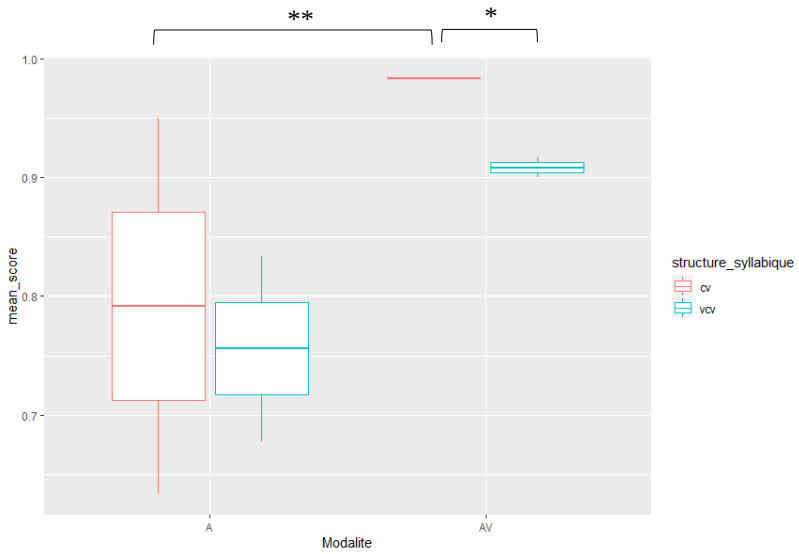

Figure 20. Pourcentage de la prononciation correcte en fonction de la modalité (A *vs* AV) et en fonction de la structure syllabique (CV et VCV)

5.3.2.2 Taux d'erreurs

Dans chaque modalité, il y a 80 items à analyser pour chaque consonne de test /f v w/ (sauf pour /w/ en modalité A avec un item de moins suite à un problème technique comme déjà annoncé).

La table 16 présente le nombre d'occurrences et le pourcentage de prononciation incorrecte en fonction de la modalité et de la consonne. De manière générale, quelle que soit la consonne, le taux d'erreur en modalité A est plus important que celui en modalité AV. Quelle que soit la modalité, /v/ est toujours la consonne qui pose plus de difficultés chez les apprenants (42,50 % d'erreurs en A et 11,25 % en AV). En modalité AV, aucune erreur n'est trouvée lors de la prononciation de /f/. Les erreurs n'apparaissent que lorsqu'il s'agit de /v/ et /w/ en modalité AV, mais avec un taux d'erreurs bien moins important qu'en modalité A.

Table 16. Nombre d'occurrences et pourcentage d'erreurs en production en fonction de la modalité et de la consonne

	Consonne /f/	**Consonne /v/**	**Consonne /w/**
Modalité A	7 (8,75 %)	34 (**42,50 %**)	13 (16,46 %)
Modalité AV	0	9 (**11,25 %**)	4 (5 %)

Les erreurs se trouvent dans les deux structures syllabiques, mais de manière générale, elles sont moins fréquentes en CV qu'en VCV pour la modalité AV (15,38 % *vs* 84,62%) aussi que pour la modalité A (46,30 % *vs* 53,70 %) (cf. figure 21).

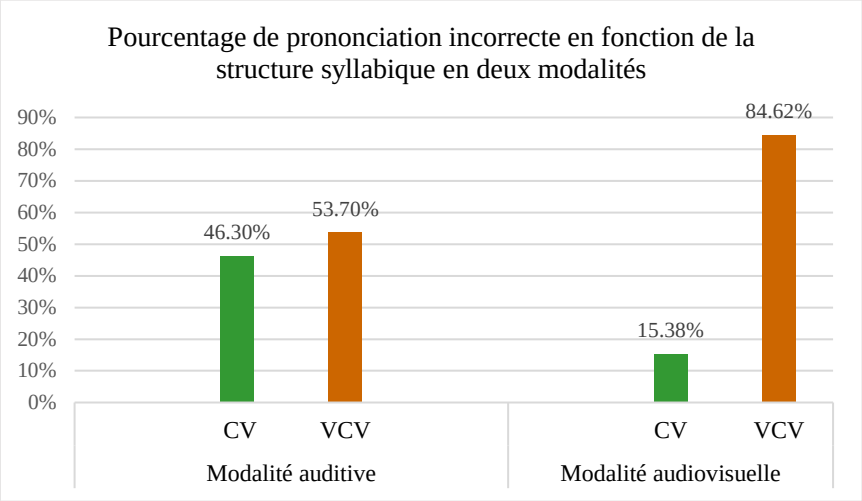

Figure 21. Pourcentage de la prononciation incorrecte en fonction de la structure syllabique (CV et VCV) en deux modalités (A et AV)

Le taux d'erreurs en fonction du contexte vocalique est présenté à la table 17. Les résultats montrent que les erreurs sont moins présentes en modalité AV qu'en modalité A, quel que soit le contexte vocalique. En modalité A, les erreurs de prononciation sont plus trouvées en contexte [e] suivi des contextes [ɛ], [a] et [i] respectivement. Les erreurs sont beaucoup moins fréquentes lors qu'il s'agit de la voyelle fermée [i]. En modalité AV, les erreurs sont plus trouvées en contexte [a], mais avec un nombre beaucoup moins important par rapport aux erreurs en modalité A.

Table 17. Taux d'erreurs en production en fonction de la modalité et du contexte vocalique.

Contexte vocalique / Modalité	[i]	[e]	[ɛ]	[a]
Auditive (Nombre des items mal prononcés : 54 items)	8	17	15	14
Audiovisuelle (Nombre des items mal prononcés : 13 items)	3	3	2	5

Lorsque les apprenants n'arrivent pas à bien produire les consonnes cibles, nous remarquons qu'il y a trois types d'erreurs suivants : la substitution, l'insertion d'un élément consonantique et la suppression (cf. figure 22). Quelle que soit la modalité de test (A ou AV), la substitution est une façon de prononciation préférée des apprenants (87,04 % en A et 92,31 % en AV). L'insertion est beaucoup moins fréquente (11,11 % en A et 7,69 % en AV) alors que la suppression est utilisée uniquement en modalité A dans un seul cas analysé (1,85 %) lors que la cible [vi] est prononcée [i].

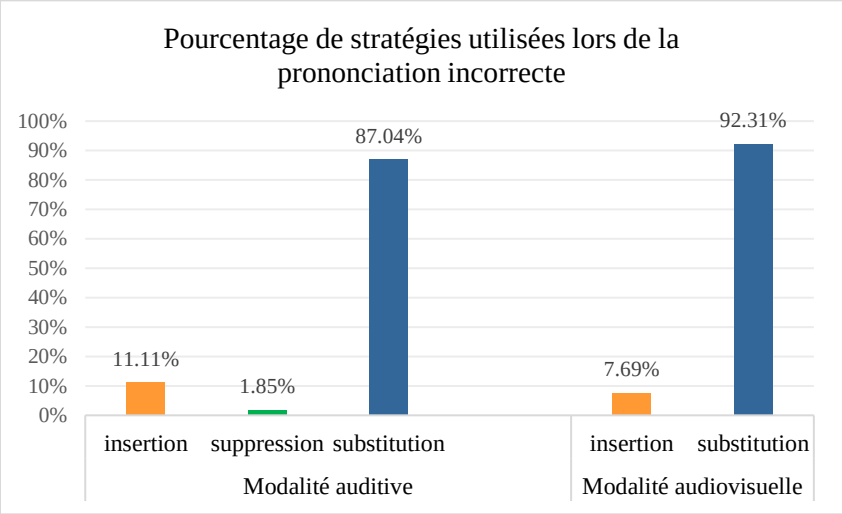

Figure 22. Pourcentage des types d'erreurs en production en fonction de la modalité (A et AV).

a. La substitution

Sur l'ensemble de la prononciation incorrecte en modalité A (54 items), les participants utilisent la substitution pour 47 items. La table 18 présente les consonnes que les participants utilisent en modalité A pour remplacer les consonnes cibles /f/, /v/ et /w/. Pour la cible /v/, les

apprenants la substituent beaucoup plus par [w] dans 24 sur 47 items analysés (soit 51 % d'erreurs), mais aussi par son homologue sourd [f] 6 fois et par l'approximante labiodentale [ʋ] une seule fois. Concernant la cible /w/, elle est remplacée par [v] (8 cas) et par l'approximante bilabiale voisée [β̬] (1 cas). La cible /f/ est plutôt bien prononcée, elle est remplacée par [s] 6 fois et [v] une seule fois.

Table 18. Nombre de cas de substitution en modalité A

La substitution des consonnes cibles en modalité A						
Consonnes prononcées / Consonnes cibles	**[f]**	**[v]**	**[w]**	**[s]**	**[β̬]**	**[ʋ]**
[f]		1	-	6	-	-
[v]	6		24	-	-	1
[w]	-	8		-	1	-

En modalité AV, les participants prononcent correctement les items qui contiennent la consonne /f/ ; aucune erreur n'est trouvée (cf. table 19). Seules les consonnes /v/ et /w/ sont mal prononcées quelques fois. La cible /v/ est remplacée par la labiovélaire [w] 5 fois et aussi par son homologue sourd [f] 3 fois. Quant à la cible /w/, elle a été prononcée [v] 4 fois.

Table 19. Nombre de cas de substitution en modalité AV

La substitution des consonnes cibles en modalité AV			
Consonnes prononcées / Consonnes cibles	**[f]**	**[v]**	**[w]**
[f]		-	-
[v]	3		5
[w]	-	4	

b. L'insertion

Quelle que soit la modalité, l'insertion est bien moins utilisée par les apprenants (11,11 % en modalité A et 7,69 % en modalité AV) (cf. figure 22). Parmi les 54 items mal produits, seules 7 sont prononcés avec l'insertion des consonnes (6 en A et 1 en AV) (cf. table 20). Il est à noter

que cette stratégie est appliquée uniquement lorsque la cible contient /v/ ou /w/ : lorsqu'un item cible comporte /v/, les participants ajoutent [w] et vice versa (cf. figures 23 et 24).

Table 20. Les cas d'insertion en deux modalités

Modalité A			**Modalité AV**		
Prononciation cible	**Prononciation réelle**	**Consonne insérée**	**Prononciation cible**	**Prononciation réelle**	**Consonne insérée**
[vi]	[vwi]	[w]	[ivi]	[iwvi]	[w]
[ve]	[vwe]	[w]			
[ewe]	[ewve]	[v]			
[ɛwɛ]	[ɛwvɛ]	[v]			
[ɛwɛ]	[ɛwvɛ]	[v]			
[awa]	[awvwa]	[v] et [w]			

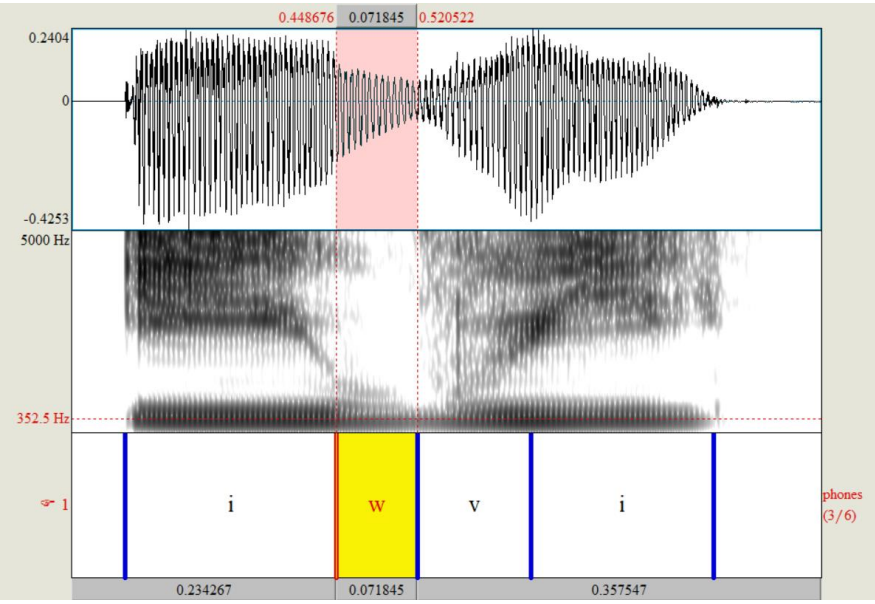

Figure 23. Exemple de l'insertion de [w] dans la cible [ivi] en modalité AV.

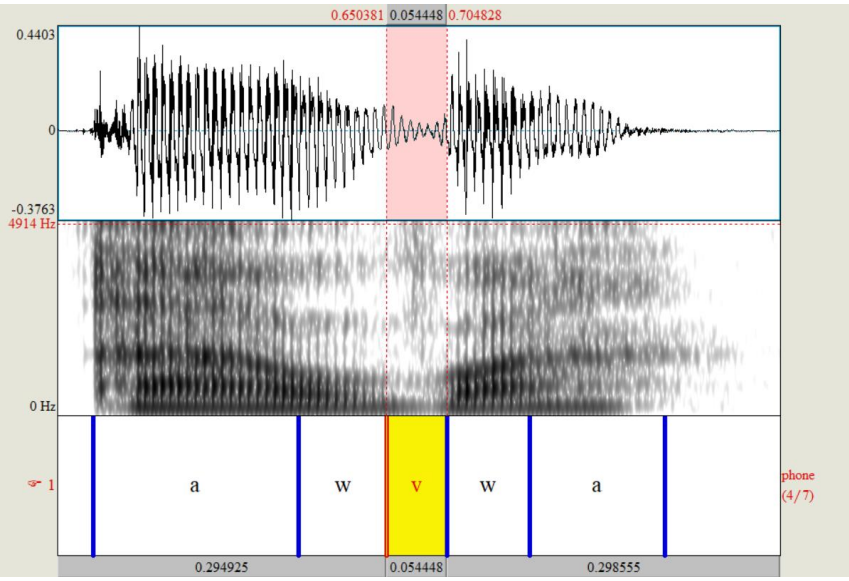

Figure 24. Exemple de l'insertion de [v] dans la cible [awa] en modalité A.

5.3.2.3 Stratégies de prononciation de /v/

Sur l'ensemble des 160 items cibles qui contiennent la consonne /v/, les participants prononcent bien [v] dans 73,13 % des cas (soit 117 items). Ils réussissent mieux en modalité AV (88,75 %) qu'en modalité A (57,50 %). Parmi les 117 items dont [v] est bien prononcé, nous avons remarqué les stratégies particulières utilisées par les quatre participants.

a. Première stratégie : le pré-voisement

Cette stratégie consiste en la mise en vibration des plis vocaux avant la réalisation de [v] (cf. figures 25 et 26). Il s'agit de la stratégie la plus utilisée parmi les prononciations considérées « correctes » de [v] par les participants (54,70 %, soit 64 sur 117 items), quelle que soit la structure syllabique (CV ou VCV), mais majoritairement en CV.

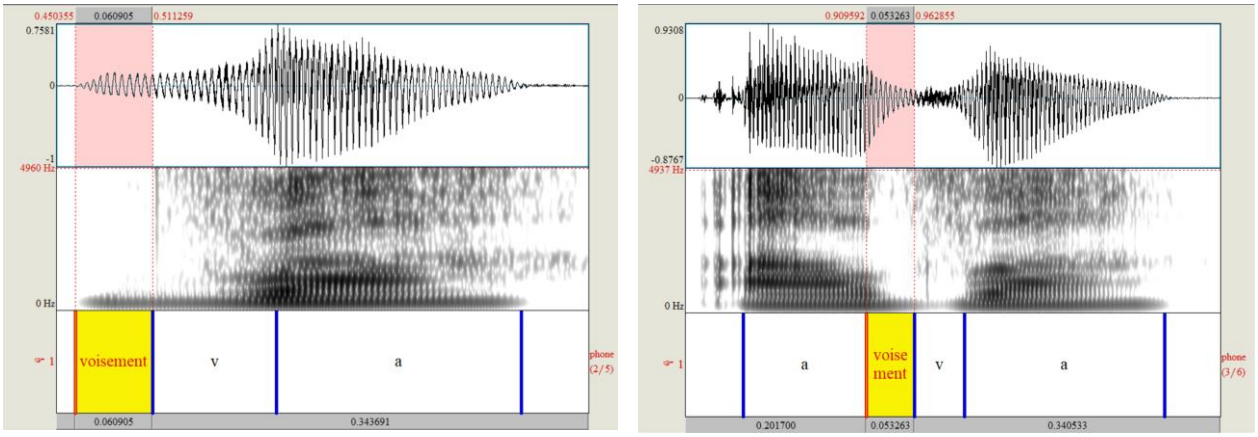

Figure 25. Exemples de la réalisation du pré-voisement avant la production de [v].
La production de [v] dans [va] (à gauche) et [ava] (à droite) d'un participant testé en modalité AV.

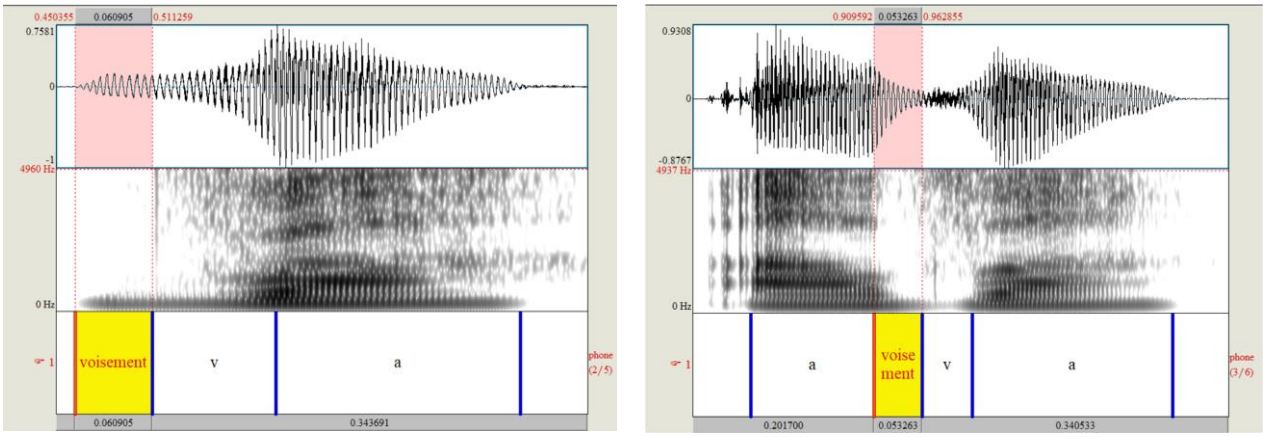

Figure 26. Exemples de la réalisation du pré-voisement avant la production de [v].
La production de [v] dans [va] (à gauche) et [ava] (à droite) d'un participant testé en modalité auditive.

Il est à remarquer également que lorsque les apprenants ne font pas le pré-voisement avant de réaliser [v], le bruit de friction est beaucoup moins intense. Comme la montre la figure 27 ci-après, l'apprenant réalise le pré-voisement (à gauche) avec un bruit de friction de [v] bien plus visible sur le spectrogramme que lorsqu'il produit [v] sans le pré-voisement (à droite).

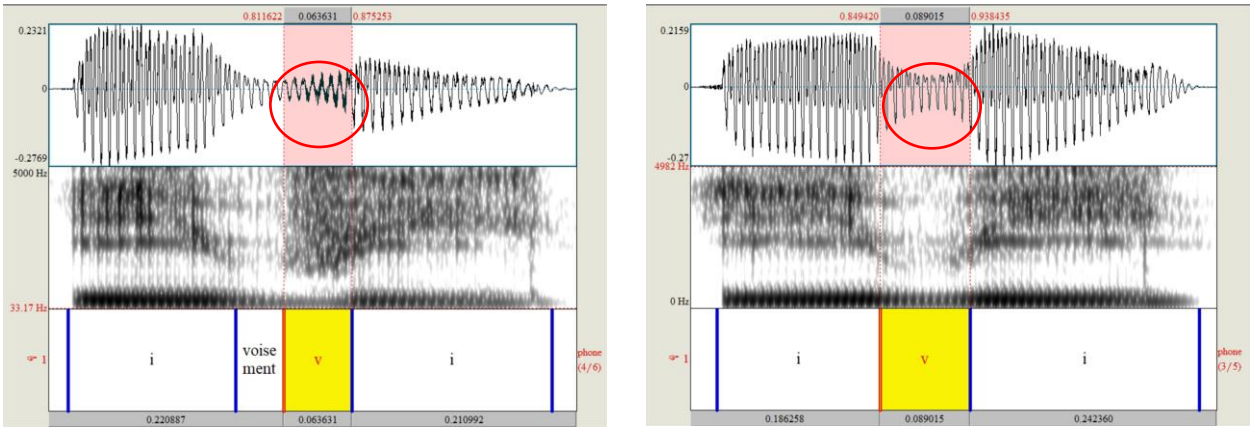

Figure 27. Comparaison des deux signaux acoustiques : à gauche, la production de [v] précédée du pré-voisement et à droite, la production de [v] sans la réalisation du voisement avant. Le cercle rouge pointe la différence de l'intensité du bruit de friction.

b. Deuxième stratégie : la décomposition des traits articulatoires

Étant une fricative sonore, la labiodentale /v/ devrait être produite avec une superposition des gestes glottique (voisement) et supra-glottique (friction). Nous remarquons que les apprenants de cette étude trouvent une stratégie de réaliser les deux gestes mais pas de manière simultanée comme il faut. Cette stratégie consiste en une réalisation du voisement avant la production du bruit de friction labiodental sourd [f], comme le montrent les figures 28 et 29 ci-après. Cette stratégie est moins utilisée que celle du pré-voisement (seulement 10 sur 117 items, soit 8,5 %) mais elle est appliquée par tous les quatre participants.

Il est à noter que nous avons fait écouter les 10 items aux deux locutrices francophones et leur avons demandé de décider quelle consonne elles entendent (sans leur montrer les signaux). Les locutrices ont tous répondu qu'il s'agit de la consonne /v/. Par conséquent, les items qui sont prononcés avec cette stratégie sont considérés comme prononciation correcte de [v].

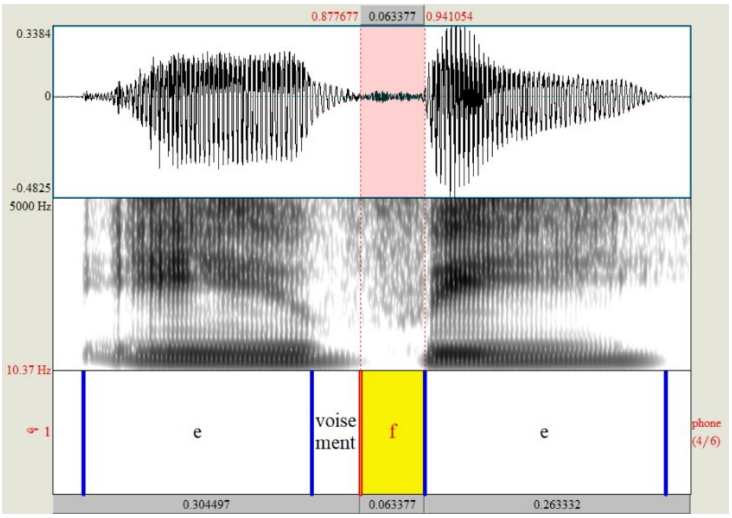

Figure 28. Exemple de la stratégie de décomposition des traits articulatoires de [v].
La prononciation de l'item [eve] d'un participant testé en modalité AV.

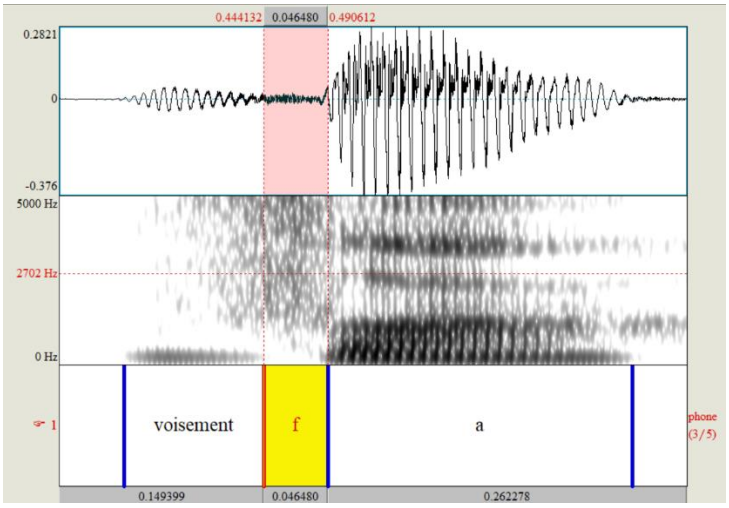

Figure 29. Exemple de la stratégie de décomposition des traits articulatoires de [v].
La prononciation de l'item [va] d'un participant testé en modalité A.

5.3.3 Relation entre la perception et la production

La figure 30 montre que, le taux de réussite en modalité AV est significativement plus élevé que celui en modalité A, que ce soit en perception (z = -2.388, p < 0.0169) ou en production (z = -2.595, p < 0.0095).

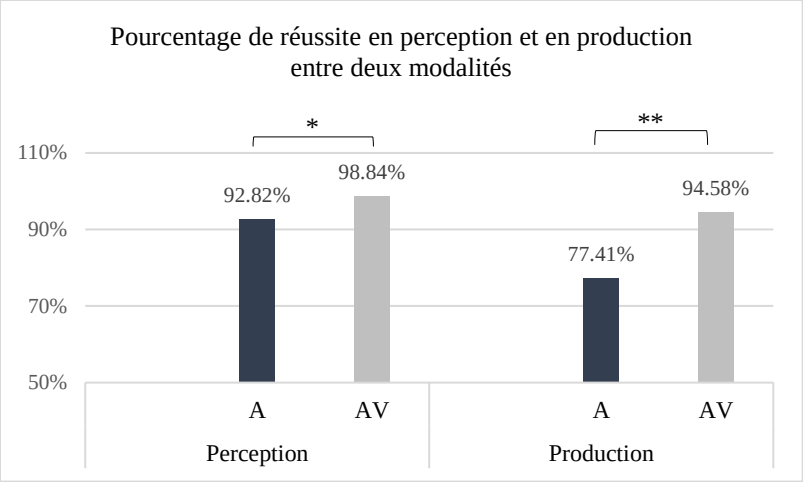

Figure 30. Taux de réussite en perception et en production en fonction de la modalité

Le taux de réussite en fonction de la consonne dans les deux tâches et en deux modalités est présent ci-après (cf. figure 31).

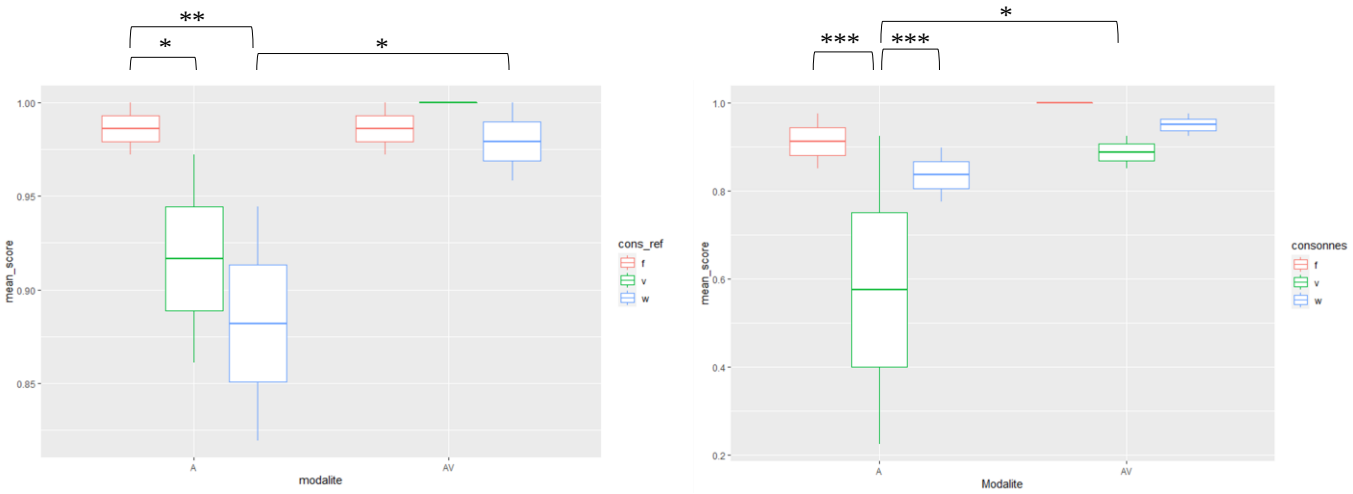

Figure 31. Taux de réussite en perception (à gauche) et en production (à droite) en fonction de la modalité et de la consonne.

La consonne /f/ est plutôt bien perçue et bien produite quelle que soit la modalité testée (98,61 % en perception et 95,63 % en production). Les apprenants ont même réussi à produire /f/ à 100 % en modalité AV. Aucune différence significative n'est observée entre les deux modalités A et AV aussi bien en perception qu'en production.

Pour la consonne /v/, le taux de réussite dans la tâche de production est significativement plus élevé en modalité AV qu'en modalité A (88,75 % *vs* 57,50 % ; z = -2.454, p < 0.0141). Cependant, dans la tâche de perception les résultats de la consonne /v/ ne sont pas statistiquement

analysables. Mais en observant le graphique, nous remarquons que la différence de la consonne /v/ en perception entre les deux modalités est assez évidente et qu'il y a plus de variabilité dans les résultats de /v/ en modalité A.

Concernant la consonne /w/, le taux de réussite dans la tâche de perception est significativement plus élevé en modalité AV qu'en modalité A ($z = -2.073$, $p < 0.0382$), alors que dans la tâche de production, les résultats de /w/ ne sont pas significativement différents. En modalité AV, /w/ est bien perçu (97,92 %) et bien produit (95 %), alors qu'en modalité A, le taux de réussite diminue dans les deux tâches même s'il s'agit de la consonne native : 88,19 % en perception *vs* 83,54 % en production.

Confusion entre /v/ et /w/

La figure 32 présente les consonnes que les participants confondent avec /v/ ainsi que le pourcentage de confusion en fonction de la tâche et de la modalité. Que ce soit en perception ou en production, le taux de confusion de /v/ avec d'autres consonnes est moins élevé en modalité AV qu'en modalité A.

Il est à remarquer que la labiodentale /v/ est toujours plus confondue avec la labiovélaire /w/ et ce, beaucoup plus en modalité A qu'en modalité AV et à la fois en perception et production (cf. figure 32). Les stimuli audiovisuels semblent mieux aider les apprenants à ne plus faire la confusion entre /v/ et /w/ en perception (aucune confusion). En production, la condition AV fait baisser considérablement le taux de confusion [v] ~ [w] de 30 % en modalité A en seulement 6,25 %.

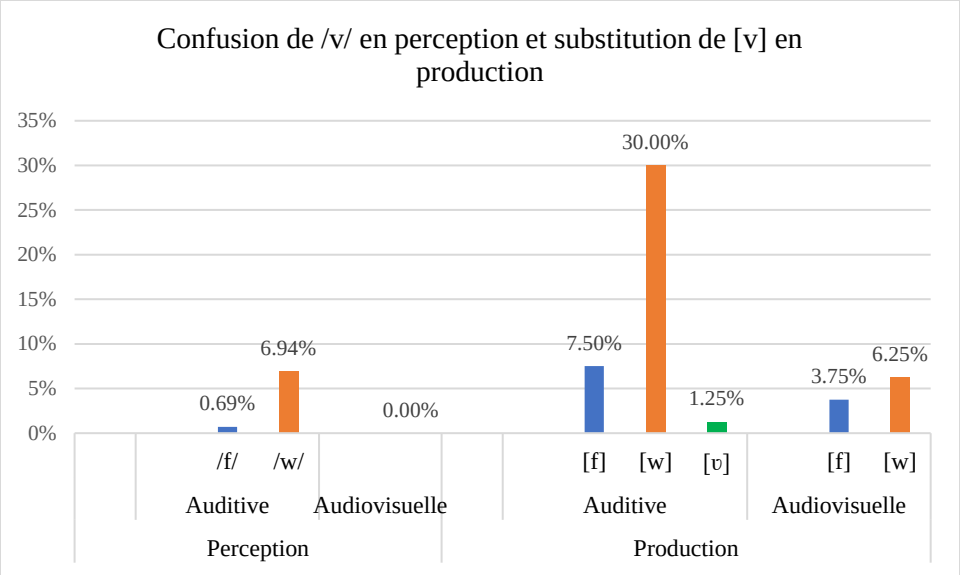

Figure 32. Pourcentage de confusion de [v] avec d'autres consonnes dans la tâche de perception (à gauche) et pourcentage de substitution de [v] dans la tâche de production (à droite).

La confusion entre /v/ et /w/ apparaît dans tous les contextes vocaliques, mais elle est beaucoup moins présente en modalité AV qu'en modalité A quel que soit le contexte vocalique (cf. table 21). En modalité A, les apprenants semblent plus perturbés en contexte de la voyelle /ɛ/, ils font plus de confusion /v/ ~ /w/ à la fois en perception (9 fois) et en production (13 fois).

Table 21. Nombre de confusion /v/~/w/ en perception et en production en fonction de la modalité et du contexte vocalique.

Tâches	**Modalités**	**Nombre total de confusion**	**Nombre de confusion selon les contextes vocaliques**			
			/i/	**/e/**	**/ɛ/**	**/a/**
Perception	A	24	8	4	9	3
	AV	3	0	1	2	0
Production	A	32	4	6	13	9
	AV	9	2	3	1	3
		Total	14	14	**25**	15

Les mêmes résultats en fonction de la structure syllabique quant à la confusion /v/ ~ /w/ sont observés entre les deux modalités (cf. la table 22). Globalement, la confusion est beaucoup moins fréquente en modalité AV qu'en modalité A. Quel que soit le type de tâche et quelle que

soit la modalité, les apprenants confondent entre /v/ et /w/ beaucoup plus en position intervocalique VCV qu'en initiale absolue CV.

Table 22. Nombre de confusion /v/~/w/ en perception et en production en fonction de la modalité et de la structure syllabique.

Tâches	**Modalités**	**Nombre total de confusion**	**Nombre de confusion selon la structure syllabique**	
			CV	**VCV**
Perception	A	24	11	13
	AV	3	0	3
Production	A	32	13	19
	AV	9	1	8
		Total	25	**43**

5.4 Discussion

Cette étude a pour objectif de tester l'apport des informations audiovisuelles dans le traitement de la consonne /v/ chez les apprenants thaïlandais du français. Deux tâches (perception et production) ont été effectuées à l'intérieur de chaque modalité testée (A et AV). Les stimuli choisis sont les pseudo-mots contenant les 3 consonnes /f v w/ qui se trouvent en position initiale de syllabe CV ou en intervocalique VCV et en contextes vocaliques non arrondis /i e ɛ a/. Quatre apprenants de niveau A1-A2 ont été recrutés. Chacun d'entre eux est testé sur une seule modalité (soit A, soit AV) avec un ordre de passage de tests contrebalancé entre participants au sein d'une même modalité. En perception, il s'agit d'une tâche de discrimination (pareil ou différent) entre un stimulus de test et un stimulus de référence. En production, les participants ont pour consigne de répéter les sons perçus.

Les résultats montrent que, le taux de réussite de /v/ augmente considérablement, que ce soit en perception ou en production, lorsque les apprenants sont exposés à la fois aux informations auditives et visuelles. En conformité avec les études antérieures de Chunsuvimol et Ronnakiat, (2001), Ngammana (2011), Le Corre (2013), Sridhanyarat, 2017 et Tusnyingyong & Tran (2022), /v/ est plus souvent confondu avec /w/. Cependant, cette confusion est beaucoup moins présente en modalité AV qu'en modalité A seule. Par ailleurs, les apprenants arrivent à mieux discriminer

/v/ de d'autres consonnes (/f/ et /w/) en modalité AV qu'en modalité A, et ce, en un temps significativement plus court, quels que soient le contexte vocalique et la structure syllabique.

L'ensemble des résultats obtenus à l'issue de cette étude permet de discuter et de vérifier les hypothèses posées au début du chapitre (cf. section 5.1).

Hypothèse 1 *(en fonction de la **consonne** et de la **modalité** de test)* :

***Consonne /f/** : (…) le taux de réussite de /f/ en perception et en production devrait être élevé, et en égalité entre la modalité A et la modalité AV.*

Les résultats montrent effectivement que les participants réussissent à percevoir et à produire /f/ avec des scores très élevés et en égalité entre deux modalités testées (cf. figure 31). Ces résultats sont en conformité avec les deux études de Wang et al. (2008, 2009) qui montrent que les locuteurs anglophones sont capables de percevoir les consonnes fricatives de l'anglais en modalité A et en modalité AV avec un taux de réussite très élevé dans les deux modalités.

Comparé aux résultats de la consonne non native /v/, les apprenants sont plus performants dans le traitement du phonème natif /f/, quel que soit le type de tâche. Ces résultats vont dans le même sens que l'Hypothèse de l'Analyse Contrastive (Lado, 1957) selon laquelle les éléments de la LE qui sont similaires à la L1 seront simples, et ceux qui sont différents seront difficiles. Puisque /f/ existe dans les deux langues (le thaï et le français), les participants ne rencontrent pas de difficultés à percevoir et à prononcer cette consonne, que ce soit en modalité A ou AV, ce qui n'est pas le cas du phonème non natif /v/ de cette étude.

***Consonne /v/** : (…) /v/ devrait être mieux perçue, mieux produite et moins confondue avec /w/ en modalité AV qu'en modalité A chez les apprenants.*

Cette hypothèse a été confirmée, le taux de réussite de la consonne non native /v/ est plus élevé en modalité AV qu'en modalité A, quel que soit le type de tâche (cf. figure 31).

En perception, le pourcentage des réponses correctes de /v/ est plus élevé en modalité AV (100 %) qu'en modalité A (91,67 %). Ces résultats vont dans le même sens que les études d'Hazan et al. (2006), de Wang et al. (2008, 2009) et de Burfin (2015) qui montrent que le taux de réussite dans la perception des phonèmes non natifs augmente en modalité AV par rapport à la modalité A. Il est à rappeler que, notre modèle statistique n'atteste pas une différence significative du taux de réponses correctes entre les deux modalités, ceci est dû uniquement à l'absence de variabilité

dans les scores de perception en modalité AV (100 %). Mais l'amélioration dans les scores est bien observée dans la perception de /v/ si en présence des stimuli audiovisuels (91,67 % en A > 100 % en AV). En production, les apprenants prononcent nettement mieux [v] en modalité AV (88,75 %) qu'en modalité A (57,50 %).

Ces résultats montrent clairement que les informations audiovisuelles de la parole facilitent la perception et la production de cette consonne non native [v] en initiale de syllabe chez les apprenants thaïlandais. Ils deviennent plus performants lorsqu'ils sont exposés aux informations audiovisuelles de la parole que lorsqu'ils sont confrontés uniquement aux informations auditives. Le taux de confusion /v/ ~ /w/ diminue entre les deux modalités A et AV en perception (de 6,94 % à 0 %) et aussi considérablement en production (de 30 % à 6,25 %). La diminution de la confusion /v/ ~ /w/ en modalité AV pourrait résulter du fait que les visèmes de la labiodentale /v/ et de la labiovélaire /w/ sont bien visibles et que le thaï possède ces deux visèmes (cf. sections 2.3 et 4.1.3). Les participants pourraient bénéficier de la visibilité et de l'existence de ces deux visèmes pour distinguer /v/ de /w/. La confusion est donc moins présente en modalité AV qu'en modalité A.

***Consonne /w/** : (…) le taux de réussite en modalité AV serait plus élevé que celui en modalité A à la fois en perception et en production.*

Il a été montré effectivement que /w/ est mieux perçu et mieux produit en modalité AV qu'en modalité A. Dans la tâche de perception, le taux de réussite de /w/ est significativement moins élevé en modalité A (88,19 %) qu'en modalité AV (97,92 %), mais la différence dans le score de prononciation correcte de /w/ n'est pas significative entre modalités A et AV (83,54 % *vs* 95 %). Même si /w/ est la consonne native, les participants rencontrent des difficultés à la percevoir en confondant avec /v/, notamment en modalité A (cf. table 14) où ils ont accès aux seules informations auditives.

Hypothèse 2 (*en fonction du **temps** de **réponse***) : *(…) la parole est multimodale et la présence des deux inputs auditif et visuel devrait donc accélérer le processus de traitement d'informations en perception, quelle que soit la consonne de test (native ou non native). Le temps de réponse en modalité AV serait moins long qu'en modalité A.*

Cette hypothèse a été confirmée, le temps moyen de réponse dans la tâche de perception étant significativement moins long en modalité AV qu'en modalité A, quel que soit le type de stimuli (natif ou non natif), quel que soit le contexte vocalique (/i e ɛ a/) et quelle que soit la

structure syllabique (CV ou VCV) (cf. section 5.3.1.3). Ces résultats sont en conformité de l'étude de Burfin (2015) selon laquelle le temps de réaction en modalité AV est significativement moins important que celui en modalité A, que ce soient les stimuli natifs ou non natifs. Le traitement du phonème natif et non natif est plus rapide lors de la présentation de double information (audio et vidéo).

Hypothèse 3 (*le lien entre perception et production*) : *(…) Si les participants rencontrent des difficultés dans la perception de /v/ puisque cette consonne non native /v/ est assimilée à la catégorie de la consonne native /w/, une corrélation entre perception et production devrait être trouvée dans les scores de bonne réponse chez les mêmes participants, quelle que soit la modalité de test.*

Nous constatons que /v/ est perceptivement confondu avec /w/ dans la plupart des cas. De même, /w/ est majoritairement confondue avec /v/ (cf. table 14). La confusion entre /v/ et /w/ est largement attestée en grande partie (30 %) en production en modalité A. Ces résultats confirment les études antérieures sur la difficulté de prononciation de /v/ chez les apprenants thaïlandais (Chunsuvimol & Ronnakiat, 2001 ; Ngammana, 2011 ; Le Corre, 2013 ; Sridhanyarat, 2017 ; Tusnyingyong & Tran, 2022) et suggèrent un lien possible entre la perception et la production : la mauvaise perception de /v/ pourrait entrainer les erreurs dans sa production, comme ce que souligne l'hypothèse de Flege (1995). Effectivement, Flege postule que la production d'un phone correspond à la représentation de la catégorie phonétique et que la majorité des erreurs de production résulte de la mauvaise perception. Les résultats dans cette présente étude montrent que la confusion /v/ ~ /w/ chez les apprenants thaïlandais existe à la fois en perception et en production. Les erreurs de prononciation de [v] pourraient par conséquent résulter de la mauvaise perception.

Nous essayons d'expliquer nos résultats concernant /v/ dans le cadre du modèle PAM (Best et al., 1988 ; Best, 1995a ; Best, 1995b) (cf. section 1.2.1) comme suit. Le phone non natif [v] pourrait s'installer dans l'espace phonologique de la L1 chez les apprenants thaïlandais. En condition auditive seule, elle serait assimilée à la catégorie native /w/ (cf. figure 33, à gauche). Cependant, la présentation audiovisuelle de la parole pourrait aider les participants à séparer [v] de la catégorie native /w/, car la confusion entre /v/ et /w/ est moins présente en modalité AV (cf. table 15). Le phone non natif [v] serait éloigné de la catégorie native /w/ à l'aide des informations audiovisuelles de la parole (cf. figure 33, à droite).

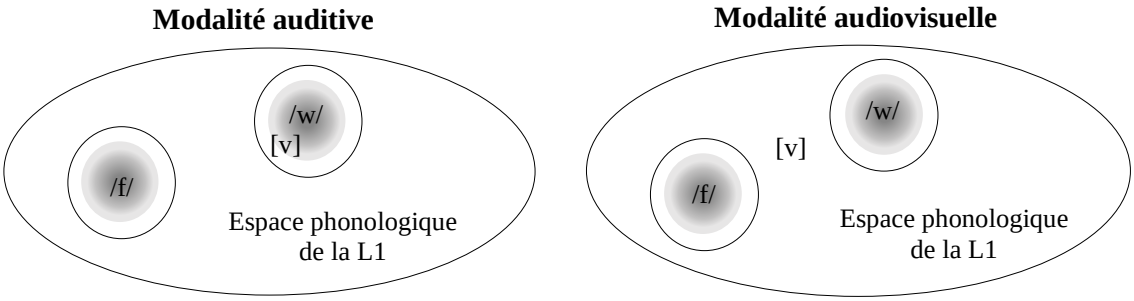

Figure 32. Représentation de la place de [v] dans l'espace phonologique de la L1 chez les apprenants thaïlandais lors de l'exposition à l'input auditif (à gauche) et lors de l'exposition à l'input audiovisuel (à droite)

Notre remarque porte sur le taux de réussite de la consonne /v/ dans la tâche de perception. Même s'il y a une augmentation du taux de réussite de cette consonne en modalité AV (100 %), le taux de réussite en modalité A reste élevé (91,67 %). Ces résultats sont en conformité des études de Hazan et al. (2006) et de Burfin (2015) sur la perception de /v/ et /b/ chez les locuteurs hispanophones. Il est à noter que /v/ n'existe pas en espagnol, mais cette consonne est allophone de /f/ en coda (Hazan et al., 2006). Les résultats de ces deux études montrent que, même si /v/ n'existe pas dans leur L1, les apprenants hispanophones sont performants en perception, et ce, en modalité A. Selon les auteurs, ce résultat pourrait résulter du fait que [v] est présent en espagnol en tant qu'allophone de /f/ (Hazan et al., 2006) et que le lieu labiodental existe en espagnol. Les locuteurs hispanophones seraient par conséquent capables de distinguer le lieu d'articulation différent entre /v/ et /b/. Dans le même sens, dans cette présente étude, le taux de réussite très élevé dans la perception de /v/ en modalité A pourrait résulter du fait que le lieu articulatoire labiodental (/f/) existe en thaï. Les participants pourraient bénéficier de l'existence du visème labiodental dans leur L1 et seraient capables de détecter les lieux articulatoires différents entre la labiodentale /v/ et la labiovélaire /w/. Cependant, même si les participants testés en modalité A perçoivent bien la consonne /v/ (91,67 %), ils la prononcent correctement seulement dans 57,50 % des cas.

Il a été démontré que la fricative labiodentale sonore est moins fréquente dans les langues du monde (Vallée et al., 1999). Selon l'Hypothèse de la Différence de Marquage (Eckman, 1977), des éléments qui moins fréquents ou moins privilégiés seront difficiles à acquérir. La fricative /v/ qui est moins fréquente dans la langue du monde constitue par conséquent un incontestable problème non seulement pour les apprenants thaïlandais, mais aussi pour les apprenants de

différentes L1s (Silverman, 1992 ; Detey, Durand & Nespoulous, 2005 ; Tri Endarto, 2015). A partir de l'analyse acoustique de la production, nous constatons que, pour produire [v], les participants commencent, dans la plupart des cas, à vibrer les cordes vocales avant d'articuler [v] (cf. section 5.3.2.3, sous-section a.). Lorsqu'ils n'arrivent pas à produire le voisement simultanément avec l'articulation supra-glottique de /v/, les traits articulatoires sont décomposés en deux : le voisement et le bruit de friction sourd [f] (cf. section 5.3.2.3, sous-section b.). Les apprenants doivent appliquer des stratégies particulières pour produire la fricative labiodentale sonore [v] moins fréquent dans les langues du monde.

Conclusion et perspectives

Plusieurs études démontrent des difficultés dans la prononciation de la consonne labiodentale /v/ en position initiale de syllabe chez les apprenants thaïlandais (Debyser, 1969 ; Chunsuvimol & Ronnakiat, 2001 ; Ngammana, 2011 ; Le Corre, 2013 ; Sridhanyarat, 2017 ; Tusnyingyong & Tran, 2022). Cette consonne, n'existant pas en thaï, est très souvent confondue avec le son [w]. La source de cette confusion semble provenir de l'association erronée entre le graphème latin < v > /v/ et le graphème thaï < ว > /w/ qui, à notre connaissance, a commencé depuis le XVII^{e} siècle à travers la romanisation du thaï et la translittération des mots d'emprunt anglais en caractères thaïs (De La Loubère, 1691 ; Pallegoix, 1854 ; Vajiravudh, 1913 ; Nacaskul, 1979 ; Tri Endarto, 2015). Les apprenants thaïlandais associent par conséquent le graphème < v > au son [w]. Lorsqu'ils sont confrontés à l'orthographe, la confusion entre [v] et [w] est plus fréquente que quand l'orthographe n'est pas présente (Tusnyingyong & Tran, 2022).

Il a été démontré que la présentation audiovisuelle facilite le traitement de la parole en LE (Hazan et al., 2006 ; Wang et al., 2008 ; Wang et al., 2009 ; Burfin, 2015). En nous basant sur ces études antérieures, nous avons effectué une étude expérimentale pour vérifier si la présentation audiovisuelle de la parole pourrait amener les apprenants thaïlandais à mieux percevoir et à mieux prononcer /v/ en position initiale de syllabe.

Deux modalités (audio seul et audiovisuelle) et deux types de tests (perception et production) ont été effectués dans cette étude. Dans une tâche de discrimination, les participants ont jugé si les stimuli de test sont pareils ou différents des stimuli de référence. En production, ils répètent les stimuli perçus au préalable, soit entendus (modalité A), soit entendus/vus (modalité AV). Deux participants ont été testés en modalité A et deux en modalité AV. Les résultats montrent que le taux de réussite en modalité AV est significativement plus élevé qu'en modalité A, quel que soit le type de tâche (perception ou production). La confusion entre /v/ et /w/ en perception et en production est moins présente en modalité AV qu'en modalité A. Par ailleurs, le temps de réponse dans la tâche de perception est significativement moins long en modalité AV qu'en modalité A. L'ensemble de ces résultats montre que les informations audiovisuelles facilitent et accélèrent le traitement de la consonne non native /v/ en position initiale de syllabe chez les apprenants thaïlandais.

Cette étude s'inscrivant dans le cadre de l'enseignement et de l'apprentissage du français montre clairement l'apport des informations audiovisuelles de la parole dans le traitement du phonème non natif /v/. Ces résultats pourraient contribuer à la réflexion des méthodes de remédiation à la perception et à la production de /v/ en initiale de syllabe chez les apprenants thaïlandais : les méthodes qui devraient aider les apprenants, qui sont bien habitués à l'association erronée entre le graphème < v > et le son [w], à déconditionner de ces habitudes acquises avec le système de la L1.

Dans le cadre restreint d'un mémoire de M2, cette étude n'a été effectuée qu'auprès de quatre participants. Cela nous encourage à élargir l'expérience avec un nombre plus important de participants. Par ailleurs, les résultats dans cette présente étude montrent que les réponses incorrectes dans la tâche de perception en modalité A sont plus trouvées lorsque les participants ont dû distinguer /w/ (stimuli de référence) de /v/ (stimuli de test) que lorsqu'ils ont dû distinguer /v/ (stimuli de référence) de /w/ (stimuli de test). Il est souhaitable d'élargir l'expérience en vérifiant si l'ordre de la présentation des stimuli pourrait impacter la perception de /v/ chez les apprenants thaïlandais.

Enfin, selon la première hypothèse de Flege (1995) qui postule qu'un même phone avec une position syllabique différente pourrait être classé dans des catégories différentes, il est donc souhaitable de tester la perception de la consonne /v/ en position finale de syllabe pour vérifier dans quelle catégorie native est assimilée cette consonne.

Bibliographie

Abramson, A. (1963). The Vowels and Tones of Standard Thai : Acoustical Measurements and Experiments. *Acta Linguistica Academiae Scientiarum Hungaricae*, *13*(3/4), 402-404.

Abry-Deffayet, D., & Chalaron, M.-L. (1994). *Phonétique : 350 exercices avec 6 cassettes*. Hachette-Français langue étrangère.

Benoît, C., Mohamadi, T., & Kandel, S. (1994). Effects of phonetic context on audio-visual intelligibility of French. *Journal of speech and hearing research*, *37*(5), 1195-1203.

Best, C. C., & McRoberts, G. W. (2003). Infant Perception of Non-Native Consonant Contrasts that Adults Assimilate in Different Ways. *Language and Speech*, *46*(2-3), 183-216. https://doi.org/10.1177/00238309030460020701

Best, C. T. (1995a). A direct realist view of cross language speech perception. Dans W. Strange (Éd.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* (p. 171-204). Timonium: New York Press.

Best, C. T. (1995b). Learning to Perceive the Sound Pattern of English. *Advances in infancy research*, *9*, 217-305.

Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *The Journal of the Acoustical Society of America*, *109*(2), 775-794. https://doi.org/10.1121/1.1332378

Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of Perceptual Reorganization for Nonnative Speech Contrasts: Zulu Click Discrimination by English-Speaking Adults and Infants. *Journal of experimental psychology. Human perception and performance*, *14*(3), 345-360. https://doi.org/10.1037//0096-1523.14.3.345

Boersma, P., & Weenink, D. (2020). Praat: *doing Phonetics by Computer (version 6.2)*. https://www.fon.hum.uva.nl/praat/

Borel, S., Vaissière, J., Lavrut, M., Noiret, A., Ambert-Dahan, E., & Sterkers, O. (2016). Quels sosies labiaux pour les voyelles nasales du français ? *Glossa*, *120*, 23-43.

Borrell, A. (1996). Parallèle entre perception et production ? Complexité du lien entre reconnaissance et production des unités phonético-phonologiques. *La Linguistique*, *32*(2), 105-116.

Brannen, K. (2002). The Role of Perception in Differential Substitution. *Canadian Journal of Linguistics*, *47*(1), 1-46. https://doi.org/10.1017/S0008413100018004

Burfin, S. (2015). *L'apport des informations visuelles des gestes oro-faciaux dans le traitement phonologique des phonèmes natifs et non-natifs : Approches comportementale, neurophysiologique* [Thèse de Doctorat en Sciences Cognitives, Psychologie et Neurocognition, Université Grenoble Alpes]. https://tel.archives-ouvertes.fr/tel-01162064/document

Calliope. (1989). *La parole et son traitement automatique* (Tubach Jean-Pierre, Éd.). Masson.

Chitkla, K. (2017). *สนทนาฝรั่งเศสลัดทันใจ [Conversation française en un instant]* (2e éd.). Tathata.

Chunsuvimol, B., & Ronnakiat, N. (2001). (V) is a really problem sound for Thai speakers. *Thammasat Review*, *6*(1), 177-195.

Danvivathana, N. (1981). *The Thai writing system* [Thèse de Doctorat en Littératures, Langues et Cultures, Université d'Edinburgh]. ERA. https://era.ed.ac.uk/handle/1842/18127

De La Loubère, S. (1691). *Du Royaume de Siam. Tome 2*. La Veuve de Jean Baptiste Coignard.

De Boysson-Bardies, B. (1996). *Comment la parole vient aux enfants : De la naissance jusqu'à deux ans*. Odile Jacob.

Debyser, F. (1969). *Difficultés phonétiques des élèves thaïlandais : Conseils pédagogiques à l'intention des professeurs de français en Thaïlande*. BELC.

Detey, S., Durand, J., & Nespoulous, J.-L. (2005). Interphonologie et représentations orthographiques. Le cas des catégories /b/ et /v/ chez des apprenants japonais de Français Langue Etrangère. *Revue PAROLE*, *34-35-36*, 140-185.

Dumont, A., & Calbour, C. (2002). *Voir la parole : Lecture labiale, perception audiovisuelle de la parole*. Masson.

Eckman, F. R. (1977). Markedness and the contrastive analysis hypothesis. *Language Learning*, *27*(2), 315-330.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, *171*(3968), 303-306. https://doi.org/10.1126/science.171.3968.303

Ferlus, M. (1997). The origin of the graph b in the Thai script. In A. Abramson, *Southeast Asian Linguistic Studies in Honor of Vichin Panupong* (p. 79-82). Chulalongkorn University Press.

Field, T. M., Woodson, R., Greenberg, R., & Cohen, D. (1982). Discrimination and imitation of facial expression by neonates. *Science (New York, N.Y.)*, *218*(4568), 179-181. https://doi.org/10.1126/science.7123230

Flege, J. E. (1995). Second language speech learning : Theory, findings and problems. Dans W. Strange (Éd.), *Speech perception and linguistic experience: Issues in cross-language research* (p. 233-277). Timonium, MD: York Press.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds “L” and “R”. *Neuropsychologia*, *9*(3), 317-323.

Götz, A., Yeung, H. H., Krasotkina, A., Schwarzer, G., & Höhle, B. (2018). Perceptual Reorganization of Lexical Tones : Effects of Age and Experimental Procedure. *Frontiers in Psychology*, *9*, 477. https://doi.org/10.3389/fpsyg.2018.00477

Hazan, V., Sennema, A., Faulkner, A., Ortega-Llebaria, M., Iba, M., & Chunge, H. (2006). The use of visual cues in the perception of non-native consonant contrasts. *The Journal of the Acoustical Society of America*, *119*(3), 1740-1751. https://doi.org/10.1121/1.2166611

Höhle, B., Bijeljac-Babic, R., Herold, B., Weissenborn, J., & Nazzi, T. (2009). Language specific prosodic preferences during the first half year of life : Evidence from German and French infants. *Infant Behavior and Development*, *32*(3), 262-274. https://doi.org/10.1016/j.infbeh.2009.03.004

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal. Biometrische Zeitschrift*, *50*(3), 346-363. https://doi.org/10.1002/bimj.200810425

Istria, M., Nicolas-Jeantoux, C., & Tamboise, J. (1982). *Manuel de lecture labiale : Exercices d'entraînement*. Masson.

Kanokpermpoon, M. (2007). A Preliminary Investigation of English and Thai Consonants. *Reflections / Sigma Theta Tau*, *10*(1), 10-21.

Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Ge, L., & Pascalis, O. (2007). The other-race effect develops during infancy: Evidence of perceptual narrowing. *Psychological Science*, *18*(12), 1084-1089. https://doi.org/10.1111/j.1467-9280.2007.02029.x

Kosawat, K. (2003). *Méthodes de segmentation et d'analyse automatique de textes thaï* [Thèse de Doctorat en Informatique, Université Paris-Est]. HAL. https://tel.archives-ouvertes.fr/tel-00626256

Kubicek, C., Hillairet de Boisferon, A., Dupierrix, E., Loevenbruck, H., Gervain, J., & Schwarzer, G. (2013). Face-scanning behavior to silently talking faces in 12-month-old infants: The impact of pre-exposed auditory speech. *International Journal of Behavioral Development*, *37*(2), 106-110. https://doi.org/10.1177/0165025412473016

Kubicek, C., Hillairet de Boisferon, A., Dupierrix, E., Pascalis, O., Loevenbruck, H., Gervain, J., & Schwarzer, G. (2014). Cross-Modal Matching of Audio-Visual German and French Fluent Speech in Infancy. *PLOS ONE*, *9*(2), e89275. https://doi.org/10.1371/journal.pone.0089275

Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science*, *218*(4577), 1138-1141. https://doi.org/10.1126/science.7146899

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, *9*(2), F13-F21. https://doi.org/10.1111/j.1467-7687.2006.00468.x

Labrune, L. (2006). *La phonologie du japonais*. Peeters.

Lado Robert. (1957). *Linguistics Across Cultures : Applied Linguistics for Language Teachers*. University of Michigan press.

Le Corre, C. (2013). *Etude de la prononciation des étudiants thaïlandais en apprentissage du français et majeure de français à l'université de KHON KAEN, Thaïlande.* Université de Khon Kaen.

Legerstee, M. (1990). Infants use multimodal information to imitate speech sounds. *Infant Behavior and Development*, *13*(3), 343-354.

Lewkowicz, D. J., & Hansen-Tift, A. M. (2012). Infants deploy selective attention to the mouth of a talking face when learning speech. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(5), 1431-1436.

Maspero, H. (1911). Contribution à l'étude du système phonétique des langue Thai. *Bulletin de l'Ecole française d'Extrême-Orient*, *11*(1), 153-169.

Mattock, K., & Burnham, D. (2006). Chinese and English Infants' Tone Perception: Evidence for Perceptual Reorganization. *Infancy*, *10*(3), 241-265.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*, 746-748.

Meltzoff, A. N., & Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. *Science*, *198*(4312), 74-78.

Meltzoff, A. N., & Moore, M. K. (1997). Explaining Facial Imitation: A Theoretical Model. *Early Development & Parenting*, *6*(3-4), 179-192. https://doi.org/10.1002/(SICI)1099-0917(199709/12)6:3/4<179::AID-EDP157>3.0.CO;2-R

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M., Jenkins, J. J., & Fujimura, O. (1975). An effect of linguistic experience : The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception & Psychophysics*, *18*(5), 331-340.

Nacaskul, K. (1979). A note on English loanwords in Thai. Dans T. L. Thongkum, V. Panupong, P. Kullavanijaya, & K. Tingsabadh, *Studies in Tai and Mon-Khmer phonetics and phonology in honour of Eugénie J. A. Henderson* (p. 151-162). Chulalongkorn University Press.

Nelson, C. A. (2001). The development and neural bases of face recognition. *Infant and Child Development*, *10*, 3-18. https://doi.org/10.1002/icd.239

Ngammana, P. (2011). *Les problèmes de la prononciation du français chez les lycéens thaïlandais* [Mémoire de Master Langue et Littérature françaises]. Université Chulalongkorn.

Pallegoix, J. B. (1854). *Dictionarium linguae Thaĭ sive Siamensis interpretatione Latina Gallica et Anglica illustratum*. Parisiis : Jussu Imperatoris Impressum in Typographeo Imperatorio.

Pascalis, O., De Haan, M., & Nelson, C. A. (2002). Is Face Processing Species-Specific During the First Year of Life? *Science*, *296*, 1321-1323. https://doi.org/10.1126/science.1070223

Patterson, M. L., & Werker, J. F. (1999). Matching phonetic information in lips and voice is robust in 4.5-month-old infants. *Infant Behavior & Development*, *22*(2), 237-247.

Polka, L., & Werker, J. F. (1994). Developmental Changes in Perception of Nonnative Vowel Contrasts. *Journal of Experimental Psychology. Human Perception and Performance*, *20*(2), 421-435. https://doi.org/10.1037/0096-1523.20.2.421

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

Premsrirat, S., Depadung, S., Suwanket, A., & Buasuang, A. (2004). *Ethnolinguistic Maps of Thailand*. Kurusapa Printing Ladphrao.

Promkesa, S. (2014). *Étude des problèmes de prononciation des consonnes fricatives du français par des apprenants thaïlandais et propositions de correction phonétique* [Mémoire de Master 2 en Français Langue Etrangère, Université Stendhal]. DUMAS. https://dumas.ccsd.cnrs.fr/dumas-01078534

Purinthrapibal, S. (2018). ปัญหาการออกเสียงสระนาสิกภาษาฝรั่งเศสของนักเรียนมัธยมศึกษาตอนปลายในโรงเรียน 14 จังหวัดภาคใต้ของประเทศไทย [Tr. Les problèmes de la prononciation des voyelles nasales du français chez les lycéens dans 14 provinces du Sud de la Thaïlande]. *Journal of International Studies*, *8*(2), 172-194.

Rose, I., MacWhinney, B., Byrne, R., Hedlund, G., Maddocks, K., O'Brien, P., & Wareham, T. (2006). *Phon (version 3.4.2)*.

Rousset, I. (2004). *Structures syllabiques et lexicales des langues du monde Données, typologies, tendances universelles et contraintes substantielles* [Thèse de Doctorat en Sciences du langage, Université Grenoble III-Stendhal]. HAL. https://tel.archives-ouvertes.fr/tel-00250154

Silverman, D. (1992). Multiple Scansions in Loanword Phonology: Evidence from Cantonese. *Phonology*, *9*(2), 289-328.

Sridhanyarat, K. (2017). The Acquisition of L2 Fricatives in Thai Learners' Interlanguage. *3L The Southeast Asian Journal of English Language Studies*, *23*(1), 15-34. https://doi.org/10.17576/3L-2017-2301-02

Sumby, W. H., & Pollack, I. (1954). Visual Contribution to Speech Intelligibility in Noise. *The Journal of the Acoustical Society of America*, *26*(2), 212-215.

Tran, T. T. H. (2011). *Processus d'acquisition des clusters et autres séquences de consonnes en langue seconde : De l'analyse acoustico-perceptive des séquences consonantiques du vietnamien à l'analyse de la perception et production des clusters du français par des apprenants vietnamiens du FLE* [Thèse de Doctorat en Sciences du langage, Université de Grenoble]. HAL. https://tel.archives-ouvertes.fr/tel-01568326

Trehub, S. E. (1976). The Discrimination of Foreign Speech Contrasts by Infants and Adults. *Child Development*, *47*(2), 466-472. https://doi.org/10.2307/1128803

Trehub, S., & Rabinovitch, M. S. (1972). Auditory-linguistic sensitivity in early infancy. *Developmental Psychology*, *6*(1), 74-77. https://doi.org/10.1037/h0032197

Tri Endarto, I. (2015). Comparison between English Loanwords in Thai and Indonesian: A Comparative Study in Phonology and Morphology. *3rd AASIC: Sustainable Development of Asian Community*, 49-52.

Troubetzkoy, N. S. (1949). *Principes de phonologie par N.S Troubetzkoy* (traduit par J. Cantineau). Klincksieck.

Tusnyingyong, S., & Tran, T. T. H. (2022). Pourquoi les apprenants thaïlandais confondent-ils souvent /v/ et /w/ du français ? Une étude pilote sur l'effet de l'orthographe. *SHS Web of Conferences*, *138*. https://doi.org/10.1051/shsconf/202213808007

Tyler, M. D., Clot, E., Villain–Bailly, M.-S., & Pattamadilok, C. (2019). Perceptual assimilation of English dental fricatives by native speakers of European French. *Proceedings Of The 19th International Congress of Phonetic Sciences*. https://hal.archives-ouvertes.fr/hal-02494288

Vaissière Jacqueline. (2006). *La phonétique*. Presses universitaires de France.

Vajiravudh, R. (1913). The Romanisation of Siamese Words. *Journal of the Siam Society*, *4*(9), 1-11.

Vallée, N., Boë, L.-J., & Stefanuto, M. (1999). Typologies phonologiques et tendances universelles. Approche substantialiste. *Linx : Revue des linguistes de l'université Paris X Nanterre*, *11*, 31-54. https://doi.org/10.4000/linx.863

Wang, Y., Behne, D. M., & Jiang, H. (2008). Linguistic experience and audio-visual perception of non-native fricatives. *The Journal of the Acoustical Society of America*, *124*(3), 1716-1726. https://doi.org/10.1121/1.2956483

Wang, Y., Behne, D. M., & Jiang, H. (2009). Influence of native language phonetic system on audio-visual speech perception. *Journal of Phonetics*, *37*(3), 344-356.

Weikum, W. M., Vouloumanos, A., Navarra, J., Soto-Faraco, S., Sebastián-Gallés, N., & Werker, J. F. (2007). Visual Language Discrimination in Infancy. *Science (New York, N.Y.)*, *316*, 1159. https://doi.org/10.1126/science.1137686

Werker, J. F., Frost, P. E., & McGurk, H. (1992). La langue et les Lèvres : Cross-language Influences on Bimodal Speech Perception. *Canadian Journal of Psychology*, *46*(4), 551-568. https://doi.org/10.1037/h0084331

Werker, J. F., Gilbert, J. H. V., Humphrey, K., & Tees, R. C. (1981). Developmental Aspects of Cross-Language Speech Perception. *Child Development*, *52*(1), 349-355. https://doi.org/10.2307/1129249

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*(1), 49-63. https://doi.org/10.1016/S0163-6383(84)80022-3

Yamlamai, N. (2017). *Etude phonémique et acoustique des consonnes finales du thaï* [Mémoire de Master 1 Sciences du Langage]. Université Grenoble Alpes.

Yeung, H. H., Chen, K. H., & Werker, J. F. (2013). When does native language input affect phonetic perception? The precocious case of lexical tone. *Journal of Memory and Language*, *68*(2), 123-139. https://doi.org/10.1016/j.jml.2012.09.004

Abréviations utilisées

A :	(modalité) Auditive
AO :	Organes articulatoires
AV :	(modalité) Audiovisuelle
CV :	Structure Consonne Voyelle
LE :	Langue étrangère
L1 :	Langue première
PAM :	Modèle d'assimilation perceptive
PAM/AO :	Modèle d'assimilation perceptive avec l'hypothèse des organes articulatoires
SLM :	Modèle d'apprentissage des langues
VCV :	Structure Voyelle Consonne Voyelle

Liste des figures

Figure 1. Représentation des scénarios d'assimilation du modèle PAM de Best (1995a) (figure repérée de Burfin, 2015, p. 20) 18
Figure 2. Détectabilité (d') du lieu d'articulation en trois modalités (visuelle, auditive et audiovisuelle) chez les apprenants japonophones et hispanophones de l'anglais (selon Hazan et al., 2006) 34
Figure 3. Pourcentage des réponses correctes chez les francophones dans la tâche de discrimination des sons /f/ ~ /θ/ en modalités auditive et audiovisuelle (selon Burfin, 2015) (figure adaptée de l'original)..... 36
Figure 4. Phrases du français translitérées en caractères thaïs avec la traduction en thaï et en anglais (Chitkla, 2017, p. 16) 46
Figure 5. Introduction de l'alphabet de l'anglais dans l'école maternelle (Le Corre, 2013)..... 47
Figure 6. Représentation schématique d'une séquence de test dans la tâche de perception en modalité audiovisuelle 59
Figure 7. Représentation schématique de la tâche de production en modalité audiovisuelle..... 60
Figure 8. Exemple de l'extraction des données 61
Figure 9. Affichage de Phon avec les données segmentées, annotées et transcrites en API 61
Figure 10. Affichage de Praat : l'analyse acoustique de la prononciation réelle du stimulus [ava] en modalité auditive (prononcé comme [awa]) 62
Figure 11. Affichage de la fenêtre session Phon avec, comme exemple ici, le traitement du stimulus [ava] : les tiers « Orthography » (entrée orthographique) et « IPA Target » (transcription cible [ava]), « IPA Actual » (transcription de la prononciation réelle de l'apprenant [awa]) et en haut à gauche, la syllabification et l'alignement entre la syllabe cible (Target Syllable) et la syllabe réalisée (Actual Syllable)..... 63
Figure 12. Pourcentage des réponses correctes dans la tâche de perception en fonction de la modalité (*A* vs *AV*) et de la consonne de référence (/f v w/)..... 65
Figure 13. Pourcentage des réponses correctes dans la tâche de perception en fonction de la modalité (*A* vs AV) et de la voyelle (/i e ɛ a/). Les voyelles < é > et < è > dans la légende correspondent aux voyelles /e/ et /ɛ/ respectivement..... 66
Figure 14. Pourcentage des réponses correctes dans la tâche de perception en fonction de la modalité (A vs AV) et de la structure syllabique (CV et VCV) 67
Figure 15. Temps moyen de réponse dans la tâche de perception en fonction de la modalité (*A* vs *AV*) et de la consonne de référence (/f v w/) 69
Figure 16. Temps moyen de réponse dans la tâche de perception en fonction de la modalité (auditive vs audiovisuelle) et de la voyelle (/i e ɛ a/)..... 70
Figure 17. Temps de réponse en moyenne en perception en fonction de la modalité (A vs AV) et de la structure syllabique (CV ou VCV)..... 71
Figure 18. Pourcentage de la prononciation correcte en fonction de la modalité (A vs AV) et en fonction de la consonne de référence (/f v w/)..... 72
Figure 19. Pourcentage de la prononciation correcte en fonction de la modalité (A vs AV) et en fonction de la voyelle (/i e ɛ a/)..... 73
Figure 20. Pourcentage de la prononciation correcte en fonction de la modalité (A vs AV) et en fonction de la structure syllabique (CV et VCV)..... 74
Figure 21. Pourcentage de la prononciation incorrecte en fonction de la structure syllabique (CV et VCV) en deux modalités (A et AV)..... 75

Figure 22. Pourcentage des types d'erreurs en production en fonction de la modalité (A et AV) 76
Figure 23. Exemple de l'insertion de [w] dans la cible [ivi] en modalité AV. 78
Figure 24. Exemple de l'insertion de [v] dans la cible [awa] en modalité A. 79
Figure 25. Exemples de la réalisation du pré-voisement avant la production de [v]. La production de [v] dans [va] (à gauche) et [ava] (à droite) d'un participant testé en modalité AV. 80
Figure 26. Exemples de la réalisation du pré-voisement avant la production de [v]. La production de [v] dans [va] (à gauche) et [ava] (à droite) d'un participant testé en modalité auditive. 80
Figure 27. Comparaison des deux signaux acoustiques : à gauche, la production de [v] précédée du pré-voisement et à droite, la production de [v] sans la réalisation du voisement avant. Le cercle rouge pointe la différence de l'intensité du bruit de friction. 81
Figure 28. Exemple de la stratégie de décomposition des traits articulatoires de [v]. La prononciation de l'item [eve] d'un participant testé en modalité AV. 82
Figure 29. Exemple de la stratégie de décomposition des traits articulatoires de [v]. La prononciation de l'item [va] d'un participant testé en modalité A. 82
Figure 30. Taux de réussite en perception et en production en fonction de la modalité 83
Figure 31. Taux de réussite en perception (à gauche) et en production (à droite) en fonction de la modalité et de la consonne. 83
Figure 33. Représentation de la place de [v] dans l'espace phonologique de la L1 chez les apprenants thaïlandais lors de l'exposition à l'input auditif (à gauche) et lors de l'exposition à l'input audiovisuel (à droite) 90

Liste des tables

Table 1. Classification de visèmes consonantiques du français proposée par Istria et al. (1982) 26
Table 2. Classification de visèmes consonantiques du français proposée par Dumont et Calbour (2002) 27
Table 3. Pourcentage des réponses correctes (identification du lieu d'articulation) en modalités auditive et audiovisuelle chez les anglophones, les apprenants coréens et les apprenants sinophones de l'anglais (selon Wang et al., 2009) (Table adapté de l'original) 35
Table 4. Phonèmes consonantiques du thaï (pour chaque lieu d'articulation, les non-voisés en colonne de gauche et les voisés de droite) 41
Table 5. Phonèmes consonantiques du français (à gauche, les non-voisés et à droite, les voisés) 42
Table 6. Graphèmes consonantiques du thaï (adaptés de Kosawat, 2003) 42
Table 7. Classification de visèmes consonantiques du thaï 44
Table 8. Stimuli utilisés dans le pré-test 52
Table 9. Trois groupes de stimuli : 24 stimuli de test, 24 de contrôle et 12 d'entraînement 54
Table 10. Exemples de séquences de test pour la tâche de perception. 55
Table 11. Exemples de séquences de contrôle pour la tâche de perception. 56
Table 12. Exemples de séquences de test et de contrôle mises en ordre aléatoire pour la tâche de perception. 56
Table 13. Exemples de stimuli de test et de contrôle mis en ordre aléatoire pour la tâche de production. 57
Table 14. Réponses incorrectes en modalité A en fonction de la consonne de référence 68
Table 15. Réponses incorrectes en modalité AV en fonction de la consonne de référence et de la consonne de test 68
Table 16. Nombre d'occurrences et pourcentage d'erreurs en production 75
Table 17. Taux d'erreurs en production en fonction de la modalité et du contexte vocalique 76
Table 18. Nombre de cas de substitution en modalité A 77
Table 19. Nombre de cas de substitution en modalité AV 77
Table 20. Les cas d'insertion en deux modalités 78
Table 21. Nombre de confusion /v/~/w/ en perception et en production en fonction de la modalité et du contexte vocalique 85
Table 22. Nombre de confusion /v/~/w/ en perception et en production en fonction de la modalité et de la structure syllabique 86

Tables des annexes

Annexe 1 Corpus (tâche de perception)..... 109
Annexe 2 Corpus (tâche de production) 114
Annexe 3 Questionnaire..... 121

Annexe 1
Corpus (tâche de perception)

4 séquences d'entraînement

N°	Items de référence	Items de test		
1	/ma/	/ka/	/sa/	/ma/
2	/si/	/mi/	/si/	/ki/
3	/eke/	/eke/	/eme/	/ese/
4	/ɛsɛ/	/ɛkɛ/	/ɛmɛ/	/ɛsɛ/

72 séquences de test

N°	Items de référence	Items de test		
1	/fi/	/fi/	/vi/	/wi/
2	/fi/	/wi/	/fi/	/vi/
3	/fi/	/vi/	/wi/	/fi/
4	/fe/	/fe/	/we/	/ve/
5	/fe/	/ve/	/fe/	/we/
6	/fe/	/we/	/ve/	/fe/
7	/fɛ/	/fɛ/	/vɛ/	/wɛ/
8	/fɛ/	/wɛ/	/fɛ/	/vɛ/
9	/fɛ/	/vɛ/	/wɛ/	/fɛ/
10	/fa/	/fa/	/wa/	/va/
11	/fa/	/va/	/fa/	/wa/
12	/fa/	/wa/	/va/	/fa/
13	/vi/	/vi/	/fi/	/wi/
14	/vi/	/wi/	/vi/	/fi/
15	/vi/	/fi/	/wi/	/vi/
16	/ve/	/ve/	/we/	/fe/

17	/ve/	/fe/	/ve/	/we/
18	/ve/	/we/	/fe/	/ve/
19	/vɛ/	/vɛ/	/fɛ/	/wɛ/
20	/vɛ/	/wɛ/	/vɛ/	/fɛ/
21	/vɛ/	/fɛ/	/wɛ/	/vɛ/
22	/va/	/va/	/wa/	/fa/
23	/va/	/fa/	/va/	/wa/
24	/va/	/wa/	/fa/	/va/
25	/wi/	/wi/	/fi/	/vi/
26	/wi/	/vi/	/wi/	/fi/
27	/wi/	/fi/	/vi/	/wi/
28	/we/	/we/	/ve/	/fe/
29	/we/	/fe/	/we/	/ve/
30	/we/	/ve/	/fe/	/we/
31	/wɛ/	/wɛ/	/fɛ/	/vɛ/
32	/wɛ/	/vɛ/	/wɛ/	/fɛ/
33	/wɛ/	/fɛ/	/vɛ/	/wɛ/
34	/wa/	/wa/	/va/	/fa/
35	/wa/	/fa/	/wa/	/va/
36	/wa/	/va/	/fa/	/wa/
37	/ifi/	/ifi/	/ivi/	/iwi/
38	/ifi/	/iwi/	/ifi/	/ivi/
39	/ifi/	/ivi/	/iwi/	/ifi/
40	/efe/	/efe/	/ewe/	/eve/
41	/efe/	/eve/	/efe/	/ewe/
42	/efe/	/ewe/	/eve/	/efe/
43	/ɛfɛ/	/ɛfɛ/	/ɛvɛ/	/ɛwɛ/

44	/ɛfɛ/	/ɛwɛ/	/ɛfɛ/	/ɛvɛ/
45	/ɛfɛ/	/ɛvɛ/	/ɛwɛ/	/ɛfɛ/
46	/afa/	/afa/	/awa/	/ava/
47	/afa/	/ava/	/afa/	/awa/
48	/afa/	/awa/	/ava/	/afa/
49	/ivi/	/ivi/	/ifi/	/iwi/
50	/ivi/	/iwi/	/ivi/	/ifi/
51	/ivi/	/ifi/	/iwi/	/ivi/
52	/eve/	/eve/	/ewe/	/efe/
53	/eve/	/éfé/	/eve/	/ewe/
54	/eve/	/ewe/	/efe/	/eve/
55	/ɛvɛ/	/ɛvɛ/	/ɛfɛ/	/ɛwɛ/
56	/ɛvɛ/	/ɛwɛ/	/ɛvɛ/	/ɛfɛ/
57	/ɛvɛ/	/ɛfɛ/	/ɛwɛ/	/ɛvɛ/
58	/ava/	/ava/	/awa/	/afa/
59	/ava/	/afa/	/ava/	/awa/
60	/ava/	/awa/	/afa/	/ava/
61	/iwi/	/iwi/	/ifi/	/ivi/
62	/iwi/	/ivi/	/iwi/	/ifi/
63	/iwi/	/ifi/	/ivi/	/iwi/
64	/ewe/	/ewe/	/eve/	/efe/
65	/ewe/	/efe/	/ewe/	/eve/
66	/ewe/	/eve/	/efe/	/ewe/
67	/ɛwɛ/	/ɛwɛ/	/ɛfɛ/	/ɛvɛ/
68	/ɛwɛ/	/ɛvɛ/	/ɛwɛ/	/ɛfɛ/
69	/ɛwɛ/	/ɛfɛ/	/ɛvɛ/	/ɛwɛ/
70	/awa/	/awa/	/ava/	/afa/

71	/awa/	/afa/	/awa/	/ava/
72	/awa/	/ava/	/afa/	/awa/

24 séquences de contrôle

N°	Items de référence	Items de test		
1	/pi/	/pi/	/bi/	/ti/
2	/pi/	/ti/	/pi/	/bi/
3	/pi/	/bi/	/ti/	/pi/
4	/pe/	/pe/	/te/	/be/
5	/pe/	/be/	/pe/	/te/
6	/pe/	/te/	/be/	/pe/
7	/pɛ/	/pɛ/	/bɛ/	/tɛ/
8	/pɛ/	/tɛ/	/pɛ/	/bɛ/
9	/pɛ/	/bɛ/	/tɛ/	/pɛ/
10	/pa/	/pa/	/ta/	/ba/
11	/pa/	/ba/	/pa/	/ta/
12	/pa/	/ta/	/ba/	/pa/
13	/ipi/	/ipi/	/ibi/	/iti/
14	/ipi/	/iti/	/ipi/	/ibi/
15	/ipi/	/ibi/	/iti/	/ipi/
16	/epe/	/epe/	/ete/	/ebe/
17	/epe/	/ebe/	/epe/	/ete/
18	/epe/	/ete/	/ebe/	/epe/
19	/ɛpɛ/	/ɛpɛ/	/ɛbɛ/	/ɛtɛ/
20	/ɛpɛ/	/ɛtɛ/	/ɛpɛ/	/ɛbɛ/
21	/ɛpɛ/	/ɛbɛ/	/ɛtɛ/	/ɛpɛ/

22	/apa/	/apa/	/ata/	/aba/
23	/apa/	/aba/	/apa/	/ata/
24	/apa/	/ata/	/aba/	/apa/

Annexe 2
Corpus (tâche de production)

4 items d'entraînement

N°	Items
1	/ma/
2	/si/
3	/eke/
4	/ɛmɛ/

120 items de test

N°	Items
1	/fi/
2	/fi/
3	/fi/
4	/fi/
5	/fi/
6	/fe/
7	/fe/
8	/fe/
9	/fe/
10	/fe/
11	/fɛ/
12	/fɛ/
13	/fɛ/
14	/fɛ/
15	/fɛ/
16	/fa/

17	/fa/
18	/fa/
19	/fa/
20	/fa/
21	/vi/
22	/vi/
23	/vi/
24	/vi/
25	/vi/
26	/ve/
27	/ve/
28	/ve/
29	/ve/
30	/ve/
31	/vɛ/
32	/vɛ/
33	/vɛ/
34	/vɛ/
35	/vɛ/
36	/va/
37	/va/
38	/va/
39	/va/
40	/va/
41	/wi/
42	/wi/
43	/wi/

44	/wi/
45	/wi/
46	/we/
47	/we/
48	/we/
49	/we/
50	/we/
51	/wɛ/
52	/wɛ/
53	/wɛ/
54	/wɛ/
55	/wɛ/
56	/wa/
57	/wa/
58	/wa/
59	/wa/
60	/wa/
61	/ifi/
62	/ifi/
63	/ifi/
64	/ifi/
65	/ifi/
66	/efe/
67	/efe/
68	/efe/
69	/efe/
70	/efe/

71	/ɛfɛ/
72	/ɛfɛ/
73	/ɛfɛ/
74	/ɛfɛ/
75	/ɛfɛ/
76	/afa/
77	/afa/
78	/afa/
79	/afa/
80	/afa/
81	/ivi/
82	/ivi/
83	/ivi/
84	/ivi/
85	/ivi/
86	/eve/
87	/eve/
88	/eve/
89	/eve/
90	/eve/
91	/ɛvɛ/
92	/ɛvɛ/
93	/ɛvɛ/
94	/ɛvɛ/
95	/ɛvɛ/
96	/ava/
97	/ava/

98	/ava/
99	/ava/
100	/ava/
101	/iwi/
102	/iwi/
103	/iwi/
104	/iwi/
105	/iwi/
106	/ewe/
107	/ewe/
108	/ewe/
109	/ewe/
110	/ewe/
111	/ɛwɛ/
112	/ɛwɛ/
113	/ɛwɛ/
114	/ɛwɛ/
115	/ɛwɛ/
116	/awa/
117	/awa/
118	/awa/
119	/awa/
120	/awa/

30 items de contrôle

N°	**Items**

1	/pi/
2	/pe/
3	/pɛ/
4	/pa/
5	/ipi/
6	/epe/
7	/ɛpɛ/
8	/apa/
9	/bi/
10	/be/
11	/bɛ/
12	/ba/
13	/ibi/
14	/ebe/
15	/ɛbɛ/
16	/aba/
17	/ti/
18	/te/
19	/tɛ/
20	/ta/
21	/iti/
22	/ete/
23	/ɛtɛ/
24	/ata/
25	/pi/
26	/pe/
27	/pɛ/

28	/pa/
29	/ipi/
30	/epe/

Annexe 3
Questionnaire

แบบสอบถาม

แบบสอบถามนี้จัดทำขึ้นภายใต้งานวิจัยปริญญาโทปีที่2 สาขาภาษาศาสตร์ มหาวิทยาลัยGrenoble Alpes ประเทศฝรั่งเศส ผู้วิจัยใคร่ขอความร่วมมือจากผู้ทดลองกรอกข้อมูลตามความเป็นจริงเพื่อจะเป็นประโยชน์ต่อการวิเคราะห์ข้อมูล ข้อมูลส่วนตัวของผู้ทดลองจะนำมาใช้เฉพาะเหตุผลเชิงวิชาการเท่านั้น และจะไม่ถูกเผยแพร่ต่อสาธารณะ ผู้วิจัยขอขอบคุณเป็นอย่างสูงที่ร่วมเป็นส่วนหนึ่งของงานวิจัยชิ้นนี้

ชื่อ ..... นามสกุล ..... เพศ ( ) ชาย ( ) หญิง

อายุ ..... ปี สัญชาติ ..... อีเมล์ .....

ผู้ทดลองพูด ( ) ภาษากลาง ( ) ภาษาเหนือ ( ) ภาษาอีสาน ( ) ภาษาใต้ ( ) อื่นๆ โปรดระบุ .....

บุคคลในครอบครัวพูดภาษาถิ่นหรือไม่ ( ) ไม่ ( ) ใช่ โปรดระบุ .....

ระดับการศึกษา .....

ภาษาต่างประเทศภาษาแรกที่เรียน .....

ภาษาต่างประเทศภาษาที่สองที่เรียน ..... ระดับ .....

ภาษาต่างประเทศภาษาที่สามที่เรียน ..... ระดับ .....

ภาษาต่างประเทศภาษาที่สี่ที่เรียน ..... ระดับ .....

เรียนภาษาฝรั่งเศสเป็นระยะเวลา ..... เดือน/ปี ประเทศที่เรียนภาษาฝรั่งเศส .....

อาศัยอยู่ที่ประเทศฝรั่งเศสเป็นระยะเวลา .....

ความผิดปกติทางการมองเห็นและการได้ยิน ( ) ไม่มี ( ) มี โปรดระบุ .....

ความยากลำบากที่พบเจอในการออกเสียงภาษาฝรั่งเศส

.....

.....

.....

.....

Tables des matières

Remerciements **5**

Sommaire **7**

Introduction **9**

Chapitre 1 La perception auditive **12**

1.1 Perception auditive de la parole : de la langue native aux phonèmes non natifs 13

1.1.1 Perception de la parole chez l'enfant pendant la première année de vie 13

1.1.2 Perception de la parole en apprentissage d'une langue étrangère 15

1.2 Perception auditive en langue étrangère : modèles d'assimilation phonologique 17

1.2.1 Modèle d'assimilation perceptive (PAM) 17

1.2.2 Modèle d'apprentissage des langues (SLM) 19

Chapitre 2 La perception visuelle **22**

2.1 Perception visuelle : la reconnaissance du visage 23

2.2 Perception visuelle dans le traitement du langage : lecture dite labiale 24

2.3 Visèmes 25

Chapitre 3 La perception audiovisuelle **28**

3.1 Perception audiovisuelle dans le traitement de la langue native 29

3.2 Perception audiovisuelle dans le traitement d'une langue étrangère 31

3.2.1 Chez les bébés 31

3.2.2 Chez les adultes 32

3.2.3 Facteurs impactant l'utilisation des informations visuelles lors du traitement d'une langue étrangère 36

3.2.3.1 Saillance du visème *36*

3.2.3.2 Organisation de l'espace phonologique de la L1 ..... *37*

3.2.3.3 Relation entre les visèmes de la L1 et ceux de la LE *37*

3.2.3.4 Performance dans la LE *38*

Chapitre 4 Confusion entre /v/ et /w/ chez les apprenants thaïlandais **40**

4.1 Le thaï 41

4.1.1 Le système consonantique 41

4.1.2 Le système d'écriture du thaï 42

4.1.3 Les visèmes du thaï 43

4.2 La confusion /v/ ~ /w/ chez les apprenants thaïlandais 44

4.2.1 A l'origine de la confusion /v/ ~ /w/ chez les Thaïlandais 45

4.2.2 L'impact direct de la translittération franco-thaïlandaise 46

4.2.3 L'impact indirect de la translittération anglo-thaïlandaise 47

Chapitre 5 Étude expérimentale sur l'apport des informations visuelles dans la perception et la production de la consonne /v/ chez les apprenants thaïlandais **49**

5.1 Objectifs, questions de recherche et hypothèses 50
5.2 Méthodologie 51
5.2.1 Participants 51
5.2.2 Constitution du corpus 53
5.2.2.1 Préparation des stimuli de l'expérience *54*
5.2.2.2 Corpus pour la tâche de perception *55*
5.2.2.2 Corpus pour la tâche de production *56*
5.2.3 Déroulement de l'expérience 57
5.2.3.1 Test de perception *57*
5.2.3.2 Test de production *60*
5.2.4 Traitement des données 61
5.2.4.1 Perception *61*
5.2.4.2 Production *61*
5.2.4.3 Analyses statistiques *63*
5.3 Résultats 64
5.3.1 Tâche de perception 64
5.3.1.1 Taux de réussite *64*
a. En fonction de la consonne de référence 64
b. En fonction du contexte vocalique 65
c. En fonction de la structure syllabique 66
5.3.1.2 Réponses incorrectes *67*
5.3.1.3 Temps de réponse *68*
a. En fonction de la consonne de référence 69
b. En fonction du contexte vocalique 70
c. En fonction de la structure syllabique 70
5.3.2 Tâche de production 71
5.3.2.1 Taux de réussite *71*
a. En fonction de la consonne 72
b. En fonction du contexte vocalique 73
c. En fonction de la structure syllabique 73
5.3.2.2 Taux d'erreurs *74*
a. La substitution 76
b. L'insertion 77
5.3.2.3 Stratégies de prononciation de /v/ *79*
a. Première stratégie : le pré-voisement 79
b. Deuxième stratégie : la décomposition des traits articulatoires 81
5.3.3 Relation entre la perception et la production 82

5.4 Discussion 86
Conclusion et perspectives 92
Bibliographie 94
Abréviations utilisées 104
Liste des figures 105
Liste des tables 107
Tables des annexes 108
Tables des matières 122

MOTS-CLÉS : consonne /v/, confusion entre /v/ et /w/, présentation audiovisuelle de la parole, traitement de la parole, apprenants thaïlandais

RÉSUMÉ

Cette étude s'inscrivant dans le cadre de l'enseignement et de l'apprentissage d'une LE porte sur des difficultés récurrentes dans le traitement de la consonne non native /v/ en position initiale de syllabe chez les apprenants thaïlandais. Il a été démontré que cette consonne en position initiale de syllabe est très souvent confondue avec /w/. Nous avons effectué une étude expérimentale pour vérifier l'apport des informations audiovisuelles de la parole dans le traitement de la consonne /v/ chez les apprenants thaïlandais du français. Deux tâches (perception et production) ont été effectuées en deux modalités (modalités auditive et audiovisuelle). Les résultats montrent que le score de réussite en modalité audiovisuelle est significativement plus élevé qu'en modalité auditive, quel que soit le type de tâche. La confusion entre /v/ et /w/ est moins présente en modalité audiovisuelle qu'en modalité auditive. De plus, le temps de réponse en perception est significativement moins important en modalité audiovisuelle qu'en modalité auditive.

KEYWORDS: /v/ consonant, confusion between /v/ et /w/, audiovisual speech presentation, speech processing, Thai learners

ABSTRACT

This study within the framework of foreign language teaching and learning focuses on the recurrent difficulty in processing the non-native consonant /v/ in syllable-initial position among Thai learners. It has been shown that this consonant in syllable-initial position is very often confused with /w/. We therefore conducted an experimental study to verify the contribution of audiovisual speech information in the processing of the consonant /v/ in Thai learners of French. Two tasks (perception and production) were conducted, and two modalities (auditory and audiovisual modalities) were set up. The results show that the percentage of success in the audiovisual modality is significantly higher than that in the auditory modality, whatever the type of task. The confusion between /v/ and /w/ is less frequent in the audiovisual modality than in the auditory modality. Also, the response time in the audiovisual modality of the perception test is significantly less than that of the auditory modality.